



OSI POLICY PERSPECTIVES

# OPEN SOLUTIONS

UNIFYING THE MEANING OF  
OPEN AND DESIGNING A NEW  
GLOBAL OPEN SOLUTIONS  
POLICY FRAMEWORK

OSI POLICY PERSPECTIVE 4  
JANUARY 2021 EDITION OPEN  
SCHOLARSHIP INITIATIVE

## ABOUT OSI POLICY PERSPECTIVES

The OSI Policy Perspectives series offers broad, common ground perspectives on key issues in scholarly communication. Each report summarizes the current state of a particular issue and what we know about it, and also attempts to articulate the perspectives and lessons of experience from all stakeholder groups in scholarly communication on this issue (particularly but not exclusively as expressed in OSI conversations) and identify what common ground might exist for building broadly acceptable policy.

OSI is not a democratic body that speaks with one voice on any particular issue. Trying to reconcile the views, intentions, and motivations of all the different actors, communities and groups in the scholarly communication space—which are very rarely entirely aligned—is challenging. We acknowledge, therefore, that these reports may be (and in fact, probably are) an imperfect reflection of the many perspectives and ideas in this group. The fact that these reports sometimes need to be published in a rush, in response to policy commenting deadlines and other pressures only makes this imperfection more likely.

We also acknowledge, however, that OSI often considers a wider range of perspectives than established policy making bodies in scholarly communication, and that our relative strength is showcasing this range of perspectives and noting how they differ, and importantly, how they share common ground. To this end, we hope it is valuable to produce these reports, however imperfect, and share them with the scholarly communication community and beyond.

© 2021 OSI. Except where otherwise noted, this report is free to use and reuse under a CC-BY-NC-ND license.

**ACKNOWLEDGMENTS:** Thank you to Roger Schonfeld, Dan Dunleavy, Shelley Stall, Tim Davies, Keith Webster and Bhanu Neupane for their valuable input into this report. Thanks as well to several others in OSI who contributed important ideas and feedback but are not named here, and to George Mason University Press.

**AUTHOR CONFLICT STATEMENT:** The lead author of this report, Glenn Hampson, is the program director and principal investigator for OSI, which receives funding from foundations, UNESCO, commercial publishers, and individual participants by way of conference fees. Funders, however, have no privileged input into OSI policy deliberations apart from being equal members of the OSI community. OSI has many voices contributing to documents such as this, and endeavors to maintain an inclusive and balanced perspective on scholarly communication issues. Mel DeSart is Head of the Engineering Library at the University of Washington; Lynn Kamerlin is a scientist and director of the Kamerlin Lab at Uppsala University; Rob Johnson is the founder and director of Research Consulting; Hilary Hanahoe is the Secretary General of the Research Data Alliance; Amy Nurnberger is the Program Head for Data Management Services at MIT; and Chris Graf is the Director of Research Integrity and Publishing Ethics at Wiley Publishing.

**DISCLAIMER:** In this report, the authors have attempted to accurately represent the perspective and ideas of the broad open solutions community, and in particular UNESCO and OSI participants, alumni and observers. However, it is possible that this attempt is incomplete and/or inaccurate. Any responsibility for errors, omissions and/or misrepresentations rests solely with the lead author. Also, the findings and recommendations expressed herein also do not necessarily reflect the opinions of all co-authors, contributors, individual OSI participants, alumni, or observers, or any of the institutions, trustees, officers, or staff affiliated with these individuals.

**PRIOR AND OTHER VERSIONS:** There are no prior public versions of this document.

**CITATION:** Hampson, G, M DeSart, L Kamerlin, R Johnson, H Hanahoe, A Nurnberger and C Graf. 2021. OSI Policy Perspective 4: Open Solutions: Unifying the meaning of open and designing a new global open solutions policy framework. Open Scholarship Initiative. January 2021 edition. doi: 10.13021/osi2020.2930

# OPEN SOLUTIONS

## UNIFYING THE MEANING OF OPEN AND DESIGNING A NEW GLOBAL OPEN SOLUTIONS POLICY FRAMEWORK

<b>INTRODUCTION.....</b>	<b>i</b>
<b>OBJECTIVES.....</b>	<b>1</b>
<b>EXECUTIVE SUMMARY.....</b>	<b>3</b>
<b>SECTION 1: EVIDENCE.....</b>	<b>8</b>
BACKGROUND .....	8
UNESCO'S APPROACH.....	8
THE GLOBAL APPROACH.....	9
OPEN ACCESS.....	11
OPEN DATA.....	13
OPEN SOURCE.....	15
OPEN SCIENCE.....	16
OPEN GOVERNMENT.....	19
OPEN EDUCATIONAL RESOURCES.....	20
PRACTICE-BASED OPEN.....	21
SYNTHESIS.....	22
STAKEHOLDER NEEDS & PERSPECTIVES.....	30
RESEARCHERS.....	31
PUBLISHERS.....	41
RESEARCH INSTITUTIONS.....	49
SCHOLARLY SOCIETIES.....	51
FUNDING AGENCIES.....	51
MULTILATERALS.....	53
OTHER OPEN-FOCUSED AGENCIES.....	54
GENERAL PUBLIC.....	56
KEY OPEN POLICIES & ACTIONS.....	57
OPEN POLICIES IN GENERAL.....	57
REGULATIONS.....	59
LICENSING.....	62
<b>SECTION 2: RECOMMENDATIONS.....</b>	<b>65</b>
THE PATH TOWARD COMMON GOALS.....	65
NARROW.....	68
BROAD.....	69
PHILOSOPHICAL.....	70
TANGIBLE.....	71
COLLABORATIVE.....	71
STRATEGIC NEEDS.....	76
TACTICAL NEEDS.....	80
DISCOVERY.....	80
EDUCATION.....	80
RESEARCHER SUPPORT.....	81
RELATIONSHIPS.....	84
MEASURES.....	84
INFRASTRUCTURE.....	85
AGREEMENTS.....	86
PILOT POLICIES.....	87
APPROACH.....	88
ACTION STEPS.....	98
PARALLEL ACTIONS.....	100
DIVERSITY AND INCLUSION.....	100
INDICATORS & DIMENSIONS.....	101
CONCLUSION.....	105
<b>REFERENCES.....</b>	<b>109</b>
ANNEX 1: DRAFT UNESCO PROCLAMATION ON OPEN SOLUTIONS.....	117
ANNEX 2: THE INTERNATIONAL OPEN DATA CHARTER.....	121
ANNEX 3: THE ALL-SCHOLARSHIP REPOSITORY.....	129
ANNEX 4: DRAFT MARKETING PLAN FOR UNESCO'S OPEN SOLUTIONS POLICY.....	131
ANNEX 5: COMMON LICENSE TYPES.....	133
ANNEX 6: OPEN DATA WEBSITES BY SECTOR/TOPIC.....	134
ANNEX 7: KEY OPEN DEFINITIONS, GROUPS & POLICIES.....	139

To everyone in OSI who has worked hard to develop a clear, fair, and complete understanding of the open space.

## INTRODUCTION

“Experiments produce new knowledge, but if that knowledge does not circulate there is little opportunity for further progress.” This observation, richly explored in David Wootton’s “The Invention of Science” (Harper, 2015), is as true today as it was 400 years ago when science was still evolving in fits and starts. There is no question that knowledge needs to be shared. The question for our age is not why, but how. Specifically, how can we share more of our knowledge in ways that are consistent with the needs and values of research, and that will create the greatest long-term benefit for both research and society?

Before the birth of science, it was normal to construct explanations that conformed with “known truths” instead of simply searching for truth—to pound the square pegs of observation into the round holes of what the church and tradition said must be true. This approach stifled learning and innovation, and kept Western civilization in the dark for 2,000 years. Today, we are using the same approach in our efforts to share knowledge. At the policy making level, we assume we know all there is to know about open research, and we are working backward, pounding square peg solutions into the round holes of researcher needs and concerns. In the process, we aren’t finding truths and unlocking the real potential of open, just creating expedient solutions that conform to our particular circumstances and ideologies. Is there a better way?

In this report, we argue that it’s important to look at the effort to share knowledge more effectively from a researcher perspective, and to be open-minded and goal-oriented in our approach to constructing open solutions that are sustainable and work globally. What are we trying to accomplish with openness? And from this understanding and agreement, what tools and systems can we put in place that will best achieve our goals? Rather than spending more time and energy debating whose square pegs are best, we should reverse our approach, searching for truth with an open mind before creating solutions. Then, we should build vibrant and sustainable open solutions together on common ground and work together to solve the urgent problems that await humanity.

This challenge may seem like much ado about nothing to outside observers. It’s research we’re talking about, after all, not world peace. But historical glamor aside, we may not always fully grasp the magnitude of the societal transformation brought about a few hundred years ago by the brave writers, philosophers, experimenters and explorers who dared to challenge the church and conventional wisdom in the world’s first search for objective truth and fact. The mindset these individuals unleashed fundamentally changed the course of human civilization. And today, the future of human civilization has never been more dependent on the future of research. The challenge of successfully transforming research into the Open Age is not one we should look upon lightly. The potential payoff is significant, but the transformation needs to be done with care, with the full involvement of researchers, and setting aside all our ideological preconceptions about what “open” should look like.

The recommendations put forward in this report are designed for UNESCO, but all agencies big and small are welcome to adopt and adapt them. I want to thank my co-authors and contributors for their help creating this report, and also the participants of the Open Scholarship Initiative (OSI), who have been debating the future of open research for the past six years. I also want to thank UNESCO for their long-time leadership on global open policy, and for being a partner who shares OSI’s belief that a robust, global, sustainable future for open research must be built together on common ground.

Finally, a central tenet of this report is that some elements of the open research community should be less judgemental in searching for open solutions. That criticism might be seen by some readers as judgemental in its own right. If you can forgive this, I hope you can hear the deeper message that the open community can do better. For the sake of all our futures, we need to do better.

Sincerely,



Glenn Hampson  
Seattle, Washington, USA  
January 2021



# OPEN SOLUTIONS

## UNIFYING THE MEANING OF OPEN AND DESIGNING A NEW GLOBAL OPEN SOLUTIONS POLICY FRAMEWORK

Society's pursuit of open knowledge solutions is disjointed. These solutions—including but not limited to open access, open data, open source, open science, open government, and open educational resources—have much in common, yet we treat them as separate entities from a policy perspective. Is it possible to unite these solutions under a single open solutions policy framework? If so, what might this framework look like, what should our open goals be, and how can we introduce this approach to the world?

### OBJECTIVES

In this report, the authors will explore combining open access, open data, open source, open science, open government and open educational resources (OER) into a single “open solutions” framework.<sup>1</sup> Exactly how to do this is an intriguing challenge since these solutions have all developed their own philosophies, challenges and constituencies over time.

Increasingly, however, open access, open data, and open science have been overlapping significantly with regard to their use of broad and overarching concepts like FAIR data inclusion and open licensing, making it both possible and desirable to consider constructing a single open solutions framework instead of continuing to see open knowledge questions and answers through the lens of multiple intersecting and overlapping jurisdictions. The main objective of this report is to explore the philosophical, pol-

1. Open scholarship isn't included in this analysis because it's already a blanket term that encompasses a variety of fields and open processes. It is, in essence, the same thing as “open solutions” except with an academic focus.

### ABOUT OSI

The Open Scholarship Initiative (OSI) is a diverse, inclusive, global network of high-level experts and stakeholder representatives working together in partnership with UNESCO to develop broadly accepted, comprehensive, sustainable solutions to the future of open scholarship that work for everyone everywhere. This document reflects the input of the individuals listed here as well as contributions from other OSI participants who are not listed. The findings and recommendations expressed do not necessarily reflect the opinions of these individuals, OSI participants, OSI participant institutions, or the agencies, trustees, officers, or staff of these institutions.

OSI is managed by the Science Communication Institute (SCI), a US-based 501c3 nonprofit charity. OSI serves as the Network for Open Access to Scientific Information and Research (NOASIR) for the United Nations Educational, Scientific and Cultural Organization (UNESCO). For more information about OSI, please visit [osiglobal.org](http://osiglobal.org).



icy and practical aspects of an open solutions framework and also go a step beyond this and prescribe a program for introducing this solution to the world. Specifically, this report will:

- 1. Provide background and conceptual analysis** of the open access, open science, and open data spaces (as well as other open solutions to a lesser extent) and describe areas of intersection and potential integration;
- 2. Identify steps that UNESCO should take** with regard to policy development, adoption and integration, as well as partnerships, bridging, capacity building and advocacy;<sup>2</sup> and
- 3. Identify indicators and dimensions** that need to be developed and/or monitored as part of this effort.

Methodologically, this report follows a rather narrow path. This is out of necessity: there are endless perspectives and tangents with an objective this broad, so we've focused on a single path that while also broad is still relatively manageable, at least conceptually.

- **Researcher-centric.** First, the evidence and recommendations presented in this report are mostly from a researcher-centric perspective. While evidence and recommendations from the perspectives of the general public, government funders, publishers, libraries, and other stakeholder groups are not ignored here by any means (and indeed, still inform the entire basis for this report), they are not the central focus of our evidence and recommendations. Providing evidence and recommendations from a multiplicity of perspectives wouldn't change the underlying facts, but it would involve accumulating and digesting a lot more facts than are already provided in the following pages. The recommendations from such a report might also differ as well (ultimately more focused, for example, on how to lower publishing costs, improve accountability, and so on). We have chosen a research-centric perspective for this report because in our particular bias, researchers need to be at the center of any effort like this that proposes to reform the future of global knowledge. Researchers are the group that generates this knowledge, they are arguably the primary consumers of this knowledge (at least the kind of knowledge we're talking about here), and their ability to access and reuse this knowledge should be the key driver in this effort. This is not to marginalize concerns about the public's right to know, or the right of funders to get a good return on their investment, or the right of libraries to have affordable access to the research that comes from their institutions—which are all vitally important and valid concerns—but fundamentally, our goal here is to focus first and foremost on whether this effort makes sense from a researcher perspective, and if so, what we need to do to improve this effort going forward.

---

2. This report will not, however, explore the impact of open solutions on international development or the UN's Sustainable Development Goals, nor will it delve too deeply into the history and foundations of these solutions, or detailed examinations of particular dynamics or policies. These are all critically important discussions—the phrase “beyond the scope of this report” is repeated multiple times on the following pages. The goal of this report, rather, is to identify the big picture, general areas of overlap where we might begin engaging with these different open efforts simultaneously instead of separately, and/or where we can begin speaking in terms of “open solutions” instead of open access plus open data plus open science, and so on. Matthew Smith and Ruhiya Kristine Seward's 2020 collection of essays, “Making open development inclusive: lessons from IDRC research” (Smith 2020) takes a similar approach to this report, looking at the intersection of open themes through the lens of “open development” rather than “open solutions.” Smith's approach is to focus on the practice-based interconnectedness of different open concepts and how these overlapping practices can guide the development of an overarching “open development” policy. For more detail on the international development angle, see Matthew Smith and Katherine Reilly's 2013 collection of essays (Smith 2013), titled “Open development: networked innovations in international development”; and also the 2019 State of Open Data report edited by Tim Davies et al (Davies 2019).

- **Interwoven.** Second, it's important to note that two key ideas are interwoven in this report and are being considered as one, even though they are separate. The first idea is that we should consider the open solutions space to be a single policy environment instead of a collection of separate environments—open access, open data and so on. The second is that we should consider common ground action to develop the future of open policies (this idea is explored more fully in OSI's Common Ground paper; see Hampson 2020). This interwoven approach is what's being discussed in this report's evidence and recommendation sections, not each idea on its own but both ideas taken together. These two ideas, however, are in fact separable. It is possible for the reader to come away convinced that a common ground approach to the future of open is important but not necessarily an approach built around open solutions. Or, the reader may conclude that the open solutions approach is valuable, but not necessarily a policy framework that supports collective action on common ground. The reason we are weaving these two ideas together is that we feel they belong together and make sense together—that they are much stronger together than as separate constructs.

## EXECUTIVE SUMMARY

The need for knowledge sharing has never been clearer and more urgent than today. From climate change to food security, HIV to COVID-19, humanity needs global solutions to global challenges. The vast international community of stakeholders involved in the creation and sharing of knowledge has been responding to this need for many years—researchers, government funders and policy agencies, private foundations, commercial publishers, universities, libraries, and more—driven largely by their own interests and judgement but also by the judgement of respected global organizations like the United Nations who have tried to define a broad picture of what kind of global knowledge sharing needs exist and what kind of global action is needed. While the amount of research being conducted in the world has steadily increased over the years, roughly doubling every 20 years,<sup>3</sup> and while this increase has led the world to remarkable new heights, our knowledge sharing practices have proven difficult to change. In the judgement of many, these practices, while growing in many parts of the world, are still inadequate to meet the urgent needs of today.

Enter the open access, open data, open science, open source, open government, and open education movements. Taken together, these movements have made a significant contribution to the evolution of our knowledge sharing practices. Each of these movements is, however, entirely separate. Each has a rich and unique history, a massive diversity of outputs, goals, tools, measures, methods, actors and stakeholders, and vibrant ecosystems of innovation. These movements also lack coordination toward common goals, which has resulted today in a lack of leadership on broad and globally workable open solutions, a lack of support for open infrastructure and other open needs, and slow acceptance and adoption of open policies.

In partnership with the Open Scholarship Initiative (OSI), UNESCO has been working since 2015 to create a common ground approach to the future of open knowledge that embraces the diversity of the open landscape and creates a sustainable and coherent approach to this future that works for everyone everywhere.<sup>4</sup> An important part of this challenge has been understanding and knitting together

3. To say nothing of the rapid and continuing increases in the number of research articles being published, the number of research journals, and the amount of data being generated.

4. OSI and UNESCO are independent from one another in this pursuit but have been actively collaborating in this challenge. UNESCO has provided approximately 25 percent of the funding for OSI operations since 2015 (the other 75 percent has come from private foundations, publishers, and OSI participants and participant institutions). In addition, OSI serves UNESCO in

the vast diversity of perspectives of all open knowledge stakeholder groups from all parts of the world, and also seeing the entire “open scholarship” landscape as being more than just about journal publishing and science, but involving a wide range of closely related interests and activities.

Where exactly do all these different perspectives, interests and activities intersect, and can a policy framework be built around this intersection that successfully merges our “meta” policies around open so we can begin to collaborate on our open goals more effectively, and in doing so enhance and streamline all open outcomes for global knowledge and society? This report explores key points of intersections under the rubric of “open solutions”—a high-level approach to open founded on a broad embrace of open instead of a multitude of disconnected approaches. The argument presented in this report is that adopting a broad open solutions focus will help policymakers develop a more effective open knowledge environment at all levels, from government to education to research, and that developing such an environment will create enormous benefits for science and society.

The final sections of this report describe an action plan that UNESCO should consider for getting from where we are now in the global open policy debate to where we need to be. A draft marketing approach for this plan is included in the annex section of this report.

### Key findings (background & conceptual analysis)

The key findings from this report, supported by many scholars in this field,<sup>5</sup> revolve around the realization that openness doesn’t have a single or set definition. Rather, open outcomes exist along a broad spectrum, and correspondingly, the definitions we apply to open and the open solutions we invent also vary quite widely. There is common ground, however, and working together on this common ground can help us come up with open solutions, approaches and policies that are much less disjointed than now and are ultimately more beneficial for the broader cause of open knowledge, which doesn’t recognize the discrete boundaries that separate data from text, government information from science information, and policymaker from private citizen.

- **Open solutions have been evolving for decades now, if not centuries.** Open is not a new phenomenon. Marking exact starting points in this evolution can lead to spurious and self-serving conclusions. Many of these efforts were initially fueled by the idealism that open information was an unalloyed public good; but over time, many have also developed a deeper sense that the reality of open is more nuanced, and that there are no one-size-fits-all solutions or interpretations.
- **Recognizing the similarities between different open solutions can help us understand these efforts aren’t necessarily separate or unique,** but instead are part of a broad, long-term, multifaceted push from many corners of many societies to make information of all kinds more open. Understanding how similar open efforts have been with regard to common challenges and approaches can help us merge open policies together in a way that wouldn’t be possible by focusing only on the specific practices of each open movement. And in-

---

an official advisory capacity as the agency’s Network for Open Access to Scientific Information and Research (NOASIR). OSI includes 450 high-level participants, alumni and observers who represent 20 stakeholder groups in scholarly communication and 250 institutions from 30 countries. Several of the authors of this paper are participants in OSI, and UNESCO’s CI sector has funded the writing of this report. See the conflict of interest statement for more details, and also [osiglobal.org](https://osiglobal.org).

5. Numerous sources are highlighted throughout this report. For instance, some scholars and organizations have recognized that: “open science” can’t be neatly defined in a way that satisfies everyone (see the STM Association’s position at <https://bit.ly/38aM1la>, or Adams 2014); open data has many practice-based definitions that defy applying an all-encompassing legal definition (see the introduction to Davies 2019, or the wide array of RDA policies); “open access” has a wide variety of outcomes (see Piwowar, Archambault, and other research); and/or that different branches of open share a lot of common ground (Smith 2020). This isn’t a contentious observation, although it is at odds with the desire of some advocates to neatly package “open” as a prerequisite to developing open policies.

tegrating these conversations can help us integrate efforts and policies in ways that might dramatically and effectively accelerate the development and uptake of open policies. We might also be able to develop common language that can help facilitate the better communication of needs, processes and outcomes. Open knowledge is a broad endeavor with a shared goals. Recognizing this will be key to transforming open solutions into more than just the sum of their separate parts.

- **Open solutions today have a wide variety of definitions and interpretations, and a wide variety of outcomes**, all expressing varying degrees of openness and experiencing varying degrees of acceptance by open advocates as being open or not open enough. Different stakeholder groups have starkly different points of view. However, all “sides” are wrong, and no one side has all the information it needs to make an informed judgement. We know open involves a spectrum of motives, outcomes and solutions. But we can no sooner pick the “right” answers from this diversity than pick the right colors from the rainbow. Each is important, and each contributes to the greater whole. By focusing instead on the whole, we can find common ground for bold and needed action, and also create mechanisms by which our debates about best practices and requirements will settle themselves over time through more engagement from various knowledge-providing and knowledge-seeking communities.
- **Over the short-term, ideologically-driven (as opposed to evidence-driven) actions will continue to create policy whiplash**, where publishers and funders don’t know what to do, researcher opinions are further marginalized, and the actions that sound brilliant today are roundly derided tomorrow. In addition, if our open reform efforts are created by the global north without adequate input and participation from all groups everywhere, then the open solutions we create and deploy are going to continue favoring those with more privilege. The richer countries will be able to fill open repositories with information, which makes their information and perspective even more dominant than now. In terms of ideologically-driven licensing rationales, the world is ready to stop pushing specific technical and licensing regimes as our core objective for open, and instead let the debate move on to other narratives like good data, common open solutions, and common goals.

### Key recommendation (steps UNESCO should take)

The key recommendation from this report is that UNESCO should host and promote an open solutions approach that focuses on achieving our common goals for open—from discrete common goals like improving science to grand challenges like tackling climate change. This goal-centric approach is built on a widely-used and well-developed model (Theory of Change),<sup>6</sup> and on the common ground that unites all open solutions and all communities in the information communication space, and that should underpin all future open solutions policies. Supporting communities of practice will still be essential—we will still need open access, open data, and open source solutions for particular fields, for example, and we still need innovation to occur at this level. But more broadly, these communities of practice should also be connected at a high level through common goals so they can innovate together to build a future of openness that advances knowledge and discovery as opposed to simply collecting open artifacts. The key to this high level framework for collaboration is inclusiveness and flexibility. This framework needs to be built by the entire community, not imposed from the outside, and especially not imposed in a way that is led by ideology instead of evidence.

---

6. We describe our approach in this report as being a “goals-based” model. Intellectually, this approach is best described by the Theory of Change model used by governments and organizations to first define long-term goals, and then map these goals backward to identify necessary preconditions. See [https://en.wikipedia.org/wiki/Theory\\_of\\_change](https://en.wikipedia.org/wiki/Theory_of_change) for more information.

- **The focus of a new, global open solutions policy should be to figure out what we want to achieve with open** and then build solutions capable of reaching these goals. Opening information is not the end goal, but only one small step in a very involved and complex process of preparing information for access and reuse. Rallying around our common goals will be an effective way to begin mobilizing the expertise we need to actually do something with open information as opposed to simply collecting it. Rather than deciding what form open should take and legislating policies to create this outcome, we need to construct a framework for dialogue and collaborative action on our common goals. Building on this framework will lead to real and robust global solutions that are inclusive, diverse, and focused on practice-based areas of common interest and common benefit.

To begin, we need a clearer and more detailed understanding of exactly what researchers want and need, what they will use, and what we hope to accomplish with open information so we can ask good questions, collect meaningful information and pursue effective solutions. It's vital to understand this wide diversity of perspectives, and also understand that one-size-fits-all solutions aren't what we should endeavor to develop. Understanding the diversity of needs and ideas in this space can also help us identify where we can construct broad support mechanisms for everyone. This approach, along with trying to better understand the open solutions space, will ultimately provide much more effective support for the future of open than creating overly-detailed open policies that no one will follow, and that will be largely inapplicable to the majority of real-world open efforts.

The more effective framework for these solutions will likely be broad, flexible, and turnkey, allowing researchers to easily comply.<sup>7</sup> Solutions will also need to be clearly beneficial to research and researchers; the incentives switch needs to be flipped or open will only grow as far and as fast as we push it. Constructing a self-incentivized solution space will allow solutions to evolve that people want and need. And critically, these are the solutions that people will end up using. Use may be key because the world already has plenty of ambitious knowledge sharing platforms that simply don't contain enough data to be useful, and/or don't get enough use or upkeep to merit further investment.<sup>8</sup>

- **Our greatest mistake as policy analysts and advisors would be to think we are smarter than this community's marketplace of ideas and outcomes**, and recommend that UNESCO attempt to impose a rigid ideological order on this diverse and deeply complex landscape. Doing so would be at best ineffectual, and at worst might fracture the global solution space instead of unite it. A global open policy needs to be more than just words. Without creating a framework for open solutions along with a framework for policy engagement through which diverse voices can be heard and community action can be pursued together, our open actions and policies will not have the truly transformative impact we need them to have, and could even be harmful. Transitioning to this environment of collective understanding and action will take time and will involve robust and dedicated efforts.

---

7. Although it's possible that the "right" solutions may also be narrow, rigid and difficult to implement; as will be noted later in this report, several of the most effective and ambitious data sharing communities do in fact utilize narrow solutions. However, a broad, flexible and turnkey model allows communities to adapt to the reality of their information-sharing needs (and build narrow solutions inside a broad framework of tools and support) rather than being forced to implement the wrong one-size-fits-all sharing solution. The point is that these solutions will need to evolve from community conversations about goals and how best to achieve these goals.

8. Institutional repositories (IRs) are the common cautionary tale here. Historically, many have eventually fallen into a state of disuse and disrepair because they are simply too difficult and expensive to maintain and don't get used enough to warrant additional investment. See Dubinsky 2014 and other surveys. However, the IR world may be improving as open policies mature and become more widely adopted—surveys from five years ago may not accurately reflect the current state of the market.

Creating UNESCO's open solutions policy framework is only part of the solution. The first step will be for a critical mass of supporters in this community to stop thinking of the open solutions universe as something that belongs to any single institution, stakeholder or interest group.

### Key assessments (needed indicators and dimensions)

Indicators and dimensions are the keys to policy evaluation and a necessary part of the policy development process. However, they are also very easy to get wrong. Wrong indicators and dimensions can lead to wrong policies and outcomes. In an undertaking like ours, where definitions are fluid and different stakeholders have different goals, it's almost inevitable that inventing indicators and dimensions to measure the current open solutions landscape will be both suspect and inaccurate. It will be better if the open indicators the global community wants and needs evolve from the goals we set together. These indicators may end up being very narrow, focusing on outcomes such as: how effectively we're addressing a particular challenge in science or government; or how quickly books in the public domain are being digitized; or how widely popular software (like Wordpress) is being used. Or they may be discipline-specific, or measure progress by institution, region, or even globally. We simply won't know until we first discuss our common goals and then develop the assessments we want and need to measure our progress toward these goals.

Outside this tight and relevant bubble, though, there are a number of dimensions indirectly related to openness that also need to be assessed. These aren't the dimensions we were looking for, but they are important nonetheless, including: forecasting what impact the fracturing of the open solutions space will have; getting a better grasp on the scope and impact of current inequities with regard to knowledge access; mapping out the stakeholder space to understand explicitly where more and better connections can be made; understanding the impact of SciHub and predatory publishing on open usage; and other detailed concerns.

# SECTION 1: EVIDENCE

## BACKGROUND

**B**efore we can weigh the pros and cons of a unified approach to open solutions and common ground open solutions policies, we need to review our key evidence. What is UNESCO's approach to open? How does this compare with the global approach? What do we know about open access, open data, open science, and other open movements and solutions that can inform our understanding of how these might merge together, and about which policy approaches and goals might work and not work?

### UNESCO's approach

Like the rest of the world, UNESCO's approach toward open has evolved over time. Different sectors and divisions have taken up different aspects of open at different times. The open access challenge—which is mainly focused on improving access to published academic articles—is the oldest of these efforts at UNESCO, beginning in the Knowledge Societies Division's Communication and Information (CI) Sector around 2001. The main focus of this effort has been to engage with global open access communities and help develop dialogue and solutions.

UNESCO's open science program is much newer, taking form in early 2019 under the direction of the agency's Division of Science Policy and Capacity Building (Natural Sciences Sector). The goal of this program is to synthesize the globally diverse approaches to open science into one standard setting approach that can help shape national laws and policies. At the 40th session of UNESCO's General Conference on October 8, 2019, 193 member states tasked UNESCO (but specifically this division of UNESCO) with “the development of an international standard-setting instrument on Open Science in the form of a UNESCO Recommendation on Open Science to be adopted by Member States in 2021.”<sup>9</sup>

This recommendation “is expected to define shared values and principles for Open Science, and identify concrete measures on Open Access and Open Data, with proposals to bring citizens closer to science and commitments to facilitate the production and dissemination of scientific knowledge around the world. The Recommendation will be developed through a regionally balanced, multistakeholder, inclusive and transparent consultation process.” In this approach to open, open science is seen as being at the center, wherein open science allows “scientific information, data and outputs to be more widely accessible (Open Access) and more reliably harnessed (Open Data) with the active engagement of all the stakeholders (Open to Society).”<sup>10</sup>

UNESCO's open solutions idea, by contrast, is an umbrella description of the organization's CI sector work that promotes various approaches to open like open education, open software, and open access. Open solutions in this depiction “targets leaders, professionals, researchers and ICT (information and communication technology) users, supporting communities of practice, encouraging empirical research and publications, and organizing key events at global, regional, and national levels to share best practices, with comprehensive programmes in: Open Educational Resources—providing teachers and

---

9. <https://unesdoc.unesco.org/ark:/48223/pf0000370291>

10. <https://en.unesco.org/science-sustainable-future/open-science> and [https://en.unesco.org/sites/default/files/open\\_science\\_brochure\\_en.pdf](https://en.unesco.org/sites/default/files/open_science_brochure_en.pdf)

learners with high-quality teaching and learning materials that allow for free use, adaptation, and distribution; Open Access to scientific information—enabling scientists and researchers to share and access scholarly information on the latest scientific advances; [and] Free and Open Source Software—providing a wide range of tools and processes for creating, exchanging, and sharing interoperable software and solutions efficiently and effectively.” This program is rooted in UNESCO’s belief “that universal access to information and knowledge is key to the building of peace, sustainable social and economic development, and intercultural dialogue,” and that advances in open movements and solutions have made this vision achievable.<sup>11</sup>

**The global movement toward open scholarship has been evolving in fits and starts for decades now, if not centuries. Marking exact starting points in this evolution can unfortunately lead toward rather spurious and self-serving conclusions...**

There are also other distinct focus points in UNESCO covering open educational resources, open coding, open government, and more. These various approaches are increasingly beginning to intersect at a policy level, although much integration work remains. At a higher level, however, the overlap is clear: At least 10 out of the 17 Sustainable Development Goals (SDGs) comprising the UN’s 2030 Agenda for Sustainable Development require constant scientific input, and open solutions are a key part of ensuring the sustainability of this input.<sup>12</sup>

As a foundation for this assessment, this report will ascertain whether UNESCO’s new open solutions approach makes sense—whether an approach like this, and particularly one driven by shared goals (as will be discussed later in this report), works conceptually, what form this kind of an approach might take, and the pros, cons and challenges of such an approach writ large. This report will not delve into details like the specific construction, alignment, funding, and capacity of each of UNESCO’s open programs, what is needed for these programs to integrate and/or develop, or what role a strengthened UNESCO open solutions program would play. This is critically important information, however, and it needs to be gathered in order for UNESCO to calculate how to enact the recommendations put forward in this report. This information is, however, beyond the ability of this expert author group to discern.

## The global approach

The global movement toward open scholarship has been evolving in fits and starts for decades now, if not centuries. Marking exact starting points in this evolution can unfortunately lead toward rather spurious and self-serving conclusions<sup>13</sup>—certain individuals and events have been more important than others. Fundamentally, the very foundation of research is built on the need to share information with other researchers and the public, and has been moderated by the limitations and tensions associated with these sharing practices. As our society lurches today toward greater expectations of openness

---

11. <https://en.unesco.org/themes/open-solutions>

12. <http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/open-access-to-scientific-information/>

13. For instance, open access advocates sometimes point to the 2002 Budapest Open Access Initiative (BOAI) as a fundamental turning point in the evolution of open access; others have noted that open access practices were evolving long before 2002 and that BOAI was just an important part of the evolution of open. It’s important to understand the full history of how open has evolved over time rather than see any single moment in this history as being seminal or beyond dispute (particularly for an event like BOAI, which involved only 17 people). This matters not just for the sake of accuracy, but also because “belief” in BOAI is often what separates open access true believers from “heretics”—an unnecessary but nonetheless real dynamic in the open access reform community. There is no doubt that BOAI’s attempt to define open (however imperfect that definition is today) brought a lot of attention to the topic and has provided a useful foundation for conversation. It is not, however, the final word in open, or an infallible definition. More will be discussed about this dynamic later in this paper. See also Poynder 2020 for a deeper discussion of open access ideology.

and our technology demonstrates what can be done with open information, it is no wonder that we also expect more openness from research and huge benefits from this openness.<sup>14</sup> In response to these pressures, different fields, regions and institutions have been taking different approaches to openness, and globally, stakeholder groups have been buffeted to varying degrees by changes in open requirements and practices—funding agencies, researchers, publishers, universities, libraries and more.

These paths are varied not just because of varying needs, but because openness itself is at once a philosophy, a practice and a goal. Different elements of these open tendrils motivate, inform, connect and differentiate various open movements (as well as actions, tools and policies) in a variety of ways. In general, the emphasis of open movements is to create openness by default, which creates tensions with cultural norms of information hoarding and sharing—subscription paywalls, copyright, secrecy, competition, national security, and so on.

All of these various movements have matured and diversified since their inceptions. All were initially fueled by the idealism that open information was an unalloyed public good; and over time, all have developed a deeper sense that there are no one-size-fits-all solutions or interpretations, that different fields and regions have different information needs, that there are pros and cons to various approaches to open, and that there are a constellation of perspectives and solutions driven less by idealism than by specialization.<sup>15</sup> The open challenge of today is to figure out how to continue to evolve our sharing and access norms and overcome some of these tensions by better understanding the root concerns involved and looking for ways to adapt.

**All of these movements were initially fueled by the idealism that open information was an unalloyed public good; and over time, all have developed a deeper sense that there are no one-size-fits-all solutions or interpretations...**

This is part of the reason why UNESCO is focusing on the big picture.<sup>16</sup> An approach that understands and embraces the diversity of opinions and definitions across research fields, regions, and different open solutions and efforts is critical for understanding open and arriving at conclusions and policies that aren't ideologically or philosophically wedded to any one open approach, policy or perspective.

Another reason for UNESCO's broad focus is that we are at something of a tipping point in open, as participants in the November 2019 Berlin Debate on Science and Science Policy noted (see Bosch 2019). While the challenges of open are immense, and the answers are wide and varied, there does seem to be broad agreement that we are at or near a unique period in history when we might be able to create a new and productive future for open by working together on our common ground.

---

14. Whether this is owing to “information capitalism,” the falling costs of transmitting information, or other factors is beyond the scope of this report. The point is that it's happening. When desktop publishing emerged in the early to mid-1990s, there was an explosion in the number of books being published and the number of publishers operating in the marketplace. There was nothing theoretical or sinister happening here—just new opportunities being explored, causing lots of market disruption as a result (and along with this disruption, the development of new opportunities). It's possible to overthink these dynamics and start inventing motives and agendas where the reality is actually much more vanilla—opportunity. Advancing technology creates new opportunities, which leads to changes in the market, which causes disruption and opportunities, which leads to new technical solutions and opportunities, and so on. This “opportunity churning” has been happening in the communication sector at a rapid clip over at least the last 30 years and shows no signs of slowing down. The forces at play here are not solely technical, commercial, legal, or societal, but a mix of all of the above.

15. For more detail, the 2019 State of Open Data report (Davies 2019) gives an up-to-date, comprehensive and illuminating view of the open data landscape.

16. And as such, UNESCO has supported and drawn on the advice of OSI since 2015 so that together, both groups can try to better understand the open space and develop global, inclusive approaches to the future of open. Jon Tennant has noted (Tennant 2019, p. 17), that OSI is probably the best arbiter of this approach due to its diversity and partnership with UNESCO.

This is also a moment when policy observers like OSI and others are beginning to note that our theoretical definitions of open are wide-ranging and do not match the reality of openness in practice; that this lack of common ground inhibits the transferability of open lessons learned; and that openness, while holding out so much potential for doing good, also has the very real potential for making rich countries richer because they are the ones who are predominately developing and populating open solutions (Smith 2020). This will continue to be the case if we ignore real-world, bottom-up open solutions in favor of open theories that are grounded in global north perspectives, and doing so slowly with open policies that are ill-fitting, lag behind the most effective cutting edge open practices, and are too often short-lived or poorly resourced (Smith 2020). The time has come to rethink our approach to how open solutions can be more unified conceptually, as well as more unified globally.

So, what exactly does this common ground look like? For the purposes of this report, we are primarily exploring open practices as they relate to research. However, all open fields and practices share some common ground so all will be mapped here to some degree (within limits, since a very deep discussion is well beyond the scope of this paper) in order to better facilitate our later discussion of what a truly broad open solutions framework and global policy might look like.

## Open access

Open access (OA) is probably the most widely misused term in the open universe. Generally speaking, OA focuses on improving the openness of research publications (for a more complete explanation, see Plutchak 2018). Where OA ventures into the realm of data or books, there is overlapping “jurisdiction” with regard to open data or OER policies. That is, for example, OA policies often call for the inclusion of datasets, and may even stipulate the exact license and repository location of these datasets, but at a granular level exactly how these datasets are constructed ventures out of the expertise of OA and into the expertise of open data.

A great many organizations and programs invoke OA principles and programs, buttressed by mountains of papers and position statements. The most popular depiction is that OA is a clearly defined concept that emerged from the 2002 Budapest Open Access Initiative (BOAI) meeting. In fact, various forms and interpretations of open access began evolving long before 2002 and have branched off in many directions over the years. Physics and astronomy researchers (and later, researchers from many different fields) began depositing their papers in the arXiv preprint repository starting back in 1991, for instance, long before BOAI concepts were codified regarding what was and was not open. Today, most arXiv papers are published with licenses that do not align with the BOAI definition of open; the same is true with bioRxiv and other preprint servers that are much newer than arXiv.<sup>17</sup> Even in PubMedCentral, which is far and away the world’s largest repository of free-to-read research papers, only about a third of the paper have a BOAI-compliant license.<sup>18</sup> The BOAI statement was an interesting and worthy take on open, but it is neither definitive nor universal. In its defense, the meeting happened in the Stone Age

**...open access today has a wide variety of definitions and interpretations, and also a wide variety of outcomes...all expressing varying degrees of “openness” and experiencing varying degrees of “acceptance” by open advocates as being open or not open enough.**

17. See, for instance, <https://blog.dhimmel.com/biorxiv-licenses> as well as other analyses.

18. From the PMC website (<https://www.ncbi.nlm.nih.gov/pmc/about/intro/>), typing “cc by license[filter]” into the search bar returns 1.9 million entries out of 6.6 million total. PMC operates under the “public access” definition of openness which only stipulates that materials be free to access, not necessarily free to access and reuse. Public access usually stipulates that in addition to access, you can do one of the “5 R’s” of open—retain (as in, keep a copy). What it doesn’t allow is the other four “Rs”: reuse, revise, remix, and redistribute.

of the Internet, long before we understood how the Information Age via the Internet would truly unfold, and it was not a large, global, or authoritative gathering (not that this would have boosted its predictive powers).

What we have seen in reality is that open access today has a wide variety of definitions and interpretations, and also a wide variety of outcomes—green, gold, diamond, bronze, black, public access, hybrid, and more (only some of which are BOAI-compliant)—all expressing varying degrees of “openness” and experiencing varying degrees of “acceptance” by open advocates as being open or not open enough. In practice, these variations have created confusion in the scholarly communication community with regard to what open is and how it should be measured. Even open access researchers don’t

agree—some will refer to all kinds of open materials as being open access, while others will claim that only BOAI-compliant information is OA. Therefore, when we talk about open growing fast or not fast enough, or certain policies being open or not open enough, we’re really talking past each other because we’re not using the word “open” in a precise manner—it means different things to different people.

**TABLE 1: ESTIMATES OF THE AMOUNT OF OPEN ACROSS THREE SOURCES**

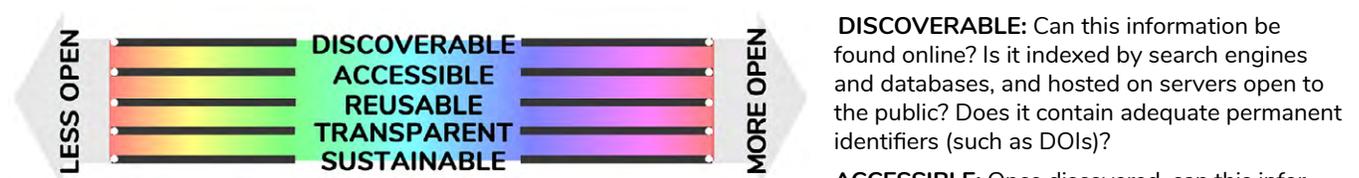
Type of open	Journal articles with Crossref DOIs, all years	Citable WoS articles with DOIs, 2009–2015	Articles accessed by Unpaywall users over 1-week period in 2017
Open (all types)	27.9%	36.1%	47.0%
Bronze	16.2%	12.9%	15.3%
Hybrid	3.6%	4.3%	8.3%
Gold	3.2%	7.4%	14.3%
Green	4.8%	11.5%	9.1%
Closed	72.0%	63.9%	53.0%

Source: Adapted from Piwowar et al. 2018

In 2016 and 2017, OSI conference participants

developed a model to help visualize how all these different open outcomes actually reside on a spectrum of outcomes (see OSI 2016 and OSI 2017b). This spectrum of outcomes—called the DARTS open spectrum—is characterized by five dimensions of openness: discoverability, accessibility, reusability, transparency, and sustainability.

**FIGURE 1: THE DARTS OPEN SPECTRUM**



**DISCOVERABLE:** Can this information be found online? Is it indexed by search engines and databases, and hosted on servers open to the public? Does it contain adequate permanent identifiers (such as DOIs)?

**ACCESSIBLE:** Once discovered, can this information be read by anyone free of charge? Is it available in a timely, complete, and easy-to-access manner (for instance, is it downloadable or machine-readable, with a dataset included)?

**REUSABLE:** Can this information be modified? Disseminated? What conditions (both legal and technical) prevent it from being repurposed or shared at will?

**TRANSPARENT:** What do we know about the provenance of this information? Is it peer reviewed? Do we know the funding source (are conflicts of interest identified)? What do we know about the study design and analysis?

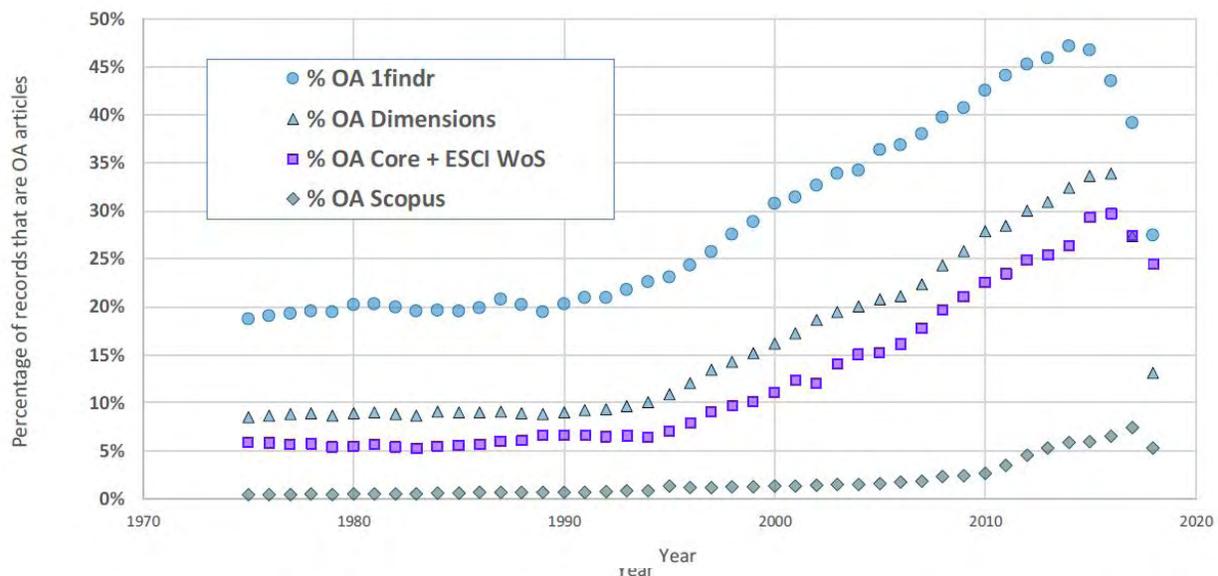
**SUSTAINABLE:** Is the open solution for this information artifact sustainable? This may be hard to know — the sustainability of larger, more established solutions may evoke more confidence than new, small, or one-off solutions.

Source: OSI 2016

The DARTS spectrum encourages more openness in scholarly and scientific communication, while also recognizing that open exists in various stages and that in some cases, optimally open may not mean maximally open. For clarity’s sake—and to avoid the appearance of usurping the OA term—we can refer to all outcomes on this spectrum as being “open” to some degree, and reserve the name “open access” only for outcomes that are at or near the right end of this spectrum.

In the figure below, researcher Eric Archambault (see Archambault 2018) demonstrates how open of all kinds (not just gold and green, but also hybrid, bronze, and read-only public access) has been growing steadily over time, with close to 50% of new materials now being published in some type of open format, and growing at around 4% per year due to “backfilling” (as materials come off of embargo).

**FIGURE 2: OPEN GROWTH RATES, 1970-2018**



Source: Archambault 2018

## Open data

Open data as we currently use the term is a newer construct than open access and open science—only about ten years young (Huston 2019, Davies 2019). The philosophical foundations of open data, however, are at least as old and diverse, with different audiences focused on different motives and outcomes, from discovery to interoperability to the moral/ethical imperative to create a “world brain” of science knowledge (Wells 1937). Like OA, open data also involves a great many stakeholders, institutions, and policies, and many different definitions and outcomes, these dispersed not so much along a spectrum of outcomes as along a variety of different perspectives and needs, such as research data, government data, health data, big data, and business data.<sup>19</sup>

19. The Open Data Handbook (<https://opendatahandbook.org>) offers this definition for open data: “Open data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike.” This implies that the data are available “in a convenient and modifiable form” such that there are no unnecessary technological obstacles to exercising licensed rights. However, many other organizations define open data differently. Also, as is discussed later in this report, a great wealth of data is open to only select users and is neither convenient nor modifiable. For all practical purposes, this is open data for the communities that have access to this data and the requisite expertise to use and reuse it. Therefore, we need to decide what “counts” as “open”—not necessarily definitionally, since certain manifestations of data (charts, graphs, tables) that present data in a new or novel way may in fact be copyrighable—but in terms of whether the

The Panton Principles for Open Data in Science, published in 2009, was perhaps the first attempt to clearly define what open data should and shouldn't look like in science. Among other points, these principles emphasized that when publishing data, authors needed to use very liberal licenses (like CC0), avoid creating restrictions like “non-commercial,” and “make an explicit and robust statement” about their wishes regarding how their data was to be used (see Murray-Rust et al. 2009 and Molloy 2011).

Seven years later, in 2016, with a focus on data accessibility, stewardship, and reuse, the FAIR Guiding Principles for scientific data management and stewardship were developed—at once a policy instrument and a philosophical goal. FAIR has proven to be popular and adaptable, stipulating in broad terms that data should be findable, accessible, interoperable and reusable. While the DARTS open spectrum is descriptive in demonstrating how open access has many different outcomes, FAIR is prescriptive in describing the ideal state for open data (achieving this state is another matter, though, as discussed later in this report).

TABLE 2: THE FAIR GUIDING PRINCIPLES

<p><b>To Be Findable:</b></p> <ul style="list-style-type: none"><li>F1. (meta)data are assigned a globally unique and persistent identifier</li><li>F2. data are described with rich metadata (defined by R1 below)</li><li>F3. metadata clearly and explicitly include the identifier of the data it describes</li><li>F4. (meta)data are registered or indexed in a searchable resource</li></ul>
<p><b>To Be Accessible:</b></p> <ul style="list-style-type: none"><li>A1. (meta)data are retrievable by their identifier using a standardized communications protocol<ul style="list-style-type: none"><li>A1.1 the protocol is open, free, and universally implementable</li><li>A1.2. the protocol allows for an authentication and authorization procedure, where necessary</li></ul></li><li>A2. metadata are accessible, even when the data are no longer available</li></ul>
<p><b>To Be Interoperable:</b></p> <ul style="list-style-type: none"><li>I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.</li><li>I2. (meta)data use vocabularies that follow FAIR principle</li><li>I3. (meta)data include qualified references to other (meta)data</li></ul>
<p><b>To Be Reusable:</b></p> <ul style="list-style-type: none"><li>R1. meta(data) are richly described with a plurality of accurate and relevant attributes<ul style="list-style-type: none"><li>R1.1. (meta)data are released with a clear and accessible data usage license</li><li>R1. 2. (meta)data are associated with detailed provenance</li><li>R1. 3. (meta)data meet domain-relevant community standards</li></ul></li></ul>

Source: Wilkinson et al., 2016

It is important to note that FAIR data and open data overlap but they are separate and distinct concepts. FAIR data aren't necessarily open, and open data aren't necessarily FAIR. As noted on the FAIR website,<sup>20</sup> “The ‘A’ in FAIR stands for ‘Accessible under well defined conditions’. There may be legitimate reasons to shield data and services generated with public funding from public access. These include personal privacy, national security, and competitiveness. The FAIR principles, although inspired

---

open community is willing to recognize and embrace these “limited open” communities of practice. Finally, it's important to note that not all data is numeric. There is also observational data, video and audio data, and other raw data from research. Some of this kind of data is potentially copyrightable because there is a significant creative element associated with gathering it.

20. <https://www.go-fair.org/resources/faq/ask-question-difference-fair-data-open-data/>

by Open Science, explicitly and deliberately do not address moral and ethical issues pertaining to the openness of data.” Conversely, a data file can be made accessible, appended to the end of a research paper, but also be mostly unusable because it lacks a clear license or sufficient explanation. Files like these may appear to “tick lots of FAIR and open boxes (e.g. persistent identifiers, basic metadata, non-proprietary file formats, etc.) but limited documentation renders the data unusable without more information on provenance, explanation of the variables, and methodology” (Higman 2019). So, like DARTS, FAIRness exists on a spectrum of outcomes. Finally, it’s important to note that FAIR principles embody best practices in research data management (RDM) that have been part of the landscape for years (Higman 2019)—the wrapper is new but the broad, overarching sentiments are not.

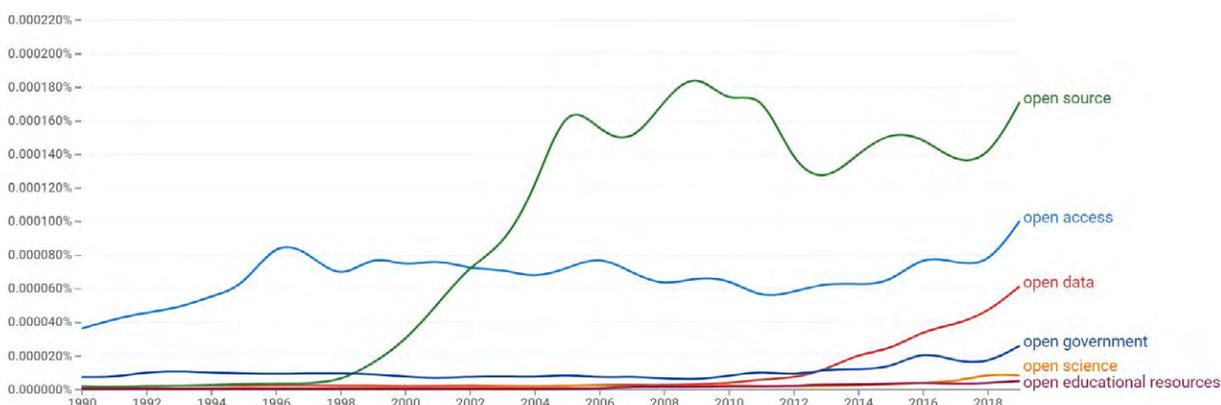
Today, moving beyond FAIR, the promise of open data has increasingly captivated publishers and funders. In its 2018 Open Science by Design report, the US National Academies of Science states that “the openness of data is seen as being critical to the progress of science, stimulating innovation, enhancing reproducibility, and enabling new research questions.” (NAS 2018)

## Open source

Open source may be the most widely used open solution term of all, or at least the open solution term that appears most often in published books.<sup>21</sup> This isn’t surprising given the huge popularity and importance of coding, and the avid open source developer community. The fundamental purpose of open source is the free reuse and resharing of computer code and software, provided attribution is given where appropriate.<sup>22</sup> The Open Source Initiative, which was formed in 1998 as an educational, advocacy, and stewardship organization, outlines the philosophy of open source code, as well as the different approved open source license types.<sup>23</sup> Essentially, as noted in the Open Source Initiative website, “Open source software is made by many people and distributed under an OSD [Open Source Definition]-com-

---

21. A Google Ngram comparison of how often various open-related terms have appeared in books over the last 30 years (see also <https://bit.ly/3fvNPr1>). This is more of a “fun” measure than a scientific one. Some of the problems with measuring word usage this way are noted in a 2015 Wired article (<https://bit.ly/36bh0Nf>). For example, Ngram measures are case-sensitive and vary when the terms are capitalized or abbreviated (open access plus Open Access becomes roughly equal to open source plus Open Source, OER is more common than open educational resources—slightly more popular than open government—and so on). Also, Ngram doesn’t do a word frequency count for academic journals, just books.



22. Requiring attribution is not allowed for free software licenses, however. According to the document for the GPL license, “While we recognize that proper citation is an important part of academic publications, citation cannot be added as an additional requirement to the GPL. Requiring citation in research papers which made use of GPLed software goes beyond what would be an acceptable additional requirement under section 7(b) of GPLv3, and therefore would be considered an additional restriction under Section 7 of the GPL. And copyright law does not allow you to place such a requirement on the output of software, regardless of whether it is licensed under the terms of the GPL or some other license. See <https://www.gnu.org/licenses/gpl-faq.en.html#RequireCitation> for more information.

23. See <https://opensource.org/> for more information.

pliant license which grants all the rights to use, study, change, and share the software in modified and unmodified form. Software freedom is essential to enabling community development of open source software.” An even more liberal open source framework is “free software” (where “free” in this context refers to freedom of reuse, not free in a monetary sense).<sup>24</sup> As defined by the Free Software Foundation, “Free software developers guarantee everyone equal rights to their programs; any user can study the source code, modify it, and share the program. By contrast, most software carries fine print that denies users these basic rights, leaving them susceptible to the whims of its owners and vulnerable to surveillance.”<sup>25</sup> In order to be compliant, free software has to fulfill the four essential freedoms<sup>26</sup>:

1. **The freedom to run** the program as you wish, for any purpose (freedom 0);
2. **The freedom to study** how the program works, and change it so it does your computing as you wish (freedom 1). Access to the source code is a precondition for this;
3. **The freedom to redistribute** copies so you can help others (freedom 2); and
4. **The freedom to distribute copies of your modified versions** to others (freedom 3). By doing this, you can give the whole community a chance to benefit from your changes. Access to the source code is a precondition for this.

In many kinds of research, open source has nothing whatsoever to do with open data. Open data is data, and open source is application. In some kinds of research, however, this distinction blurs—particularly when dealing with large datasets that require programming to understand. Without this accompanying code, there would only be a large jumble of information—no organization, analysis, or understanding. So, both open data and open code are required to maximize the integrity and replicability of this kind of work.

## Open science

The virtues of and need for openness are woven into the very fabric of science. In the mid-1600s during the birth of experimentalism, best practices in knowledge creation, sharing and preservation were widely debated. These best practices were also widely abused, even by some of the most prominent scientists at the time. There was rampant plagiarism, piracy, and usurpation of ideas, and partly because of this, a true crisis of fact (indeed, “facts” weren’t even a concept yet)—not knowing for certain which experimental knowledge was real and which was fake, which government laws and statistics had been accurately transcribed and which had not, and who invented the knowledge being communicated (if indeed this information was factual). The genesis of “open science” at this point in history was to communicate facts within a closed group of virtuosi who could then vouch for this information and put their imprimatur (such as that of the Royal Society) on it. Open most definitely did not mean sharing

The very concept of “discovery” didn’t exist until the late 16<sup>th</sup> century. Up to this point, the formal pursuit of knowledge consisted of studying ways to justify Aristotle’s world view; the assumption was that all knowledge was already known, and the role of higher education was to “rediscover” what had been forgotten.

24. As an interesting aside, in computational chemistry there is a big divide between quantum chemists, who typically use and develop proprietary and commercial code; and molecular simulation people who tend to go much more down the open/free software route (there is still some commercial/proprietary). Arguably there should be a shift toward free software approach, since researchers are unlikely (in most cases) to earn big money from academic code, whereas by making it free researchers can push science forward. See Elofsson 2019 for additional reading.

25. <https://www.fsf.org/>

26. GNU Operating System, What is Free Software? <https://www.gnu.org/philosophy/free-sw.html>

knowledge with the general public quickly or in a way that might jeopardize the intellectual property rights of inventors.<sup>27</sup>

Concurrent with this evolution, science itself evolved as an intensely competitive field. The very concept of “discovery” didn’t exist until the late 16<sup>th</sup> century; “facts” didn’t emerge until 150 years later. Up to this point, the formal pursuit of knowledge consisted of studying ways to justify Aristotle’s world view; the assumption was that all knowledge was already known, and the role of higher education was to “rediscover” what had been forgotten. A rapid-fire series of events in the late 15<sup>th</sup> century and mid-16<sup>th</sup> century changed this attitude, beginning with the discovery of the new world by Christopher Columbus, Tycho Brahe’s cataloging of a supernova, and Galileo’s discovery of moons orbiting Jupiter. These events in particular, and others soon to follow, uprooted the deeply held societal notions that the universe was perfect and that all knowledge was static.

A downhill race for discovery soon ensued, marked by intense competition to be first (and followed soon thereafter by the practice of eponymy—naming discoveries after discoverers). There weren’t any established methods yet to record discovery—patents were limited and local; copyright was still an unimagined concept; and publishing was still very much in its infancy—so “scientists” (this name wasn’t used yet either) would often go to great lengths to ensure their discovery was recorded, such as sharing secretly encoded messages describing their discoveries with close friends and known fellow “natural philosophers” in Europe and North Africa. As publishing matured, the race began to be first to publish—Galileo hurriedly submitted his early observations of Jupiter, suspecting that others wouldn’t be far behind (since telescopes were becoming increasingly popular at the time). Newton and Leibniz fought a years-long legal battle over who invented calculus—both invented calculus independently but Newton didn’t publish his discovery; and the Royal Observatory’s director, John Flamsteed, fought the Royal Society director Robert Hooke for years for stealing his astronomical data to include in lucrative almanacs. Competition for discovery was at the very root of science, and arguably, without this competition there would have been no science, no facts, and no discovery.<sup>28</sup>

Perhaps the next great shift in the evolution of science occurred in the US after World War II, when there was a rapid increase in spending on research, followed by a rapid increase in government calls for accountability in spending. Where was all this research spending going, what benefit was it creating for society, and how could the public be protected from spending fraud and abuse? This focus eventually culminated in the systems we recognize today, including competitive funding, institutionalized peer review, and a proliferation of journals to satisfy the publishing and disclosure demands of a vast and growing body of global research (Baldwin 2018).

Overlapping this path to transparency has been the same push toward open that we’ve seen in every other part of society. In science, the combination of these forces have manifested in a wide variety of ways. Fecher & Friesike (2013) postulate that there are now five main schools of thought with regard to open science:

- 1. Democratic school:** Believing that there is an unequal distribution of access to knowledge, this area is concerned with making scholarly knowledge (including publications and data) available freely for all;<sup>29</sup>

---

27. For a detailed account of this history, see Adrian Johns’s “The Nature of the Book” (Johns 1998) and also David Wootton’s “The Invention of Science” (Wootton 2015).

28. For a detailed account of the historical origins of science and discovery, see “The Invention of Science” by David Wootton (Wootton 2015).

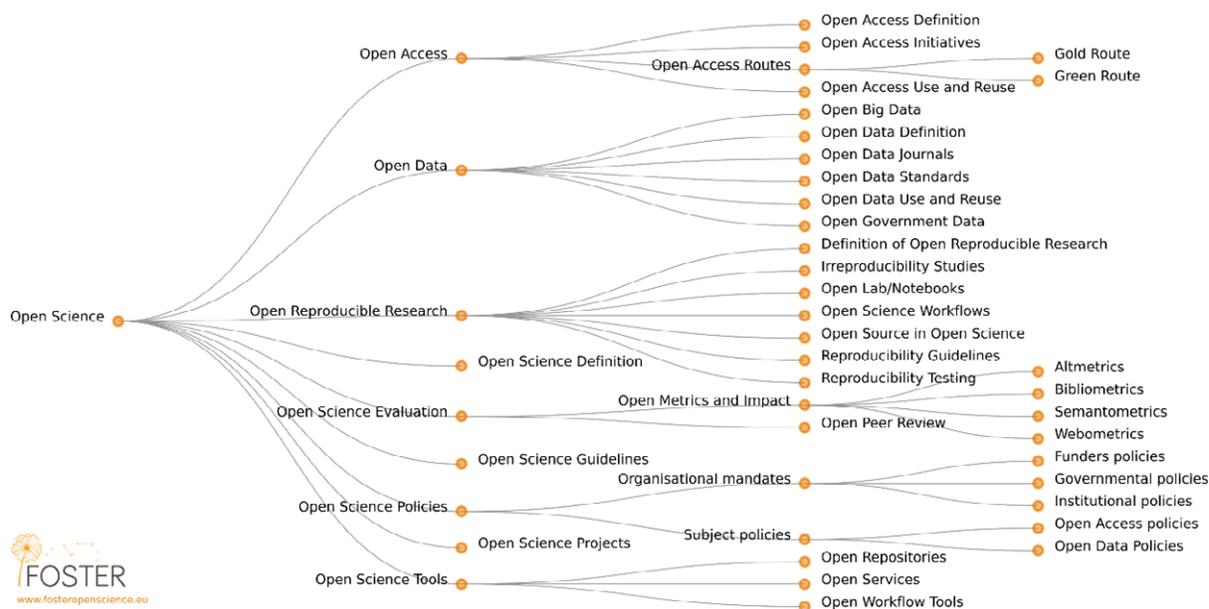
29. These schools of thought aren’t necessarily separate and distinct; combinations and overlaps certainly exist.

2. **Pragmatic school:** Following the principle that the creation of knowledge is made more efficient through collaboration and strengthened through critique, this area seeks to harness network effects by connecting scholars and making scholarly methods transparent;
3. **Infrastructure school:** This thread is motivated by the assumption that efficient research requires readily available platforms, tools and services for dissemination and collaboration;
4. **Public school:** Based on the recognition that true societal impact requires societal engagement in research and readily understandable communication of scientific results, this area seeks to bring the public to collaborate in research through citizen science, and make scholarship more readily understandable through lay summaries, blogging and other less formal communicative methods; and
5. **Measurement school:** Motivated by the acknowledgment that traditional metrics for measuring scientific impact have proven problematic (by being too heavily focused on publications, often only at the journal-level, for instance), this strand seeks “alternative metrics” which can make use of the new possibilities of digitally networked tools to track and measure the impact of scholarship through formerly invisible activities.

From this varied background, a number of different definitions of open science have arisen, all of them—as we think the body of evidence in this space clearly indicates—“wrong” in their own way. Joroen Bosman and Bianca Kramer created a useful framework (Bosman & Kramer 2017) for assessing the diversity of definitions. In their analysis, they conclude that there are at least five different categories of definitions for open science, including:

1. **Broad definitions** that use selective interpretations (like from the EU and US), to serve as broad open science policy wrappers;
2. **Maximal definitions** like the open science taxonomy, below, that accommodate all processes relevant to making science and knowledge more open;

FIGURE 3: OPEN SCIENCE TAXONOMY



Source: Knoth & Pontika 2015

3. **Practical definitions** that denote a core of concern while leaving room for many different implementations (e.g., the FAIR principles, which are discussed later in this paper). In practical use on a university campus, open science often is a shorthand umbrella term for many activities (e.g., open licensing of data or posting open access preprints) that stem from many different principles (e.g., transparency, inclusivity, access, social justice);
4. **Personal definitions** that combine many themes and elements from an individual point of view; and
5. **Catchphrase definitions**, like the popular “Open science is science done right.”

## Open government

Open government is a concept at least as old as the need for science to be open. Originating during the Age of Enlightenment, Locke and Rousseau in particular envisioned the need for a partnership between people and the societies they inhabited, involving individual rights and consent of the governed. An accurate accounting of what governments were doing was aspirational at best during this period in history, however; reliable printing practices were still in their infancy during the mid-17<sup>th</sup> and early 18<sup>th</sup> centuries, and, as mentioned, societies throughout Europe were experiencing crises of fact—there was so much rampant disinformation being printed that it wasn’t even clear whether laws were being accurately recorded.<sup>30</sup> Still, the foundation had been laid for expectations of accountability in government, and this foundation persisted across history and a wide array of governments.

**Open government is a concept at least as old as the need for science to be open. Originating during the Age of Enlightenment, Locke and Rousseau in particular envisioned the need for a partnership between people and the societies they inhabited, involving individual rights and consent of the governed.**

Fast forward to the mid-20<sup>th</sup> century, contemporaneous with the evolution of other open philosophies and movements,<sup>31</sup> governments in the 1950s and 1960s began instituting laws intent on fostering more transparent, accessible, and accountable governments. This work has been nurtured and spurred on by organizations like UNESCO, the World Bank, the Open Society Foundation, the World Wide Web Foundation, and many other state-run and nonprofit organizations.

The Open Data Barometer is one of several organizations that attempts to monitor the current global state of open government, which most groups in this space interpret variously but generally mean the degree to which government data is openly available. The ODB’s surveys are longitudinal and cover a “range of questions about open data contexts, policy, implementation, and impacts and a detailed

---

30. See Johns 1998 for a detailed account. One might argue we’re experiencing something similar today as well. However, today we’re experiencing a contest between truly authoritative information and blatantly false misinformation (think Q-Anon). At least we know what’s true and what’s false, whereas in the mid-17<sup>th</sup> century there was no such anchor in reality. However, Wootton (Wootton 2015) notes that in the early years of printing, before the concept of “discovery” emerged, there also wasn’t a cultural sense of “intellectual property.” Hence, some of the piracy and misinformation that was occurring was simply a result of demand—that printers wanted to share new knowledge and make money in the process for doing so—whereas today the motivation isn’t to share knowledge but to knowingly spread disinformation.

31. Some of this can be traced to the open society work of Karl Popper. UNESCO has also championed freedom of information since its inception and has been a global leader in advancing information access as a universal human right.

dataset survey completed for 15 kinds of data in each government, which...[touch] on issues of data availability, format, licensing, timeliness, and discoverability.”<sup>32</sup>

Many groups are engaged in the open government advocacy side, as mentioned. With regard to policy formulation, the ranks are thinner. The Open Data Charter (ODC) is one such organization, growing out of the 2013 G8 Open Data Charter. ODC is an ambitious, non-binding pledge, currently signed by 22 national governments, to make their key data open by default where ever possible and to make this data accessible and reusable. While terms like OER, open access and open science aren’t explicitly mentioned in the charter or in letters of support received from governments (nor is this information tracked by projects like ODB), the charter does highlight the need for better data in order to inform better evidence-based policy making on issues like climate change. The charter can also be modified going forward to include more specific (“policy plus”) detail on issues such as open science. ODC is an important example, therefore, of an implicit understanding that already exists throughout the world of the connection between open data, open science, and open government.<sup>33</sup> In this case, open data is being used as proxy to describe a wide range of open solutions outputs.

Open government goes far beyond data, of course, and also merges with issues of social justice, accountability in government, human rights, and beyond. Also included in open government statistics, for example, is the fact that 112 countries around the world have freedom of information laws, although much work remains with regard to improving the awareness and implementation of these laws and access to this information.<sup>34</sup>

The World Justice Project (WJP) tracks global data on “the extent to which a government shares information, empowers people with tools to hold the government accountable, and fosters citizen participation in public policy deliberations,” including “whether basic laws and information on legal rights are publicized and...the quality of information published by the government.”<sup>35</sup> Much of this information is outside the scope of this report, but there is a general overlap with regard to the accessibility of government funded research data. That is, although WJP doesn’t break out research data specifically, governments around the world have increasingly instituted measures to make the results of publicly funded research publicly available, and the world’s most open governments—Western Europe, Canada, Australia, New Zealand, and the United States—also happen to be countries that rank highest in terms of the percentage of their research published in open format (Science-Metrix 2018); vice versa, more closed (in terms of open government) countries like India, China and Russia have much lower open research outputs. There are wide differences everywhere in how open government and open government research data policies are being implemented, and also interesting dichotomies, with governments in Latin and South America ranking relatively low on the open government scale but high in terms of the percentage of their research published in open format.<sup>36</sup>

## Open educational resources

Open educational resources (OERs) are another major presence in the open solutions space. OERs can take many forms, from open courseware (like free and open licensed classroom materials), to books, to massive open online courses (MOOCs) and knowledge resource banks like Wikipedia and the Khan Academy.

---

32. For more information, see the ODB website at <https://opendatabarometer.org/>

33. See <https://opendatacharter.net> for more information. Other global open data (within the context of open government) policies have come out of the G8, G20, and OECD.

34. For more information, see UNESCO’s 2018 report on world trends in freedom of expression (UNESCO 2018).

35. See <https://worldjusticeproject.org/> for details

36. For example, Brazil leads the world with 74% of its research content published in open format (Science-Metrix 2018), but ranks 31<sup>st</sup> with regard to open government measures

OERs overlap with academia in a major way as schools (including university level open courses for the general public) seek more and more legal ways to keep costs down for students and keep teaching materials up-to-date. There is also overlap with regard to usability in particular—with regard to figuring out how to make use of all the open access, open data, open science, open government, open source/code, and other open solutions materials in the world. Once these materials are licensed in an open format, what then? OERs are one practical application of this information availability.

OERs are the newest entrant into the open solutions space. The Hewlett Foundation took the early lead on inventing this concept in 2002, motivated by the need for more accessible and reliable educational resources around the world and the potential of the Internet to provide these resources. By around 2008, a number of other significant OER projects had sprouted up, like MIT's OpenCourseWare, Carnegie Mellon's Open Learning Initiative, the Open University UK's OpenLearn, and Rice University's OpenStax. Together with other relatively new organizations like OER Africa, the Institute for the Study of Knowledge Management (OER Commons) Creative Commons, the Open Society Institute and the Shuttleworth Foundation, these groups formed the beginnings of the OER movement and created a global infrastructure for open licensing.<sup>37</sup>

### Practice-based open

It's also important to mention the several "practice-based" types of open that exist in this conversation, such as open collaboration, open recognition/reward, open publishing practices, and so on. These are open-based approaches to the many side issues in publishing that need to be addressed as part of any comprehensive reform effort in scholarly communication, such as peer review and evaluation. Insofar as these deal with the practical "how" of open research more than the conceptual "what" (such as journal articles, data or code), open practices are at the cutting edge of innovation and tend to flow across conceptual boundaries:

- **Open collaboration:** These are tools and processes that support inclusive and networked research practices. For instance, there are many different virtual research environments (which vary by discipline; see also the Center for Open Science's Open Science Framework); collaborative writing platforms like Authorea and Overleaf; massively open online papers, or MOOPs (see Tennant 2020a); reference management tools like Mendeley and Zotero; annotation tools like PubPeer and Hypothes.is; data/code project collaboration sites like GitHub; and academic social networks like ResearchGate.<sup>38</sup>
- **Open recognition and reward:** These are tools and processes that help improve researcher identification and evaluation. Tools like CRediT (Contribution Roles Taxonomy), for example, allow for a standardized description of each author's individual contribution to an article; using ORCID helps permanently connect researchers with their research activities; the San Francisco Declaration on Research Assessment (DORA) defines how we can move away from journal based evaluations; and the Leiden Manifesto offers guidance on how to use research evaluation metrics more responsibly.
- **Open publishing practices:** These are tools and process that help increase the transparency and reproducibility of research processes and products. For instance, protocols can be regis-

---

37. For more information, see the Hewlett Foundation's report on origins of the OER sector (Atkins 2007).

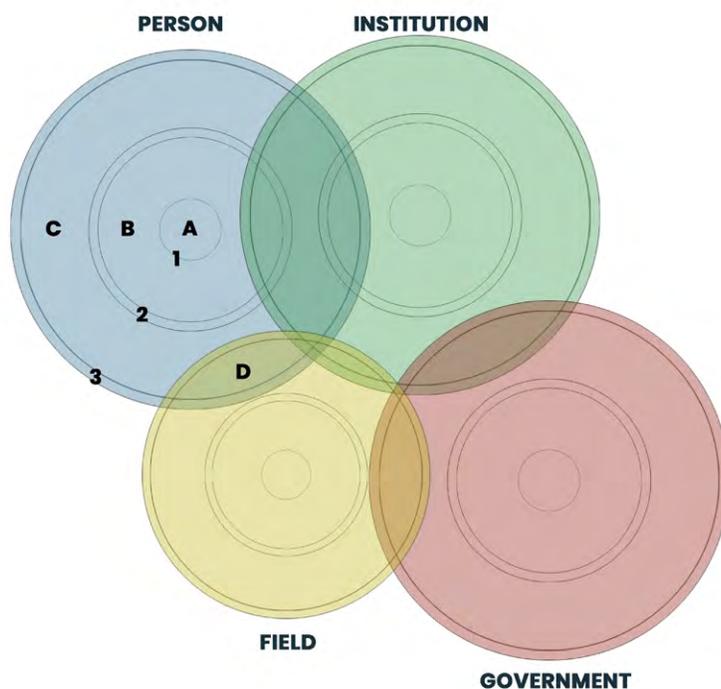
38. ResearchGate is a for profit and it's unclear what they are going to do with all the data they gather or whether their copyright violation practices will end. The network isn't mentioned here by way of endorsement, just example since they are currently the largest academic network (unless we also count organizations like AAAS). Academia.edu is another prominent company in this category.

tered through Registered Reports or ClinicalTrials.gov; open notebooks can be used to track research step by step; open peer review processes can help daylight research critique; using pre-prints can provide an important source of feedback as the final version of an article is developed; and open publishing standards like xml can increase the usability and visibility of documents.

## Synthesis

How do all these concepts overlap with regard to the information we deliberately share?<sup>39</sup> Let's briefly consider four possible approaches. First, for all practical purposes—and this will become more evident as additional evidence is presented in this report—it's important to note that there are essentially an infinite variety of information seeking behaviors and information granting solutions in the open space and beyond, describing the interaction between two given researchers, a researcher and her particular

FIGURE 4: INFORMATION BEHAVIORS



Different entities have different issues with regard to access, depending on their relationship and the nature of information being sought. These issues intersect in an essentially infinite variety of ways. Sometimes, access is easy, expected, and/or of obvious interest/need. At other times, none of the above apply. Take for instance the intersecting information clouds of a person, an institution, a field of study, and a government. At the core of each information cloud (region A) is closely-held information (think financial information, for instance). Surrounding this information is an access barrier (wall 1). In the next information ring, region B, we find information that is both accessible and usable—in the case of individuals, for instance, banking records that are accessible to both individuals and institutions. Both access and usability barriers encompass region B, in both directions. Incoming access barriers (wall 2) are defined by discovery, findability, technical feasibility (like having an Internet connection) and so on; the outgoing usability barrier is defined by issues like format, ease-of-use, language, and so on. The next information region (area C) is the information we “shed” either voluntarily (by publishing, for example) or involuntarily (by using a web browser). Here again, this cloud is encased by usability and access

barriers (wall 3). In this case, the incoming access barrier might be defined by paywalls, and in the incoming usability barrier might be defined by whether the data is clean or standardized. Every user—every individual, institution, government, or other entity—has a different information cloud profile, with different ways in which information gets shed into region C, different motives and methods for access, and different barriers for actually utility information instead of just accessing it. Indeed, for every information seeking interaction, a different information cloud profile will exist. Region D represents overlap, which can take on many different forms. In some cases, this overlap might result in the creation of an information silo, for example, and this silo may attempt to make the access and usability barriers lower in the region of overlap.

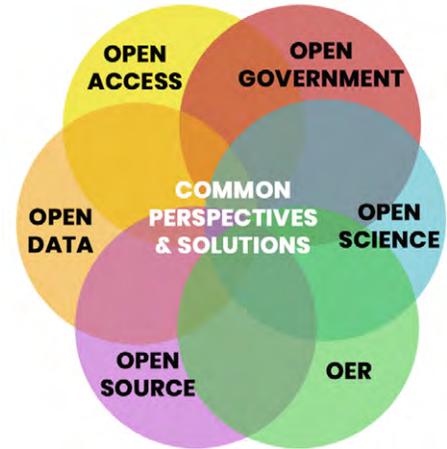
39. Here, the word “deliberately” is used because there is also a great deal of personal, non-research information we involuntarily share in the normal course of a day via the Internet, or other personal information like health and financial that is shared but not with the intent to reuse publicly. This analysis focuses on just a subset of information that has to do with knowledge creation. Also, it's important to note that the dynamics of voluntarily sharing information probably differ in important ways from the dynamics of sharing information due to regulations and mandates. This difference hasn't been explored yet to the best of our knowledge, but one hypothesis is that there will be more diversity in the types of and solutions for freely shared information versus information shared via mandates and regulations (at least strictly constructed ones).

institution, between a particular institution and a government, and so on (see Figure 4). This space is much broader and more complex than we've discussed so far in this report. It's important to keep this in mind when considering a synthesis model that describes the entire open space and not just the parts we have labeled. All of this activity is facilitated by and gives rise to a diaspora of actors, policies, programs and concepts.

The first and most obvious approach to understanding the overlap between these behaviors and solutions is to use brute force, cataloging the entire corpus of open solutions definitions, tools, policies, programs and actors, arranging them on a white board and looking for points of intersection (Figure 5). This approach is tempting but it ultimately misses the point that these interchanges and intersections represent a search for information, not a category of open effort. At best, this approach just reveals a dispiritingly complex taxonomy of open that is too broad to be useful, not any fundamental truths about open solutions.

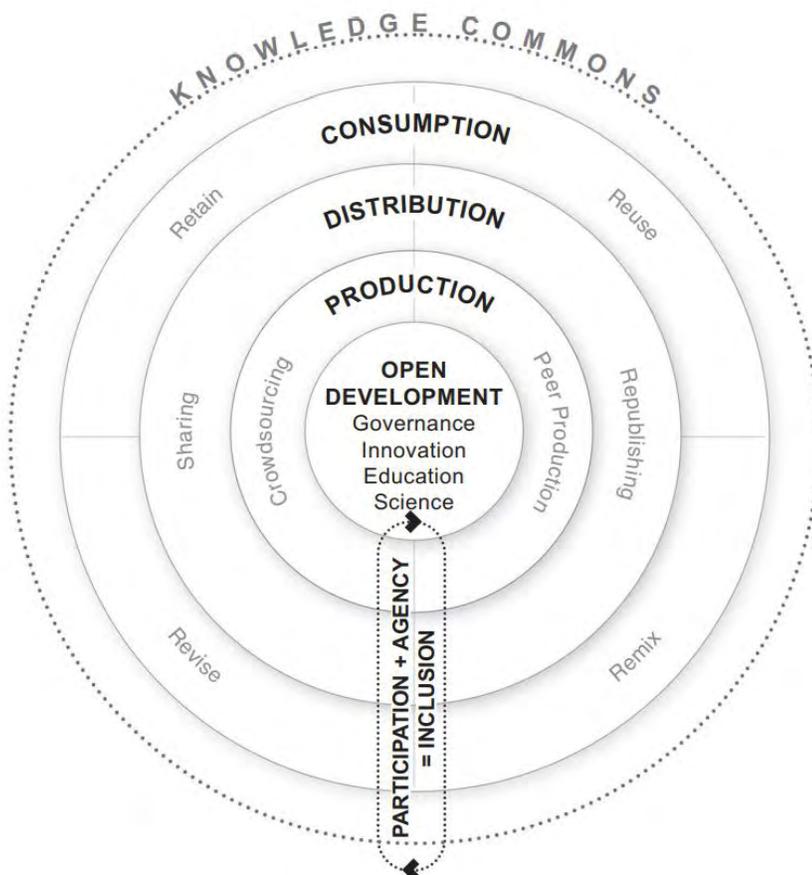
The second approach is similar but broader. In their 2020 work, "Making open development inclusive: lessons from IDRC research" (Smith 2020), Matthew Smith and Ruhiya Kristine Seward conclude that open efforts are so interrelated at the practice level that they can be best described as "open development" practices (very similar in thinking to this report's "open solutions" nomenclature). Surrounding

FIGURE 5: OPEN SOLUTIONS OVERLAP



and supporting these common practices are consumption, production and distribution systems that also share many common features (Figure 6).

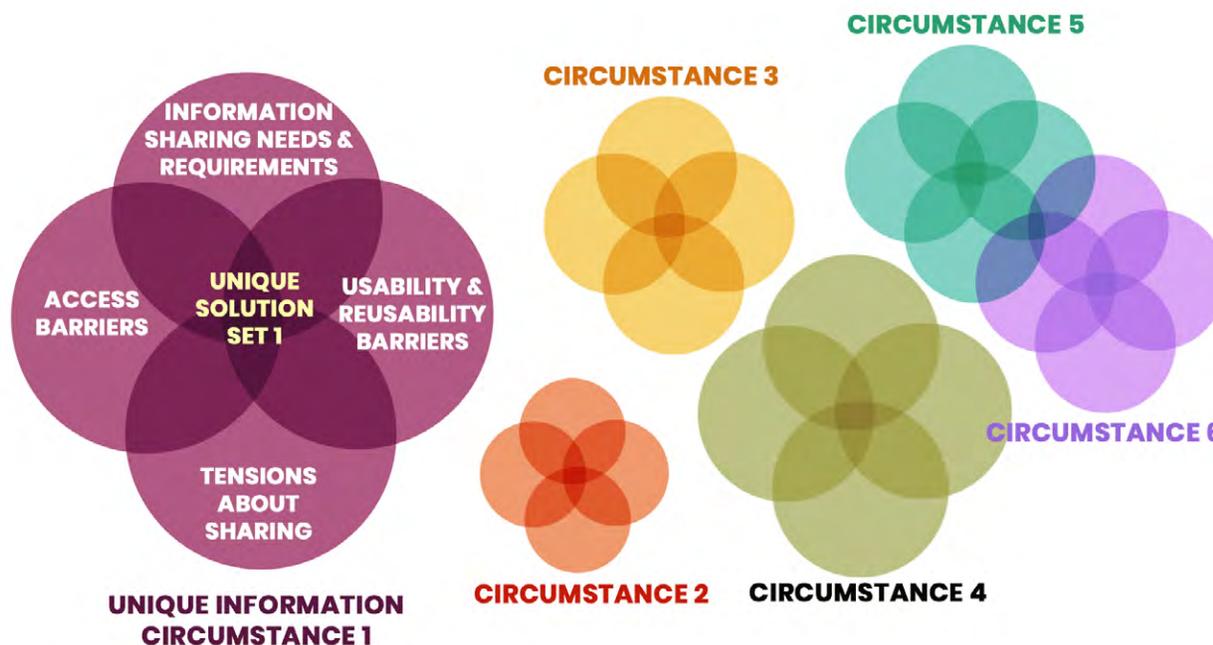
FIGURE 6: OPEN SOLUTIONS SYNTHESIS



Smith and Seward's approach avoids the temptation of the first approach to simply organize all phenomena in the open universe, but their model may still be too specific to reveal any fundamental truths about open solutions. That is, it still aims to organize this universe on the basis of "what" is being observed (empirical truths) instead of digging for the more fundamental conceptual and theoretical truths that explain "why" the open universe operates the way it does. Also, from a policy perspective, it may not be helpful to simply note that all open solutions have similar practices. There is clearly overlap. The question is how do we translate this common ground into action? In this case, open solutions do indeed share many similar prac-

Source: Smith 2020

FIGURE 7: OPEN SOLUTIONS PHILOSOPHIES



Philosophically, every search for open information can be characterized by four factors: needs, access, tensions and usability. These factors intersect in unique ways for each open information search. Diversity is a fifth common factor, but this is expressed as the various unique solutions and solution sets and a wide variety of unique circumstances. At each point of intersection for each of these circumstances, a spectrum of unique open solutions exists. For instance, at the intersection of access and need for one particular circumstance, a variety of solutions might exist—say to address the availability of affordable STM journals in Nigeria. But there are far fewer solutions that also address the tensions and usability concerns of this one circumstance. Also, all of these solutions may be inapplicable to the availability of affordable HSS journals in Nigeria, or the usability of open data in chemistry research in Belgium, or the tensions around using CC-BY in bioinformatics journals in China. Each circumstance has its own unique intersections of concerns, and the solutions for each of these intersections aren't necessarily applicable to other circumstances; however, many circumstances are similar enough that they can share the “templates” of solutions and in some cases even the solutions themselves can interoperable. This is why an understanding of and respect for the diversity in this space is essential to coming up with solutions that work, instead of looking for one-size-fits-all solutions.

tices, but these practices are also very specific. Here, we want to avoid the temptation to invent one-size-fits-all open policies based on observations alone (e.g., all grass grows, therefore we only need to make one kind of lawn mower).<sup>40</sup>

The third approach is philosophical (see Figure 7). Considering what we know about the purpose, means and tools of each open solution, the philosophical approach reveals that all solutions are simply different manifestations of a limitless combination of information seeking needs, tensions, barriers and usability issues:

- 1. The need, desire or requirement to share information.** Needs and desires exist for any number of reasons, within and between each open solution, to create valuable new spaces for information sharing: to rectify inequities in how information is currently being shared, to create the potential for new interactions, to facilitate growth and seamless interchange in a field, for

40. This ties into an important point, brought up later in this report, that an effective open solutions model needs to be grounded in inclusion, and built and maintained by the information communities it serves, rather than imposed from the outside.

moral-ethical or social justice reasons, or even to make money. Requirements can include complying with government, institutional and/or funder mandates for sharing, or sometimes legal requirements (like the US Freedom of Information Act).<sup>41</sup>

- 2. Tensions about sharing.** These tensions take the form of seeking assurance that authors receive proper credit for their ideas (especially in cases where discovery is involved), that ideas are not misused or misrepresented by others, that privacy concerns (with regard to health research) and security concerns are respected (with regard to sharing business or government data). In the open access, open data and open science policy conversations to date, these tensions have frequently been downplayed by policy makers. Surely, so goes the argument, researchers understand that knowledge belongs to society. Therefore, research generated with the public purse should be willingly given to the public, on demand. This argument often runs roughshod over the history of competition in science and simply assumes that science will continue to thrive in an environment where discovery is no longer the driving incentive, replaced instead by sharing—where preliminary data should be released immediately to other scientists and the public so that anyone anywhere can use this knowledge at once. This is a bridge too far in the minds of many scientists, and may be the primary reason why there is pushback against the current way of thinking about open.<sup>42</sup> In their mind, is this kind of open worth the risk?
- 3. Access barriers.**<sup>43</sup> Simply making information available doesn't necessarily ensure access. Across fields, the most common access barriers are cost and findability (or discoverability—which is to say, is the information being shared properly coded for discovery on the internet or stored in the right places?). Other access barriers include timeliness (embargoes in the print world), political interference, and technical limitations (internet bandwidth, for instance, or lacking the right hardware or software, especially for activities like high-throughput data processing). Although not technically a barrier to access in the same sense as these other factors, awareness and motivation also keep researchers from using open resources. This lack of awareness and lack of motivation limits the growth of content in open platforms and limits what can be accomplished with them.
- 4. Usability and reusability barriers.**<sup>44</sup> Getting access to information is only part of the battle. The next step is being able to adequately use or reuse this information. Common barriers include a lack of common standards (from technical to formats to naming conventions, units, and measures, all of which goes to the issue of interoperability), a lack of context (posting a dataset without also including a complete set of notes for how to interpret this dataset, which can be dangerous and lead to misuse), language differences, insufficient transparency to satisfy integrity and reliability questions (researchers cannot trust that data is truthful and replicable if they don't have enough accompanying analysis to satisfy their concerns), and licensing formats.<sup>45</sup>

---

41. With regard to requirements, as will be noted later in this report, a sizable percentage of researchers often request exemptions from sharing for any number of reasons (for instance, to protect the confidentiality of ongoing research work).

42. Survey data exploring this dynamic is discussed later in this report.

43. To address the open access goals of printed research, access and usability concerns are summed up by the DARTS Open Framework, with DARTS covering discoverability, accessibility, reusability, transparency, and sustainability. Open data goals are often summarized by FAIR, requiring that data be findable, accessible, interoperable and reusable.

44. Ibid.

45. Here, Fair Use (also known as Fair Dealing) laws allow text in any licensing format, including assigned copyright, to be used within reason. As long as large chunks of research papers aren't reprinted verbatim (and also considering the character of the use, the nature of the copyrighted work, the percent of the work being reprinted and the effect on the potential market for or value of the copyrighted work), there is generally no restriction on citing this kind of work, or even mining copyrighted research papers for content. The most serious limitation with regard to licensing formats in this particular genre of knowledge may have to do with the inconsistent or improper licensing of datasets. The ideal licensing format for data is CC0, or public domain, since this format doesn't require attribution. When data is licensed in a more restrictive format such as CC-BY (which is a liberal license for print but not for data), attribution is required, which is difficult to do and track when datasets are mixed

Other barriers can include differences in goals—what is suitably open for one group may be less than suitable for another group. Even these open goals themselves may pose a barrier if they are incompatible with the degree to which information can be shared (for instance, if there’s a requirement to immediately post study data to the web but a typical dataset in your study is far too large to upload or download).

- 5. A diaspora of diversity.** All open fields have a vast amount of activity and diversity. This diversity involves a wide variety of stakeholders and institutions, all with different motives, perspectives, goals, policies and measures. There are agencies upon agencies in every field working together to coordinate activities and streamline policies, although this is generally just a lot of siloed collaboration (see Davies 2019 for instance). Most notably, there is generally no widespread and coordinated action to connect different fields of open—organizations tend to specialize only in open data, for instance, or open access (repository groups are an exception, straddling both of these fields; major research universities can also be hubs of local integration). These diasporas are all struggling with the same challenge that open is growing—how fast is difficult to quantify because there are so many differing perceptions involved (e.g., what counts as “open”), a wide variety of dimensions/indicators/metrics for open growth and impact (e.g., open citations), large gaps in the understanding of needs and best practices, and large capacity gaps (for infrastructure development, education/training, and monitoring/evaluation).

Each of these five common concerns can be unique for each information challenge. With regard to the first challenge of assessing the need for sharing, this need may not exist in all situations, or even if it does, the reasons for sharing may be outweighed by other concerns like the practicality of doing so (for instance, if the data in question is too large to share, has national security implications, or is protected by intellectual property rights). In other cases, usability may be impacted by timeliness issues—data that takes five years to clean and prepare for public consumption may not be as usable as data that is immediately usable. The questions and challenges are at once varied, broad and deep, and the solutions we invent display a spectrum of outcomes for each point where the Venn diagram of needs, tensions, barriers and/or usability issues overlaps.

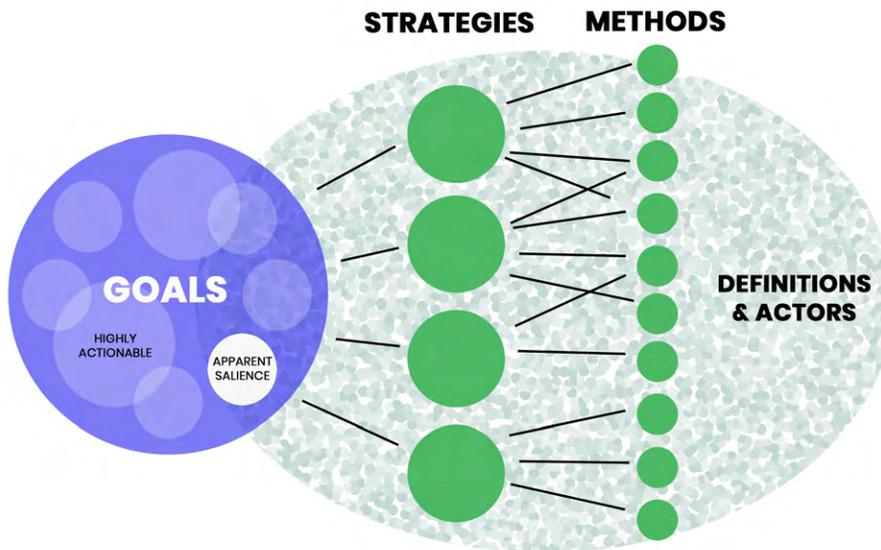
The fourth and maybe most usable approach (at least from a policy development perspective) to understanding the overlap between various open solutions is the goal-based approach (Figure 8). Fundamentally, all open efforts share one common goal—to make information more open. To achieve this goal, many different strategies and methods are employed. In this approach, the diaspora of actors and definitions that surround this activity are largely irrelevant in terms of modeling. They are just a reflection and manifestation of the open strategies and methods employed.

- 1. Goals:** The overarching goal of every open solutions approach is to make information more open. Inside this big tent are a number of more specific goals, some outcome oriented and others more organizing in nature. Outcome-oriented subsets include (depending on the circumstance) making information immediately open (no embargoes, making it open subject to specific licensing conditions, making it freely open in perpetuity, and so on. Organizing subsets include focus points like improving science, improving transparency in government, improving the equitable distribution of research knowledge around the globe, or improving the interoperability of data. All of these subsets are vital for mobilizing activity and expertise, but they can also be distracting because they are often promoted as ends in themselves, blinding us to the fact (due to their apparent salience) that the common goal of all of these efforts is openness, not a particular flavor of openness. This is true even if subsets don’t overlap. That is, for instance, even if the sep-

---

and merged for reuse. This becomes particularly problematic when large sets of data are accumulated. One bad apple can spoil the bunch—one restrictive contribution can render the entire large dataset unusable from a legal perspective.

FIGURE 8: OPEN SOLUTIONS GOALS



In this model, goals beget strategies, which beget methods. It's important to note that some subgoals may be more "actionable" than others because they share a common goals and beliefs (for example, improving science). Locating and building upon the most actionable of these common ground subgoals is important and will be discussed later in this report. Other subgoals, however, may only appear to be more salient because of the degree to which they are engaged with strategies and methods (for instance, having more stakeholder or policy connections). The "apparent salience" of these subgroups doesn't necessarily make them more actionable—that is, they are more fertile ground than common ground. Also, this apparent salience can blind us to what our larger common ground goals look like.

arate goals of improving the equitable distribution of knowledge around the world and improving research don't overlap (though they do), they all share a common goal of improving the openness of information.

2. **Strategies:** Because the goals space is so diverse, different open solutions necessarily involve different strategies for advancing openness. For open access, the primary strategy is to engage with publishers and academic libraries—publishers as the interlocutors of research knowledge, and libraries as the primary buyers of the journals in which this knowledge is communicated. For open data, the primary strategy is to engage with the communities (business, academic, etc.) who are generating data to understand what they are doing and what needs exist. For open government, the primary strategy is to engage with governments and citizens to advance legislation that will free up information for a variety of needs and purposes. Subsets of strategies can involve organizing philosophies like DORA that capture a particular ideal of open and enlist a particular audience to advance this ideal; and also the organizing principles of various efforts and actors—to reform publishing, to advance science, to improve access, and so on.
3. **Methods:** In visualizing how the open solutions space is connected, imagine the methods phase as the portion that is connected in name only. Flowing from a common general goal and a largely disconnected variety of strategies, a vast array of methods has been deployed—open publishing solutions, data repositories, a wide variety of tools and best practices, and so on. Subsets of methods can include data tools, open publishing platforms, legislation, agreements, and more—whatever instruments work to advance the cause of open. Some of these methods overlap, many do not. It isn't important to focus on finding overlap here because, practically speaking, each open solution needs to maintain its own methods—we needn't concern ourselves with trying to find methods that work for both open source and for open access. The important thing to note is that, conceptually, these methods are "fringe" considerations, not core. They are all working to advance various offshoots of open.

Figure 8 illustrates this relationship. Note again how in this approach the details we often focus on—our definitions of open, the different kinds of open solutions, and the variety of tools in this space—are all really minor players in the overall scheme. The overall goal—to make information more open—dominates this perspective.<sup>46</sup>

This goals-based approach has much in common with the Theory of Change approach that is commonly used by governments and other organizations to first identify the long-term change that is desired, and then work backwards to map out the actions and policies needed to create this change. See Table 3 on the following page for details.

In summary, then, we can conclude that there is a vast variety of potential interactions between researchers, of overlap between concerns, and of potential solutions that address these interactions and overlaps. What we tend to notice most often in this space is only the overlap between usable policies (like FAIR, for instance), or agencies that are involved in multiple open solution spaces, but in fact, the contours of the open solutions space are all defined by the same foundational characteristics. From a philosophical perspective, these contours are the need, desire or requirement to share information; tensions about this sharing; access barriers; usability barriers; and a diaspora of diversity. From a goal-based perspective, these contours are goals, strategies, and methods. The rest is just noise—important noise, but also noise that distracts us from seeing the connections and common ground in this space.

Matthew Smith and Katherine Reilly’s 2013 collection of essays (Smith 2013), titled “Open development: networked innovations in international development,” supports this generalized approach to open. In their work, the authors identify several crosscutting themes between open movements:

- 1. Openness is a means, not an end.** OSI has been preaching this concept for years. Smith and Reilly note that across open movements, openness is seen as a way to solve problems and improve benefits—as a means to end, not the end itself;<sup>47</sup>
- 2. Open is layered.** The root object upon which open is layered, like technology, can be enhanced by openness of varying degrees, but it is not a benefit created from nothing and unto itself. This is an important concept insofar as making sure that open evolves and isn’t pursued as a complete disjuncture;
- 3. Open is disruptive.** Open models can be very disruptive both in technical and social terms;
- 4. Structure is needed.** Harnessing the power of open requires structure, and does not happen by chance;
- 5. The ideal is never the reality.** No open model is ever universally and completely open. OSI describes this variation in open access publishing with the DARTS open access spectrum (described earlier);
- 6. Openness requires a critical perspective.** If we truly want open to succeed, we cannot let it be an attractive facade that masks inequities or causes unintended consequences. We need to keep our open efforts grounded and honest, and question our cultural and ideological assumptions; and

---

46. As an organizing idea, however, we may find it insufficiently motivating to rally around the flag of openness in general. Therefore, rallying around a common goal is an important approach to consider, as described later in this report.

47. Arguably, though, this is precisely what many open access advocates are trying to do today, pushing for open as an ideological goal as opposed to working to develop open solutions that will make research better.

- 7. Openness is a complex process, not a state.** The reality of open is fluid and uncertain and extends in at least three dimensions: openness of content, openness to people, and openness in process. When we typically consider open paradigms, we only deal with the content and people aspects. But as an open model becomes more and more open along these dimensions, it also opens up along the process dimension.

TABLE 3: THEORY OF CHANGE MODEL

### Description

Theory of Change (ToC) is a specific type of methodology for planning, participation, and evaluation that is used in companies, philanthropy, not-for-profit and government sectors to promote social change. Theory of Change defines long-term goals and then maps backward to identify necessary preconditions.

Theory of Change explains the process of change by outlining causal linkages in an initiative, i.e., its shorter-term, intermediate, and longer-term outcomes. The identified changes are mapped – as the “outcomes pathway” – showing each outcome in logical relationship to all the others, as well as chronological flow. The links between outcomes are explained by “rationales” or statements of why one outcome is thought to be a prerequisite for another.

The innovation of Theory of Change lies (1) in making the distinction between desired and actual outcomes and (2) in requiring stakeholders to model their desired outcomes before they decide on forms of intervention to achieve those outcomes.

Theory of Change can begin at any stage of an initiative, depending on the intended use. A theory developed at the outset is best at informing the planning of an initiative. Having worked out a change model, practitioners can make more informed decisions about strategy and tactics. As monitoring and evaluation data become available, stakeholders can periodically refine the Theory of Change as the evidence indicates. A Theory of Change can be developed retrospectively by reading program documents, talking to stakeholders, and analyzing data. This is often done during evaluations reflecting what has worked or not in order to understand the past and plan for the future.

### Applying the model

An important first step in the process is identifying a workable long-term goal and long-term outcomes. The long-term goal should be something the initiative can realistically achieve and that everyone involved understands. A trained external facilitator is best to lead the group to consensus and specificity in this process. Once a long-term goal is identified, the group then considers: “What conditions must be in place for us to reach the goal? Any such necessary conditions should be shown as outcomes on the Theory of Change pathway, underneath the long-term outcome. These outcomes act as preconditions to the long-term outcome. The process of identifying preconditions continues, drilling down the pathway by posing fundamental questions such as: “What has to be in place for this outcome to be achieved?” and “Are these preconditions sufficient for the outcome to be achieved?” In these sessions, participants may use markers, sticky notes, and chart paper to identify and organize outcomes, surface assumptions, develop indicators, and so on. The messy group work is then usually captured by the facilitator in digital form, through which the content can be expanded, edited, printed, shared, and otherwise managed as the theory continues to be developed.

### Growth

The use of Theory of Change in planning and evaluation has increased among philanthropies, government agencies, development organizations, universities, international NGOs, the UN, and many other major organizations in both developed and developing countries. This has led to new areas of work, such as linking the Theory of Change approach to systems thinking and complexity. Change processes are no longer seen as linear, but as having many feedback loops that need to be understood. Consequently, Theory of Change is strengthening monitoring, evaluation and learning. They are also helping to understand and assess impact in hard to measure areas, such as governance, capacity strengthening and institutional development. Innovations continue to emerge.

Source: Excerpted from Wikipedia ([https://en.wikipedia.org/wiki/Theory\\_of\\_change](https://en.wikipedia.org/wiki/Theory_of_change))

...recognizing these similarities can help us understand that these efforts aren't necessarily separate or unique, but are also part of a broad, long-term, multifaceted push from many corners of many societies to make information of all kinds more open. Understanding how similar the open efforts have been with regard to common challenges and approaches can help us merge open policies together in a way that wouldn't be possible by focusing only on the specific practices of each open movement.

Can depictions like these help us create a policy framework that advances “open solutions?” Certainly it helps to note that different open efforts are similar when viewed from a high level. And recognizing these similarities can help us understand these efforts aren't necessarily separate or unique, but also part of a broad, long-term, multifaceted push from many corners of many societies to make information of all kinds more open. Understanding how similar these open efforts have been with regard to common challenges and approaches can help us merge open policies together in a way that wouldn't be possible by focusing only on the specific practices of each open movement. Also, if we can identify the most salient areas of overlap, then from a policy perspective we might begin to address these different open efforts simultaneously instead of separately. Once we find these “hooks” we can begin building more substantial connections over time.

also be seen as a reaction to a specific set of circumstances, and a unified theory of open will only reflect a synthesis of the different interests at play as opposed to cross-cutting policy solutions that will effectively address these circumstances. This is an important aside. Many forms of open are, after all, uneasy alliances of interests with minimal points of agreement, whose specific interests won't necessarily be better served by joining into a collective push for openness rather than by continuing with their solitary pursuit of open solutions.

One might argue, however, that even though these open movements share common features, each can

This may be true, but as we will discuss later, these two approaches shouldn't be mutually exclusive. We can and should “allow” the open access community to continue developing solutions specific to open access needs, and so on. Since broader alignment does exist, however, between open goals, movements, needs, solutions and more, we can use this alignment as a framework for mobilizing action on grand challenges, for working more closely together to advance the unified cause of “open solutions,” and also for addressing common ground open solutions needs such as infrastructure, standards, and outreach. More will be discussed about this approach in the next section of this report.

## STAKEHOLDER NEEDS & PERSPECTIVES

The needs and perspectives of researchers, publishers, funders, universities, libraries, and other groups active in the open solutions space are incredibly diverse. Fortunately, there is growing understanding of these needs and perspectives, and also a growing normalization of open. Part of this normalization is simply due to a maturation of the open effort—more is being done, more is being published,<sup>48</sup> more best practices are emerging, and there is more evidence of the benefits of open.<sup>49</sup> In addition, the open environment is also supported by more open laws and policies, and there is more widespread use of open licensing instruments.

48. As Archambault 2018 notes, over half of all new academic journal articles are now carry some kind of open license.

49. Awareness of these benefits is an important topic that will be touched on later in this paper.

While these diverse needs and perspectives often align, in many cases they do not. Different stakeholder groups can have starkly different points of view about what solutions are best or even needed; there is often a lack of awareness amongst researchers of open processes and benefits; there is often conflict between regional needs, and the specific needs of different fields; and even differences over how to define open. In this environment we see a constant, contentious debate over who gets to decide the path forward; unfortunately, this debate often takes the form of whose view has the highest moral ground.

These debates are illuminating for understanding the totality of the open debate, but from a policy perspective, they are also hugely complex, requiring a search for more information and broader perspectives than are being presented. As the body of evidence in this space clearly demonstrates, all “sides” are wrong, and no one side has all the information it needs to make an informed judgement. And yet judgements have rained down on the open solutions space nonetheless, particularly in open access.

Therein lies the rub. This space has become so judgemental that many people and groups who need to be heard from simply don't feel like talking any more.

## Researchers

The most important people in this entire conversation are researchers. This is also the most underrepresented group, although not for lack of trying. Researchers are busy and have their priorities. Getting involved in a 20-year-long debate about the future of open isn't the highest priority for most.<sup>50</sup> Also, researchers aren't a homogeneous group. Just because we may hear a steady stream of open opinions from a few scientists and science membership groups doesn't make these voices—valued as they are—representative of all science, or even their fields.<sup>51</sup>

This said, what we can conclude from the multitude of researcher opinions that have been voiced, and also from a large body of survey research in this field and from market outcomes, is that the needs and perspectives of researchers with regard to open solutions fall into six main categories: impact, confusion, trust, access, effort and regional differences.

## Impact

Researchers are increasingly less suspicious of open (owing to the increasing normalization of open, as mentioned) and think of open as a tool that can be used for personal benefit—to help their research get into more hands, get more citations, get more reuse, and to make more of an impact on science and society. These sentiments have been measured repeatedly over the years and in a number of different ways—asking about attitudes toward open access, publishing in general, open data, and so on.

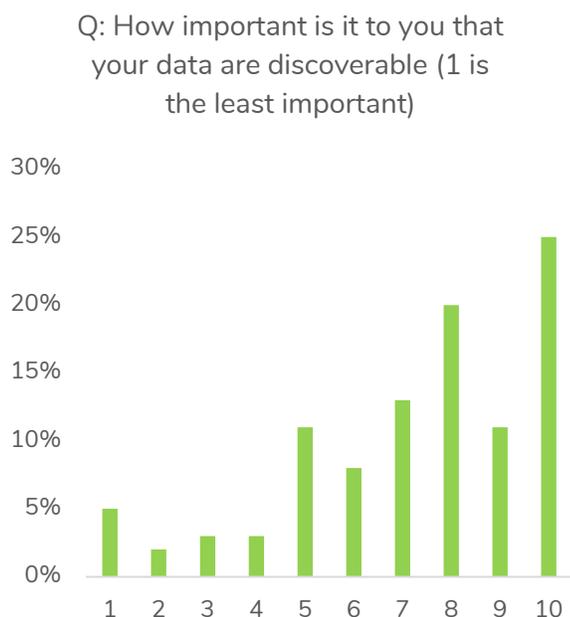
---

50. This fact is probably the most fundamental problem of all in the open solutions space, where researchers are not treated by policy makers as being active stakeholders, and even see themselves as being outside the policy making process. Where involved, they are involved via researcher organizations who send representatives who may or may not represent the views of the organization as a whole (the larger the membership body the less likely this is to be the case due to diversity of opinion), but certainly not the broader research community. Unless it comes in the forms of mandates, researchers in general tend to be unaware of what is going on in the open space, or the implications of open for their work. How does one bring ordinary researchers as stakeholders to the table? That is one of the biggest barriers to uptake at the moment.

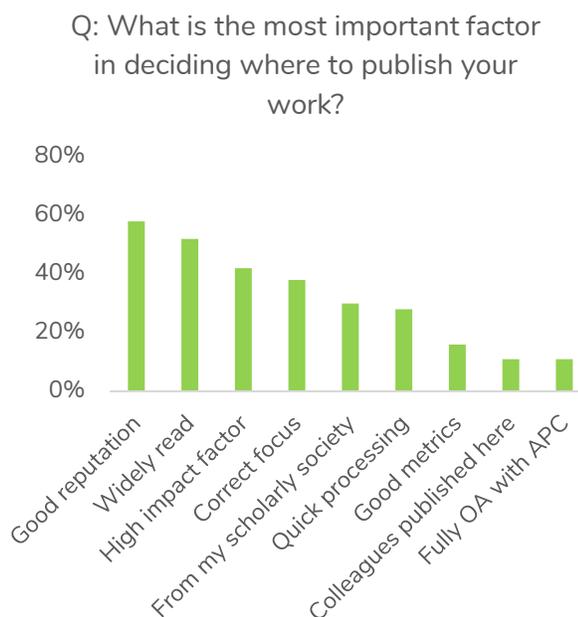
51. It's also important to note here that although researchers come from everywhere, the most highly cited researchers are concentrated in the US, EU and China, together accounting for over three-quarters of the total (with the US accounting for 42%; see Clarivate 2020). More broadly, adding up the full top-10 list, about 85% of highly cited researchers are coming from just 10 countries. By institution, 15 of the top 20 most highly cited research institutions are in the US, two are in China, two are in the UK, and one is in Germany. Therefore, arguably, any open solution that doesn't have buy-in from these most influential research countries, and particularly the US, won't be able to change the global calculus with regard to open.

According to one of the more recent studies on open attitudes among researchers (Wiley 2019a), for example, the top reasons why researchers chose to publish in an open format are to improve visibility and impact, improve public benefit, and improve transparency and reuse. However, these answers need to be considered in context, understanding that not all researchers chose to make their work open. In the context of publishing attitudes in general (see the 2019 Taylor & Francis survey), researchers primarily choose where to publish based on journal impact factor and reputation. Open access is a secondary concern, but it is still valued (see the chart footnotes below for more detail).<sup>52</sup> There are other reasons why researchers want to share their research, but most often, personal benefit incentives are the most important.

**FIGURE 9: RESEARCHER ATTITUDES TOWARD OPEN**



Source: SpringerNature 2018 survey of researchers (n=7656; Stuart 2018). These percentages are aggregate figures. There is slight variability by discipline, with biological, earth, and medical sciences ranking higher than physical and “other” sciences. There is also slight variability by region.



Source: Taylor & Francis 2019 survey of researchers (n=2755). There are a number of omitted responses here related to cost.<sup>53</sup>

52. According to one of the more recent studies on open data attitudes (see Stuart 2018, the 2018 Springer Nature survey of researchers), around seven in 10 researchers see data sharing as being very important.

53. These responses have been omitted because with open access publishing, about 70% of published articles are paid via APCs (Parsons 2016), and in developing countries, these APCs are mostly paid by authors themselves (Scaria 2018). Therefore, the most accurate cost reply is the last one—how authors feel about OA publishing paid via APC. By comparison, 53% think a journal being free to publish in is the most important factor, but there are very few choices like this—called “diamond” OA—where the publishing costs are covered by funders or foundations (measured by number of articles published). Similarly, 39% of survey respondents said that open access publishing was the most important factor if there was no APC charge, but again, this kind of arrangement is a distinct minority of the number of articles published per year, and is an almost invisible slice of the more highly cited research work. See Hampson 2019, OSI’s critique of Plan S, which cites an extensive body of literature on this topic. Finally, a recent analysis of 319 journals that flipped from subscription to APC format (Khoo 2019) shows that authors or their funders are, in practice, more willing to pay high APCs than they admit in surveys (which is not a good thing since APC prices are increasing a 2.5-6 times the rate of inflation; also, studies like Scaria 2018 show that Global South authors are in fact excluded from participating in science with high APCs).

Surveys also show that most researchers value choice and flexibility, and most funders and publishers support these choices rather than mandating and restricting choice, by, for example prescribing exactly where data should be shared, or which journal should be published in. Evidence suggests that researchers choose open practices when options that meet their needs are provided to them, from opting into open peer review to electing to publish their work as a preprint first, to making their data FAIR and choosing a licensing format that works best for their needs (see Wiley 2019a, Wiley 2019b, Tan 2020, Taylor & Francis 2019, Tenopir 2017 and other research).

These are generalizations, of course—there is significant variation in these impact estimates. Different fields have different open needs, outcomes and conventions (see Archambault 2018; the difference between STM fields and arts and humanities fields can be particularly large); researchers at different career stages make different use of open publishing options (see Nicholas 2017, Tenopir 2017, and many author surveys);<sup>54</sup> and different regions and institutions have different open conventions.<sup>55</sup>

There are also conflicting incentives at play here, adding tension to the decision about whether it's more important to value research impact, or societal impact. From the Taylor & Francis survey (Taylor & Francis 2019) we learn that the overwhelming majority of researchers surveyed think diversity, collaboration, interdisciplinary work, and support for early career researchers and researchers in lower income countries are all important for the future of research. Most (77 percent) also want more funders and institutions to support open access financially. But an even larger percentage (84 percent) want the freedom to submit their work to the journal of their choice regardless of funding model. In the end, at least for now, perceived research impact trumps societal impact. A successful policy model will need to align these two incentives.

## **Confusion**

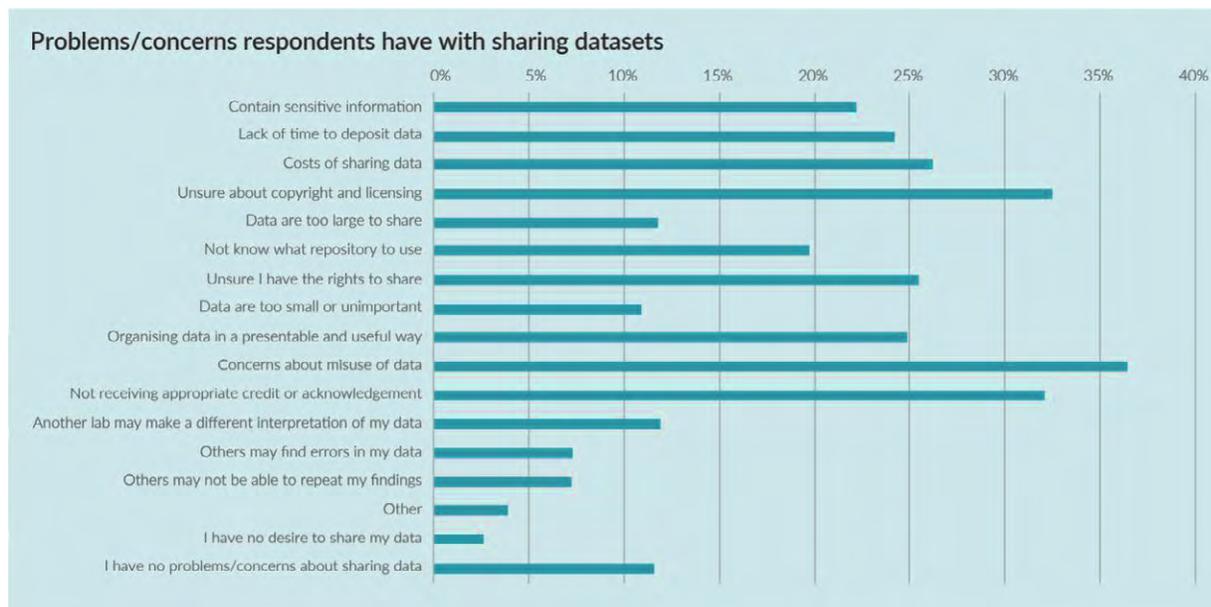
The open landscape can be confusing. Because of this, researchers can be confused (and concerned) about what open means, how to do it, and what impact it will have on their research and careers. For example, the culture of communication in academia doesn't normally recognize, reward, or incentivize data sharing when evaluating researchers for promotions and grants. So why bother? In addition, researchers may have concerns about how to organize their data in a useful way, which repository to use, how embargoes are handled, which license to apply (and widespread dislike of the CC-BY license format; see T&F 2019 among other surveys), the time and costs involved in sharing, data sensitivity issues (for instance, whether confidential data or data requiring special sharing agreements can be openly shared, archived or cited), whether their discoveries will be "scooped" or whether their data will be misused or misinterpreted, and even how data citations should be formatted (see Stuart 2018, NASEM 2020 and Cousijn 2018 for fuller discussions). Even the premise that unfettered access to data by the general public is an unvarnished good isn't always agreed upon (Chen 2018 and Stuart 2018). In particle physics, for example, the "challenges of scale and data complexity mean that there are certain necessary reproducibility use cases that are better served by a tailored tool rather than an open data repository. Such tools need to preserve the expertise of a large collaboration that flows into each analysis.... Confidentiality might aid this process so that the experts can share and discuss in a protected space before successively opening up the content of scrutiny to ever larger audiences..." (Chen

---

54. One explanation for this is that early career researchers understand the need to establish themselves and impress tenure committees by getting their work published in high impact journals, whereas established, tenured researchers are less concerned with the need to impress. In general, open publishing options are seen as being less prestigious (and they generally have lower impact factors) than closed, subscription publishing options (Hampson 2019).

55. Through the work of SciELO, South America—Brazil in particular—leads the world in the amount of research published in open formats. Other regions have varying degrees of uptake (see Science-Metrix 2018). Notably, China—which now leads the world in research publishing output—also trails the world in open. This is by design, not accident. China's approach to open is different than in Europe and South America (Lee 2020).

FIGURE 10: RESEARCHER TRUST



Source: Digital Science 2019

2018). The same is true in biomedical research, where concerns exist that making preliminary data open poses an unnecessary risk to public health, as well as a risk to research integrity and intellectual property rights (NASEM 2020).

So, in general, “confusion” over open covers a lot of ground, from impact to trust and effort (discussed below) and beyond; this confusion is a major roadblock to open adoption. Making open “simpler” would help; researchers are often at a loss about where to start (see T&F 2019). For example, from a researcher perspective, if you have a PAR agreement and just need to click a button to request a voucher, this is less of a barrier to making things open and increases OA uptake. Or, if researchers can get data or code uploaded to GitHub or Zenodo or something similar with just a few clicks they are more likely to do so. Removing barriers (and educating researchers about how to use easy one-click open options) will play an important part in reducing confusion and getting researchers motivated and engaged.

**Trust**

Closely related to confusion is the issue of trust. This is highlighted by the Digital Science 2019 State of Open Survey, which revealed that “The biggest barrier to research data sharing and reuse seems to be a matter of trust, and in particular trust in what others may do with researchers’ data if it is made openly available. The...survey revealed that over 2,000 respondents had concerns about misuse of their research data.” As mentioned earlier in this report, these concerns shouldn’t be surprising. Some concerns (like usurpation and misuse) are as old as science itself, and were even animating reasons for why science formed the way it did. As also mentioned earlier, while openness was seen as a virtue in the earliest years of science in order to encourage civilized debate about experimental findings, at no time did openness mean making methods and data available to the public or even to other researchers (which early on resulted in the theft of ideas). Openness was a necessary framework for debating the merit of scientific ideas, and was never intended as a vehicle to put all science knowledge immediately and forever into the public domain. In parallel, competition for discovery is a driving force in science, perhaps even the single most important driving force.

In our quest to invent the future of knowledge, we must not simply assume that our modern theories of how knowledge should be shared are superior to the actual knowledge sharing practices that have evolved over time and that have made scientific research such a revolutionary, successful and vitally important force for society over the last 400 years. In order for open to succeed, the issue of trust needs to be thoroughly explored and concerns about trust need to be thoughtfully and honestly addressed. Solutions here will need to look for common ground. What do we need from open? Eventual access to clean data that will help researchers, or immediate access to all data that may undercut discovery and threaten privacy? We need to make sure that our primary goal, first and foremost, is to help research succeed, and build our policies on that foundation rather than from the foundation that solutions promising quick and total openness are necessarily the most effective and desirable.<sup>56</sup>

**“The biggest barrier to research data sharing and reuse seems to be a matter of trust, and in particular trust in what others may do with researchers’ data if it is made openly available...”**

## **Access**

There is a lack of research on whether researchers have access to all the information they need. Anecdotal there’s certainly an access gap for smaller research institutions and institutions from lower resourced regions.<sup>57</sup> Also, considering access to all kinds of information everywhere, there are plenty of walls and barriers of all kinds (from lacking access to certain government datasets to being outside the access bubble for data that can only be used by researchers inside a given network; more will be discussed about this later).

In general, though, journal access is the most widely acknowledged type of access problem experienced by researchers, mostly owing for now to the high cost of subscription journals and the vast number of these journals that libraries need to purchase (which has been increasing rapidly in recent years).<sup>58</sup> This cost of subscriptions has been an animating force behind the open access movement, but there are also other forms of access problems that stem from bad discoverability, transparency, sustainability and more. For instance, only two-thirds of data files generated by researchers are published as supplementary files or deposited in repositories (Stuart 2018), and even for data that are submitted, long term access can be a problem due to issues like link rot (Good 2016), publisher insolvency,<sup>59</sup> and lax data archiving policies (Laakso 2020).<sup>60</sup>

Data also becomes obsolete (at rates that vary by field), researchers move on to other studies, and attempts to access data too long after the fact are effectively impossible because the data is either no longer relevant, has not been stored in a findable or retrievable manner, is not detailed enough to reuse, or the principal investigator is the only one with the “keys” because the study’s data management team long ago disbanded (NASEM 2020).

---

56. This isn’t to say that researchers alone should decide in advance what potential reuse exists for their work, but that our approach to open solutions should be led by better understanding the actual needs and concerns of researchers.

57. For the latter, publishers support Research4Life, which provide discounts to help close this gap.

58. We say “for now” because, as will be mentioned later in this report, we may be on the cusp of a new access problem where high APC costs will severely limit the ability of many researchers to publish. Access barriers to published materials are known as paywalls; access barriers to the ability to publish are known as playwalls.

59. For example, see Laakso 2020 for an analysis of how hundreds of open access publishers are at risk of disappearing (in addition to the 176 which have already disappeared between 2000 and 2019).

60. With regard to link rot, Good’s article links to various reports showing how an alarming amount of information on the web is no longer accessible through its original link, which has caused an information access crisis in the legal field. With regard to publisher issues, Laakso reports that in addition to many open access publishers going out of business, around two-thirds of publishers listed in the Directory of Open Access Journals (DOAJ) do not have clear data archiving and preservation policies.

## **Effort**

The effort required to participate in open hasn't often percolated into survey results, but analysts who look at the researcher community from the outside have long noticed this factor. Wiley's 2018 survey of researchers (Wiley 2018a) noted that "even those with a neutral or positive attitude towards open scholarship find that it's a challenge to translate the FAIR principles into their own discipline's commu-

**“[E]ven those with a neutral or positive attitude towards open scholarship find that it's a challenge to translate the FAIR principles into their own discipline's community standards and practices....”...Removing barriers (and educating researchers about how to use easy one-click open options) will play an important part in getting researchers motivated and engaged.**

nity standards and practices.... Recently, an enthusiastic proponent of open research remarked privately 'Sharing just slows research down.'”

Part of this attitude is a reaction to open policies that simply don't fit well. Some fields, for instance, generate data that are difficult or far too large to share.<sup>61</sup> Time-poor researchers in other fields might be overwhelmed by requirements to fully describe their data in order to maximize its utility to others, compounded by a lack of clear benefit for doing so, and the need to fill out diverse submission forms for repositories and journals (Jeffries 2018). In other fields, there are inconsistencies in data that make it hard to reuse (Faniel 2020).

At the same time, even those working in well-resourced institutions experience problems. Notes open data expert Fiona Murphy (Wiley 2018b), "I've worked with climate scientists whose departments are regularly hacked, and who receive a steady bombardment of 'Freedom of In-

formation Act' requests from climate skeptics which puts them in a real ethical quandary: they know that the data should be made available, for transparency and trust, but they are also keenly aware that these attempts to hack and requests for information are motivated by the desire to discredit their work and professional reputations. I often hear people say 'This isn't what I signed up for.'”

## **Regional differences**

As noted in this section and throughout this report, there are many differences between researchers (involving field of study, career stage and more) that any comprehensive open solutions policy needs to account for. Regional differences are often given short shrift, however, because open conversations tend to be dominated by US and Western European opinions. It is critical, therefore, to recognize that not all regions of the world can approach open solutions the same way. For example, efforts like OA2020 (the global flip) and Plan S started out by putting a huge emphasis on favoring the APC path to open. In the developed world, this may be an acceptable strategy—around 70% of gold articles are paid for by APCs (Bjork 2018), and for US-based researchers, these fees are paid by research institutions about half the time (Parsons 2016). In the developing world, however, these fees are mostly paid by authors themselves (Scaria 2018), and open publishing fees that work in the US and EU—often

---

61. Poynder 2019 notes that "some of the costs of forcing openness on the research community are clear to see.... Considering open data specifically, it may turn out that the sheer quantity of data now being generated in the research process makes sharing data more expensive and difficult than any of the perceived benefits it provides. As Spinal Cord Injury researcher Vance Lemmon has put it, 'Recently, with development of fluorescent light sheet microscopes that can image very large volumes at high resolution, image stacks have become enormous, from 5 GB to 1 TB. This kind of data is very hard to share.'”Also, it is important to provide the training and support for this kind of work.

higher than \$2,200 per article<sup>62</sup>—are utterly unreasonable in most of the developing world.<sup>63</sup>

Solutions for open government, open educational resources, and open data can also have significant regional differences, varying by budgets, politics, internet connectivity and many other factors. Focusing here just on open data, in her 2018 interview with Wiley (Wiley 2018b), Fiona Murphy noted the following: “There is such a wide range of experience [with regard to data], but something that really struck me when I visited Botswana last year, for International Data Week, was the realization that local and regional researchers were at even more of a disadvantage than I had thought. Often people have to collect and manage their data at their own expense and in their own time. This means that they greet the idea of then opening their data up, potentially for better resourced researchers who might have superior computing facilities to mine to analyze, with a justifiable sense of unfairness. This problem came up over and over again.”

**In order to find common ground between these solutions, and solutions that adequately address global concerns and also properly align incentives, we really need to understand what the totality of the open landscape looks like.**

### **Synthesis: Be honest and realistic**

There is much more detail to go into but this exploration is unfortunately well beyond the scope of this paper.<sup>64</sup> Suffice it to say that understanding the depth and breadth of author concerns and motivations is vitally important when piecing together open solutions that have different goals and origins. In order to find common ground between these solutions, and solutions that adequately address global concerns and also properly align incentives, we really need to understand what the totality of the open landscape looks like.

As mentioned earlier, we know that self-interest is a big mover—being the first to discover, making an impact, getting noticed, getting funding, and getting tenure. So too is the interest in complying with funder mandates, responding to publisher choices, and the influences exerted on researchers by their field, institution and region—norms and preferences, technical capabilities and the like. But there are other influences at work as well. Maybe most significant among these is that “open” simply means different things to different people. For instance, as noted in OSI’s open science policy recommendations to UNESCO (Hampson 2020b), at least three other major categories of variation exist with regard to the definitions, motives and goals that underlie research attitudes and actions on open solutions (see the table below for what this variation looks like just with regard to open science).

---

62. See Budzinski 2020 for a detailed analysis of APC’s from major publishers, which are typically much higher than \$2,200, and also Siler 2018. Hampson 2019 also discusses this topic at length and provides additional references. The full range of APC prices varies widely—in the US\$5,000 to US\$10,000 range for the highest quality journals.

63. Co-author and researcher Lynn Kamerlin notes that one reason many researchers are skeptical of open is that it has typically been implemented through mandates that stipulate gold OA as the only way to be compliant (not as much in the US, more so in Europe, and long before Plan S) and this approach typically has to be paid by individual researchers out of their research grants. These grants have not increased to compensate for the extra cost of publishing, while at the same time, APC costs have been going up quite rapidly (Wellcome 2018) and researchers are feeling the crunch as a larger and larger percentage of their grants go to the cost of disseminating rather than doing research (co-author and library expert Mel DeSart notes that US agencies are actually now encouraging grant applicants to include a line item for OA publication in their applications, so OA publications costs are now a fairly routine line item in application budgets). Transformative agreements may help in this regard. While they are controversial among some in the OA community who see them as another version of big deals, researchers lucky enough to be covered by them are often grateful. Moving away from the APC, if possible, is going to play an important role in both increasing ease of OA uptake and also rehabilitating researcher relationships toward OA.

64. Infrastructure and big data are discussed later in this report. For additional reading on the multitude of other issues, please see OSI’s publications on Plan S, common ground and open science, all listed in the reference section (Hampson 2019, 2020, 2020b).

TABLE 4: CATEGORIES OF VARIATION REGARDING RESEARCHER PERCEPTIONS OF OPEN SCIENCE

**Definition.** Different groups define “open” differently. Because of this, OSI uses an “open spectrum” to visualize the wide variety of open outcomes in the marketplace. These outcomes vary along five dimensions: discoverability, accessibility, reusability, transparency and sustainability (DARTS). Many different combinations of these five dimensions are considered open, not just combinations that are “completely open” (see the DARTS figure in this report). Proponents of open can advocate improving read-only-access to data, read-and-reuse access to print materials, or making journal articles free to read or simpler “freer than now.” All support open science, but all refer to different kinds of open outcomes. Achieving these different outcomes involves different policy choices, so our words matter.

**Motives.** There are a variety of motives for trying to make science information more open. This variation affects the kinds of solutions we prioritize, and the degree to which our solutions are optimally incentivized and aligned with our community’s needs and goals. For some stakeholders in science, their primary motive for doing open science is centered around improving collaboration. For others, it involves increasing interdisciplinary work and discovery, or on improving reproducibility, transparency, accountability, access, or equity. Fecher & Friesike (2013) posited that “open science” consists of five broad motives, or “schools of thought” (as noted in this report).

**Goals.** Some actors in open science see open as a goal unto itself without identifying specifically what it will accomplish. Some see open as a pathway to achieving particular goals, including specific research and societal goals. And for some, open science is a catch-all phrase like STEM (discussed later in this section), used to describe a future that will cure many of science’s problems (e.g., reproducibility will increase, discovery will accelerate, society will gain more value from research, etc.; OSI subscribes to this latter philosophy, but we also support creating the foundation and policies needed to make this philosophy more than just words). Indeed, we see an a-la-carte uptake of open science drawing from four main categories of elements: organizing, evidence-based, prescriptive, and practice-based (as noted previously in this report).

Source: Hampson 2020b.

As if all this diversity wasn’t confusing enough, it’s also important to note that open science and high-integrity science are not necessarily synonymous. In our marketing hype for open we often claim that these two concepts are linked, but in truth, while they intersect, open science isn’t necessarily good science, nor is closed science necessarily bad (at least from an integrity perspective). The best practices for access, reproducibility and transparency espoused by open can, if properly applied, lead to better science, but simply thinking of open as an end goal (instead of a tool for producing better science) can also lead to journals and preprint servers that publish anything at all without regard to merit. Science as an honor system conflicts with pressures to publish frequently, and to maximize visibility and impact. And open science is also wholly insufficient by itself to address other integrity needs in science like peer review and proper research methods and analysis.<sup>65</sup> So, while the goal of open writ large should be to improve science, the reality is there are many reasons why authors choose open, and many things that open alone cannot do. It’s important to be realistic in our assessments.

This isn’t to say that we shouldn’t hype open. We just need to be honest about what we know and don’t know (yet), and utterly scientific in our approach to learning more and developing the best solutions for researchers. Take the open access citation advantage (OACA) as another example. Exactly how big this advantage is (or if it exists everywhere) has not been definitively

**[O]pen science isn’t necessarily good science, nor is closed science necessarily bad....The best practices for access, reproducibility and transparency espoused by open can, if properly applied, lead to better science, but these practices can also be foiled by bad gatekeeping mechanisms...**

65. This isn’t a particularly revelatory sentiment that requires footnoting—a number of articles have been published to this effect. The recent VI BRISPE conference, for instance, focused entirely on this theme (see VI BRISPE 2020). This sentiment does, however, not get mentioned often enough in debates about open.

established yet.<sup>66</sup> The hypothesis is that open journals will be easier to read and therefore cite, but the reality is that most highly cited articles still come from subscription journals (which corresponds with the fact that the overwhelming majority of the highest impact journals are still subscription based).<sup>67</sup> These realities act as sort of a persistent anti-incentive in the system against a faster transition to open. Still, the “underlying” hypothesis here that visibility helps is certainly sound. One fly in this ointment, though, may be that visibility alone appears to be insufficient. Preprints, for example, may (in some cases) lead to a visibility bump in the final published versions of articles due to the enhanced attention these articles get in the media and on Twitter (Klar 2020). And recent evidence suggests there is also a citation bump for SciHub articles (maybe SciHub articles are easier to find and access?).<sup>68</sup> But at the moment at least, based on the evidence we have so far, this citation bump doesn’t necessarily translate across all fields for all OA publication types.<sup>69</sup> If the evidence were clear and convincing, researchers would be flocking to OA. The facts appear to be more nuanced.

The same is true for other depictions of open benefits of open data—skeptical policy audiences need more data and evidence. For instance, the Research Data Alliance’s “The Data Harvest: How sharing research data can yield knowledge, jobs, and growth”<sup>70</sup> describes how open will benefit citizens, entrepreneurs and scientists in different ways, but this publication doesn’t describe the tensions these audiences feel. While it is true that open data could indeed result in “more accountable, efficient and effective businesses and government[s]” that are “empowered, [and] have the information they need to make decisions in all spheres of life,” it is also true that open data has resulted in a marked lack of privacy, the emergence of identity theft and a culture that expects free information despite the fact that this freedom has costs (think of how newspapers everywhere are collapsing, for example). Similarly, while no one disputes that “open data is a source of inspiration for entrepreneurs and provides the raw material for new products and services,” or that for scientists, open data “will make their work easier and faster, as more data and tools are put within reach,” the reality is that “open” is just part of the puzzle. How to interpret data is a huge and largely unmet challenge to date because raw data doesn’t come with a guide book that shows how it should be interpreted, integrated and reused, and this downstream use is fraught with peril for misunderstanding, misuse, and misappropriation. So, it’s important to be optimistic in our forecasts for open solutions but also realistic regarding the pros and cons of the path ahead.

The potential cost of not sharing is also sometimes noted as a reason to support open—products that don’t appear, jobs that aren’t created, medicines that don’t save lives, efficiencies in research that don’t happen and slower progress in understanding disease, and in addressing environmental and energy challenges. Here again, these are hypotheticals we need to approach with enthusiasm but also honesty and humility. It isn’t that we should believe less in the potential of open solutions, just that we need to keep grounded so we don’t end up pinning all our hopes on these solutions or pushing solutions that may create unintended consequences instead of benefits.

---

66. See Lewis 2018 for a good discussion of the shortcomings of OACA studies to date. This information corpus is changing rapidly though.

67. See Scimago’s journal impact rankings at <https://www.scimagojr.com/journalrank.php>. Subscription journals tend to have a higher impact factor than open journals for two primary (and related) reasons: because the majority of the subscription journals have been around for decades and because the publishers that publish them have been around for decades, or even over a century at this point, especially where learned societies are concerned. As a result of that longevity, both the journals and the publishers have had all that time to build a strong reputation, where the vast majority of open journals are less than 20 years old. Admittedly, there is still a subset of the research community that simply doesn’t trust that these new-fangled open journals are as high a quality as traditional journals, even when the empirical evidence may show otherwise. (Hampson 2020)

68. A recent study of SciHub citations—<https://arxiv.org/abs/2006.14979>—noted a measurable citation increase for SciHub downloads over paywalled versions of articles. While intriguing, it’s not clear whether this is an outburst of pent-up citation demand (for long-closed articles) or a true advantage. There is also pushback in the research and publishing communities about “whether sharing links to Sci-Hub could in itself be considered illegal”—see <https://bit.ly/3mp3PxV>.

69. Holmberg 2020 concludes that there is an open access advantage in some fields, and a disadvantage in others.

70. See RDA 2014. This is a critical passage and doesn’t reflect what good and important work RDA is doing in the open solutions space. RDA’s work is highlighted later in this report in a more complementary light.

The COVID era provides a good example of this dynamic. The need to share global health data has long been appreciated, and mechanisms for sharing this data effectively have evolved over time, in part through the efforts of the World Health Organization but also many other government and nonprofit actors. With the emergence of COVID-19 early in 2020, open science activists were quick to note how the large increase in preprint submissions was heralding a new and long overdue era of information sharing in science. Publishers also opened their archives to provide free access to researchers looking for information to inform their search for a COVID vaccine. While this is a good story and a good example of working together in a crisis, we also need to understand the details in context. For instance, preprint submissions only account for a small fraction (around 3%) of total research, and because of the lack of gatekeeping mechanisms for most preprint servers, science has experienced an “infodemic” of research.<sup>71</sup>

**[T]he majority of research is funded by corporations, not governments....If we end up creating a playing field between publicly- and privately-financed research that is too uneven with regard to sharing requirements, we may inadvertently create an environment where university researchers are second class citizens...**

This infodemic has had pluses and minuses—a rapid sharing of COVID data, but also a rapid sharing and publicizing of bad research, as well as misdirected research funds (as of June 2020, for example, the majority of COVID research was dealing with the discredited hydroxychloroquine cure being promoted by Donald Trump).<sup>72</sup> In addition, new research papers aren’t sharing granular data that has commercialization potential—proprietary information about vaccine formulas, for example, is not being shared. Many open advocates have also pointed to this burst in sharing as evidence of the way things should be in science, but here again, the reality is that our current peer review systems were not designed for this pace of work, publishers cannot maintain this pace of rapid publishing without imposing additional costs, the integrity of science is being jeopardized, and so on. So when it comes to promoting the benefits of openness to researchers, we need to see the good and the potential here, but also be honest about the downsides and risks and discuss all of this in a way that is more evidence driven than now.<sup>73</sup>

Finally, we should remember that globally the majority of research is funded by corporations, not governments.<sup>74</sup> While we can and should debate the imperative of opening data whose discovery is funded by the public purse, privately funded research faces no such questions.<sup>75</sup> In business, research

---

71. “We’re not just fighting an epidemic; we’re fighting an infodemic,” said Tedros Adhanom Ghebreyesus, Director-General of the World Health Organization (WHO) at a gathering of foreign policy and security experts in Munich, Germany, in mid-February 2020, referring to fake news that “spreads faster and more easily than this virus.” See <https://bit.ly/3ftaHHP>.

72. One of the big risks is that on this topic the preprint space has exploded (relatively speaking, anyway; preprints still only account for a small fraction of total published research) with highly politicized research that can support pretty much any ideological stance conceivable on this issue. For experts, most of this literature is dismissable due to its frequent severe technical and other flaws, but unfortunately this non-gatekeepered research is also being picked up by the media, hyped, and in extreme cases being used to influence policy decisions. There has been tremendous benefit to quickly sharing information during this pandemic, but also critical weaknesses in the system are being exposed.

73. See Cai 2020 for a fuller exploration of publishing dynamics during the COVID crisis this year. Of particular interest here is that formally published articles have far outpaced preprints, even though preprints have received more than their fair share of media buzz. Of the 87,500 articles published so far in 2020 on COVID, two-thirds (58,000) were not preprints. Also of interest is the international profile of this work: 92% of the researchers affiliated with formally published work have come from Europe (41% including the UK), the US (33%) or China (18%). So at least in this particular case, open as a gateway to massive international collaboration has been the dream, but the reality is that countries spending the most on research publish the most research papers (and this being the case, it’s a bit surprising that China’s numbers haven’t been higher since they outspend the US; is this because more of China’s researchers have published their COVID-19 work in preprint format?).

74. In the OECD area, this figure is around 60 percent (see OECD 2015). In the US, it’s around 75 percent (NSF 2020).

75. This dichotomy is broad and doesn’t adequately consider the vast amount of research that is done in partnership between research institutions and industry, and the very important role of the research institution in framing expectations for the work conducted under their auspices. The point here is to be wary of the research classes we may be creating.

can be shielded from public view until funders have squeezed every last ounce of value from it (and even then they are under no obligation to share; knowledge is a commodity, and there's no motivation to share it unless it's to gain from it somehow—sell, attract new customers, etc.).<sup>76</sup>

So, we need to be careful what we wish for, lest we inadvertently create an environment where the intellectual property expectations for universities are completely inferior to those enjoyed by industry. In such a case, open policies could end up being at odds with legislation everywhere that has long focused on trying to maximize the return on government investment in university research (in the US, through the Bayh-Dole Act). Confusing this focus will set off policy fights everywhere, and might also conceivably incentivize researchers to sit on their data even longer than now to make sure they're completely finished using it before sharing; seek exceptions to publishing mandates (which already happens in significant numbers; see Research England 2018 for a recent example); or turn down research funding that comes with the wrong kinds of publishing strings attached.<sup>77</sup>

All of these considerations and more are woven into the fabric of researcher concerns with regard to open. In order to build greater participation and trust from the research community, it is imperative that we understand and appreciate the full spectrum of this community's needs, ideas and concerns, and deal scientifically and responsibly with assuring that our efforts to build and promote new open solutions and policies are aligned with the best interest of researchers and include ample input from this vital community.

## Publishers

Publishers were instrumental to the very founding of science and remain an important part of science today. Without a way to record and share discovery, the scientific revolution would not have taken root in the mid-1600s. Today, as Keith Yamamoto said in his closing remarks to OSI2017 conference delegates (see OSI 2017a), “If you don't publish your experiment, it is exactly like not doing it.” But what is publishing? Two blue-ribbon panels at OSI's 2016 conference dug into this question, thinking through how we define publishing today, where it's heading, and what it might look like in the future. A summary from one of these panels is contained in the box on the next page.

The short explanation is that there are many different kinds of publishers and publishing services in the research ecosystem. These publishers and services have a lot in common, but there are also important differences—for-profit, nonprofit, large, small, university based, scholarly society based, discipline specific, broad, subscription, open, and so on. Each of these different entities focuses on the same general end goals, but often in service to different audiences, and with different priorities, interests and capabilities.

Perhaps the most important distinction between these different publishing entities is that approximately half of the 3 million journal articles published every year—as well half of all scholarly journals, articles and citations (Larivière 2015) and the majority of all open research articles (Piwowar 2018, Rodrigues

**[E]very stakeholder plays an important role in the open knowledge ecosystem and that there can be no real and sustainable improvements to this system without everyone's involvement.**

---

76. Granted, corporations then own the rights as opposed to researchers—“work for hire” laws apply in this case where they do not in the case of university researchers working with grant funds.

77. Of course, turning down funding is only an option in large wealthy countries where there may be multiple sources of funding available. In many countries there is typically only one major funder, so these researchers will not have much choice if they want to get funded. For a broader discussion of intellectual property licensing trends at universities, see Caviggioli 2020, and also <http://wapo.st/357R8AE>, and <http://bit.ly/38fIMu1>.

**TABLE 5: WHAT IS PUBLISHING?**

Henry Oldenburg, who founded *Philosophical Transactions* in 1665, defined the four functions of scientific publishing as registration, certification, dissemination, and archiving. Registration (establishing who first made a discovery) and certification (review of the findings by one's peers) remain fundamental functions of scholarly publishing today. In fast-moving areas of science, in which competition for grants is fierce, days matter, and the pressure on publishers to publish quickly, to establish a researcher's precedence, is intense. In certification, we are seeing many experiments in post-publication review and commentary, but pre-publication peer review remains the norm in the vast majority of disciplines. Dissemination remains a critical function. In a digital environment, discovery requires high-quality, reliable metadata and new tools and services such as CrossRef, Open Funder Registry, XML-coding, semantic indexing, and indexing by search services. Archiving is a cooperative undertaking, between publishers, libraries, and third-party services.

Oldenburg might not recognize many of the digital elements of today's scholarly publishing, and would likely be astonished at the scale, but he would probably still recognize the two principal forms of output of scholarly research, the journal article and the monograph. In each case, publication is essentially defined by a single final output, which is largely still text-based and consumed in printed, or PDF, form. Also, even though open access has changed how we think of publishing today it has not changed the two core functions of registration and certification. Traditional and open access publishers alike develop and curate journals and monographs for specific scholarly communities, manage peer review, produce and index articles/monographs and their metadata, and distribute them digitally. While open access has seen the launch of a small number of so-called 'mega-journals,' many researchers choose to publish in specialist journals which directly address their communities, and driven by systems of academic recognition and reward, prefer to publish with the strongest brand that will accept their submission.

While journal publishing is now largely a digital, and digital first, enterprise, even if the final output would be recognized by Oldenburg, monograph publishing is still to a large extent a 'print first, digital second' undertaking, with only a few publishers experimenting with open access models. All too often the digital book is little more than a PDF of the print edition, with the occasional ancillary supplement of data and multimedia. We are beginning to see some publishers, especially in the sciences, producing digital monographs which take advantage of the potential of the format—full html, embedded multimedia, interactive charting, and so on—and some university presses are experimenting with open access and digital monographs, but these experiments should be conducted far more widely, and across more areas in the arts, humanities and social sciences.

### **Publishing in the future**

Scholarly research is an increasingly diverse, complex, and interdisciplinary endeavor, with growing importance assigned to incremental stages of evidence and argument. We therefore envision a future publishing paradigm that is networked, open, and significantly more dynamic than the traditional model. Emerging signs point to a shift from the current "event-driven" model that focuses primarily on the publication of a print (or PDF) article or monograph to an ongoing process-driven, digital model that reflects more of the research lifecycle through a scholarly record that is comprised of much more than text. The boundaries of what constitutes a book or article are becoming blurred, and we are increasingly recognizing the value of other content. Over time, we anticipate that publishing will regularly encompass a richer and more interconnected range of scholarly content, including data, lab and field notes, software, preprints, social media posts, video—performative and experiential—and multimedia, with possibilities for interactivity among the various elements. We will likely see additional forms of content and new methodologies introduced through the use of virtual reality, gamification, and other innovative technologies— all of which are intended to help construct and enhance meaning and have the potential to make scholarship more relevant to society. By sharing, collectively assessing, and interconnecting this diversity of scholarly output, the research community will recognize these results as valued, first-class objects of research. The ability to continually access, assess, and interact with these many assets will be key markers of successful research as well as successful publishing.

As more of the research lifecycle is shared, we anticipate that researchers will experiment widely with formal and informal channels of disseminating their scholarly results. Through this experimentation, researchers and others will learn what is most valued and needed from the established publishing process versus what is best suited for alternative methods. In contrast to today's practices, tomorrow's myriad scholarly artifacts are likely to be released incrementally through

Source: The above text is, with a few minor changes, a verbatim copy of the report from the second OSI2016 workgroup which explored the question of what is publishing (see OSI 2016a). Two such reports were written, both worthy of sharing. To read the second report or other reports like this, see the OSI website at [osiglobal.org](http://osiglobal.org).

diverse global systems. They may represent smaller (or larger) unbundled components, distributed at a much more timely pace with varying degrees of openness. Linkages among the discrete parts would be established in standardized ways to retain, expose, and build on the inherent intellectual relationships. Assessment and certification of all elements would be considered essential, and would go well beyond today's more limited approach to peer-review. Disciplinary experts and data curators, for example, would be collaboratively assess data in order to ensure methods are appropriate, uncertainties are well described, documentation is complete, and standards are followed. Software assessment would include expert reviews as well as test-driven methods such as continuous integration. There would be mechanisms for ongoing community assessment of all published artifacts, and the major assessment criteria (beyond those of originality and research merit) would be reproducibility and reusability. The published "book" or "article" would not be a final event, but part of an ongoing scholarly conversation.

This model assumes a general trend toward openness, with restrictions based on ethical norms rather than proprietary considerations, as they are currently defined. Given the differences in disciplines, we would expect that the timing of openness might vary widely among different communities. At present, humanists are much less inclined to share their in-progress work, whereas many scientists are accustomed to working in teams and distributing repetitive drafts for quick review and feedback. In addition, there may be variations in the degree of openness with which scholarship is released. For example, a virtual lab notebook may initially be shared only among a small team of researchers and curators and later released to others in the research community. Even then, the entire notebook might not be fully open, but instead, be released selectively, on as-needed basis, to provide evidence or provenance for another publication. The timing of release may also be dictated by intellectual property concerns, such as patent application filing.

This dynamic and diverse publishing ecosystem will involve changing roles for today's key partners in the world of publishing, including researchers, publishers, universities, and funders. In addition, new partners will be required if we are to successfully deliver, disseminate, and preserve future research. Data curators, data scientists, software developers, and designers would work with researchers, librarians, editors, and publishers to develop a network of interconnected scholarly work supported by a variety of institutions including: data repositories, research labs, libraries, publishers, network providers, standards organizations, software providers, and professional societies from public, private, as well as nonprofit sectors. These new, shared roles and the resulting scholarly output will require different types of reward and recognition that go well beyond the traditional measures of scholarly impact.

Finally, it is important to note that the future of publishing is likely to be a mix of today's formal and informal methods of scholarly dissemination. Questions remain whether these two approaches will continue to develop in parallel; become combined, complementary systems that incorporate the best of both models; or evolve into entirely new alternatives. Regardless, it seems clear that the open, global network will increasingly be the primary means and method of distributing scholarly communication. The network will also serve as the dominant forum in which scholars interact with their peers, and will increasingly become the preeminent platform for building audiences (and collaborators) that extend scholarship beyond academia.

## How to get there

Conversations about where publishing is today, and what it might become in the future, inevitably turn to the purpose of scholarly works. At present, these outputs form a large part of how scholars and institutions are evaluated. That means any significant change will require thought and action on the part of those who conduct this evaluation—funding agencies, tenure committees, scholarly societies, accrediting bodies, and others. While no one solution or player will effect systemic change, each can contribute with its own experiments. Instead of rewarding only the peer-reviewed and published paper, evaluators may assign value to post-publication contributions by others, data sharing, collaboration, and other important steps in the scientific process. In short, we want evaluation metrics that reflect how we want science to work.

Funders of scholarly activity, particularly scientific research, are eager to support innovative work that pushes the boundaries of basic knowledge and maximizes translational possibilities. In the past decade, initiatives such as NIH's Pioneer Awards, the HHMI Investigator Program, Wellcome Investigator Awards, and the joint HHMI/Gates/Simons Faculty Scholar Program have encouraged the unfettered imagination of early career investigators, the spark of cross-disciplinary collaboration, or the risk-taking of established scientists willing to abandon the familiar for untried approaches. During the same period, compelled to evaluate the impact of the supported work on scholars and the general public, funders discovered systems of publication out of step with their goals, characterized by restricted access to output, restrictions on redistribution, peer review ill-disposed to novelty, and community journals closed to cross-disciplinary work. And they found few means of measuring the consequences of their funding decisions. Their responses, from open access man-

dates and the funding of new breeds of journals, to the instigation of new metrics, are at the heart of the present turmoil in scholarly communication and their voices will surely continue to be heard.

Collecting, evaluating, and disseminating a wide range of research artifacts will affect scholars, librarians, and publishers in a variety of ways. Scholars will have to adapt to changing expectations from funders regarding distribution of scholarship resulting from the funding. If universities change promotion and tenure evaluation requirements, scholars at those universities will shift their publishing and dissemination practices accordingly. It is logical that all stakeholders will focus their efforts on what is rewarded. As publishing becomes a more continual process, as described above, scholars will be expected to follow standards and provide complete documentation, including ensuring that uncertainties are well described for data sets and other works that are not currently a routine part of published research results. Documenting and communicating the related outputs of the research process will also fall to scholars, just as we have citation norms in the current system.

Librarians have already expanded their role from assisting scholars with the research process, to assisting scholars with research outputs. For example, librarians help scholars comply with funder requirements, such as the NIH Public Access Policy, and this role will expand with new funder requirements. Librarians also continue to play a role in assisting scholars organize, apply metadata, and preserve their research outputs, such as data sets. The collection of these materials will need to be standardized. The growth of digital humanities centers in libraries has brought new tools to bear on the research process, with librarians serving as guides to scholars on applying the most appropriate technology to their research question. On a broad scale, services like the Research Center of HathiTrust, which provides text and data mining across the corpus of works in the HathiTrust digital archive, will continue to leverage library collections and librarian expertise to provide new research opportunities for scholars. These new roles will require new skills and new ways of collaborating with scholars and third parties at other universities or in the private sector.

Publishers also assist authors with meeting funder requirements, such as depositing manuscripts in PubMed Central. Publishers continue to play a vital role in promoting standards and best practices that facilitate the ecosystem of available research results. In addition, publishers will need to establish new products and services, partner with new service providers, and invest in talent and training as the industry shifts away from the manuscript-centered, broadcast approach and toward an approach that reflects the way users both consume and interact with Web content and engage with other users. All of this will require visionary proactivity, continuous user monitoring and analysis, and swift reactions when the unexpected inevitably emerges.

The new, dynamic publishing ecosystem we envision will require new types of business models and expanded partnerships and collaborations to succeed and be sustainable. It is important to note that this new ecosystem relies on a complex infrastructure of many different organizations, technologies, and professionals. When infrastructure works well it is transparent and often taken for granted. As such, it is often difficult to garnish the attention and funding necessary for ongoing support. As we see with the evolution of traditional infrastructures like the power grid, many different funding models will need to work together across multiple scales in a decentralized network. Governments will continue to play a foundational role in establishing appropriate regulatory frameworks and support of public goods, but private funders, universities, commercial and nonprofit labs, technology companies, scholarly societies, libraries, and publishers will also play critical roles. Governments and institutions need to take a lead in establishing new public-private partnerships, private finance initiatives, and local-level partnerships that coordinate and sustain necessary publishing services. Again, specific approaches are likely to vary across disciplines.

Ensuring that the needs of the researcher remain at the heart of these multi-stakeholder partnerships will be fundamental to the success of this future vision of publishing. Researchers are increasingly confronting time and resource challenges that distract from the core mission of discovery and scientific advancement. In response, governments, institutions, and the private sector must work jointly to establish a friction-free environment for researchers to share scholarship. This includes both infrastructure and policy elements. Infrastructure solutions require a seamless user experience based on platforms and tools that are interoperable and in constant communication. Policy solutions require coordinated efforts that will drive and incentivize behavioral change, including short-term efforts to recognize and measure individual contributions in the research process and the longer-term systemic changes to academic evaluation mentioned previously. The future of publishing is enabling researchers to accelerate pre- and post-publication discovery and increasing the discoverability and impact of their scholarly work.

Our team identified several recommendations to inform future OSI efforts. The three primary recommendations focus on aspects of open scholarship and open access publishing that require more data, evidence, and discussion.

**[T]he correct approach to an open solutions future is to focus less on excluding certain parties from participating or on legislating specific ways of doing business, and more on figuring out what we want to achieve with open and building the capabilities of open solutions.**

2020)—come from just a handful of commercial publishers. Consequently, most of the concerns from policy makers and advocates about the future of open research have been directed at these companies (primarily Elsevier, but also Springer/Nature, Taylor & Francis, Wiley and Sage). Concerns expressed as criticism have been fueled in large part by the narrative that the profit margins of these publishers are too large.<sup>78</sup> This depiction is omnipresent in the open access space—open access marketing literature and open access scholarship are both filled with this profiteering charge. For whatever reason, these publishers haven't pushed back publicly.<sup>79</sup> After decades of pressure from open access lobbying groups (and now, as mentioned later in this report, major funders as well) the relationships between several key stakeholder groups in academia have been strained. It is very common to hear from activists and government policy makers alike that profit-seeking has no place in the scholarly communication ecosystem and should be removed from it—that any acceptable solution to the future of open needs to be built around a different set of incentives.

The frustrations are completely understandable, of course. Globally, costs have long been a major barrier to access, and the budgets of even the richest universities have been strained by the costs subscribing to a growing number of journals. Today, as the world pivots from paying subscription costs to paying APCs, this strain is going to continue to be felt (if not by universities then by funders who will be called upon to pay these charges).

Still, the reaction of universities and funders to the cost of publishing is somewhat untethered from the reality that these institutions also pay for a myriad other services related to research; from the fact that the volume of research being conducted and published has expanded dramatically over the last 20-40 years (see again sources like Piwowar 2018); and from the fact, demonstrated repeatedly by author surveys over time, that researchers want and depend on a high-functioning publishing apparatus. For every critic who claims the scholarly publishing is broken, there are many supporters in research who think the scholarly publishing system is serving their needs just fine. The argument over library budgets and research funding allowances has spilled over into a critique on how information is created and the important role—active and historic—that publishers have had and continue to have in vetting, editing, cataloging and sharing new knowledge. To the extent this debate can help us improve the publishing process, then it can be productive, but too often the debate more closely resembles an attempt to unilaterally reform publishing without regard for how this will impact everything else in the research ecosystem.

As untenable as the cost barriers are, OSI has taken the position that every stakeholder plays an important role in the open knowledge ecosystem and that there can be no real and sustainable improvements to this system without everyone's involvement. This position is based partly in reality—the major

---

78. The library serials crisis of 2015-16 was a major flashpoint in this debate, fueled by rising journal costs and declining library budgets.

79. Privately, Elsevier officials explain that their balance sheet is not being read accurately—that their actual net profit is closer to eight percent. Elsevier is part of RELX, and therefore its balance sheet can't be accurately evaluated without knowing exactly how much revenue and expense is "hidden." They also note that costs are being driven by far more research being conducted and published today than in years past, but that the cost per unit published is lower today; that researchers like the current publishing system and don't want to throw it away in favor of something new; and that publishers are being singled out for criticism in an ecosystem where high cost is an issue on many fronts (from room and board charges to campus overhead costs; see College Board 2019). The fact that some major publishers are now charging extra for an open access license on top of the APC fees (ACS does this, for example)—a practice that isn't all that surprising given that these publishers are making a calculation of "lost" future income from downloads—only adds to this air of distrust toward publishers.

commercial publishers are providing a valuable service to research and are responsible for publishing the majority of research articles—and partly in the realization that open is going to cost something. The question is how much. To date, the answer is “at least as much as the current system.” Policy recommendations that suggest we will somehow be able to develop a world of research communication where there are no costs, and/or where credible and usable information flows directly from researchers to the public are unrealistic. The approach that profits of any kind are a deal breaker simply lacks business sense. Whatever entities we create to manage information flows in research will need to have budgets and personnel, and will need to be responsive to the needs of customers and strive to make continual improvement.

History has demonstrated time and again that markets are the most effective way to achieve these outcomes—not government control but systems that serve the public good while also retaining attractive investment incentives. We can’t and shouldn’t try to legislate profit from the open solutions ecosystem—nor can we legislate away competition and metrics—but we can work to create systems that better serve our needs and provide a better return on our investments.

There is a concern at the moment that major publishers will consolidate their position and end up controlling too much of the future market of scholarly communication (“vendor lock-in”), to which the answer is to encourage more competition in the marketplace (e.g., encourage more profit-making, not demand that philanthropy-funded non-profit solutions are the only welcome market entrants). If the demand for open products and services increases substantially then providers of all kinds will compete to serve this demand and create products that serve the open environment well. If providers don’t see this demand but instead only respond to individual funder mandates for open, or settle one-off disputes with library systems with PAR agreements,<sup>80</sup> then publishers will not be engaged to their full potential and the open solutions we develop will be just good enough to satisfy our bare minimum requirements.

Given this, the correct approach to an open solutions future is to focus less on excluding certain parties from participating or on legislating specific ways of doing business, and more on figuring out what we want to achieve with open and building the capabilities of open solutions. By making open solutions a welcoming, vibrant and effective environment, with tools that can exploit the potential of open and be used to tackle big and pressing issues, we then make open a clear choice, and researchers (and societies) will expand their participation in open because it’s in their best interest to do so. The publishing community can help lead the charge toward this future by developing products and services that simply cannot be invented from thin air by open activists, and sustained in a way that ensures these solutions will be effective over the long term.

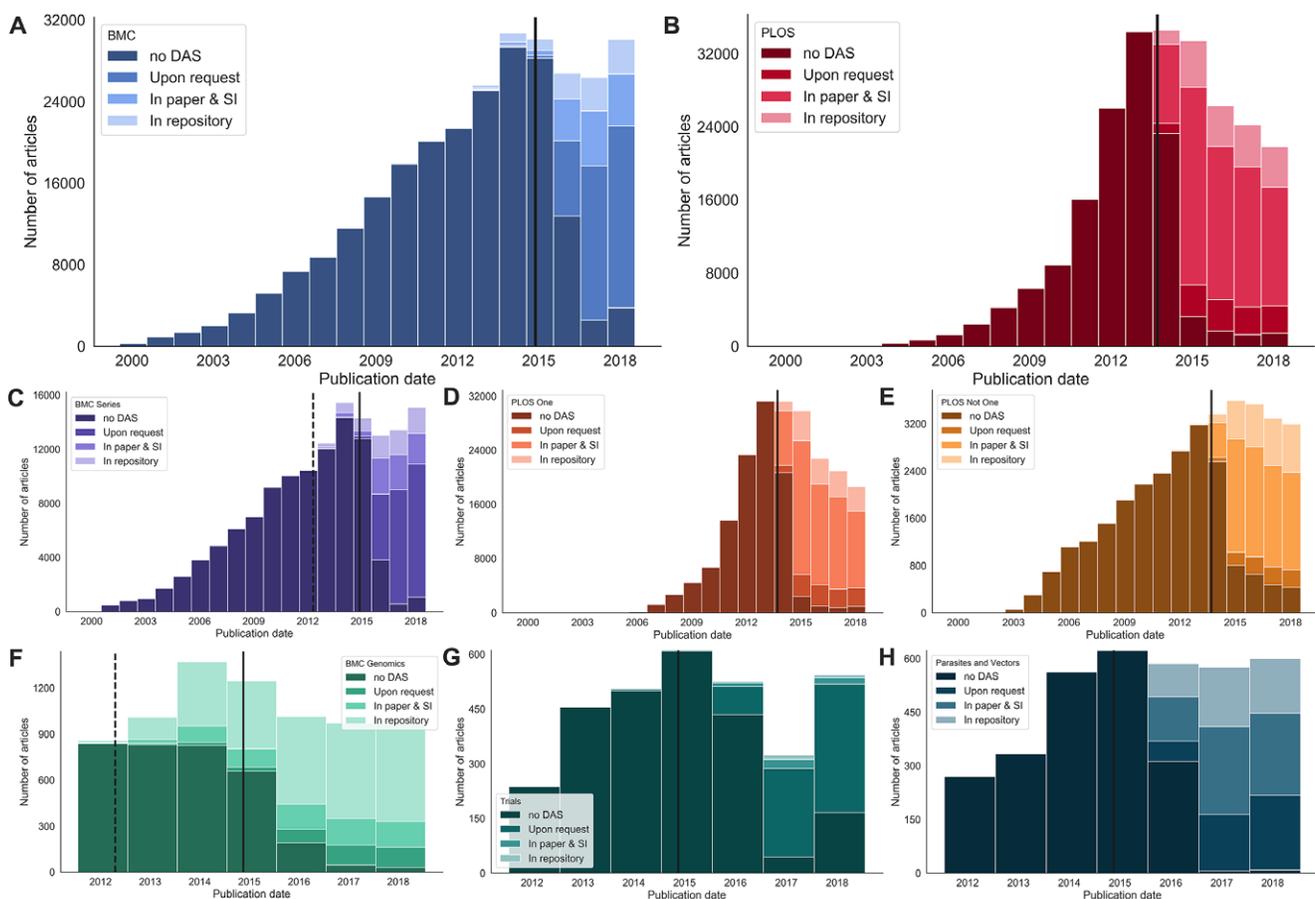
This public-private tension exists to a much lesser degree in open data—there is plenty of debate about who should control what and about what kind of data should be publicly available—but no similar debate about whether publishers should have a role in data management. In open source, open government and OER, there are also no comparable public-private tensions. The open source community, for example, regularly debates issues of vertical applications—how software will be used in sectors like health care and higher education (as opposed to horizontal applications, which have to do with operating systems, web servers and the like). But as mentioned in the open source section, this field of open has a very robust private-public relationship.

In the areas of open access and open data, we are learning from publisher experience that open practices are adopted most rapidly when policy and implementation plans align (White 2020). There is also strong evidence that well-implemented data sharing policies enable and drive change in data sharing

---

80. These types of agreements (PAR or RAP) generally stipulate that for a set price, universities will not only get access to a publisher’s journals, but that university researchers will be able to publish in these journals without charge.

FIGURE 11: IMPACT OF JOURNAL DATA SHARING POLICIES ON DATA SHARING



All the histograms above show the number of publications from specific subsets of the dataset and classify them into four categories: No DAS (0), Category 1 (data available on request), Category 2 (data contained within the article and supplementary materials), and Category 3 (a link to archived data in a public repository). The vertical solid line shows the date that the publisher introduced a mandated DAS policy. A dashed line indicates the date an encouraged policy was introduced. The groups of articles are as follows. A: all BMC articles, B: all PLOS articles, C: all BMC Series articles, D: PLOS One articles, E: PLOS articles not published in PLOS One, F: articles from the BMC Genomics journal (selected to illustrate a journal that had high uptake of an encouraged policy), G: articles from the Trials journal (published by BMC, selected to illustrate a journal that has a very high percentage of data that can only be made available by request to the authors), H: articles from the Parasites and Vectors journal (selected to illustrate a journal that has an even distribution of the three DAS categories). Articles are binned by publication year.

Source: Colavizza 2020

behavior and that this data sharing may also increase citation impact (Graf 2019, Colavizza 2020; see Figure 11 for details).

Industry wide, all major publishers signed an initiative to make 2020 the “STM Research Data Year” (see STM 2020 for details), drawing on evidence-led, community driven outputs from the Research Data Alliance, Center for Open Science, FORCE11 and other groups to work with publishers and other partners to boost effective sharing of research data by:

- **Sharing:** Increasing the number of journals with data policies and articles with data availability statements;
- **Linking:** Increasing the number of journals that deposit data links to SCHOLIX; and

- **Citing:** Increasing citations to datasets in accordance with the Joint Declaration of Data Citation Principles.

To help improve outcomes in the publishing space, publishers are working together through initiatives like this (and also position statements like in STM 2016<sup>81</sup>) and through efforts like FAIRsharing.org’s “Data Repository Selection: Criteria That Matter” project, which was inspired by a discussion between the Force11 Data Citation Implementation Pilot (DCIP) group and the joint Force11 and Research Data Alliance (RDA) FAIRsharing Working Group on the need to develop a shared list of recommended data deposition repositories.<sup>82</sup>

Publishers have also actively responded to the research community’s demand for data citation principles, and a number of publishers have endorsed the Joint Declaration of Data Citation Principles to ensure researchers get proper credit for their work.<sup>83</sup> In the policy framework for all journal publishers, developed by an RDA Interest Group with contributions from publishers and other stakeholders, publishers are also working to improve:<sup>84</sup>

- **Citizen science:** Publishers are developing new ways of engaging both researchers and the wider community in science. Some examples include the publication of lay summaries of articles analyzing important research (e.g., through programs such as PatientInform), and the opening up of scholarly communication to wider public audiences (e.g., through the Public Library Initiative in the UK);
- **Infrastructure:** One focal point is to expand the use of Digital Object Identifiers (DOIs) and other persistent identifiers for all research outputs and objects (like process identifiers—PIDs); and
- **Research integrity:** The outcomes needed for open science include not only improving visibility and impact through open measures, but integrity as well.<sup>85</sup>

In terms of existing open policies, the open data policies of major publishers can be grouped into six main tiers (Hrynaszkiewicz 2020), highlighting the wide range of approaches currently being employed with regard to key data considerations. This variation has probably contributed to confusion and uncertainty about open data, but taken as a whole, there is wide variability in this space and the general trend is toward more robust and more effective open data policies from publishers.

Going forward, specific actions that publishers can implement in the open solutions ecosystem are explored in a 2020 NIH presentation by Varsha Khodiary, titled “What Role Can Publishers Play in the Open Data Ecosystem” (Khodiary 2020). Some of the ideas mentioned in this presentation include inventing new ways to identify data objects, and building single entry-point solutions and systems that can help ensure there’s one name for institutions, authors, book titles, and so on. It is worth noting here that broad solutions like the All-Scholarship Repository solution (ASR), mentioned in places throughout this report (but not yet considered a serious policy idea), address concerns like these by creating a single repository for all data.

81. [https://www.stm-assoc.org/2016\\_10\\_31\\_STM\\_Open\\_Science\\_Position.pdf](https://www.stm-assoc.org/2016_10_31_STM_Open_Science_Position.pdf)

82. FAIRsharing.org is leading a group of publishers to define the characteristics associated with data sharing repositories that really make a difference for publishers (the benefits of which publishers can pass on to researchers). Draft criteria are published on the Open Science Framework. See <https://doi.org/10.17605/OSF.IO/N9QJ7>

83. <https://doi.org/10.25490/a97f-egyk>

84. <https://doi.org/10.6084/m9.figshare.8223365.v1>

85. See <https://www.wiley.com/network/researchers/blogs/doing-whats-right-trust-and-accountability-in-peer-review>

TABLE 6: RESEARCH DATA POLICY FEATURES BY TIER (1-6)

	Policy 01	Policy 02	Policy 03	Policy 04	Policy 05	Policy 06
Definition of the research data	<input type="radio"/>					
Exceptions to policy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Embargoes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Supplementary materials	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data repositories	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data citation	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data licensing	<input type="radio"/>					
Researcher/ author support	<input checked="" type="radio"/>					
Data availability statements		<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Data formats and standards				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Mandatory data sharing (specific data types)				<input checked="" type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Mandatory data sharing (all papers)				<input type="radio"/>	<input checked="" type="radio"/>	<input checked="" type="radio"/>
Peer review of data				<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
Data Management Plans (DMPs)				<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

<p><input type="radio"/> Provide information</p> <p>The text for the policy feature will be included in the policy template but it is clear that the feature will not be enforced and checked as part of the publishing or peer review process</p>	<p><input checked="" type="radio"/> Provide information and action</p> <p>The text of the policy feature is included and makes clear where applicable that the feature will be checked and enforced in the publishing or peer-review process</p>
--	--

Source: Hrynaszkiewicz 2020

There is much more to discuss here, but a common refrain in this paper is that going into more detail is beyond the scope of this paper. At this juncture, the main takeaway is that publishers are an important part of the knowledge ecosystem. We will bear this in mind later when discussing what kind of policies to develop.

### Research institutions

It's an odd fact of science communication that non-university researchers conduct far more research than university-based researchers, but publish far less.<sup>86</sup> Therefore, when we speak of the needs of

86. Here, we use the word “publish” in the academic journal article sense, not articles that are less likely to be findable, accessible, reviewed and cited, such as internal studies or white papers. Part of this dichotomy is explained by the fact that university-based researchers conduct more basic research, while non-university researchers conduct more applied research. Basic research is generally seen as a public good and is more likely to be government-funded than applied research. Basic research can therefore be practiced with maximum openness, whereas applied research is generally more monetizable. Also,

“research institutions” in the scholarly communication ecosystem, we need to recognize the perspectives not only of universities but of government-owned research centers, non-university research institutions like Cold Springs Harbor Laboratory, the Max Planck Institute, the Fred Hutchinson Cancer Research Center, and a myriad industry-based research centers, pharmaceutical companies, defense companies, technology companies and more. Because of the outsized contribution of university-based researchers to science literature and the tight relationship between publishing and academic tenure, we tend to focus our concerns only on university research, but this isn’t sufficient. In fact, non-university research institutions also publish a lot, and some even have highly influential open publishing programs (like the Max Planck Digital Library).

**Because of the outsized contribution of university-based researchers to science literature and the tight relationship between publishing and academic tenure, we tend to focus our concerns only on university research, but this isn’t sufficient.**

Also at the university level, there are a wide variety of perspectives about open solutions. First, not all are major research centers—not all have the prodigious scholarly output of the University of California system, the research reputation of Harvard, or the research budget of Johns Hopkins. Most embrace research as part of their academic environment, but they aren’t necessarily driven by the demands of or are heavily dependent on the revenue from the research grant pipeline.

**[T]he debate at these institutions over open solutions—particularly open access—has been mostly managed by university libraries to date, and library interests don’t always align with how university provosts see the world.**

Second, for the most part the debate at these institutions over open solutions—particularly open access—has been mostly managed by university libraries to date, and library interests don’t always align with how university provosts see the world (Murray 2018).<sup>87</sup> Also, like universities themselves, libraries come in many different shapes and sizes. Most major US and EU university libraries share similar concerns about the cost of access to research journals and are actively looking for solutions. But there are also many university libraries outside the US and EU, or smaller libraries that are not housed in major research institutions, who do not feel their interests are well being represented by the philosophies and solutions of major US-based university libraries. These are institutions who don’t have the buying power to negotiate customized publish-and-read agreements with publishers, or don’t have the financial resources to afford solutions that rely on authors to pay for journal publishing costs out of their own budgets.<sup>88</sup>

---

this statistic shouldn’t be confused with impact, where some studies have demonstrated the vitally important role the private sector has had in discovery in fields like pharmaceuticals. For more detail about the relative R&D budgets of the private and government sectors, see the US Science and Engineering Indicators website (NSF 2020) and also OECD 2015. For more detail about (for example) the institutional affiliation of authors who published articles in the journal Nature during 2018, see <https://www.natureindex.com/annual-tables/2019/institution/all/nature-science>.

87. In addition to the survey data reported in Murray 2018, the disconnect between provost and library perspectives has also been reported to OSI through private channels with regard to the need for libraries to adopt open solutions that are grounded in the needs and concerns of the research community instead of ideological arguments. Unfortunately, there doesn’t appear to be any published corroboration of these reports. It is worth noting that in almost all of the cases where OSI reached out to universities to assign representatives to work with OSI on the issue of open, provost offices appointed their library heads to this role, even though their perspectives may differ. Also, it is important to note that there have been several high-profile cases of alignment between provosts and libraries (at Harvard, MIT, and UCLA, for example); concurrently, much discussion has also been happening at the library level about how to best represent and meet the information needs and interests of university researchers in the future (see <https://sr.ithaka.org/publications/its-not-what-libraries-hold-its-who-libraries-serve/> for example).

88. In the EU and US, author publishing charges are most often covered by institutional budgets. However, in the developing

And third, as was mentioned previously in this report (in the section on researcher perspectives), we should be wary about the prospect of creating different classes of rights and requirements for different classes of researchers based on where they work. We're already doing this to a large extent due to large differences in open solutions between fields, regions and institutions. But if we truly want to create an open solutions world where policies work for everyone everywhere, we should be mindful of this distinction and work instead to create open solutions that are portable across institutional and geographic boundaries. This necessarily means that we need to stop thinking of "institutions" as primarily universities and libraries, but instead think much more broadly and inclusively.

## Scholarly societies

Scholarly societies are major stakeholders in the scholarly communication ecosystem, representing and serving the needs of researchers. Most societies rely on meetings, publications and member fees to fund their activities. With regard to their publication revenues, some of the more prominent open reform plans of late would outlaw the type of publication societies primarily rely upon—so-called "hybrid" open, where some articles in a journal are free to read and reuse and other articles are available only to subscribers.

While all societies have been generally supportive of open efforts, they are also generally (with a few exceptions) not at all supportive of open requirements that impact their bottom lines.<sup>89</sup> In order to survive in the landscape being created by Plan S, for instance, some societies will need to (and already have) abandon their in-house publishing capacity and subcontract this out to major commercial publishers (which, ironically, increases the very power of commercial publishers that Plan S wants to keep in check), move to APC gold versions of their most desirable journals (shifting the cost of these publications from subscribers to the society's research base), or negotiate separate open access deals with individual customers.<sup>90</sup> Here, the need exists to engage societies, but in a way that isn't as threatening and inflexible as now.

## Funding agencies

According to a recent Wiley survey of authors (Wiley 2019a), about 41% of funders are currently encouraging (23%) or requiring (18%) that authors make their articles openly available. Exactly what this means in practice varies widely, however. While CC-BY licensing is the norm in such cases, there are also broad allowances for authors who can show that publishing in a subscription journal is their best choice; also, none of these funder provisions require that the final version of record be published in an open format, only the author's accepted manuscript.<sup>91</sup>

**[M]any of the major nonprofit funders in the open space have aligned themselves with the ideological position that open means something very specific.... Government funders haven't been as particular about ideology..**

---

world, most institutions don't have this kind of budget and most authors are left to pay for APC's out of their own pockets (Scaria 2018). This flipping of access barriers in science world from subscriptions to author charges—from paywalls to playwalls—has consequences for researchers in the developing world: If they can't afford to publish, their research will simply go unnoticed.

89. See McNutt 2019 for an example of this sentiment.

90. For example, ACM signed a transformative agreement with four major research universities in January 2020 (<https://osc.universityofcalifornia.edu/2020/01/acm-open/>). On its face this is a good development for open. But to the extent it portends the need for other scholarly societies to also negotiate such agreements—at least with their largest customers—in order to simply remain competitive, the potential for suboptimal outcomes here is real (beginning with the administrative burden for all parties involved). Some have even lamented the rise of transformative agreements as unintentionally making the open landscape worse off by creating a world of locked-in contracts that primarily benefit large, wealthy institutions. See <https://bit.ly/3eUTxSV>.

91. This is a complex process, but generally speaking, the author's accepted manuscript (AAM) is the final version of an article that is sent to publishers, after peer review but before any formatting of the article has taken place. The final formatted article

Still, funders are increasingly considered the tip of the spear in the charge for more open. Indeed, many of the major nonprofit funders in the open space have aligned themselves ideologically with the position that open means something very specific—generally CC-BY licensed material that is immediately available for free, includes a dataset, and is deposited in the right repository (which also needs to satisfy specific requirements). These funders have implemented publication policies for the research work they fund that aligns with these narrow open requirements, and have also supported community-wide initiatives that align with these goals.<sup>92</sup>

Government funders haven't been as particular about ideology, and their funding goals haven't been as ambitious either, focusing mostly on small-scale undertakings like building discipline-specific collaboration networks or repository integration solutions. Governments have not seen the open solutions challenge as central to science, and have certainly not taken bold or ambitious steps to fund these efforts in any significant way.<sup>93</sup>

This said, the interest and engagement from funders is real and substantial. Funders are becoming more interested in good data management practices. Still, only a portion of the research data generated today is reliably archived, shared and reused, and this represents a loss of knowledge that leads to a repetition of research and a waste of funder resources.

Here, an education approach is needed. Funders are only one part of the scholarly communication ecosystem but they currently exert an outsized influence on it, and they distort open outcomes by creating and supporting programs that only align with their perspectives. It is important for funders to see a bigger open picture than they currently do now. Whether this is possible is another matter. Funders are run by people, and the people in positions of power in these agencies determine what their agencies believe in. The open program leaders at funders like Gates, Arcadia, Wellcome, and OSF (among others) all frequently and publicly pine for a world where the large commercial publishers (mostly but not solely Elsevier) aren't part of the scholarly communication ecosystem, so barring a change of heart by their superiors, it is unlikely that these funders will become more open minded any time soon.<sup>94</sup>

But change does happen once funders start allowing themselves to be led by evidence. After initially supporting Plan S, for example, the Swedish Foundation for Humanities and Social Sciences (Riksbankens Jubileumsfond) removed its name as a signatory to this agreement six months later and decided to create its own open science plan and not sign any other declarations. Ultimately, they decided, serving the needs of Sweden's researchers turned out not to be a simple issue with a one-size-fits-all solution.

---

is the official version of record (VoR). Changes in fact can take place between the AAM and VoR stage, which means that the VoR is the preferred version to cite. However, most all open requirements stipulate that only the AAM be deposited in an open format (such as a green repository), and allow the VoR to still be published in a closed format. So in other words, an author can deposit their AAM in arXiv, publish their official citable article—the VoR—in Science, and still be entirely compliant with a funder's open access policy.

92. They have also, with a few exceptions (like the Sloan, Arnold and McDonnell Foundations) generally eschewed supporting broader work like OSI's that try to find common ground approaches, especially those that include listening to the concerns of commercial publishers. This observation about funders is based on the work of Glenn Hampson in leading OSI. For an extended and broader look at the ideology of the open access movement, see Poynder 2020.

93. Ibid.

94. Ibid.

**What is needed at this juncture is an agreement to agree—an agreement between the multilaterals who have either expressed an interest in open solutions and/or who have already developed such solutions, to come together and develop a coherent global open solutions policy framework that will work for all multilaterals and their constituents.**

Following from this, in July 2020, the European Research Council (ERC) announced their decision to withdraw as a supporter of cOAlition S and instead follow an independent path towards OA implementation, citing as motivation for this equity concerns for European researchers as well as the importance of respecting researchers' publishing needs.<sup>95</sup>

## Multilaterals

At the multilateral level, open solutions are increasingly being integrated into other important agendas such as the UN's Sustainable Development Goals. This integration has led to a certain level of policy maturity for open amongst multilaterals, wherein it is now important to figure out how to allow open to reach its full potential by moving from dialogue to implementation, and by developing more open standards, tools and practices that are effective—a difficult task.<sup>96</sup> Finding open access solutions that work for everyone everywhere has been a challenge for many years now; finding global open data solutions is a relatively younger challenge but proving to be every bit as elusive.<sup>97</sup>

Still, all this hasn't stopped agencies like the UNESCO, OECD, WHO, the EU and the World Bank from charging ahead with open solutions—not in unison, but separately, pursuing policies that address needs and concerns as these organizations perceive them. This is both good and bad—good in the sense that open solutions are being taken seriously, bad in that there is no coordination between multilaterals about what open solutions should be adopted.<sup>98</sup> What this means is that the approach to open the World Bank has developed and promoted is not also being used by UNDP or UNIDO; the open philosophy and efforts supported by WHO are not also supported by the FAO and UNEP; and what UNESCO is trying to develop now will not necessarily also be adopted by OAS, the EU, OECD, ASEAN, or other multilaterals.

What is needed at this juncture is an agreement to agree—an agreement between the multilaterals who have either expressed an interest in open solutions and/or who have already developed such solutions, to come together and develop a coherent global open solutions policy framework that will work for all multilaterals and their constituents. Why? Because no single agency wields that much influence with the entire world—especially in the economically more developed countries—and because most countries in the developing world engage with a variety of multilaterals, not just one. Granted the UN has awarded UNESCO the authority to develop an open science policy, but this policy only invites member states to adopt government policies that align with UNESCO's recommendation, and only with specific regard to open science.<sup>99</sup> It does not also call on UN member organizations like WHO and IBRD to adopt

**Just as has happened with the influence of a wide variety of researchers, multilaterals and publishers, the constellation of agency interests and perspectives has led to the development and deployment of a wide variety of overlapping, intersecting and conflicting open policies over the years**

95. See <https://erc.europa.eu/news/erc-scientific-council-calls-open-access-plans-respect-researchers-needs>.

96. See the chapter on multilaterals in "The State of Open Data" (Davies 2019) for a fuller discussion of these challenges.

97. This challenge is complicated in no small measure by the fact that open uptake is not very robust in many parts of the world—not just open access and open data, but open government as well. Open government data use by citizens is still low in developing countries—the majority of citizens around the world can't access or use this data. (Davies 2019)

98. In the course of OSI's work, extensive outreach was made to UN agencies regarding their open efforts. Most of these agencies didn't have a solid point of contact regarding open policies, and none were aware of each other's work in this regard. All, however, expressed an interest in learning more and in collaborating. So, while there is fragile hope for working together, the reality of doing so hasn't been explored yet—the UN can be an extremely bureaucratic organization, with layers of approval needed for the simplest tasks.

99. See [https://en.unesco.org/sites/default/files/en-unesco\\_osr\\_first\\_draft.pdf](https://en.unesco.org/sites/default/files/en-unesco_osr_first_draft.pdf).

UNESCO's recommendations, or to modify their preexisting definitions of open and their preexisting support for various open efforts like Plan S. And importantly, it does not protect the world from wrestling with a "bad" open recommendation, should this be what emerges from UNESCO's effort.

UNESCO's effort is an important step, but it's just a first step. In order to be truly effective in the multilateral sphere, it must form the core of a much larger, much broader partnership, not only involving other multilaterals but also universities, funders, publishers, and other government agencies who are willing to adopt and strengthen UNESCO's approach.

### Other open-focused agencies

Other agencies of all kinds—for profits, nonprofits, special interest groups, activist groups, solutions providers, lobbying groups, library trade groups, publishing trade groups, higher education policy groups, and more—have been very active in developing open policies over the last 20 years. The focus of this activity ranges from narrow to broad, regional to global, single field to multidisciplinary, advocacy to product development, open access to open solutions. A representative list of some of the more active groups in the open space is included in the annex.<sup>100</sup>

Just as has happened with the influence of a wide variety of researchers, multilaterals and publishers, the constellation of agency interests and perspectives has led to the development and deployment of a wide variety of overlapping, intersecting and conflicting open policies over the years. There is considerable collaboration at the margins—many agencies support and build on each other's work and principles.

However, deep rifts between agencies have also formed along ideological lines, pitting: the capabilities and interests of rich EU and US libraries against poorer ones;<sup>101</sup> scholarly societies against the universities where their researchers work;<sup>102</sup> and agencies who support open against publishers who are also working to develop open solutions.<sup>103</sup> One of

**The challenge of meeting these needs is manifold...from privacy and policy barriers to paywalls, Internet access, and usability issues.... There are also real-world cost-benefit calculations to make.... What is the right balance between the public's need for (and potential benefit from) openness versus the real-world costs of generating and storing this information and making it suitable for public consumption?**

100. It is beyond the scope of this paper to provide too much detail here. The open space is very active, and a great many organizations have contributed to its development.

101. One emerging issue, for instance, is how major libraries and library systems are able to negotiate exclusive "transformative" agreements with major publishers (typically publish-and-read agreements, or read-and-publish agreements; see Hinchliffe 2019 for details). The gist of these agreements is to create a one-stop-shopping solution for universities so that paying one fee covers both authoring charges and the costs of making published research free for all to read. These agreements cause the APC payment system becomes institutionalized, however, and universities who lack the resources to negotiate such deals are left needing to pay for APCs on a case-by-case basis (they can now read more published research for free, but have less ability to publish their own research). These fees can be extraordinarily out of reach for most of the world's authors. Also, these agreements shift costs. They're based on publication output, so large research universities with a prolific researcher population will pay much more under these agreements. These cost increases are not marginal in some cases; there is a huge assumption here that large research institutions will be willing and able to permanently shoulder the cost burdens of this approach.

102. The financial health of many scholarly societies relies on revenues from conferences and publishing. With regard to publishing, most society journals are either subscription based or hybrid (where part of the content is open and part is subscription). Plan S originally proposed banning subscription and hybrid journals, which put scholarly societies in financial jeopardy.

103. For those who follow the debates in the scholarly communication space, bias against the involvement of commercial publishers is a defining characteristic. Major funders and advocacy groups often portray commercial publishers as an unwelcome and illegitimate presence. While this portrayal has softened over the last few years, at least from high level actors in this space, there is still a deep lack of trust between many groups borne of years of division and acrimony. See Poynder 2020.

the foundational recommendations of OSI's Plan A, therefore, is for these agencies is to take a step back from their disagreements over the details and look instead at the bigger picture of what they're trying to accomplish with open, and in doing so, figure out where they can work together on common ground toward common interests.<sup>104</sup>

These groups can find this common ground by talking to each other, of course. They can also work together through goal-oriented organizations like the Research Data Alliance (RDA). As a global open science/data organization, one of the aims of RDA is to reduce or remove the barriers between stakeholders, domains, and geographic borders, thereby facilitating open science efforts and interoperability. RDA does not use or endorse a specific definition of open science, given the range of approaches and foci for individual RDA members. The key concept of making the results of research “as open as possible, as closed as necessary” is often used in the RDA community as a shorthand description, along with the goal of sharing data, software, publications, and research samples whenever possible. RDA also emphasizes the research lifecycle (planning, collecting, processing and analyzing, publishing and sharing, preserving, reusing) from the beginning of the research to the long term archive, and that open science approaches apply throughout the lifecycle.

**...even though these policies share a lot of common language and goals, most are not global or broad, but instead are narrowly tailored to fit specific circumstances by region, field or institution, based differing interpretations of what openness entails (or should entail).**

RDA is a bit of a unicorn amongst scholarly communication agencies, however.<sup>105</sup> Most agencies have a narrower focus on just one particular open solution set (like COAR), or one particular region (like AmeliCA) or field (like AGU). Therefore, this siloed expertise tends to result in perspectives and actions that are tailored to specific audiences and not always easy to scale or integrate.

While these tailored solutions are necessary and practical in terms of visualizing the contours of global policy, the general rule seems to be that the agencies with a broader focus tend to see the complexity of the global issue space better than agencies with a narrow or more advocacy-centric focus. RDA, for example, doesn't endorse any single definition of open science because it experiences firsthand that at a global level there is a tremendous amount of diversity in the open science arena, whereas the EU's endorsement of a single definition is driven by an ideology of what open science should be for Europe. Similarly, the STM Association takes a neutral, analytical view of the open space even though the individual publishers in its ranks have a variety of views on open; and scholarly societies as a whole prefer a broad, inclusive approach to open (see McNutt 2019, for example) even though some societies are very much in favor of a Plan S approach.

These broad and inclusive viewpoints from high-level organizations don't reflect expediency or a lack of knowledge about or commitment to open, but the reality that outside our individual and closed-group advocacy bubbles there exists a wealth of different perspectives on open solutions. If we can begin embracing this big picture as the true reality of our existence, rather than our own microcosms of reality, then we will be able to go far, and quickly. There is a tremendous depth and breadth of energy and interest in this space.

---

104. Plan A is discussed later in this report. See <http://plan-a.world> for more information.

105. Other similar actors include OSI and FORCE11—organizations whose purpose is to foster dialogue and greater understanding.

## General public

The general public is another important stakeholder in this conversation but like researchers, there is no one “public” so it’s hard to speak with a single voice on this stakeholder group’s relationship with open solutions. Included in the “public” are ordinary citizens, citizen scientists, journalists, investors, and businesses of all kinds (primary consumers of open data are IT and research companies; see Davies 2019), all dealing with issues ranging from health care to the environment to mining, urban development, transportation, telecommunications, and beyond, with wide variation by global region and development status.

The needs of general public subgroups like businesses and citizen scientists may be more discrete and therefore more easily addressed than the needs of the general public writ large, who want, need and deserve information, transparency and accountability. They also want, need and deserve easy access to this information—particularly knowledge that impacts their health and economic well-being. In this regard, for example, public right-to-know laws are critical. They allow consumers to make better choices, protect our health, allow investors and government policy makers to make better decisions, and allow markets to function more efficiently. These laws also promote democratic decision making and the power of ordinary citizens.

The challenge of meeting all these needs is manifold, though, from privacy and policy barriers to paywalls, Internet access, and usability issues (language, readability, format, and more). There are also real-world cost-benefit calculations to make. With regard to research data, for example, how important is it that everyone have access to exabyte-sized data dumps from CERN versus just the researchers that request this information and know what to do with it? How vital is it that government research data be copyright free versus simply free to download? What is the right balance between the public’s need for (and potential benefit from) openness versus the real-world costs of generating and storing information and making it suitable for public consumption? This has similar overtones to debates about universal health care—if we truly feel that universal knowledge is a right, then how are we going to pay for it all, and what will it look like (and in this case, does universal knowledge apply to everything or should we focus on just a few things the public really needs most of all and that we can provide really well)?

We are at still at the cutting edge of exploring most of these questions. Our default position for now is that more is better, but we haven’t really analyzed which open knowledge practices are most needed and effective, and which information slices are in highest demand.

For open government data, the evidence so far isn’t clear and compelling. Scattered evidence shows that uptake and usage of open government data isn’t very impressive as a whole. The barriers aren’t generally cost-related, but have to do with downstream matters like usability, ease-of-access, and relevance—that this isn’t the kind of information citizens are looking for (see Davies 2019 and Ruijter 2020). With open access, on the other hand, cost is a major barrier, and open access materials tend to be more heavily downloaded than closed materials. Even so, most of the people who are seeking this information are researchers or “halo” participants in the research community.<sup>106</sup> In research, open data

---

106. Families searching the medical literature for cures and treatments is an important exception here, as well as citizen scientists. Other than these populations, it’s difficult to come by definitive data on the demand for research by the general public. While it’s obviously critical for families in need to be able to access research, and it’s an important asset for citizen scientists to be able to do the same (as well as inventors and businesses), we don’t have a clear idea of how much of this kind of usage is actually happening. In November 2020, Springer Nature (see Springer Nature 2020) released the findings from a survey of 6000 users to some 37,000 SDG (Sustainable Development Goal)-related materials it had published over the preceding 10 years (not only journal articles but books and conference proceedings as well). Around 60% of access was from people in research; of the remaining 40%, about 15% was from non-researchers with a scientific purpose (such as government agencies, physician, pharmaceuticals companies and science journalists). The remaining 25% were a long tail of users from a variety of

improves access from the general public, but we also know that access to open research data is generally restricted to qualified researchers. Open source may be the standout exception here. Can we learn lessons from the explosive public uptake of open source (think cell phone apps) with regard to what kind of information the public wants and needs and how the public can most effectively consume this information?

These questions and issues concerning the general public's access to research information are critically important to the open solutions space, but a full exploration of these questions and issues is beyond the scope of this report. This is one of the key areas where more understanding is needed in order to provide a firmer foundation for effective global open solutions policies.

## KEY OPEN POLICIES & ACTIONS

In this section, we will take a quick look at open policies in general, and a slightly more detailed (but still high level) overview of regulatory and licensing regimes. A much more detailed exploration of these issues is needed, as noted, before attempting to craft an actual open solutions policy that can successfully integrate with existing policies and regulations around the world.

### Open policies in general

Even though definitions of open vary widely, as discussed previously, this hasn't stopped a proliferation of open policies from developing anyway, based on independent definitions. These policies have emerged from governments as well as a great many institutions and organizations—too many to list here. By and large, even though these policies share a lot of common language and goals, most are not global or broad but instead are narrowly tailored to fit specific circumstances by region, field or institution, based on differing interpretations of what openness entails (or should entail). Plan S is an outlier in this regard, aiming for a global audience and defining open very narrowly as requiring CC-BY licensing, zero embargo, and FAIR data deposit in an approved repository.

By contrast, in the US the 2013 OSTP Memo (Holdren 2013), which established the US Public Access program, only requires that publicly financed research be made available—embargoes and traditional copyright are allowed. The target audience is US agencies and researchers who receive grant funding from the government. The result has been some of the most prolific open solutions in the world, such as PubMedCentral, DOE's PAGES, and NASA's Technical Reports server.

At the institutional level, major research universities like Harvard and MIT use well developed but flexible language around their open policies (and while these policies apply only to university-affiliated faculty,

---

backgrounds with no particular research need in mind. The survey also noted that most in this group had difficulty accessing the information in question; if that group had easier access, their use would likely have been higher. This is a limited sample, though, of a limited set of materials. We don't have solid numbers about broader demand. We have the same uncertainty about open data that is made available for "qualified" users (research databases that allow accesses to qualified researchers only—generally meaning researchers with the requisite expertise and connections who can also provide a scientific rationale for needing access)—we don't know yet whether making data available in even these restricted settings is a dreamy "build it and they will come" approach or whether the effort will pay dividends over time. Even various analyses of SciHub usage patterns, for example, note how most download activity is centered around universities, but there is no specific category for "unaffiliated" use that may be coming from private citizens. Anecdotally, the evidence from a variety of sources suggests that the general public's demand for open access, open science, and open data is low outside of these specific subgroups of the public. This isn't to downplay the need for or importance of public access to research—just to note that at present, the uptake of open by the general public isn't clear, and open policies aiming to make everything available to everyone are more being aspirational than responsive to actual market demand for this information.

they have also served as templates for other universities),<sup>107</sup> whereas universities like Oxford simply encourage the use of open formats and provide resources to help authors do so.<sup>108</sup>

Funder policies also vary widely. The Gates Foundation uses a Plan S-compliant policy that predates Plan S, requiring CC-BY licensing and no embargo period as a condition of funding, as well as data inclusion; Wellcome uses flexible rules similar to Harvard, requiring that the AAM be CC-BY licensed (CC-BY-ND is acceptable) and deposited in a free repository like PubMedCentral;<sup>109</sup> and the Sloan Foundation simply requires that grantees do their best to make information products open.<sup>110</sup>

Overall, open access, open data, open science and other open policies are still mostly separate entities that don't exist together under a common framework leading to the impression that, for example, open data is just an afterthought in most open access policies.<sup>111</sup> Attitudes here are changing though—we are beginning to see several cases where policy integration is being pushed and/or the shortcomings of existing policies are being actively addressed.<sup>112</sup>

This effort is vital, but it's also limited by the current state of standards—integration is important, but its pace is limited by different terms, definitions, and so on. At present there generally aren't any open data standards, for example, although we know they're needed. Experience shows that “what seems to work best is when groups of stakeholders, such as funders and publishers, identify common objectives and are able to work together towards achieving them” (Wiley 2018a). With open government data, data released by governments can be on many different websites, in many different formats, using many different units of measure, covering different time periods (often with little to no historical data), all published with varying frequency and levels of accuracy and completeness. The same is true for research data, even in the same field, with a few notable exceptions (like HIV/AIDS vaccine data). But at least more people are beginning to see the need, the connections, and the potential. Overall, the utility and cross-comparability of all this data is certainly exciting but it's also mostly aspirational at present.

Which approach to open access and open data is correct isn't for this author group to decide. OSI wrote a detailed critique of Plan S in early 2019 that concluded this particular plan was too hasty to work—it may in the end be just fine, but it's important to put more thought into it and include more perspectives than have been included to date.<sup>113</sup> What we're trying to do in this report is see how open solutions policies intersect.

---

107. Harvard's open access policy requires authors to grant the university nonexclusive rights to publish the pre-publication (author's accepted manuscript) version of their research using a CC-BY-NC license, allowing flexibility with regard to the timing of this deposit, and with waivers available on request. See <https://osc.hul.harvard.edu/policies/> for more details. No mention is made of embargoes or data inclusion, and authors can still publish the final version of their articles in any journal, open access or subscription.

108. See <http://openaccess.ox.ac.uk/home-2/open-access-at-oxford>.

109. See <https://bit.ly/2SVEwGz>.

110. From the Sloan Foundation's website (<https://bit.ly/3ITFuzg>), “While not everything can or should be free, maximal and appropriate openness remains a core value of the Sloan Foundation. In this Information Products appendix, potential grantees are asked to attend to the outputs their research will create and how those outputs can best be put in service to the larger scientific community.”

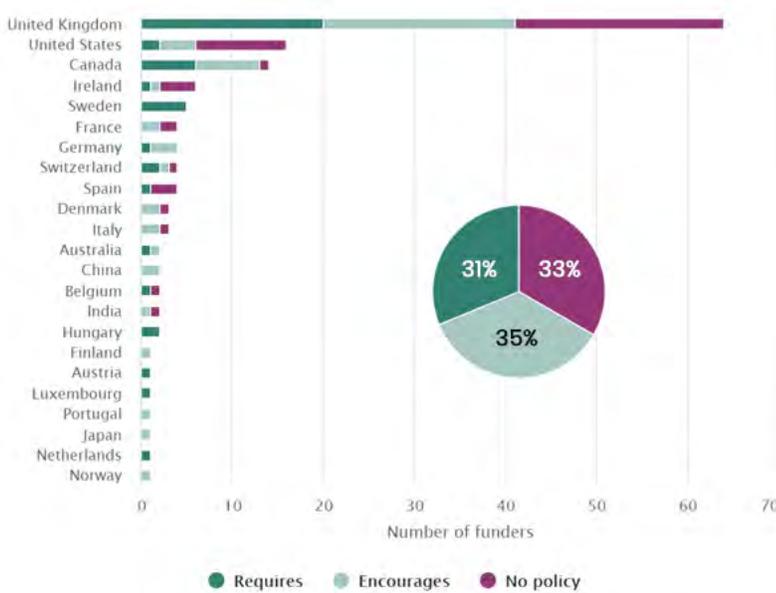
111. Here, it's important not to confuse data management plans with open data policies. The former have long been commonly required; the latter are a rarer breed.

112. Plan S is a good example of a policy effort that is trying to take a next generation integrated approach to open policies. The movement to create data citation standards is another example. In “A data citation roadmap for scientific publishers,” a FORCE11-generated study, the authors note how “Over the past several years many authoritative science policy bodies have recommended robust archiving and citation of primary research data to resolve problems in reproducibility, robustness and reusability (Cousijn 2018). Studies by CODATA, the U.S. National Academy of Sciences, the Royal Society, and other groups recommend that scholarly articles now treat the primary data upon which they rely as first class research objects” (Cousijn 2018).

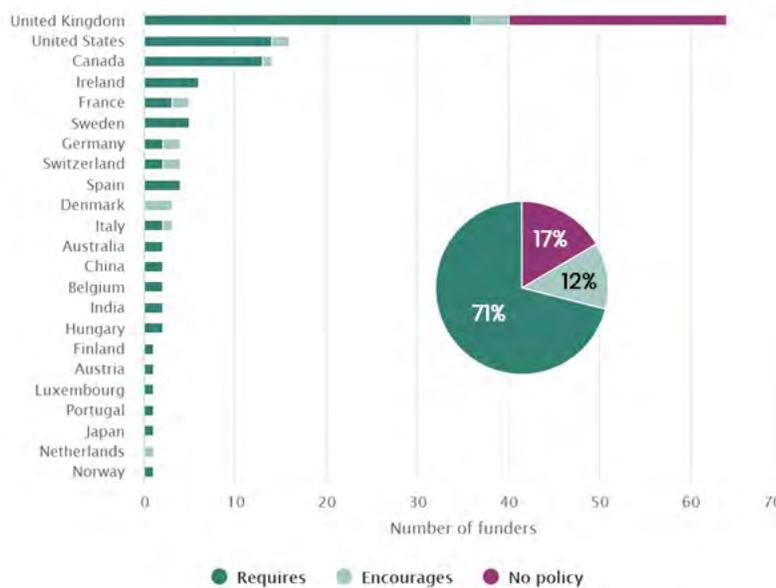
113. For more information, see Hampson 2019.

FIGURE 12: FUNDERS WITH OPEN ACCESS POLICIES

**PUBLISHING MANDATES**



**ARCHIVING MANDATES**



Source: Motter 2020 (derived from Sherpa-Juliet on October 21, 2019)

The answer is that there is a lot of common ground. Unfortunately, this reality may not be enough to build upon since, as noted earlier, there is also so much distrust and confusion in this space. Convincing some universities and publishers who have been engaged in a bitter dispute in recent years to now work together on their common ground for the future of open may be a pipe dream (granted, a few have been able to negotiate separate PAR agreements with publishers but these are the exception, not the rule). As a community, we can either work together as one to lead everyone toward a better future that focuses on our shared values and needs, or we can let the loudest voices dictate the terms and conditions of this future. We can advocate for this former outcome—working together—but can’t make it happen without the leadership and support of a global organization like UNESCO.

**Regulations**

Creating open solutions is one thing. Transforming these into regulations is quite another. The details of designing workable regulations work extends far beyond the scope of this report, involving a constellation of concerns such as forms, processes, agencies, legislation, mandates, reporting, administration, interest groups, standards, and more.<sup>114</sup> Hopefully, UNESCO’s open solutions effort will reach a point in the near future where all of these factors can be considered. For now, from a very high level, we need to first examine the landscape of existing regulations to see what’s needed and what fits where. In terms of just legislation, different laws

114. See [https://thefdp.org/default/assets/File/Presentations/Sept\\_2020%20-%20Expanded%20Clearinghouse.pdf](https://thefdp.org/default/assets/File/Presentations/Sept_2020%20-%20Expanded%20Clearinghouse.pdf)

around the world already regulate the access and reuse rights concerning a wide variety of research information. Additionally, with open access and open data in particular, a wide variety of mandates are in effect across the globe at the funder and institutional level.<sup>115</sup>

How all these regulations will interact with each other on the issue of open solutions is still a work in progress. This isn't a simple matter of how one local regulation affects research practices in a predictable manner, but of how the global network of laws intersects, particularly noting how these laws can be uncoordinated even within a given country, let alone globally, and can interpret the meaning of open in different ways and aim for a variety of different outcomes.

One example of this lack of coordination comes from the US Environmental Protection Agency, which in 2019 sought comment from the public on a proposal for “strengthening transparency in regulatory science.” As OSI and other organizations commented on this proposal,<sup>116</sup> EPA’s regulation was fraught with problems, such as confusing “open” with “transparency” (requiring the disclosure of private health information used in studies as opposed to simply requiring these studies be transparent, as all good science is already), and administrative overreach (EPA’s regulation would have effectively mandated that all science in the US be conducted in the “open” way they proposed, thereby affecting research funded by NIH, NSF, DoD, and other agencies). Also pending in the US is a proposal from OSTP to eliminate embargoes from the US Public Access policy.<sup>117</sup> This proposal would align the US with Plan S in terms of requiring that all research material produced with US government funds be made immediately available, but in doing so mandate an approach to open science that is inconsistent with the approach currently used by US government agencies that fund science, and also override the concerns and policies of many US researchers, publishers and scholarly societies.<sup>118</sup>

Other US government policies that impact on open science include the 2019 OPEN Act (Open, Public, Electronic, and Necessary Government Data Act),<sup>119</sup> which requires US federal agencies to publish their information online using machine-readable data formats; the 2020 GREAT Act (Grant Reporting Efficiency and Agreements Transparency Act), which calls on US federal agencies to use open data (i.e., data standards, transparency and access) to modernize federal grant reporting; FOIA (the Freedom of Information Act); and HIPPA (the Health Information Protection and Portability Act).

Not all such policies created to serve the needs and interests of a particular country stay within its borders. The EU’s GDPR is a good example of this. Because of its global reach, concerns have been expressed that GDPR will clash with open data policies from other countries, funders, journals, and institutions. According to one recent in-depth review of GDPR (Staunton 2019), “There is little insight or guidance contained within the GDPR as to the appropriate safeguards that must be in place...[and this lack of clarity] may render the research unethical and not in line with individuals interests.” The Staunton paper explores the specific policy deficiencies of GDPR in detail, and in addition provides a cursory overview of some of the myriad other policies that current inform open science practices, such as the Helsinki Declaration (defining modern rights of patient consent) and the Taipei Declaration (defining data access to biobanks). Here again, unfortunately, a full examination of all of these policies is well beyond the scope of this report.

---

115. Going into too much detail here is beyond the scope of this paper. However, the annex section of this paper lists some of the higher profile laws.

116. See <https://www.epa.gov/osa/strengthening-transparency-regulatory-science>

117. See <https://www.federalregister.gov/documents/2020/03/05/2020-04538/request-for-information-public-access-to-peer-reviewed-scholarly-publications-data-and-code>.

118. It’s likely that both of these initiatives are dead in the water as of January 2020 due to the change of administration.

119. This is included in the Foundations for Evidence Based Policymaking Act (H.R.4174) as Title II.

China and India are also critically important outliers as two of the world's most abundant sources of research papers. China's view of open is unique, not at all tethered to EU definitions of open but focusing instead on read-only open and geared toward developing an internal capacity for information sharing instead of sharing with the world (Lee 2020). Meanwhile, India recently announced its intent to try to reach a nationwide agreement for subscription journals, which is also completely at odds with the Plan S approach.<sup>120</sup>

What ripple effects will this burgeoning geopolitical split with China continue to have on science? What might happen to science as more and more countries start limiting the access of China's student population (on whom university finances depend),<sup>121</sup> or as they start pushing back on openly sharing scientific information with China due to the country's record of intellectual property theft? Not just with regard to China but more broadly, what happens if a free rider situation develops in basic research where only a few countries contribute the bulk of this information, being made freely available to the rest of the world through robust open policies? Will pressures develop to restrict access to the countries that fund this work (similar to a plan floated a few years ago for a "regional" solution to OA primarily intended for the EU)? How might all these dynamics cut at the growing movement for open science (see Schonfeld 2020 for a deeper discussion)?

Overall, the big picture with all of these global regulatory policies is simply to note that they exist, in abundance, and that UNESCO needs to take these policies into account in order to avoid conflict. How to do this is a massively complex question, and to the best of our knowledge, no one has ever explored the common ground that unites open solution regulations around the world. This exploration is important though, since UNESCO's intent is to recommend that governments implement its open solutions policies. This recommendation simply won't be implemented if it conflicts with HIPPA, GDPR, or the major open policies of national funders or major research institutions.

And then, of course, it's vital to recognize that creating these policies is only the beginning of the process. Much work remains afterward to make everything workable—all the nitty gritty detail of weaving these policies into national, regional and institutional frameworks will take even more time and flexibility. UNESCO can't just drop a new policy on the world and declare mission accomplished.<sup>122</sup>

One final note on the regulatory front is more philosophical than legal or operational. It goes without saying there is an urgent need at this point in history for more government transparency and collaboration regarding research. COVID is only the most obvious example; climate change research must also not be forgotten. But in the US, the Trump administration is not releasing enough data to allow

**Overall, the big picture with all of these global regulatory policies is simply to note that they exist, in abundance, and that UNESCO needs to take these policies into account... [UNESCO's new open solutions policy] simply won't be implemented if it conflicts with HIPPA, GDPR, or the major open policies of national funders or major research institutions.**

---

120. See <https://www.nature.com/articles/d41586-020-02708-4>.

121. At many universities around the world, students from China often comprise a significant portion of the student population. Many of these universities are more financially dependent on their undergraduate students than graduate students, but more dependent on international grad students to staff research labs.

122. From an academic perspective, science is global and the lack of international streamlining on mandates and policies can create challenges in international collaborations where different participants are bound by different rules and regulations. These different rules can also distort choices, insofar as encouraging researchers who are subject to fewer rules and restrictions to seek out collaborators who are similarly less encumbered.

researchers to adequately track the spread of the disease and devise effective countermeasures.<sup>123</sup> In addition, the US has pulled out of WHO and the Paris Climate Accords, buried federal efforts on climate change, politicized the US Centers for Disease Control and Prevention, and dismantled years and years of carefully considered environmental policies. The US isn't alone here—all governments are susceptible to stifling open government efforts for their own reasons.

But the fact that this is so easy to do, and will have such profound repercussions, is chilling. What can be done? What can UNESCO do? A pending change in US administrations will help matters, hopefully, but the damage that has been done and time lost ultimately illustrates the case that open solutions depend on the incentives and good will of the stakeholders involved. If these evaporate, what's left?

The long-term solution may be to try to create stronger international knowledge repository systems (perhaps something akin to an All Scholarship Repository, or the European Open Science Cloud) that make data archiving simpler and more manageable; that are less prone to tampering than one-off reporting systems designed and controlled at the government and agency level; that put knowledge archiving and management solutions directly into the hands of researchers instead of intermediaries like publishers, libraries, funders or, dare say, government officials; and that are ultimately more scalable and sustainable. An open solutions system that is robust, vibrant, interconnected and deeply rooted within research communities will be better able to withstand the vagaries of shifting policy and political winds.

## Licensing

Open licensing practices vary widely by field of research and type of open—open access versus open data, for instance. In the print world where open access licensing norms apply, more than half of all new research articles being published today use some version of the Creative Commons (CC) family of licenses.<sup>124</sup> The primary goal of these licenses is to enable easier use and reuse of materials than is allowed under traditional copyright.

While all licenses require attribution, different licenses stipulate different kinds of reuse permission. The most liberal of these, and the license generally preferred or mandated by many research funders (often as a precondition for receiving the grant)<sup>125</sup> is the CC-BY license, which allows others to distribute, remix, adapt, and build upon work, even commercially, as long as attribution is given for the original creation. For researchers concerned about unauthorized commercial reuse of their work, more restrictive versions of this license can be applied. For instance, the so-called CC-BY-NC-ND license allows others to reuse their work, except no commercial uses are allowed (NC means “non-commercial”) and no derivative works based on the original can be created (ND means “no derivatives”).

**[L]icensing formats have long been a huge point of contention among authors. Survey after survey has shown over the years that authors dislike and distrust the CC-BY requirement by a wide margin...**

Open licensing formats have long been a huge point of contention among researchers. Survey after survey has shown over the years that researchers dislike and distrust the CC-BY requirement by a wide

---

123. See <https://www.nytimes.com/2020/10/23/opinion/coronavirus-data-secrecy.html?action=click&module=Opinion&pg-type=Homepage>.

124. See <https://creativecommons.org/licenses> for a description of these licenses. See Archambault 2018 for data on the percent of open scholarly journal articles. See Piwowar 2018 for an estimate of the percent of the scholarly record which is still “dark.” Considering the full body of published research—new plus historical—most of the research articles currently available to read in repositories like PubMedCentral are held under traditional copyright (Hampson 2019).

125. By allowing authors to hold copyright instead of publishers. See for example “Publishing Your Research Open Access,” <https://www.vr.se/english/applying-for-funding/requirements-terms-and-conditions/publishing-open-access.html>

margin (see T&F 2019 for a recent example), preferring some other solution instead. Even assigning copyright to the publisher is preferred to CC-BY, according to survey data. The top reasons for this distrust—which have actually increased over the last few years—are intellectual property and confidentiality concerns (55%); ethical concerns (37%); concerns about misrepresentation or misuse (28%); and worries about being scooped (23%; see Wiley 2019b). Compounding this confusion, around two-thirds of researchers may be completely unaware of what kind of licensing format is being applied to their work (Wiley 2019a).

There are workarounds to this issue, such as creating new ways to officially timestamp new ideas, creating new licenses specifically tailored to academic research, or continuing to move away from the practice of assigning copyright to publishers,<sup>126</sup> but the dogged focus of

**[O]pen source today has achieved a level of successful commercial integration and adaptation that may preview what the open access and open data markets will look like ten years from now...**

many open advocates on the strictly CC-BY approach (versus options like CC-BY-NC-ND) has created tensions in the solution space. More flexibility and creativity is needed—namely an approach that listens to the concerns of researchers and tries to develop solutions that address their concerns rather than continuing to force CC-BY as a requirement for open.

Data is another matter entirely. Data itself cannot be copyrighted—it exists in the public domain. Formatted data can be copyrighted, however, and by “formatted data” we mean databases, data tables and the like. Therefore, when we speak of open data we’re really speaking of making the container for data open and accessible, not the data itself, which is open by default (Korn 2011).

Far and away the preferred license for open data is CC0 or its equivalent—public domain.<sup>127</sup> Anything more restrictive means data reuse can be jeopardized. For instance, if data is CC-BY licensed, thereby requiring attribution, then an untenable situation develops when a CC-BY licensed dataset is merged with a public domain licensed set. Sussing out which data elements going forward need to be attributed and which do not can become unworkable, particularly for large datasets (think genomics, for example).<sup>128</sup>

The particulars of software licenses (open source and open code) are somewhat different.<sup>129</sup> As noted earlier, there are two key concepts in this field: “open source software” and “free software.” These terms are very similar and are often used interchangeably, although there are some key differences, mainly in that free software uses an even more aggressively open licensing framework than open source software. Open source and code are intertwined with open science on many levels, most visibly at the big data level where huge datasets need to come with tools attached to allow analysis. Without these tools,

---

126. Typically, Fair Use provisions allow research materials to be reused in ways that are adequate for science. For instance, papers can be cited and excerpted, data tables can be repurposed, and so on. What isn’t permitted under traditional copyright is for entire works to be reproduced and/or adapted in their entirety for any purpose whatsoever. Mass reproduction for money making purposes is what authors are trying to prevent by selecting a CC-BY-NC (“non-commercial”) license; unauthorized alteration (derivatives, adaptations, updates or other transformations) is what’s covered by adding a “no derivatives” (ND) clause. This issue of remixing and reuse is particularly important for researchers in the arts and humanities, where a few, long form manuscripts and books are the normal research output as opposed to many short research papers and data tables.

127. See the annex section of this report for licensing examples. Several governments and international organizations have also developed bespoke licenses, such as the UK Open Government License, the World Bank Terms of Use and the French Government License Ouverte. Standard licenses (like CC and the Open Database License), however, can provide greater recognition for users, increased interoperability, and easier compliance. See Korn 2011 (page 7) for a look at different open data license types and the pros and cons of each.

128. The same situation is true for free software. For addition reading, see Korn 2011 and Hendler 2012.

129. See Open Source Initiative, Licenses and Standards, <https://bit.ly/376iYgR>. Accessed September 1st 2020.

there's just a huge jumble of unusable information. If the tools are openly available, then the information we're making available becomes actually usable.<sup>130</sup>

This evolution of open source in the marketplace of ideas is worth examining in more detail because this development has happened quickly compared to open access, and the impacts have been so widely embraced. Open source and open code have been huge commercial successes, spawning ecosystems of innovation that have advanced the capability and security of our computerized world and opened up ideas, solutions and best practices across a world of needs, applications, platforms and devices. Indeed, arguably, open source today has achieved a level of successful commercial integration and adaptation that may preview what the open access and open data markets will look like ten years from now, namely that commercial players control the product but the community benefits. In open source, 90% of the code is now written by commercial companies instead of by the public, but the public serves a vital role as “product manager”—they can see the code and help debug it, answer questions from the public, and develop new ideas and applications. They can also get free use of lesser or freemium versions of the software, while the company can develop paywalled versions for more advanced users and applications.<sup>131</sup>

**[F]undamentally, do we throw up bulwarks to more and faster progress on open solutions by legislating who can participate in the development of these solutions and exactly what these solutions should look like? Or do we level the playing field so everyone can participate and all outcomes are welcome?**

In open science, this might be the road we're on, where raw data can be open to the public, but the refined data and the applications for processing this data are paywalled. Is this a bad outcome? The answer may depend on how comfortable we are with the notion that business (and with business, the desire for profit) drives innovation. Refining and developing data is going to take a massive investment of time and money. It's not likely this kind of investment—at the necessary scale and sustainability—is going to come from UNESCO, governments or foundations. So, fundamentally, do we throw up bulwarks to more and faster progress on open solutions by legislating who can participate in the development of these solutions and exactly what these solutions should look like? Or do we level the playing field so everyone can participate and all outcomes are welcome? Can the marketplace, with support from governments to help correct imbalances, help create the best outcomes?

To most people in the modern policy world, the answer is clearly to allow market mechanisms to function—this really isn't a point of debate. But in the open solutions space, this is a central point of debate—that commercial involvement in science is immoral and/or the prospect of commercial “lock-in” of data solutions is something that needs to be avoided at all cost. To the extent these arguments are widespread and deeply held, it's important to understand why this is the case and whether such positions create an environment where it's hard to come together on common ground. We argue later in this report that finding common ground is still possible—indeed, that this ground is actually quite fertile.<sup>132</sup> Still, this version of the future will involve something akin to what happened with open source and open code—an unleashing of boundaries and a cultivation of private sector innovation and involvement.

---

130. For a good overview of big data, see Hariri 2019. Big data is particularly salient in fields such as genomics and high energy physics. Datasets here can be genuinely massive, plus noisy and unfiltered. Currently, datasets that reside in the exabyte (EB) or zettabyte (ZB) ranges are generally considered to be big data. One EB is equal to a thousand petabytes (where one petabyte, or PB, is a thousand terabytes), and one ZB is a thousand EB. Walmart collects 2.5 PB of data every hour, and the Internet processes 1.8 EB of data per day (Hariri 2019).

131. See <https://techcrunch.com/2019/01/12/how-open-source-software-took-over-the-world>.

132. For a deeper discussion, see OSI's Common Ground paper (Hampson 2020).

## SECTION 2: RECOMMENDATIONS

### CHARTING THE PATH TOWARD COMMON GOALS

As described in the first section of this report, the open solutions universe is a rich, diverse, and complex mix of history and technology, needs and concerns, ideas and policies, tools and strategies, institutions and regions, grand ambitions and multifaceted challenges. And while there is a significant amount of common ground between various open solutions, there are also differences. Still, it is clear that a more unified policy embrace of open solutions will help these solutions become more than the sum of their individual parts, and that this approach should be founded on our common goals.

How do we move forward? Let's first quickly recap some of the key points from the previous section. UNESCO is trying to come up with a workable global approach to the future of open. The vast majority of other open policies being explored and implemented—involving a great many actors from across the ideological, stakeholder and geographic spectrum—have narrower, more localized goals. Open access, open data, open science, open source, open educational resources, and open government all have their own separate histories, constituencies and language, but also share many of the same general open tools and practices. Information sharing philosophies, behaviors and barriers are also similar. A number of thoughtful analyses of the open space have concluded that openness is a complex process, not a state—a means, and not an end—and that openness exists along a spectrum of outcomes for open access materials, open data, and open source.

At a global level, there is no significant policy integration between these various approaches to open.<sup>133</sup> The policy, regulatory and licensing regimes for open are a testament to this fact, at once uncoordinated, inadequate and overreaching, reacting to rapid changes in the information marketplace but at the same time struggling to create broad reforms that work as intended and don't conflict or get incorrectly interpreted. In a similar vein, there are no successful one-size-fits-all solutions in any open field. The diversity of needs and interests is simply too great—by field, region, institution, funder mandate, and more.

Researchers are central to this diversity and to our analysis in this report. There are a variety of reasons why researchers might want to make their information more open, beyond the fact that openness is an animating ethos of knowledge creation. The most important of these has to do with ensuring proper benefit—making their research visible and transparent, improving the value of their research to society, and making their work reusable. In the broader context, though, these motives take a back seat to making sure their research is published in high quality venues with high impact where it can be shared with colleagues and accrue benefit for science, society, and their own careers, regardless of whether these venues are open access or subscription based. Researchers can also be confused by open, distrust their information will be used fairly and properly, and put off by how much effort can be involved in complying with open requirements.

---

133. As mentioned elsewhere in this report, large research institutions typically do a good job of integrating various open streams, particularly open access and open data. But on a broader scale, different open movements have their own logic and constituencies, and the actors like RDA and FORCE11 who are actively trying to integrate these conversations on a global stage are few and far between. This said, there are signs that the history of silo-ization is starting to change at the margins, as, for example, publishers move into regulating data as well as access to publications. We're still at the leading edge of this dynamic. For now, and at scale, there has been very limited movement because of the lack of common ground—that is, there is too much variation in everything from audiences to stakeholders to terminology, policies, and so on, for coordinated action to get much policy traction.

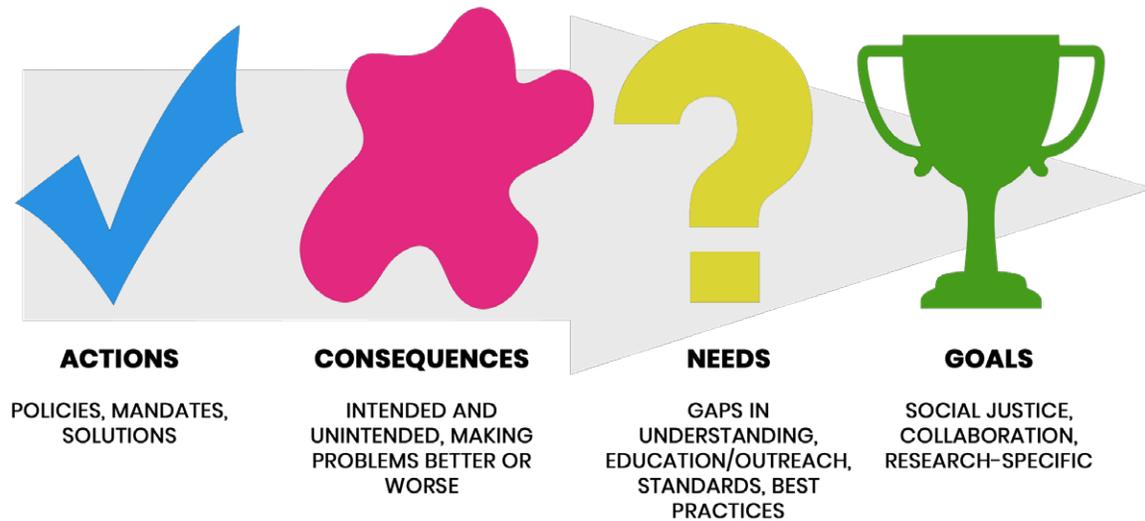
Among the other major stakeholders we've looked at in this paper, publishers are the gatekeepers to whether we're sharing quality research or just junk; together, the major publishers account for the majority of all articles published every year, open and subscription, and they are also taking a leading role in developing a more robust future for open data. Yet they're also vilified in the open access space for profiteering (not so in the other open spaces where the state of private-public collaboration is more mature). As for research institutions, this is a much larger group than just universities, but we tend to focus mostly on universities in this conversation. Scholarly societies serve the needs of the research community and yet the open solutions we're developing are threatening the revenue streams that keep these societies afloat. Major private funders and government funders are key players in the open universe and are driving the most significant open reform efforts at the moment, but many have begun taking an ideological approach to open instead of an approach that is driven by evidence. Multilaterals aren't leading in open policy development at the moment (although UNESCO was a pioneer in this movement and continues to be deeply involved), and aren't on the same page with regard to the future of open policies. The general public demands a right to know (and has a right to know) but access and usability are two different challenges; there may be lessons we can learn from how the public actually uses information that can help us design more effective open architectures. And finally, the constellation of interests and perspectives of other organizations in this space, from service providers to non-profit advocacy groups, has led to the development and deployment of a wide variety of overlapping, intersecting and conflicting open policies over the years.

So, taking a deep breath then, here is our \$64,000 question: What are we trying to achieve? The answer can't be provided by any one institution or stakeholder group working alone from a siloed perspective:

- **Who and what?** “Open” is a very diverse, very aspirational vision. Should we aspire to make everything available to everyone, everything available to some, some things available to everyone, or some things available to some? The choices we make will drive the systems and practices we develop;
- **Why?** Building on the “who and what” question, we need to ask ourselves why we want open? Is our goal to help communities of practice work more effectively together? Do we want to make research more transparent so we can improve replicability and better the impact of our investments in research? Do we want patients to have better access to information regarding their ailments or treatments, and teachers to know the newest and best information available to pass along to their students? Are we moved by the need to improve access to knowledge around the globe? What if our answer is “all of the above”? Here again, looking at this challenge from the perspective of researchers first and foremost, the choices we make and the priorities we set will identify the concerns we need to address, and the solutions we zero in on.
- **How?** Do we build one silo or a network of silos? Do we simplify and incentivize systems for sharing, or do we mandate sharing (at least where it can be realistically mandated, such as daylighting government-generated or government-funded data)? Do we allow for a range of open outcomes and licenses, or do we require only the most liberal licenses? Do we mandate the sharing of all research information immediately or do we allow researchers suitable time to analyze their data before sharing it (not an infinite amount of time, but at least enough to assuage their concerns about getting scooped)?

The reality of the open solution space is that these fundamental questions have been answered in a dizzying variety of ways because we have developed solutions first, and the rationale for these solutions second. Along the way, we have discovered that our solutions have unintended consequences, and that we lack adequate understanding of certain outcomes and dynamics in this space.

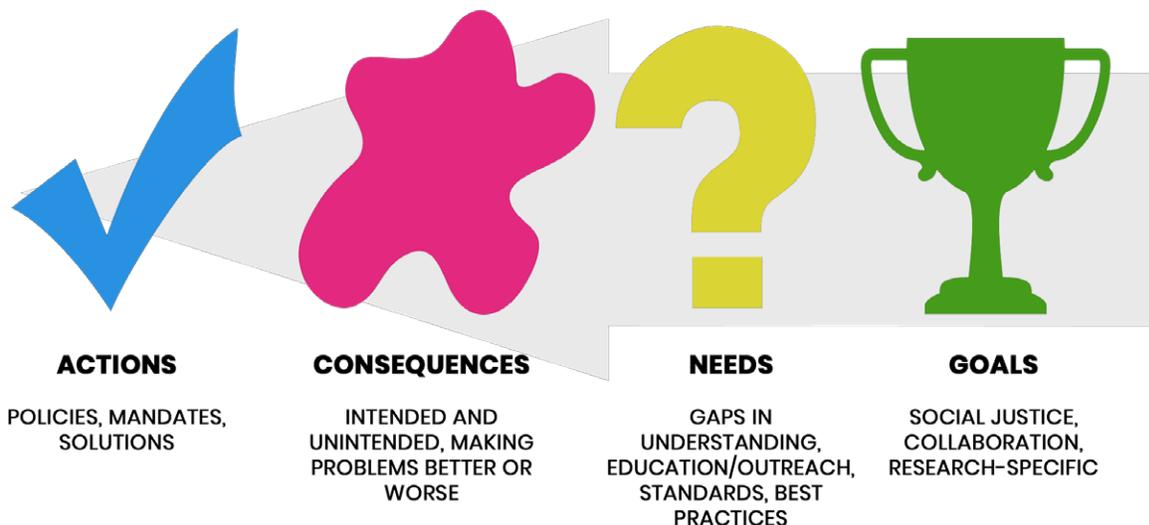
FIGURE 13: OPEN SOLUTIONS ACTIONS FIRST, GOALS LATER



“Solutions” like FAIR standards, for example, didn’t emerge before we embarked on the open data path, but only after as a beacon to help remind researchers that sharing is good. These standards aren’t detailed or practical enough to require interoperability or describe at a practical level how open can be guaranteed (recalling from earlier that open data isn’t necessarily FAIR and vice versa). Indeed, recall that “reusability” is difficult to achieve even in the most homogeneous data environments, and ensuring “discoverability” can be expensive for the kinds of massive image-based dataset found in some fields, or impossible for datasets that include private health information.

Therefore, by the time we get to finally thinking about goals, we are trying to achieve outcomes that our path (our actions and lack of understanding, combined with setbacks due to unforeseen consequences) has not prepared us to achieve. Hence, the idea of starting with common goals instead of actions makes sense from a strategic point of view.

FIGURE 14: OPEN SOLUTIONS GOALS FIRST, ACTIONS LATER



This approach makes sense in every other facet of life, from planning a vacation to building a house to running a business or a government agency. We don't buy a truck full of wood, hammer it together, and then assess whether we can live in this structure, what's missing, and what we need to do to bring it up to code. The Theory of Change model described earlier in this report focuses on identifying goals first, and only then moving toward actions. Since we're focusing a lot on the needs of researchers, how does this approach play out in knowledge creation endeavors like science? The same. While there are plenty of entertaining stories in science about accidental discoveries, the scientific method itself is premised on setting out goals and planning first, then experimenting (besides which, this kind of rigorous planning is essential for winning grant support).

The practices of research data networks shows how this goals-directed process works. In a recent white paper from Sage Bionetworks describing the governance mechanisms for open science work (Mangravite 2020) the authors note that “the ‘right’ system of governance is determined by first understanding the nature of the collaborative activities intended”—by setting out goals for this collaboration. Then and only then are governance structures developed, as described in the table on the following page. The Sage paper goes on to describe various licensing, data ingest, user qualification, data de-identification, and data transition (from trusted to open) standards which are also critically important in the real world of data consumption. By beginning with a common goal, this network moved to create workable policies that addressed the most salient issues with regard to data use, and has built up considerable experience and expertise in this area (as have other research networks).

By comparison, the conversations we normally hear in open solutions debates have to do only with standardizing information licensing format at the point of ingest, but as you can see from the Sage experience (which is only representative of the complexity of this challenge), this is only a tiny part of the equation because data in the real world exists in a number of different databases and formats, subject to a wide variety of usage restrictions and governance structures. Real world data isn't simply open, end of conversation. The lesson here is that even if we create new systems, processes and policies to guide the future of open, actually transitioning to this environment will take many years and will involve robust and dedicated efforts to include existing materials in this new framework. The complexity of this task cannot be understated. Simply creating a common goal framework is only part of the solution.

The same sort of nuances and complexities have been encountered in other open efforts, as evidenced by the flexible policies at Harvard and elsewhere (described previously); the shift away from CC-BY to CC-BY-NC-ND licensing; the slow growth curve of strictly BOAI-compliant open access; the experience of RDA in engaging with the complexity of the open data world; the public-private integration of open source solutions; and so on. Based on this understanding of the open solutions space, there are at least five approaches to charting our path toward common goals: narrow, broad, philosophical, tangible, and collaborative.

### Narrow

Let's say our global goal is something like “to improve the usability of research and the value of research to society.” Great. How? We can create a flexible and inclusive framework for participation in open. But which mechanisms would be most helpful? Which would actually work in practice? Sage is an example of the latter. Sage data is only available to Sage researchers. Similarly, research databases like GenBank, DataSphere and DataSpace collect research data on specific focus areas—genetics, cancer and HIV/AIDS vaccines respectively. All make data publicly available, but

**[Narrow approaches] make sense because this kind of environment lends itself well to the excruciatingly detailed work on standards and oversight required to clean the data, integrate it, ask the right questions to fill in gaps, allow researchers to see how their data is being used, and so on.**

TABLE 7: GOVERNANCE STRUCTURES AND THEIR ATTRIBUTES

Governance structure	Number and linkage of parties	Degree of data Availability	Degree of freedom to use data	Challenges common to the governance success	Primary governance design pattern
Pairwise	One-to-one	Medium/High	Medium/High	Uneven status of parties, value of data	Informal or closed contract
Open Source	One/some-to-many	High	High	Rights permanently granted to user	License
Federated Query	Many-to-many, via platform	High	Medium/Low	Defection of creators	Contract and club rules
Trusted Research Environment	One/some-to-many	Medium/Low	Medium/Low	Users agree to be known, surveilled	Data transfer and use agreements
Model-to-Data	One-to-many	High	Low	Not all who apply can use data	Restricted analyses, data curation
Open Citizen Science	Many-to-many	High	High	Capacity for analysis is uneven	Contract or license
Clubs, Trusts	Some-to-some	Medium/Low	High	Easy to create things governed more liberally. Trusteeship can be revoked.	Club / Trust rules
Closed	Many (to none)	Low	High	Fundamental limits to collaboration	Public laws, security protocols
Closed and Restricted	Some (to none)	Low	Low	Fundamental limits to collaboration	Public laws, security protocols

Source: Mangravite 2020

only to registered users. In the case of DataSphere, users can only be “bona fide researchers” and use is restricted<sup>134</sup>—DataSphere owns the data and no patents can be claimed on discoveries made via this data. These are all closed systems where researchers in one field are sharing data with each other in a trusted environment for a limited purpose. This make sense because this kind of environment lends itself well to the excruciatingly detailed work on standards and oversight required to clean the data, integrate it, ask the right questions to fill in gaps, allow researchers to see how their data is being used, and so on. But this approach also results in information silos, and these silos may or may not fit with our vision of truly open knowledge (creating long approval delays, limiting access, requiring membership in a network, restricting downloads, prohibiting cross-platform use, and so on).

## Broad

Most broad approaches to the future of open solutions are still in the realm of science fiction in this community’s policy debates. Still, at the risk of detracting from the credibility of this report, it’s import-

134. See <https://data.projectdatasphere.org/projectdatasphere/html/registration>. The same is true for clinical trials platforms like CSDR and Vivli, which are searchable databases of several thousand studies. Researchers must submit requests to access data from studies of interest, and also approval from their institution’s review boards. Most of these data access request are approved, but many are not for a variety of reasons, including inadequate scientific rationale or methodology, or inadequate qualifications of the research team (NASEM 2020).

ant to note that these broad approaches do have purpose and appeal and shouldn't be disregarded out of hand as being without merit (see, for example, the sentiments expressed in NASEM 2020 and NIH 2020), even if the exact solutions are still elusive. In principle, one such approach would be to build something akin to an “All Scholarship Repository” (ASR).<sup>135</sup> The closest thing we have to this at the moment may be Zenodo, a data cloud created by the EU's OpenAire program and operated by CERN, or the European Open Science Cloud (EOSC).<sup>136</sup> The general vision of these systems to varying degrees, is to create one interface through which we can search for complete data and text without encountering barriers of any kind (not just registration or paywalls but also technical and interoperability barriers like differing standards, data sharing agreements, consents and metrics).<sup>137</sup> With such systems, we can focus our efforts less on finding and accessing information and more on reuse.

Where the ASR approach differs from the EOSC approach is in creating a single silo instead of linking together existing silos through metadata or artificial intelligence tools. The benefits of such a single silo approach may be manifold (again, emphasizing that this approach currently has zero policy traction)—in theory, allowing us to create something akin to a single preprint server from which the world of research could then flow into overlay journals (in whatever format), or into customized analytics platforms that will vary tremendously by field, vendor, institution, and so on. Innovation would become entirely focused on the user—in most cases, the researcher—and not on what funders want or what publishers are willing to develop. In addition to the operational advantages of such a system, it's also possible that the process of data sharing will become easier when there is more data to work with and greater overlap between research communities and perspectives. In such a world, as noted in the NASEM 2020 report, “data generators might collect and steward data in a way that better facilitates sharing, and data users might be able to propose secondary analyses that make the best use of the available data.”

## Philosophical

As discussed earlier in this report, there are a wide range of motives for open. Some of these overlap, many do not. There is no solid bedrock of opinion across the open solutions universe that unites everyone in common cause that, for instance, all information should be free. This variation exists between user communities, but especially between open solutions. For instance, open government, open education and open source have a very public servant ethos—there is no question that open is intended to benefit everyone everywhere. Open access has a similar ethos but, at least when it comes to research, much of the information being daylighted is highly technical and written in a highly impenetrable manner, so the primary audience is going to be other researchers (which isn't to say there isn't a non-researcher audience for this material; see the Nature 2020 survey). With open data, again at least at the research level, this is even more constricted than open access for similar reasons, and also due to concerns about misuse and unauthorized reuse.<sup>138</sup> Here, protecting the integrity of research is a primary consideration that trumps sharing beyond a circle of known experts.

---

135. See annex section for more details. Again, this is hypothetical and not yet a model that is getting serious policy attention.

136. Zenodo covers a wide range of disciplines and data types, but it isn't heavily populated at the moment and data uploads are limited to 50GB, a woefully inadequate size for datasets in many kinds of research. The issue of big data is mentioned at various points in this paper. “Data” means different things to different communities, and the mechanisms for sharing this information are similarly varied. The discussion around open data often ignores that in some disciplines, a dataset can be hundreds of terabytes or even larger. This is not something that can be just uploaded to a website—bandwidth becomes a huge limiting factor in data sharing, and in some cases it's more economically and logistically reasonable to fly somewhere with hard drives than try to transfer data remotely. Some communities like high energy physics already have solutions in place for this, but for other communities that don't have an “open” culture, figuring out how to make such data sets open will be a huge and serious problem. In general people are moving towards bigger and bigger data in other disciplines too. How should we handle this?

137. See NASEM 2020 for a discussion of these barriers with regard to the sharing of clinical trials data. See the annex section for more detail on the ASR.

138. These concerns and the resulting licensing variations are discussed in more detail later in this report.

TABLE 8: OPEN SOLUTIONS ETHOS

Ethos	Open gov	OER	Open source	Open science	Open access	Open data
Everyone should have access to everything	Yes	Yes	Yes	To the extent possible. TBD.	To the extent possible. TBD.	No—this is neither practical nor desirable.
Everyone will be able to make use of everything	Yes (although the reality is much different)	Yes	Yes	To be determined	Not likely	No (especially big data)

## Tangible

There is a wealth of common ground in research for working together on tangible goals like curing cancer, combating climate change, or more broadly, improving science. These types of goals may ultimately be the most actionable—they are tangible enough without being overly broad, and they don't try to embrace disparate philosophical motives. They are also, however, woefully unprepared, as the experience of Sage Bionetworks shows. That is, we think our solutions have everything to do with simply “opening” information when in fact this is only one small step in a very long, involved and complex process. Rallying around common tangible goals would be an effective way to begin mobilizing the expertise we need to actually do something with open as opposed to treating open as a goal unto itself. It will also help shift the conversation away from the “orthodoxies” of whose approach is right and whose is wrong, and toward the practical matters of figuring out what's needed to achieve our goals, beginning with establishing goals and then working forward.

## Collaborative

Our goals for open will likely continue to be multi-pronged: Different fields of open, different fields of research, and even different governments, regions, institutions, and research networks will prefer to pursue goals that are tailored to their particular needs and audiences. Certainly, funding and support is much easier to generate for narrowly tailored goals of clear and demonstrated benefit to specific communities than for broad, global goals of the “build it and they will come” variety.

But does this mean that from a policy perspective, our common ground goal should simply be to acknowledge a constellation of uncoordinated and disconnected policies?

Yes and no. We should encourage the marketplace of ideas and products to participate in figuring out how to advance our common goals for open, and work together to support each other's efforts, especially where there is overlap between these efforts. But at the same time, we need to provide guidance, developed by the full open community, that describes what a common commitment to open looks like in terms of broad, flexible, actionable policy. Collaboration works best if we're all rowing in the same direction, not so much if we have different destinations in mind.

OSI's Common Ground paper (Hampson 2020) describes one version of this collaborative approach to the future, identifying a wide swath of activities in this space that constitute common ground in pursuit of open access. This paper's conclusions are generalizable to open solutions as well, particularly the broad vision of what the global solution space looks like. As noted in this report, a common ground

**[W]e think solutions have everything to do with simply “opening” information when in fact this is only one small step in a very involved and complex process. Rallying around common tangible goals would be an effective way to begin mobilizing the expertise we will need to actually do something with open as opposed to treating open as a goal unto itself.**

foundation will have multiple, irregularly-shaped points of intersection on multiple points of common interest. These multiple points of intersection might look like this:

- **Work together to get all research materials somewhere onto the DARTS open spectrum.** Seventy percent of the world’s research is closed and entirely off the open spectrum. What if we work together to get this number down to ten percent in ten years? One easy way we can do this is by valuing all open outcomes instead of prejudging which outcomes are superior (evidence to this effect will accumulate over time). Step one is to first get as much research as possible somewhere onto the open spectrum.
- **Work together to improve all open outcomes.** Getting more information onto the open spectrum is just a first step. From there we can work together to improve open outcomes (for instance, an institution or an information artifact can begin its open journey at one open level and improve over time). From this inclusive and non-judgmental approach, open adoption will become the norm and improvements over time will incentivize change and adoption, which will incentive more improvements and more adoption.
- **Work together to immediately improve access where it’s most needed.** What kinds of outcomes are wanted by researchers and where? Where are improvements needed and why? The access holes we’re looking to fill and the outcomes we’re looking to improve may be fairly discrete—for instance, improving access to medical research for low resource institutions. Can these needs be addressed quickly and effectively through targeted reforms instead of slow burning systemic changes?
- **Work together to improve open clarity and standards.** What’s the simplest way to participate in the future of open research? We need solutions that are easier for researchers to understand and value and easier for universities to implement. We also need better standards. What are the neon bright guidelines that all researchers and publishers should know and follow with regard to open research?<sup>139</sup>
- **Work together to address urgent needs.** There are many such needs to choose from but none more urgent and global than climate change. Many of the research disciplines connected to climate science are too closed. What if the international open community—including commercial publishers—worked together to not only open climate research but to actively integrate this work, make connections, and facilitate discovery?<sup>140</sup> We can prove the concept of open and at the same time work together to save our planet.
- **Pilot open solutions.** Let’s build things with open—combine, curate and standardize data, make new connections, bridge the gaps between disciplines, see new fields, make new discoveries—in short, do work that proves open is the future.

---

139. There are international standards in the open solutions space referencing copyright law, universal digital object identifiers, and so on, but no international standards describing, for instance, how journals should conduct peer review, or what constitutes a legitimate and credible journal. Not all emerging open formats are created equal; standards can help ensure a baseline of quality and reliability. COPE and other organizations have created strong first drafts of this kind of work (see COPE 2018). The next step is for the international community to review (and modify as needed) these proposals and build the capacity of publishers worldwide.

140. There are already programs and procedures, both at the publisher level and the international level, to help researchers respond to global health emergencies like Ebola, Zika and the Coronavirus. See, for example, NIH’s Emergency Access Initiative, or Elsevier’s information resource centers (Reller 2020). It’s important to note here that we’re not suggesting daylighting private health information from studies or discounting studies where this information can’t be publicly evaluated. These are both bad ideas, and don’t do anything to help science or science policy.

- **Look beyond.** As a community, we are just now beginning to look beyond the journal article to figure out what we really need. What tools and systems should we build? To what end (specifically)? What role will artificial intelligence have in being able to synthesize research? What forms of research communication will be the most important in tomorrow’s research environment (the answer will differ from one field to the next)? Rather than spending so much time and effort debating how to turn our research communication horse and buggy into a rocket ship, maybe we should just build a rocket ship?

**Instead of relying on one-size-fits-all approaches we can create an inclusive open movement that is informed and empowered by diversity and opportunity.**

So—narrow, broad, philosophical, tangible, and collaborative—or maybe something else or some combination of these paths, as long as all paths lead to Rome. The journey is certainly possible, if we have the will. But, we may ask, how is this approach different from our current approach? It’s different because there are exactly zero policy agencies and instruments in the open solutions universe today that incorporate a broad, inclusive, and globally diverse set of views and perspectives, and that are informed and empowered by the vast diversity and opportunity in this universe. Instead, we mostly rely on ideologically-driven one-size-fits-all approaches that act first, and ask questions later—that build the house first without any plans, and then struggle to make the house habitable and bring it up to code.

From an outcomes perspective, a common path, common goal approach might start small by picking the low-hanging fruit in order to build confidence, and then over time could move on to more complicated and challenging collaborations.

FIGURE 14: THE FUTURE OF OPEN (GOOD VERSION)



Source: Hampson 2018

After 15 years of working together, what does this full potential look like?

- Open is clearly defined and supported
- Open is the standard output format
- Open solutions are robust, inclusive, broad, scalable and sustainable
- Almost all knowledge is discoverable
- The global access gap is nonexistent<sup>141</sup>
- Solutions for the arts and humanities are built-in
- Connected issues are resolved
- Incentives are aligned so scholars embrace open because they want to
- Open is simple and clear so scholars know what it means and why they should do it
- Predatory publishing is defeated so it no longer threatens knowledge integrity<sup>142</sup>
- Standards and global guidelines are clear for all journals, which helps the market
- The marketplace remains competitive so open products remain cutting edge
- Repositories are integrated, not just connected
- Data standardization is widespread and robust.

All of this leads to an Open Renaissance where many kinds of improvement happen to research, the research ecosystem grows exponentially more powerful, new fields and directions emerge based on easier and more robust interdisciplinary work, funding efficiency improves, and discovery accelerates. The social impacts of research surpass today (including improved literacy, public engagement, and public policy impact), knowledge becomes more of a global public good, and society reaps the benefits.<sup>143</sup>

It's impossible to imagine what future might lie ahead on this path, but "exciting" is probably an understatement. Before the birth of what we now call science, societies had no word for "discovery," and no sense that the knowledge of tomorrow would be superior to the knowledge of yesterday. The changes brought about by experimentalism, natural philosophy, and the search for objective truth shook the world to its foundation—indeed, shook the world off its foundations—and led to a steady climb out of darkness and toward enlightenment.<sup>144</sup> What will happen if we can share this miracle of knowledge more effectively, not just within science but within society? What new discoveries will we make then? What unseen connections will become apparent? What problems will research be able to solve for society?

And what if we don't work together on the challenges ahead? Maybe we'll reach our goals more slowly, maybe we won't reach them at all, or maybe the solution space will fracture. Continuing with our go-it-alone approach, for example, may eventually result in competing regional solutions

**Before the birth of what we now call science, societies had no word for "discovery," and no sense that the knowledge of tomorrow would be superior to the knowledge of yesterday... What will happen if we can share this miracle of knowledge more effectively, not just within science but within society?**

---

141. At least in terms of what can be controlled. For people without access to a computer, the internet, or even 24 hour per day electricity, making everything open doesn't by itself close the access gap.

142. See Anderson 2019 for a fuller discussion of this issue.

143. Most of this section is verbatim from Hampson 2018.

144. See Wootton 2015.

where we end up with one open future for China, another for the EU, and still other futures for South America, Africa, and other regions, each working to solve their own unique concerns and perspectives. This approach may also force changes across diverse disciplines that may not work well (for example, open solutions that work in physics generally don't work at all in history), causing researchers in some fields to lose interest in an open future altogether. Or it may lead to unintended consequences that don't necessarily benefit research, again causing a drop in interest.<sup>145</sup>

A go-it-alone approach also fails to address the significant concerns in government offices around the world that there are intellectual property and security ramifications of a vastly more open research world (see Poynder 2019 for a lengthy list of examples)—not just sharing data freely but collaborating on research projects and even allowing certain foreign nationals to study at certain universities. Can we proactively address concerns like these by working together more effectively, or do we wait and react to future legislation that directs researchers to collaborate and share on the basis of nationality rather than merit?

There are larger, distinctly modern currents at work here that have the potential to utterly reshape our answers to the many questions posed by open research. If we work together, our ability as a community to deal with these currents will be informed, unified and strong. If we are a fractured community, however, where every country and stakeholder group is just in this for their own benefit and is pursuing their own national agenda and vision of the future then there will be no bulwark against these nationalistic tides and the global effort to make knowledge more open may suffer as a result.

In summary, good reasons exist for working together as a global community on the many challenges of open solutions— from a wealth of common ground interests to a need for common ground solutions to systemic problems; from making open solutions more attractive and coordinated, to aligning incentives, removing obstacles, better understanding national needs and interests and charting a course for a much more exciting and robust open future. Still, there are those in the open solutions community who disagree with the necessity or desirability of this approach—experts who believe limited solutions are the best we can hope to achieve; open advocates who think trading one evil (like subscription prices) for another (like author fees) will produce the greater good; or observers who believe our disjointed system as it's currently evolving will eventually get us to the right point without the need to deliberately seek broader solutions. These perspectives are all valued and valuable. Many such perspectives inform this debate—there are no black and white answers. Indeed, there are a wealth of questions that have no answers at all.

**The potential that an open future holds for knowledge and society is vast. It behooves all of us to work together to develop this future the right way. Exactly how we do this is the question we should be trying to answer now.**

And this is precisely why, considering what's at stake, it is so critically important that we put aside our differences in this community and summon the will to look thoughtfully and carefully at how we are approaching the common challenges we face. Are we certain our current efforts are truly the best we can do as a community or are some of our approaches more expedient than thoughtful, inclusive, robust, effective and sustainable? And if these efforts are simply expedient, then we need to ask ourselves whether our shortcuts are wise. The potential that an open future holds for knowledge and society is vast. It behooves us all to work together to develop this future the right way. Exactly how we do this is the question we should be trying to answer now.

---

145. One example here is that if we replace subscription paywalls with "play-walls" where authors need to pay to have their articles published, this will arguably be a worse outcome since we're then dealing not just with research that's hard to access, but also with research that doesn't get published in the first place.

## STRATEGIC NEEDS

In terms of strategic planning in this space, we can start by focusing on the three P's of open: Philosophy, policy, and practice. These three P's are epitomized in the goals-based perspective of open solutions discussed earlier. Open solutions flow from the broad to the specific, from a broad purpose to narrow perspectives and narrower toolsets—from philosophy to policy to practice (except when we're building our house first and figuring out how to live in it second).

The philosophy of open is generally benign, except, in the case of research, when it narrowly (and incorrectly) claims that open is defined by certain policies and practices.<sup>146</sup> Such interpretations don't accurately describe the broader philosophical landscape of open, or the needs, norms and practices that many research groups would recognize, but instead are hybrid philosophies built to express certain ideological preconditions. In doing so, these philosophies exclude consideration a wide variety of needs, perspectives, solutions, adaptations, and evidence (see Poynder 2020 for a fuller discussion of this dynamic). Apart from this, there are also a variety of open philosophies (as discussed earlier) but not any single “meta-philosophy” that we all agree on and that motivates us all to action.

The policy of open can also be contentious, although again, at least with regard to research, the general policy of sharing and transparency is not only agreeable to most researchers but essential to the proper conduct of research. The same is generally true for open data, open source and open governance. Policies that stipulate openness are on their face not contentious, but can start to become so when they run up against real concerns and constraints on issues like privacy, intellectual property, or security.

Alas, the practice of open is where our discussion really gets complicated. Compounded by the fact that practices are built on conflicting, divergent and shifting philosophical and policy foundations, the practice-based reality is that every discipline defines its own best practices and tools, including whether or not it is good to share research outputs, and deciding when, how, and what to share. Even so, there are many practice-based areas of overlap.

Strategically, from a policy formulation perspective, it's important to engage on all three of these levels. Most of the really transformative common ground in this space will happen through our engagement at the philosophy level—understanding and agreeing on common cause for why we want open solutions—whereas specific policies and practices will continue to evolve in a more customized manner, including specific solutions by field, region or institution as necessary.

We don't necessarily need to actually agree on this philosophical change in order to make progress, though. The approach being taken by the Research Data Alliance is to work with members on their evolving policies and practices for open science and wrangle them into alignment with a new philosophical approach such that they increasingly:

- **Work toward a world that is open by default**, but closed as necessary (based on valid arguments);
- **Maintain a healthy tension between the commercial and “public good”** aspects of open science;

---

146. Such as but not limited to the Budapest Open Access Initiative (BOAI), which defines “open access” in narrow terms. Plan S defines open access even more narrowly than BOAI.

- **Never losing sight that our common ground is built on values that are critically important to science**, like transparency, integrity, and recognition;
- **Tear down the perception that open science is totally new**, and that science of the past does not matter; and
- **Understand that open science creates opportunities but also requires changes from many parties, particularly researchers**, and therefore it is important to include them in dialogue about open; that change like this needs strong political commitment and the financial investment to back it; that lower and middle-income countries may (or may not) lack sufficient political will and financial resources, and/or may have different priorities from higher-income countries; that diverse and representative stakeholder groups are critical to finding lasting and workable solutions; and that each country or region will want to execute research policy and investment (and open research policies within those) following the direction and speed that they believe is right.

Most of the really transformative common ground in this space will be found at the philosophy level—understanding and agreeing on common cause for why we want open solutions—whereas specific policies and practices will continue to evolve in a more customized manner, including specific solutions by field, region or institution as necessary.

OSI's perspective, as described in its open science recommendations to UNESCO (Hampson 2020), takes a similarly broad, bottom-up approach. These recommendations—which work for all kinds of open and not just open science—state that inclusive, effective, sustainable open policies and practices intended to benefit all researchers everywhere must be:

- 1. Researcher-focused.** Research communication tools, services and options need to be developed with heavy input from the research community, with solutions and approaches driven by researcher needs and concerns. In this case, where are the needs and gaps in current practices related to open science? For instance, where are the current data sharing needs most urgent, and what are the roadblocks to wider use and uptake (e.g., systems, standards, etc.)?
- 2. Collaborative.** Successful and sustainable solutions will require broad collaboration, not just to ensure that all perspectives are considered, but also to ensure there is broad ownership of ideas.
- 3. Connected.** There are a great many interconnected issues in scholarly communication. We can't just improve the openness of information without also addressing issues such as impact factors, peer review, and predatory publishing. Reforming scholarly communication will require a systemic approach.
- 4. Diverse and flexible.** There are no one-size-fits-all solutions to scholarly communication reform. Instead, there are many different pathways to reform, likely including many that have not yet been conceived or deployed. Diversity, creativity and flexibility in this undertaking should be encouraged, at the same time noting that we should try to maximize adherence to the other principles represented here.
- 5. Informed.** We need a better understanding of key issues in scholarly communication before moving forward. For instance, what is the impact of open research? The more accurate and

honest our assessments, the more accurate and honest our reform efforts can be, the easier these efforts will be to promote, and the more successful they will be.

- 6. Ethical and accountable.** We need enforceable, community-developed/driven standards to ensure the integrity of journal publishing, repositories, and other related activities/products, and to ensure that unethical approaches are not embraced.
- 7. Common goal oriented.** We must discuss and plan for what the future of scholarly communication means, beyond just having access. For instance, we need to identify precisely what we plan to do with open information, where we will need data interoperability, what tools and procedures we will need to achieve this interoperability, and so on. By doing this, we focus on and strive for our community's common goals.
- 8. Equitable.** Researchers everywhere need to be able to access and contribute content to the global body of research information with minimal barriers. To the extent practicable, research information—particularly information central to life and health—should not be unreasonably constrained by issues such as high access and distribution costs,<sup>147</sup> poor journal indexing, and a lack of capacity-building programs (e.g., programs that can help small, niche publishers in developing countries adopt best practices with regard to peer review, archiving, and so on).
- 9. Sustainable.** Scholarly communication reform approaches need to be sustainable, which flows from all the other elements in this list. That is, the reform solutions we design need to be achievable, affordable, popular, effective, and so on.
- 10. Transparent.** This community needs to maintain as much transparency as possible in this effort in order to address the trust issues that have plagued this space for so long.
- 11. Understandable and simple.** This community needs to agree on a few simple, high-level, common-ground goals for scholarly communication reform—not anything specific with regard to publishing requirements, for example, but a general set of goals that are understandable, achievable, and adaptable. By setting out general goals that can be easily achieved, participation can be made simple and easy, with low barriers to entry.
- 12. Beneficial.** In the end, these reforms need to benefit research first and foremost. While the argument to improve benefits to society is central, these benefits need to be matured carefully, deliberately, and realistically in order to ensure that societal benefits are indeed being conveyed as intended, and that research is not being harmed in the process.

So in other words, rather than trying to first reorient the philosophical framework for open before acting, RDA's and OSI's strategic approach has involved trying to realign policies and practices such that they lead to a new philosophical framework as a consequence. Once this philosophical framework becomes better established over time, then building new open policies and practices in and on this framework will lead to even more solutions that are inclusive, diverse, and of common interest and common benefit to the future of open.

From a researcher perspective instead of a systems perspective, there are also strategic issues that need to be considered at the researcher level. It is important, for instance, for researchers to understand what impacts their decision to publish openly will have on decisions earlier in the research

---

147. We frequently focus just on the cost of accessing data, but the high costs for APC publishing and data sharing can also segregate researchers into producers and consumers of knowledge, depending on their financial situations.

workflow that will enable actions like open data publication—from how data are organized and licensed to whether protocols are registered, open notebooks are maintained, marketable intellectual property is protected, and so on. This will require the research community to take a broader and longer-range view of their workflows than they are accustomed at present (Dallmeier 2017).

Similarly, the pursuit of open will require that researchers get out of the mindset that openness simply means tacking on data as an afterthought at the end of a scientific endeavor. In her 2018 *Nature Physics* article, “Open is not enough,” Xiaoli Chen (Chen 2018) and her co-authors note that in order to ensure reproducibility and reusability, data needs to be accompanied by software, workflows and explanations. Indeed, the entire research lifecycle process needs to be geared more toward the idea of “meshing seamlessly” with existing research procedures, “encouraging the pursuit of reusability as a natural part of researchers’ daily work.... In this way, the generated research products are more likely to be useful when shared openly.” The authors note that the CERN Analysis Preservation and reuse framework relies on three strategic pillars—that researchers:

- 1. Describe:** Adequately describe and structure the knowledge behind a physics analysis in view of its future reuse. Describe all the assets of an analysis and track data provenance. Ensure sufficient documentation and capture associated links;
- 2. Capture:** Store information about the analysis input data, the analysis code and its dependencies, the runtime computational environment and the analysis workflow steps, and any other necessary dependencies in a trusted digital repository; and
- 3. Reuse:** Preserve analysis assets and computational workflows in the cloud to allow for their validation or execution with new sets of parameters to test new hypotheses.

“All of these services,” notes Chen, “developed through free and open source software, strive to enable FAIR compliant data and can be set up for other communities as they are implemented using flexible data models. For all these services, capturing and preserving data provenance has been a key design feature. Data provenance facilitates reproducibility and data sharing as it provides a formal model for describing published results.”

Chen summarizes CERN’s strategic guiding principles for open data as follows:

- **Define your reproducibility goals**
- **Incorporate best practices early in your research**
- **Build on what is there**
- **Structure your knowledge**
- **Capture your content**
- **Capture your workflows**

**[R]ather than prescribing exactly what needs to be done in order to improve the future of open, RDA’s and OSI’s strategic approach has involved constructing a framework for dialogue and collaborative action. Building on this framework will lead to solutions that are inclusive, diverse, and focused on practice-based areas of common interest and common benefit.**

- **Raise awareness**
- **Embrace openness whenever possible**
- **Enable liberal and fair reuse.**

## TACTICAL NEEDS

The strategic framework is the global community’s general reason and pathway for action. Within this framework, specific “tactical” actions need to be taken to ensure that our lofty plans end up being more than just words. Many of these actions will be very specific in nature, others more general. Our tactical needs for developing the global future of open solutions can be grouped into several key categories: discovery, education, researcher support, relationships, measures, infrastructure, agreements, and pilot policies. Each of these categories is vital and has its own ecosystem of actors and actions.

### Discovery

While there has been a good deal of research into open over the last 20 years, considerable work remains. There is still much we don’t understand, such as exactly which open solutions work best, where and why; what influence predatory publishing is having on the market, on science, and on author behavior; whether embargoes are financially necessary for publishers; what researchers really want and need (anticipating huge variation by field); the full scope of needs and challenges of researchers from developing countries; the unique needs and concerns of researchers in the arts, humanities and social sciences (“unique” in the sense that most reform policies and proposals for open research are STM-centric and don’t work well or at all outside of STM); and so much more.<sup>148</sup>

To find these answers, we need to enlist the best of the best in the open research community—perhaps chief among them Caroline Wagner, Jason Priem, Heather Piwowar, Eric Archambault, Mikael Laakso, Michael Mabe, Vincent Larivière, Carol Tenopir and Rob Johnson.<sup>149</sup> Unfortunately, there has been a great deal of research in this space that has been ideologically directed, justifying a predetermined conclusion instead of evaluating facts, or that misinterprets the open space and assumes all open outcomes are the same. With regard to this second error pattern, these measurements of the open space fail to account for the diversity of outcomes and proclaim, for example, that open access is growing remarkably fast when in fact the only thing growing are the kinds of OA that “don’t count” because they are ideologically impure (see again Piwowar 2018 for a good overview of how fast different kinds of open are growing). Support for more research is needed, as well as for more solid research.

### Education

The open community has overestimated the degree to which researchers are informed and convinced about open solutions. There is, in fact, a great deal of misinformation and lack of information in this space which is hindering progress, as discussed earlier. In order to make more and faster progress on open reforms, our community needs to be better informed.

---

148. See OSI’s Plan A (<http://plan-a.world>) for more detail on the kinds of knowledge gaps that exist and learning projects that can help close these gaps.

149. Apologies for not including more names on this list. There are probably at least a half-dozen others who belong here. The point is that not all open researchers are equally knowledgeable—some are more thoughtful, objective and rigorous than others.

Researchers, funders and institutions need to learn more about open options and processes; policy makers need to learn more about the diversity and complexity of this issue; and people actively working in the open space need to learn more about each others' work so they can find more ways to collaborate on achieving common goals. A broad learning effort is necessary on many levels to begin creating the open changes we seek.

Our community also needs better systems for listening to stakeholder feedback, and for creating and adjusting to solutions accordingly. Of particular focus on the listening front, we need a clearer and more detailed understanding of exactly what researchers want and need, what they will use, and what we hope to accomplish with reforms so we can make sure to ask good questions, collect meaningful information and pursue effective solutions.

Of course, many organizations have long contributed to this education effort—most notably university libraries. Often, however, the training provided focuses entirely on open access solutions, and only the version that adheres to narrow ideological definitions.<sup>150</sup> University libraries in particular will be essential partners in the open solutions education effort, so it will be important to provide librarians with the background and training materials they will need.

### Researcher support

Different open solutions require different kinds of support; support also varies by field, region, open goals and more. Here, it's important to understand the wide diversity of needs, and that one-size-fits-all support solutions aren't what we should endeavor to develop. As with any serious and successful undertaking, it's important first and foremost to understand our market and our customer base.

This said, there is abundant literature available on the general kinds of support that should be made available in the open solutions space, and we can build on and adapt these observations. With respect to research publishing, for example, Wiley Publishing posits that the most important training needs of early career researchers includes language support, grant writing, manuscript preparation, peer review and deciding where to publish (Woodward 2017). "Language support" comes in at least two different forms—helping authors navigate the complexities of publishing their work in English (still the current lingua franca of science; companies like Editage/Cactus specialize in this kind of assistance),<sup>151</sup> and supporting more translation work so science is available both internationally and to vital local audiences. Investment in both kinds of linguistic support should be a priority component of research grant funding.

The European Commission, in an extensive report assessing the support needs for open science (see EC 2017), has summarized these needs as follows—that institutions should:

---

150. See, for example, the OA toolkit provided by the US Association of College and Research Libraries (ACRL) at <https://acrl.libguides.com/scholcomm/toolkit/openaccess>.

151. The lingua franca of science has changed over time from Latin to French to German, and now English. Is Chinese next? Looking at the global citation map of science, the US is still firmly in the center of highly cited research so it's not likely that a switch from English will happen any time soon.

**Researchers, funders and institutions need to learn more about open options and processes; policy makers need to learn more about the diversity and complexity of this issue; and people actively working in the open space need to learn more about each others' work so they can find ways to collaborate on achieving common goals.**

- **Provide the technical infrastructure** for open science (like high-speed data centers, data repositories and virtual platforms);
- **Provide the technical tools** to facilitate researchers in doing open science (like software for data creation, storage, and sharing);
- **Provide professional support staff** for general and specialist support for researchers (like data stewards, IT technicians, data scientists, legal experts, discipline specific data managers and librarians);
- **Implement and promote the use of data management plans** in all research projects; and
- **Ensure a legal framework is in place** for the secure, legal, and ethical sharing of data.

FORCE-11 (in Cousijn 2018) notes that for open data, key needed areas of support are to:

- **Revise editor training** and advocacy material
- **Revise reviewer training** material
- **Provide guidance** on author responsibilities
- **Specify a policy** for data citation
- **Ask authors** for a Data Availability Statement (DAS)<sup>152</sup>
- **Specify how** to format data citations
- **Provide guidance** around suitable repositories
- **Provide specific guidance** on citing groups of datasets reused in meta-analyses, and
- **Consider licensing formats** and implications.

In her 2017 paper, “Connecting research data to workflow,” Sünje Dallmeier-Tiessen (et al) comes up with an equally helpful to-do list that would help support researchers in their efforts to make data more open, including: “Assignment of persistent identifiers (PIDs) to datasets, code, models etc; creation of metadata to support data citation and discovery; adoption of recognised metadata standards; data documentation e.g. describing data using both domain-relevant and generalised terminology so that others may understand how and why the data, code, models, etc were produced; linking research data documentation to author PIDs (e.g. ORCID) and, where relevant grant information; linking research data documentation to other research products e.g. data management plan, data paper, journal article; technical review, e.g. describing cleaning, de-identification, or quality assurance; peer review of data, e.g. by researchers or by editorial reviewers. While many of these tasks have been researched and practiced for many years in the data preservation and open access repository communities, there have been few empirical studies of them in the data publishing context.” (Dallmeier 2017)

**...it's important to understand the wide diversity of needs, and that one-size-fits-all support solutions aren't what we should endeavor to develop. As with any serious and successful undertaking, it's important first and foremost to understand your market and your customer base.**

152. Data availability statements and data management plans play an important role because they make researchers think about data from the beginning of the research pipeline. There are great tools and resources online to help with this; data officers in academic settings can also provide training and assistance.

Springer Nature’s 2018 survey of 7700 researchers from around the world (Stuart 2018) distills the support needs of researchers into a desire for support that makes open easier and more of a turnkey operation. Specifically, the survey concludes that “researchers’ efforts to archive, publish and share data continue to be hampered by time constraints and a lack of knowledge around data standards, metadata and curation expertise, repository options, and funder requirements. Subject and regional differences do exist, suggesting where targeted approaches may be helpful. But there are common global challenges that require concerted attention: the provision of more education and support for researchers, and faster, easier routes to sharing data optimally.” (Stuart 2018)

There’s a fascinating amount of detail here, as well as overlap. Clearly, many different groups have been thinking for a long time about how to support the growth of open solutions (and we haven’t even touched yet on support perspectives regarding open source, open government, and OER). Suffice it to say that researcher support is needed and essential. It’s also critical to remember when reviewing this variety of ideas and perspectives that the most workable, most detailed open solutions are going to continue to be developed by individual fields (and within these fields, specific research projects and groups),<sup>153</sup> institutions, countries and regions. Understanding the diversity of needs and ideas in this space, then, can help us identify where we can construct broad support mechanisms for everyone (like education and infrastructure). This approach, along with trying to define the open solutions space, will ultimately provide much more effective support for the future of open than creating overly-detailed open solutions plans (for science, open access, open data, and so on) that may not be followed, and that may end up being largely inapplicable to the majority of real-world open efforts.

The fundamental question for support, after all, as Vint Cerf posited at the OSI2017 meeting, is one of incentives (see OSI 2017a). How can open solutions of all kinds evolve because doing open is in our own best interest, not because we’re required to do open? Once we develop this understanding and construct (or allow the marketplace to construct) solutions that begin to better align incentives, then our job of providing support shifts from desperately encouraging a single seedling to grow through constant watering and care, to one of managing a grocery store chain with thousands of locations worldwide.

Constructing a self-incentivized solution space will allow solutions to evolve that people want and need. And critically, these are the solutions that people will end up using. Use is key because the world already has plenty of ambitious knowledge sharing platforms that simply don’t contain enough data to be useful, and/or don’t get enough use to merit further investment.

For now, as a community, we need to support emerging open frameworks and their practitioners and customers—developing best practice models, expanding education and training, making inroads into more data standards and reuse, and tackling big, real world challenges with our emerging capabilities. The key is to cultivate an ecosystem where we nurture and develop solutions working together toward a world where open accomplishments not only happen but are clearly

**Understanding the diversity of needs and ideas in this space can help us identify where we can construct broad support mechanisms for everyone.... This approach, along with trying to define the open solutions space, will ultimately provide much more effective support for the future of open than creating overly-detailed open solutions plans...that no one will follow, and that will be largely inapplicable to the majority of real-world open efforts.**

---

153. For instance, see the US Institute of Medicine’s 2015 plan to improve the sharing of clinical trials data—IOM 2015 and NASEM 2020. A great many other fields, networks and projects have specific (and not necessarily open) data sharing plans that fit their norms and needs, as discussed earlier.

benefiting all involved in a variety of ways (from better research to better policy to good profits).

## Relationships

There are many institutional, regional, and global approaches to open science, and a vast array of actors and stakeholders behind these actions. One of the easiest and most valuable “tactical” actions we can take to support a future of open is to build bridges between these existing ideas and efforts, build on our successes in this space, and learn from our lessons of experience. Indeed, community built and run mechanisms can offer considerable value, giving a voice to discipline-specific researchers in close collaboration with data professionals, and generate meaningful, actionable and useful outputs. It is fundamental to support these types of mechanisms. There are a few very successful organizations in this space already, like FORCE11, RDA, and NISO, whose work can be built upon and expanded.

It will be particularly worthwhile to approach this task with an eye toward finding common ground in the open solutions space. Most of the groups in this space from across the regional and stakeholder spectrum recognize and respond to many of the same challenges and issues. This commonality exists both within and between stakeholder groups. As a broad, global community, though, we have never taken time to work through our differing perspectives and identify specific ways we can work on these challenges and issues together, globally and at a high policy level.<sup>154</sup> What are our common goals for the future of open writ large? What can we learn from open movements about what we’re trying to accomplish in academia and where do we ultimately want these efforts to lead us? Are there specific common ground solutions we can move forward with right away? Building on the common ground we have in this community, we have a better chance of developing the right detailed solutions together, in the right order, and for the right reasons, and these solutions will have a better chance of being adopted, sustained, and bearing fruit.

## Measures

There are a great many “measures” the open community can implement (or at least continue working on together), from data standards to definitions to systemic issues like peer review and impact factor reform. To focus briefly on just a few of these, making more open databases compliant with FAIR standards will be important. So too will be implementing data citation standards in order help bring data out of the closet and make it into the first class research object the US National Academies of Science would like it to be.<sup>155</sup>

---

154. There have been many instances of limited sharing and collaboration, but nothing approaching a global movement to work together. OSI conference delegates have done this kind of work—their ideas and perspectives are summarized in OSI’s Common Ground policy paper. These ideas and perspectives might be helpful seeds of a broader, global conversation.

155. As noted in Cousijn 2018, “Data citations are formal ways to ground the research findings in a manuscript upon their supporting evidence, when that evidence consists of externally archived datasets. They presume that the underlying data have been robustly archived in an appropriate long-term-persistent repository. This approach supersedes “Supplemental Data” as a home for article-associated datasets. It is designed to make data fully FAIR (Findable, Accessible, Interoperable and Reusable).” Publishers implementing data citation can provide domain-specific lists of acceptable repositories for this purpose, or guide authors to sites that maintain these lists.

**Constructing a self-incentivized solution space will allow solutions to evolve that people want and need. And critically, these are the solutions that people will end up using. Use is key because the world already has plenty of ambitious knowledge sharing platforms that simply don’t contain enough data to be useful, and/or don’t get enough use to merit further investment.**

Taking a closer look at data metrics is another approach. These metrics are an important indicator of impact by researchers and other stakeholders, second only to data citations, but they can't currently fill this role due to the lack of standardization on how usage metrics should be created, collected and reported (Counter 2018). This said, doubling down on a code of practice for research data usage (and there is such a code)—while this does help data repositories better manage and demonstrate the value of research data—also creates yet another roadblock to making open data use simple (i.e., the detail and complexity is a disincentive). This is all the more reason to think in terms of building infrastructure solutions like the All-Scholarship Repository (mentioned below) that make open data easy and rewarding.

## Infrastructure

The global open community needs to invest more in infrastructure (particularly systems that are publicly owned, to the extent possible)—the technical systems, like data repositories, that are related to the access, use, stewardship and preservation of research information— in order to ensure the interoperability and long-term sustainability of these systems, and also to help encourage, achieve, sustain and monitor progress in this space.<sup>156</sup> Our community should develop these things together, and reasonably quickly, so reforms can be more easily adopted and the open landscape can be more quickly and easily improved and maintained. Accelerating the development and adoption of more domain-specific standards is another approach that can help improve the return on investment from these systems (which is to say, to help transform them from simple storage depots for information into a more dynamic spaces where this information is actively mined, transformed, and reused).<sup>157</sup>

Francine Berman's take on open data infrastructure is worth noting here (see Berman 2020):

The problem of insufficient and inadequate data infrastructure is particularly pressing in the academic research community. Infrastructure used by researchers often has a comparatively small "market" of users compared with infrastructure products and services in the private sector or targeted open-source efforts for commonly used programs and systems. Research data infrastructure is often developed and utilized by individual researchers, specific projects, or domain communities as a means to an end (new results, scholarly literature). Responsibility for its maintenance, upgrade, and support may be unclear or inconsistent. Researchers are incentivized to move on when their work evolves to focus on new results, often deprioritizing the maintenance and support of previously useful infrastructure.

Part of the difficulty in creating and maintaining research data infrastructure in academia is recognition and funding. Infrastructure often plays second fiddle to new exploration as a funding priority. It is challenging to obtain funding to maintain or improve infrastructure for the purpose of keeping it going or making it more useful to a larger user base. Moreover, while research results advance the reputation of researchers and their institutions (often leading to greater opportunities and funding), developing or maintaining effective working infrastructure rarely has this outcome. In the academic environment, this generally translates into less recognition and resources for the professionals who enable data-driven research than for their faculty research colleagues who conduct the research.

---

156. Noting that there are already a lot of infrastructure tools and services available, but these are fragmented. Our work should begin with a survey of existing tools across disciplines and an assessment of their merit and interoperability, gaps that need to be filled, priorities for filling these gaps, and so on.

157. A number of agencies have long been involved in this line of work as well, such as CODATA, Dryad, COAR, RDA and others.

**As a broad, global community... we have never taken the time to work through our differing perspectives and identify specific ways we can work on these challenges and issues together at scale. What are our common goals for the future of open? Can we learn from the open movement writ large what we're trying to accomplish in academia and where we want this work to ultimately lead us?**

These problems are exacerbated because what makes research notable and what makes research infrastructure notable is exactly the opposite: Research is noteworthy for its successes; infrastructure development and deployment is noteworthy for its failures. You may not think about the fact that hot water is available in your building until the plumbing breaks or that your commute home is smooth unless there are potholes in the road. Research infrastructure for the access, use, stewardship and preservation of data is similar. Until data is corrupted, lost, unusable, or becomes inaccessible, support for the infrastructure that delivered it, or what is needed to maintain and support it, may be deprioritized. In some cases, even the loss of data may do little to change priorities.

**Until data is corrupted, lost, unusable, or becomes inaccessible, support for the infrastructure that delivered it, or what is needed to maintain and support it, may be deprioritized. In some cases, even the loss of data may do little to change priorities.**

Broad approaches like the European Open Science Cloud (EOSC) or ASR, discussed previously, are one possible approach to the infrastructure question, involving building new systems from scratch that try to address many if not all research infrastructure shortcomings in one fell swoop. However, this “one database to rule all” approach isn’t discussed as often as other infrastructure solutions (or more accurately, really isn’t discussed at all at high levels), such as simply investing more in our existing infrastructure tools and systems (especially systems that can serve multiple needs and constituencies).

One important caveat to keep in mind with this discussion is access. Specifically, in order for infrastructure solutions to bring value to everyone everywhere—instead of making the digital divide even worse—access to the Internet and to robust computing environments are needed. This access can be limited in low and middle-income countries, so we can’t talk about infrastructure solutions in a vacuum. We also need to consider how researchers are going to access these solutions and whether they will be able to benefit from them without also investing heavily in internet and hardware needs (Newton 2020).

## Agreements

Joint agreements, knowledge exchange, and transparent policies support the creation of open practices, reduce the reinvention of the wheel, and decrease the potential overlap of open efforts and investments. There are obviously a multitude of open solutions agreements currently in existence, many working at cross purposes, many focusing just on one aspect of open, many that are “exclusive” arrangements between a set group of participants, and so on. There are also many agreements that are simply statements of purpose on what the future of open should look like—FAIR and DORA, for example.

While these agreements are an important part of the open landscape, their impact can be positive or negative—positive if they bring together diverse voices in search of common standards, for example, and negative if they divide stakeholders into camps of right and wrong, good and bad. Open access, open data and open source all have many of the former kinds of agreements; open access also has several of the latter. Open science, open government and OERs are more hybrid in nature, incorporating open access, source and data principles in a variety of ways.

Adding even more agreements to the open solutions landscape is probably inevitable, including UNESCO’s forthcoming open science recommendation. It would be preferable if these policy instruments could be more general and uplifting going forward, reflecting the vast diversity in this space and focusing on what the world hopes to achieve with open by working together. For instance, effective new agreements in this space might take the form of a Paris Climate Accord on open solutions, where broad global goals to increase open are set out without defining the specific means for doing so (attaching a specific and compelling reason for this openness will help, as noted; also, as with the Paris accords, subsequent rounds of negotiations can begin zeroing in on specific solutions).

OSI's general framework for these types of policies was noted previously (wherein it's important that they be researcher focused, equitable, sustainable, and so on). The Research Data Alliance posits that constructive and structured dialogue with diverse and representative stakeholders is a requirement, with the goal of creating flexible and inclusive policy that enables choice and innovation. CERN's general guiding principles also align with OSI and RDA, advocating an approach that is goal-driven and best practice oriented.

## Pilot policies

Taking everything into account with regard to what an open solutions policy should look like and how it should develop strategically and tactically, one final important tactical step is to actually build and pilot a few such policies at the institutional level—not a global call to action of the sort UNESCO might issue, but a hands-on guide to implementing this kind of policy at universities and other research institutions. There are many ways to go about this task, of course.

Ideally, such a product will develop from the ground up (as opposed to being mandated from the top down), bringing together existing efforts, addressing common goals, and responding to the needs and concerns of researchers. The steps involved in this process might look something like this:

- 1. Assemble a working team** comprised of leaders who are fluent in all current open policies at a particular institution (for journals, data, code, OER, and other institutional matters—administrative data, funding, etc.).
- 2. Catalog all of the institution's open policies.**
- 3. Look for pluses and minuses**—areas where openness is doing well, and where it is struggling;
- 4. Assess goals and needs:** What common goals exist that can be built on? For instance, is there a common desire to better promote the work of university researchers and make this work freely available (at least inside the university)? Where would a common approach or common tools make sense or help? Where is more outreach, education and training needed?
- 5. Pilot small, comprehensive solutions** that focus on specific goals. For instance, such a goal might be to make all natural sciences research at an institution as open and integrated as possible. Design approaches that are broad, flexible, and turnkey, allowing researchers to easily comply without having to hire an attorney to interpret regulations or spend weeks cleaning their data first. Importantly, compliance should be clearly beneficial; the “incentives switch” needs to be flipped or else open will only grow as far and as fast as we push it.<sup>158</sup>

**Design a solution that is broad, flexible, and turnkey, allowing researchers to easily comply without having to hire an attorney to interpret regulations or spend weeks cleaning their data first.... This [solution] should be clearly beneficial; the incentives switch needs to be flipped or else open will only grow as far and as fast as we push it.**

---

158. These policies can be as simple as Harvard's, requiring that some reasonable approximation of the final version of research work be made open (still allowing for researchers to choose where they publish the final version of record). Or they can be more ambitious, providing clear paths and incentives for every kind of information artifact without labeling these as open access, open data, open government, and so on. Another significant pilot might involve funding an “open research” team at a university (perhaps serving just one department or field) that sits between the university and an ASR, and is responsible for ensuring that all research gets translated and preserved in an open format. This function of such a team is described in more detail in the ASR section of the annex.

6. **Improve upon and expand these solutions and practices over time** based on a record of accomplishment and growing familiarity with open information needs and networks.
7. **Be part of a global network of pilot open solutions programs**, communicating regularly to compare best practices and lessons of experience.

This recommendation is all, unfortunately, still a bit too broad to be usable. The reality is going to be informed by each institution's current state of open. Institutions without existing open policies will simply be able to adopt UNESCO's new global open solutions policy as is. For institutions that already have robust open policies, "unwinding" these may not be feasible if too much financial and political capital has already been invested in describing exactly what open access means to that institution's libraries and researchers. In this case, a "rider" policy approach may be needed. For example, this new global policy might acknowledge at the outset that even though a particular university adopting the new open solutions framework already has a suite of existing open policies, they are welcome to adapt this new global policy for use in their institution, whether by amending their existing policies or revising the UNESCO policy at the margins (with these margins specified so the core tenets remain unifying) so their research community can get the maximum benefit from these new global efforts. Whether an institution like Harvard would ever choose to do something like this depends on how the field of open solutions develops. Best practices and lessons of experience may eventually give researchers and managers of these policies ample reasons to consider revisions. Even so, reevaluation isn't the goal here, but embracing the diversity of open solutions and approaches and letting ourselves be guided by needs and evidence.

## APPROACH

The 2019 "State of Open Data" report (Davies 2019) notes that we may currently be at a point in the evolution of open where we are beginning to think of open data less as a movement and more as a normative tool—that researchers now "perceive their role as someone who engages with open data as one tool among many." This shift in perceptions is core to understanding the adjustments we need to make in the years and decades ahead.

The same can be said of the open access movement. It has matured to the point where the solution space is robust and diverse, and most new articles are being published in some kind of open format. And yet there is still much ideological tension in the space, driven in large part by a persistent depiction that there is a "true" kind of open access and a "false" kind, rather than a spectrum of outcomes.

Getting beyond this is key—letting researchers decide what kind of licensing works best for them, what kinds of journals, what kinds of peer review, and so on. Given time and space to experiment and evolve, the open access space will do well. If it is painted into a corner however, with regard to requiring CC-BY licensing or requiring publication in certain kinds of journals, outcomes will likely continue to be suboptimal—possibly growing more slowly, but likely splitting into different paths, with different regimented outcomes for different regions and fields, which will affect the interoperability of these solutions as well as the window for creating policy that embraces diversity and nurtures common ground solutions.

This isn't to say that libraries, funders and other stakeholders should avoid taking a leadership role in pushing for change, just that these efforts should be the starting point for conversation and not the

final word. Librarians instituting cost-cutting measures for their universities (like shifting from print to digital issues of journals) is entirely appropriate; librarians instituting reforms that also mandate how researchers university-wide (or even beyond) will be required to publish share their data is more problematic, at least in the absence of extensive planning and consultation involving the international research community (which has not occurred). Of course, since researchers are a diverse body, often with entrenched interests, waiting for them to arrive at a consensus on what to do isn't practical either, nor is expecting this body to have strong opinions about the future of open (which they generally don't). But this engagement deficit is a challenge, not license to create sweeping reform policies without consultation, particularly when so many deep and different interests and concerns are affected. More effectively engaging researchers in this debate will be challenging but it is critically important. In the meantime, the frustrations are palpable on all sides. As one library leader in OSI noted, "Until [researchers] realize they have the ability to drive the bus, either someone else needs to drive it or it's either not going anywhere or will only crawl forward slowly."

**...ideologically-driven (as opposed to evidence-driven) actions will continue to create policy whiplash, where publishers and funders don't know what to do, researcher opinions are further marginalized, and the actions that sound brilliant today are roundly derided tomorrow.**

The danger of miscalculation is real as well. To the extent that the many of the unilateral reform actions of governments, libraries and funders are ideologically-driven as opposed to evolving from engagement with the research community, these actions will continue to create policy whiplash, where researchers, publishers and funders don't know what to do, researcher opinions are further marginalized, we get a "good enough" open environment instead of one that truly robust and incentivized, and the actions that sound brilliant today are roundly derided tomorrow.

APCs are one such case in point. OA2020 and Plan S were originally premised on shaky evidence showing that the world had plenty of resources to support a rapid transition from subscription publishing to an author-pays model. These two efforts developed a subscription and hybrid ban approach, only to soften this approach a few years later as evidence mounted that APCs were in fact poised to make the access situation worse in the developing world, not better, because most researchers outside of the US and EU simply can't afford to pay US\$2,200 and up to publish a journal article.

The same may end up being true of PAR agreements, the latest flavor of the month OA solutions. These agreements work for the large, rich research institutions, but not for others who are excluded. The Matthew Effect kicks in—the rich get richer and the poor poorer.<sup>159</sup>

So, rather than being led by evidence and learning from the marketplace about what works and what doesn't, the open access movement has been led instead by ideology, and the result has been an almost politically-charged dynamic where people are clearly working together on solutions, yet also hold wildly divergent ideas about right and wrong, and for the most part don't really express these opinions (except for the advocacy leaders) because doing so exposes the deep ideological rifts in this space.<sup>160</sup>

In comparison to the soap opera that is open access, open science has been debating and devising practical open solutions that are needed and that work. Indeed, publishing industry guru David Wor-

159. Of course, one can argue that Plan S is in fact learning from the marketplace and adjusting accordingly. But this argument endorses an approach to policy making that is rooted in trying to make ill-considered ideology more palatable (in the vernacular, putting lipstick on a pig). Ideally, we should develop the right policies first rather than trying to dress up bad ones.  
160. And this can get very uncomfortable. Look no further than Twitter to see how much cyber bullying happens in this space as a handful of open access advocates deride anyone whose opinion diverges from theirs.

**...integrating these conversations can help us integrate efforts and policies in ways that might dramatically and effectively accelerate the development and uptake of open policies. We might also be able to develop common language that can help facilitate the better communication of needs, processes and outcomes.**

lock predicts that as open science goes, so will go the future of open and of publishing. The urgent and highly salient needs of open science—from rigorous peer review, to metrics and evaluation to data analysis, huge data files, quick scanning of papers via artificial intelligence, and more—are driving many of the conversations we’re having about open reforms. He likens the search for open solutions to a game of musical chairs between the five big stakeholder groups—researchers, institutions, librarians, funders and publishers—where, when the music stops, someone is going to be left standing. We therefore need to construct our open science policies and strategies very carefully because these are going to have an outsize influence on the future of other open policies and scholarly communication stakeholders (Worlock 2020).

As mentioned earlier, the open source movement has probably already reached an equilibrium point, where there is a robust mix of private and commercial involvement that leads to strong and integrated product development as well as a large ecosystem of user support and new product development. There is nothing policy makers can “do” on this front—open source already clearly belongs to the world.

Open government and OER policies, also as mentioned earlier, draw on and incorporate the experiences and best practices of open access, open data and open source. Policies here will be greatly affected by what happens upstream—that is, if open access and open data policies become more internationally streamlined, and particularly if the incentives of open become more self-evident, this will facilitate the more rapid adoption and deployment of useful and sustainable open government solutions, and a wider availability of OER resources of all kinds.

Understanding, then, that each open field on its own is evolving toward (or has already reached) a state where “openness” can be broadly interpreted; that open access publishing solutions increasingly involve open data requirements; that open data and open source are already inseparable when it comes to large data sets; and that research-related open government and OER policies simply derive from best practices in open access, open data and open source; then what are the pros and cons of simply starting to think in terms of “open solutions?” On the pro side, integrating these conversations can help us integrate efforts and policies in ways that might dramatically and effectively accelerate the development and uptake of open policies. We might also be able to develop common language that can help facilitate the better communication of needs, processes and outcomes, and avoid one-size-fits-all approaches that try to shoehorn STM solutions into HSS settings or EU-centric payment schemes into regions of the world that can’t afford US\$10,000 APCs.

On the con side, these different open efforts have developed their own specialized concerns and constituencies over time. Would it make sense to anyone to try to lump free software definitions together with open access definitions? Probably not, although that needn’t be our approach. Each open pursuit at each research level should and will continue to be very diverse, but policy-wise, integrating these broad conversations can lead us to a more comprehensive understanding of open needs, and facilitate the recognition and development of common ground approaches to standards, infrastructure, sustainability and more. The open access, open data and open source communities can continue on as before, but at the UNESCO level, “open solutions” can be a new, comprehensive policy approach that recognizes these three tracks (along with open government and OER) are really all different verses from the same song, and are much more effective and meaningful when performed together.

Overall, the world may be ready to stop pushing specific technical and licensing regimes as our core objective for open, and instead let the debate move on to other narratives like “good data” or even, as we’re discussing here, common open solutions or common goals.<sup>161</sup> Because there’s still a paucity of hard data showing the benefits and impact of open, though, and because the solution space is so diverse and fragmented, it’s going to be challenging to get a critical mass of participants on board with this approach, at least at first. One of the most effective approaches, then (as described earlier in this report) may simply be to pick urgent problems and solve them—dispense with the ideology and practices altogether and start working with open solutions on common goals, and in doing so, prove the worth of openness, develop the appropriate best practices, and advance science, all in one fell swoop.

Overall, the world may be ready to stop pushing specific technical and licensing regimes as our core objective for open, and instead let the debate move on to other narratives like “good data”... or...common open solutions or common goals.

If we take this approach, there are three key barriers we need to overcome in designing a global “open solutions” approach to open:

- 1. Goals:** Without common goals, the open solutions marching band is playing five different songs and marching in five different directions. UNESCO can help create an environment where we can begin working together on common ground toward common goals. Doing this will help orient this community’s actions in such a way that our efforts and solutions can have more overlap and orientation.
- 2. Standards:** Without consistent standards and formats, open is just a jumble of information of little use to anyone, and prone to misinterpretation and misuse. We should focus our efforts first and foremost on what we want to do with open (point 1, goals), and secondly on what our data needs to look like in order to truly be useful and reusable. These are going to be field-level conversations at first, but over time we can begin looking at interdisciplinary connections and standards.
- 3. Silos:** At present, the world is creating all kinds of data silos. One risk here is that we’re making the open problem worse, not better. These silos operate under different licenses with different standards, both for operating and for the data they contain. Generally, they only communicate via metadata—a search across silos will pull in this metadata and give viewers a snapshot of what they’re searching for, but not a full-text or full-database search of the entire record. For this reason, broad ideas like EOSC and ASR (or their equivalents) should be given adequate policy attention.<sup>162</sup>

These are the three key barriers, but not the sum total of where we need to focus strategically and tactically, as discussed earlier. They are simply the major roadblocks that we’re going to encounter, sooner or later, as we set forth on this journey.

Of course, as we set forth, language will be another barrier, but it’s one we’re more likely to overcome without needing massive coordination and investment. Arguably, we’re at the cusp of a new era in research, matched only by the era that Francis Bacon tried to describe in the late 16<sup>th</sup> Century when

---

161. The extensive analysis in Davies 2019 reaches this same conclusion (although not necessarily with regard to using an open solutions approach). Smith 2020 advocates the open solutions approach and also increased community engagement in the development of open policies. Both reports represent a synthesis of open research conducted over many years by Canada’s International Development Research Centre (IDRC).

162. Here again, this isn’t an endorsement of ASR, but of the need to explore the possibility of broad solutions.

the new terms “science” and “discovery” were just emerging into the vocabulary and perceptions of society, increasingly used but with no common definition. Then as now, there were vast similarities and overlap in usage and experience surrounding these terms, but also significant differences. And then as now, recognizing it’s all the same general endeavor was and is vitally important. It was key then to transforming science into what we recognize today as the pursuit of knowledge, and it’s key now to transforming open solutions into more than just the sum of their separate parts.

With one sustained and effective push that embraces and nurtures the diversity in this space, we can truly usher in a new era of discovery and transformation. This embrace of diversity is critical. We don’t know where this adventure will take us. It’s important that we not decide in advance that our destination is across town when it might well be across the universe.

Only after we’ve built the open solutions ship and set sail should we worry about how to promote it widely. Some of this promotion will need to take place in advance in order to socialize this approach globally, but once the effort is underway, growth will be organic: once enthusiasm and interest in the broader open solutions approach begins to build, this approach will sell itself.<sup>163</sup> This principle is illustrated on the World Bank’s Open Data Toolkit page,<sup>164</sup> showing how to kick-start the virtuous cycle of open data, where more (and better) data encourages more data uptake and innovation (apps, etc.), which in turn promotes more (and better) data. This same principle was enunciated by OSI with regard to the emergence of an Open Renaissance, as discussed earlier. Once a global, inclusive effort is underway, it can advocate for actions like making data deposits simple (e.g., through EOSC), and working together on major challenges (e.g., starting with climate change).

Let’s illustrate the differences between this future world, on course to an Open Renaissance, with what would be required today if we were to try to mobilize scientists around the world to tackle climate change (beyond what the IPCC has been doing to gather data and make the case that climate change is actually happening). Such an effort would be (and has been) effectively impossible—requiring appeals to share information in formats that aren’t familiar (recalling that most researchers have never heard of terms like CC-BY, DORA and FAIR), using license formats that don’t inspire trust, and repository solutions that are as diverse as the data. Contrast this to an approach that sports a clear and unified goal, leaving aside all the details of exactly what can be shared and how it can be integrated (see next figure).

---

163. Assuming, of course, that enthusiasm and interest in this broader open solutions approach will actually begin to build. At this point we see plenty of evidence, including the perspectives and sentiments noted throughout this report, but we would need to assess this more carefully over time based on the global reaction to UNESCO promoting this idea more widely at the government level.

164. See <http://opendatatoolkit.worldbank.org/en/demand.html>

**Arguably, we’re at the cusp of a new era in research, matched only by the era that Francis Bacon tried to describe in the late 16<sup>th</sup> Century when the new terms “science” and “discovery” were just emerging into the vocabulary and perceptions of society, increasingly used but with no common definition. Then as now, there were vast similarities and overlap in usage and experience surrounding these terms, but also significant differences. And then as now, recognizing it’s all the same general endeavor was and is vitally important. It was key then to transforming science into what we recognize today as the pursuit of knowledge, and it’s key now to transforming open solutions into more than just the sum of their separate parts.**

FIGURE 15: ADVERTISING THE FUTURE OF OPEN



**share your data, save our planet**

- ✓ All science and government information welcome in any format
- ✓ All information entered into the All-Scholarship Repository---no hassle, no barriers
- ✓ Specialists work to integrate and share data
- ✓ Many collaboration opportunities available (research, writing, inventing)

See [climatechange.UNESCO.int](http://climatechange.UNESCO.int) for details.

This kind of approach grows through best practices and iteration. Left behind are all the questions about whose definition of open is correct, whose solutions are preferred, which standards are needed, what embargo period applies, what mix of paywalled and free content is needed, and on and on. Open is as open does. By dispensing with the debate and moving straight to action, we can let best practices and iteration sort out the details of what’s needed and what works, rather than first requiring global agreement on all these predicative elements, which is unlikely to happen anyway. We know that open is a spectrum of motives, outcomes and solutions. We can no sooner pick the “right” answers from this diversity than pick the right colors from the rainbow. Each is important, and each contributes to the greater whole. By focusing instead on the whole, we can find common ground for bold and needed action, and also create mechanisms by which our debates about best practices and requirements will settle themselves over time through more engagement from the research community.

The final recommendations from RDA’s 2014 Data Harvest Report are very much in line with this thinking, and while already six years old they are still very valid and relevant today. RDA’s recommendations for the future of open data (and this is generalizable to all kinds of open solutions) are to:

- 1. Require a data plan, and show it is being implemented.** We want a system to let researchers around the globe gather, store, share, re-use, re-interpret and act upon each others’ data – a global digital commons for science.
- 2. Promote data literacy across society, from researcher to citizen.** A massive programme of human engineering will be needed—in professional training, general education and cultural attitudes. Data sharing provides the foundation for a new branch of science. It must be acknowledged as such, and ranked alongside other major disciplines. This means fundamental changes in the incentives, career paths and academic status of what has, hitherto, been a fairly low profile discipline. At the same time, data science needs more professional bodies, and internationally transferable qualifications. This is homework for university administrators, education ministries and learned societies. Training in the use, evaluation and responsible management of data needs to be embedded in curricula, across all subjects, from primary school to university.
- 3. Develop incentives and grants for data sharing.** An informal estimate is that the infrastructure and operation of a truly effective data sharing system could cost on the order of 5 per cent of total research budgets. Scaled up globally, it seems massive. But so, too, the economic return. Funding should come from both private and public sectors—and for that to happen, proper incentives will be needed, some financial, some institutional.

4. **Develop tools and policies to build trust and data-sharing.** For one scientist to share data with another, there must be trust. Forget for the moment all the technical details: Sharing is a human behavior, and one that only happens under the right conditions. People have to trust that, by sharing data with others, they will gain more than they lose.
5. **Make sharing more turnkey.** There are several ways to go about this. One is to look into repository reform, whether this means building or supporting the construction of a single global repository with a single set of standards, or vetting a list of recommended repositories for data sharing. Many publishers already maintain such a list. The Registry of Research Data Repositories<sup>165</sup> is a full-scale resource of registered repositories across subject areas. Re3Data provides information on an array of criteria to help researchers identify the ones most suitable for their needs (licensing, certificates & standards, policy, etc.).
6. **Support international collaboration.** This is a global effort. Collaboration within nations is important, but the biggest benefits will come from cross-fertilisation with other regions, cultures and economic systems.
7. **Don't regulate what we don't yet understand.** One of the greatest risks in government is for politicians to act too soon. Until the world's scientific community has more experience with the costs and benefits steps towards regulation should be seriously considered. Issues such as privacy and ethics should be handled in consultation with the data and scientific communities, as well as with society at large.

With regard to RDA's recommendation number seven, we've already mentioned here how the APC flip approach has rapidly fallen into disfavor. Consider the recent history of the tech industry, as well, which has developed a number of "laws" over time to help better explain how information technology in the wild really develops, lest we think we're smart enough to control it. Some of the laws that apply to the open solutions debate include:

- **Amara's Law** (credited to scientist and futurist Roy Amara): This law states that "We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run."<sup>166</sup> In this case, there is no shortage of hyperbole in the open advocacy world—proclamations that open will result in better science, better access to science, and trickle down benefits for society. However, open solutions are only part of what's needed to improve science and its impact. We also need to look at issues like research integrity, peer review, impact factors, and so much more—issues that intersect with open (which is to say that open solutions can help address some of these challenges), but for which open solutions alone aren't sufficient. On the other end, "improve science" is probably a woeful underestimation of the power of open solutions to change the world—to truly usher in an era where we can make new connections between disciplines, discover new cures and make new insights due to the vast merging of data, and create science that inspires public involvement and confidence in ways that we've never seen before.
- **Thackara's Law:** "If you put smart technology into a pointless product, the result will be a stupid product." In 2005, tech critic John Thackara noted the tendency of designers to incorporate technology into products just because they can without asking whether they should.<sup>167</sup> We're

165. Re3Data, <https://www.re3data.org>

166. From Susan Ratcliffe, ed. 2016. Roy Amara 1925–2007, American futurologist. Oxford Essential Quotations. 1 (4th ed.). Oxford University Press. doi:10.1093/acref/9780191826719.001.0001.

167. See <https://amzn.to/37kPWdR>

doing the same thing with open solutions, trying to apply one-size-fits-all global solutions to research problems that don't exist. For instance, yes, there is a compelling case for making critically needed global health research more discoverable and interoperable. But does the same apply across all fields of research?<sup>168</sup> We sometimes characterize the open challenge as the need to develop free and immediate access to all research everywhere, but in fact, as discussed at length in the first part of this report, there are actually many different motives for open and a staggering array of different open needs and methods that vary by field. What if we solve each of these open research challenges on their own merits rather than trying to solve everything everywhere with one-size-fits-all solutions? Broader, systemic solutions may evolve with time and experience—generalizable best practices, for instance. We can start by tackling the challenges that researchers and policy makers want to tackle and are ready to tackle and build from there.

- **Reed's Law:** "The Value of a Network Increases Dramatically When People Form Subgroups for Collaborations and Sharing."<sup>169</sup> In 1993, Robert Metcalfe postulated that the value of a network is proportional to the square of the number of users of the system (for example, two people with telephones can only make one connection, whereas more people with telephones can make many more connections—the growth is not linear). In 2001, computer scientist David Reed postulated that the number of possible connections is even larger than Metcalfe guessed because the reality of information sharing on the web is that people often share in groups, and these groups often share information as well. Reed's Law helps explain why social networking sites work ("everyone's on Facebook because everyone's on Facebook."). In the case of open solutions, as long as we're looking at open as one-to-one transactional sharing, we're not envisioning the reality of how information is actually shared. We need to enable information sharing environments that allow diversity to take root, and that allow users and user groups to drive needs, solutions, and connections. In this kind of a robust and thriving environment, we're much more likely to realize the full potential of open than by pursuing solutions that see "open" as an end state in itself—by fixating on requirements that simply make information open in a particular way instead of focusing on how that information will actually be shared and used.

We know that open is a spectrum of motives, outcomes and solutions. We can no sooner pick the "right" answers from this diversity than pick the "right" colors from the rainbow. Each is important, and each contributes to the greater whole. By focusing instead on the "whole," we can find common ground for bold and needed action, and also create mechanisms by which our debates about best practices and requirements will settle themselves over time through more engagement from the research community.

We sometimes characterize the open challenge as the need to develop free and immediate access to all research everywhere, but in fact, as discussed at length in the first part of this report, there are actually many different motives for open and a staggering array of different open needs and methods that vary by field.

168. "Yes" is an acceptable answer. But as noted in the first section, there are many different open motives and methods.

169. See David P. Reed. 2001. The Law of the Pack. Harvard Business Review. <https://hbr.org/2001/02/the-law-of-the-pack>

- **Matthew Effect:** This principle stipulates that the rich get richer. It isn't a tech-centric policy per se but it is applicable in tech with regard to how better tech resources beget better outcomes, which lead back to better tech resources. In research, it is used to explain how top researchers accumulate more grant money and prestige and thereby maintain their status as top researchers and are hard to dethrone (Bol 2018). If our open reform efforts are created by the global north without adequate input and participation from all groups everywhere, then the open solutions we create and deploy are going to (and already doing so) favor those with more privilege. The richer countries will be able to fill open repositories with information, which makes their information and perspective even more dominant than now; these privileged countries will also be able to afford to buy or create the processing facilities needed to make use of that open information.

**If our open reform efforts are created by the global north without adequate input and participation from all groups everywhere, then the open solutions we create and deploy are going to (and already doing so) favor those with more privilege. The richer countries will be able to fill open repositories with information, which makes their information and perspective even more dominant than now**

Now, recall all these laws in the context of the history of the Internet. Who thinks it would have been wise in 1993 (when Metcalfe postulated that the value of a network is proportional to the square of the number of users) to regulate what the Internet should look like in the year 2020? In 1993, we were still struggling with how and why to build websites. Microsoft Outlook and AOL had just been launched. And yet here we are now, laying the groundwork for widespread open solutions by prescribing exactly what rules this movement needs to follow across all regions and fields

**Who thinks it would have been wise in 1993...to regulate what the Internet should look like in the year 2020? ...We risk suffocating the enthusiasm for open, the potential uptake of open, and the diversity and inclusion of open, under a stiff, unimaginative and inflexible blanket of definition and regulation.**

of study based on innocent guesses (like BOAI) made five, 10, and even 20 years ago, defining what inputs and outputs are acceptable, and what open itself means. There is no denying that we need best practices, standards and infrastructure—these were essential to the growth of the Internet as well. But there's a difference here, in that standards like html and TCP/IP are to the web what solutions like EOSC and xml are to open—platforms and solutions where all open information can be easily accessible and commingle in any licensing format. Licensing is not the common denominator that will set open free. This said, licensing is valuable for ensuring the reusability of large scientific datasets, but it may be a bridge too far for now (remember Thackara's Law) to also expect that all government datasets everywhere can be CC0 licensed, or that all history treatises henceforth need to be published in CC-BY format. We risk suffocating the enthusiasm for open, the potential uptake of open, and the diversity and inclusion of open, under a rigid and ill-fitting blanket of definition and regulation.

The alternative approach we're advocating here is to build a global framework—buttressed by education, support, tools, incentives, and more—for making all information available and accessible to the extent possible. The easier we make compliance, and the more inclusive our effort, the more information we collect; the more information we collect, the more usable our open effort becomes and the more used it gets (kicking in Reed's Law). And then, over time, as the benefits of this collaborative enterprise grow and become more apparent, the world of users will be self-incentivized to share in new

and better ways that are most appropriate for their needs and goals. Eventually through this continued engagement and dialogue we get to Amara’s Law: There’s really no predicting what will come of this newly opened universe, except to predict that the benefits are likely to exceed anything we can predict.

In the conclusion to the “State of Open” report (Davies 2019), the authors echo these sentiments and also OSI’s conclusions (mentioned later in this section), wherein “the question now for the broad open data movement is how to scale up and sustain the stakeholder engagement, infrastructure building, governance processes, capacity development, and cross-community networking that appear central to successful long-term open data initiatives, while not losing sight of the value of simply making data available and progressively enhancing its usability and usefulness over time.”

There are many paths that UNESCO can follow at this point:

- 1. Hold off on creating any new policies for now, but instead begin to convene** the many organizations, ideas, and policy initiatives that already exist in this space. It may be best to unify this space behind collective action first instead of creating an entirely new policy framework and hoping others will follow.
- 2. Create a very simple and broad policy framework for collaborative action**—one that embraces the diversity in this space and uplifts the global community to work together on the future of open. See Annex 1 for a draft UNESCO proclamation announcing this kind of approach.
- 3. Follow UNESCO’s open science roadmap recommendations**, whether or not these adequately address the concerns expressed in this report and the diversity of open opinions inside UNESCO. Over time, engage with this policy from the inside to help modify it as needed.

- 4. Adopt OSI’s Plan A** outright or as a blueprint for an eventual United Nations policy.

Regarding this final point, OSI’s Plan A encapsulates recommendations that have emerged from OSI’s five-year-long examination of the scholarly communication landscape. Over this period, high-level experts among OSI’s participants have shared, analyzed, promoted, critiqued and debated a wide array of perspectives via conferences, meetings, reports, and email correspondence.

Based on its five years of analysis, Plan A prescribes that, moving forward, the international community should:

- **Discover critical missing pieces** of the open scholarship puzzle so we can design our open reforms more effectively;
- **Design, build and deploy** an array of much needed open infrastructure tools to help accelerate the spread and adoption of open scholarship practices;

“...the question now for the broad open data movement is how to scale up and sustain the stakeholder engagement, infrastructure building, governance processes, capacity development, and cross-community networking that appear central to successful long-term open data initiatives, while not losing sight of the value of simply making data available and progressively enhancing its usability and usefulness over time.”

- **Work together on finding common ground perspective solutions** that address key issues and concerns (see OSI's Common Ground policy paper for more detail); and
- **Redouble our collective efforts to educate and listen** to the research community about open solutions, and in doing so design solutions that better meet the needs of research.

In pursuing these actions, the international community should:

- **Work and contribute together** (everyone, including publishers);
- **Work on all pieces of the puzzle** so we can clear a path for open to succeed;
- **Discover missing pieces of information** to ensure our efforts are evidence-based;
- **Embrace diversity.** No one group has a perfect understanding of the needs and challenges in this space, and different groups have different needs and challenges.
- **Develop big picture agreement on the goals ahead** and common ground approaches to meet these goals; and
- **Help build UNESCO's global open roadmap.**

Plan A also recommends that the community's work in this space be common-goal oriented, accountable, equitable, sustainable, transparent, understandable, and responsive to the research community. With regard to this last point, our open solutions must be developed carefully and in close collaboration with the research community. By doing so, we will ensure that research is protected from rash and ill-considered reforms during this transition and that it is well-served by the outcome of our efforts.

### Action steps

The specific action steps we need to take will vary widely by field because there are no one-size-fits-all solutions. Also, these steps will involve “layers” of action—some that are geared more generally toward improving the common ground of open solutions, and then very critically, other steps that are narrowly tailored to address the specific needs and challenges of different kinds of research. These specific plans are critical to ensuring that our open efforts aren't just pie-in-the-sky wishful thinking—we need to be able to connect our grand vision with the tough reality of implementation. This is akin to the so-called “last mile” in networking. It's one thing to build a nationwide 5G network, and quite another to ensure that consumers are able to use this network—to make connections to cell phones and homes that actually allow people to benefit.

These action steps should be geared toward recognizing the vast interconnectedness of the open space, and also toward rearranging how we approach the future of open. That is, instead of pursuing myriad disconnected open policies, each seeking to address one small piece of the open universe, and then lamenting how these policies are inadequate, leaving many needs unaddressed, and even creating unexpected adverse outcomes, we can flip the script. As discussed earlier (recall Figures 8, 13 and 14 showing how our open solutions efforts should flow from goals to actions instead of vice versa), let's think first about our long-term goals and needs for open, then figure out how existing efforts and processes can knit together to meet these needs and goals, and finally, work together to fill in the gaps by building new systems and processes.

To kick off this approach, beginning in mid-2021 the global open community should come together to build a strategic framework for collaborative action on open—a general framework describing common ground perspectives and needs in this space and upon which common ground actions can move forward, as described earlier. This framework will support many kinds of action as also described earlier in the “tactical” needs section of this report.

The timing of these subsequent actions will become clearer as the global open community develops more trust, more evidence, and a record of working together successfully. Timing will also depend on where we start—on what common ground we identify, what alliances we build, and how deeply torn we allow ourselves to become by policies that divide and conquer instead of unify and build—by Plan S and other approaches of this ilk that prosecute the case for open, versus collaborative approaches that learn and build together.

The actions that happen under this new global policy will speak louder than words. If a critical mass of governments, funders, institutions and researchers can come together behind an open solutions plan that is built on studies, infrastructure development, common ground work, and education/outreach; and if these groups can begin working together sooner rather than later on setting out the mileposts for this new path; and if such actions start to generate robust and measurable results, and start to develop effective and sustainable mechanisms for input, funding and growth; then this kind of policy approach will take root, and will gradually take the wind out of the sails for the prosecutorial approach. But if it doesn't take root, it will just be seen by open advocates as an attempt to subvert “true” open efforts (again, going back to the ideological overtones of this debate).

In order to avoid this fate, UNESCO should announce a broad alliance of partners in this policy at the same time the policy is announced—governments, funders, institutions and researchers—as well as a sustainable mechanism for management and support. None of this should be a difficult task or a high bar. Most governments, funders and institutions have not taken sides in this debate and are awaiting guidance; many have already worked with UNESCO on this issue over the years; and the time is right for action. Four months of consultations between March and June of 2021 might be sufficient to create a powerful initial list of signatories. Also, as we've discussed previously, most researchers already favor open solutions approaches that are broad and flexible.

A rough timeline for this work might look like this, assuming efforts can get underway in early 2021:



1. **March–June 2021:** Promote the open solutions approach amongst government policy makers and secure informal commitments for supporting this approach, including funding;
2. **March–June 2021:** In parallel, develop an organizational plan for this effort, including staffing and funding requirements, and accountability mechanisms;

...instead of pursuing myriad disconnected open policies, each seeking to address one small piece of the open universe, and then lamenting how these policies are inadequate... we can flip the script.... [L]et's think first about our long-term goals and needs for open, then figure out how existing efforts and processes can knit together to meet these...and finally, work together to fill in the gaps by building new systems and processes.

3. **July–September 2021:** Soft-launch this partnership and begin the process of identifying priorities and specific action plans;
4. **September–December 2021:** Begin working on key action items (studies, infrastructure, etc.);
5. **2022–2025:** Generate a track record of success with a wide range of open solutions projects, reporting regularly to all member groups and to the non-member international community;
6. **2026–2030:** Take the next step in leading the global charge on major key issues in open—from open standards efforts to metrics, tenure reform, and globally-integrated repositories (or a single ASR). Formalize global agreements that emerge from this robust working group; and
7. **2031 and beyond:** Usher in the Open Renaissance, and reap the rewards of a multitude of actions in this space that will be continually improving science and the value of science to society.

### Parallel actions

Plan A also proposes that, in parallel to these main points of action, the open community should work together to improve the relevance of open solutions to researchers and society by:

- **Picking a grand challenge like climate change and piloting an open solutions approach to solving it** (but making this challenge narrow enough to benefit from this approach—for instance, daylighting, integrating and promoting all academic, industrial and government research into the science and technology of large-scale carbon-dioxide and methane removal);
- **Creating zero-embargo compassionate use access portals** for patient families and for researchers combating health crises (whether through a new program or by strengthening and expanding the existing Emergency Access Initiative);
- **Creating a more robust Research-4-Life program** to improve access for lower-resourced regions and institutions; and
- **Considering how to modify current openness programs** to improve researcher uptake and engagement.

### Diversity and inclusion

Any narrowly tailored or one-size-fits-all approach is unlikely to succeed. Even just considering open science, an open future will necessarily need to be an inclusive movement that is informed and empowered by its vast diversity. This will lead knowledge and society to new discoveries and opportunities, globally and across all disciplines.

For open science, this broad vision includes:

- **Fostering open science literacy** through the development and promotion of open science curricula within educational programs, particularly ethics and AI (e.g., OS MOOCs);
- **Growing open science support** through the establishment of an assortment of open science roles throughout institutions (e.g., increased Dutch national support for open science);

- **Unpacking research papers** and providing discovery and credit mechanisms for research objects/components including software, data, workflows, etc. (e.g., FAIR, SCHOLIX);
- **Creating common assessment mechanisms** to track open science progress (e.g., OA Monitor, OA Switchboard);
- **Expanding funding and support for better connections** between research and the public (e.g., communicating science to public, citizen science);
- **Broadening computational support for data processing** and analysis-intensive research (e.g., EOSC, Binder);
- **Working in parallel to improve the integrity of science** (through open peer review, increased transparency, incentives/credits for reviewers, improving the quality of preprint services, and more);
- **Gathering lessons of experience** from standards and approaches used in related fields and endeavors (e.g., Schema.org);
- **Tending to the needs of open science culture** via early career communities, and offering advancement opportunities (e.g., RDA early career, Carpentries training); and
- **Continuing to think through national plans for open science** (e.g., the US National Academies' Open Science by Design, the European Commission on Open Science).

Our recommendation is for UNESCO to embrace the diversity of not just the open science space but the entire the open solutions spectrum. This ethos of diversity and inclusion is the foundation upon which a truly robust open future will be built.

Our greatest mistake as policy analysts and advisors would be to think we are smarter than this community's marketplace of ideas and outcomes, and recommend that UNESCO attempt to impose a rigid ideological order on this diverse and deeply complex landscape that could at best be ineffectual and at worst fracture the global solution space instead of unite it. Rather, we recommend a common-sense, collaborative, experience-driven open solutions policy that unites the disparate elements in this space, listens to all communities, embraces diversity, nurtures growth and innovation, and leads the way to an Open Renaissance that will truly transform the future of our society's relationship with knowledge.

### Indicators and dimensions

In the multilateral policy world, think of dimensions as the broad targets of reform, like education or health care, and indicators as the specific measures that show the progress of efforts to improve outcomes for that target—in the case of education, for example, years of schooling, attendance and literacy rates. Indicators and dimensions are keys to policy evaluation and a necessary part of policy development. However, they are also very easy to get wrong. Wrong indicators and dimensions can lead to wrong policies and outcomes. In an undertaking like ours where definitions are fluid and different stakeholders have different goals, it's almost inevitable that inventing indicators and dimensions to measure the current open solutions landscape will be both suspect and inaccurate. The future isn't hopeless, of course. It may eventually help, for example, to visualize the range of open outputs along the DARTS spectrum, as described earlier in this report. Still, for now, it will be better if the open indi-

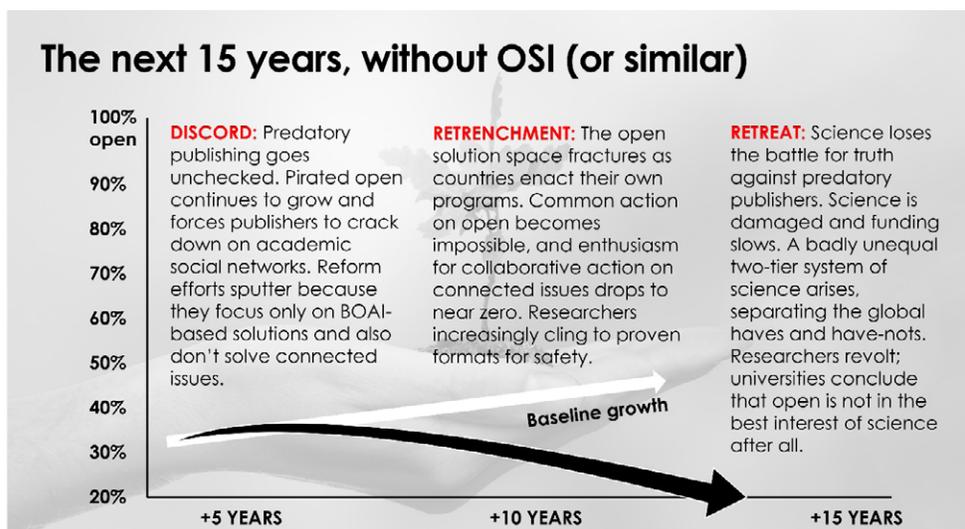
cators the global community wants and needs evolve from the goals we set together. These indicators may end up being very narrow, focusing on how effectively we're addressing a particular challenge in science or government, or how quickly books in the public domain are being digitized, or how widely popular software (like Wordpress) is being used. Or they may be discipline-specific, or measure progress by institution, region, or even globally. We simply won't know until we first discuss our common goals and then develop the assessments we want and need to measure our progress toward these goals.

Outside this tight and relevant bubble there are a number of dimensions indirectly related to openness that also need to be assessed. These aren't the dimensions we were looking for, but they are important nonetheless, including:

1. **Forecasting the impact of the fracturing of the open solution space** (as noted earlier in this report). What if we don't work together on the challenges ahead? As noted in OSI's Common Ground paper, continuing with our current go-it-alone approach may eventually result in competing regional solutions, where we end up with one open future for China, another for the EU, and still other futures for South America, Africa, and other regions, each working to solve their own unique concerns and perspectives. This approach may also force changes across diverse disciplines that may not work well (for example, open solutions that work in physics generally don't work at all in history), causing researchers in some fields to lose interest in an open future. Or it may lead to unintended consequences that don't necessarily benefit research, again causing a drop in interest. A go-it-alone approach also fails to address the significant concerns in government offices around the world that there are intellectual property and security ramifications of a vastly more open research world (see Poynder 2019)—not just sharing data freely but collaborating on research projects and even allowing certain foreign nationals to study at certain universities. Critically examining the impact of the current state of

**Our greatest mistake as policy analysts and advisors would be to think we are smarter than this community's marketplace of ideas and outcomes, and recommend that UNESCO attempt to impose a rigid ideological order on this diverse and deeply complex landscape that could at best be ineffectual, and at worst fracture the global solution space instead of unite it.**

FIGURE 16: THE FUTURE OF OPEN (BAD VERSION)



Source: Hampson 2018.

geopolitics on the future of open is a similarly important concern. Roger Schonfeld has written about what this means in terms of the ostracizing of China (Schonfeld 2020), as mentioned earlier in this report. Also worth noting are what impact rising nationalism in general might have on open attitudes, from the US to the UK to Brazil and elsewhere. As the world, and the academic world in particular, reels from the economic devastation of COVID, what impact will this have on how we view the imperative to share information funded by taxpayer dollars? Can we proactively address concerns like these by working together more effectively, or do we wait and react to future legislation that directs researchers to collaborate and share on the basis of nationality rather than merit? As noted earlier, there are larger, distinctly modern currents at work here that have the potential to utterly reshape our answers to the many questions posed by open research. If we work together, our ability as a community to deal with these currents will be informed, unified and strong. If we are a fractured community, however, where every country and stakeholder group is just in this for their own benefit and is pursuing their own national agenda and vision of the future, then there will be no bulwark against these nationalistic tides and the global effort to make research more open may suffer as a result.

- 2. Focusing more deeply on the current global inequities regarding access to knowledge.** Different regions of the world have markedly different needs and experiences with open solutions. While a global approach has its advantages, so too do local approaches that can target specific needs and shortcomings. More thinking is needed with regard to how well (or whether) a global open solutions approach might be able to specifically address regional and local inequities (compared and contrasted with current approaches). How can we better assess the differences that currently exist, as well as the impact of these differences, and also evaluate the impact that following different open roadmaps will have?
- 3. Situating these recommendations in the context of the UN’s Sustainable Development Goals.** An SDG integration analysis has already been conducted separately for open access, open data, open science, and open government (Digital Science 2019). What hasn’t been explored yet (and needs to be explored) is how a unified open solutions approach would interact with SDGs, and how to measure these interactions.
- 4. Resolving questions related to UNESCO’s internal decision making process with regard to open.** It is imperative that the different sectors inside UNESCO which have been developing open policy speak with one voice. Otherwise, this internal difference of opinion will cause confusion worldwide: Should people believe UNESCO’s Natural Sciences sector about open science, or UNESCO’s Communication and Information (CI) Sector about open access? The open definitions and approaches of these two sectors are completely different. This conflict will also put OSI (which was appointed by the CI Sector in 2018 to serve as UNESCO’s open access advisory body, and also provided input into UNESCO’s open science policy development process) in an awkward position of publicly disavowing UNESCO’s open science policy, at minimum. On a related note, this report has also left unaddressed specifically what is needed for UNESCO’s various open programs to integrate and/or develop, and what role a strengthened UNESCO open solutions program can realistically play.
- 5. Explicitly making connections between the many actors in the open space and identifying areas of common ground where they can begin working together.** Some of these connections (mostly with regard to open access) are explored in OSI’s Common Ground paper (Hampson 2020) but much more work is needed in this regard. Specifically, as noted in this report, while it is helpful to think of the search for open information as a constellation of unique circumstances that have four factors in common—needs, access, tensions and usability—it is also important from a policy development perspective (in order to facilitate

building bridges and making connections) to make a complete accounting of all the key actors, programs, policies and concepts used in open access, open data, open source, open science, open government, and OER, and understand how they overlap. A very rough list of the key actors, programs, policies and concepts is included in the annex section of this report.

- 6. Understanding the full impact of SciHub on open usage patterns.** SciHub continues to grow and thrive, and its continued use is affecting the open access and open science spheres in ways we don't fully understand yet. Napster is a parallel to this situation, fundamentally changing the music industry in the late 1990s while the music industry sued over copyright infringement and failed to quickly adapt to modes of more easily sharing and downloading music. Where is SciHub taking open access, can the publishing industry adapt, and what will (or should) the future of open access look like as a result?
- 7. Understanding the growth of predatory publishing and the impact of this practice on scholarly publishing.** Like the advent of desktop publishing in the 1990s, the growth of open has created a market opportunity that is being exploited. But this growth is not even, and not regulated.<sup>170</sup> There are a huge number of journals now that have fake credentials and publish anything for a fee. How can this market be regulated to ensure that research isn't damaged? Should it be regulated (and if so, by whom)? What impact are predatory journals having at present?
- 8. Detailing specific answers.** This report has painted a broad picture of how our current open solutions and policies can be developed going forward. However, it doesn't go into detail about exactly what these policies should look like. This is a deliberate omission. Solutions in the information world need to emerge from community needs and innovation and not be imposed. This is why many of the current open solutions, particularly in open access, are so defective: These solutions weren't developed by the user community, but for the user community. Solutions developed by the community might, for instance:
  - be responsive to the long-running concerns about the highly unpopular CC-BY license.<sup>171</sup> Perhaps the community might recommend developing new licenses instead like CC-EDU, CC-STM or CC-HSS that address the needs and concerns of specific audiences while still working toward open goals;
  - focus on a single repository solution like ASR or the equivalent—one turnkey location for all preprints and data managed by a new class of information managers worldwide (funded by a small overhead charge on all grant funding);
  - recommend a massive shift toward preprints and overlay journals;
  - focus more on making data a first class research object; and
  - focus more on the potential for artificial intelligence to scour papers and provide summaries, leads and insights.

---

170. If it happens, this regulation might take many different forms. For instance, the US Federal Trade Commission ruled in 2019 that predatory publishing giant OMICS had deceived its customers and should refund their publishing fees (in the amount of US\$50 million; see <http://bit.ly/3q13DRw>). In 2018, India started cracking down on predatory publishers operating within its borders (see Priyadarshini 2018). Some other manner of regulation (perhaps in the form of some kind of accreditation) that originates from universities, scholarly societies, or publishing associations might also be possible.

171. See, for example, Taylor & Francis 2019 for a recent survey of license popularity. Of course, there are some audiences who like the CC-BY license. But as with so many other policies in this space, the pushback against a one-size-fits-all licensing solution has more broadly led to a variety of workarounds like CC-BY-NC-ND, rather than a conversation about what open license features researchers want and need.

Monitoring the solution space is important and needs to be part of UNESCO's policy plan. However, this needs to be done in order to lay the groundwork for collaborative action on the future of open, not decide what this future will look like. Defining what the future will look like and then legislating options that restrict innovation and redirect investment will create barriers that in other eras would have prevented the invention of air travel, space exploration or nuclear energy.

We are, to a fault, incredibly bad at forecasting the future (how many of us are still using jet packs?). It's infinitely preferable to set broad goals for the future, and then engage the power of researchers and society to reach these goals without specifying exactly how this needs to be done. This power of innovation is the magic sauce that no policy maker, however smart, talented, well-intentioned or educated, will ever be able to replicate.

## CONCLUSION

To begin, let's summarize this conversation in terms of the key objectives of this report. We've covered a lot of technical and philosophical ground, but our first priority is to ensure that we've clearly answered the key questions we set out to answer.

- **The key findings (background & conceptual analysis)** from this report, substantiated by a wealth of research from other scholars, revolve around the fact that openness doesn't have a single or set definition. Open outcomes exist along a broad spectrum, and correspondingly, the definitions we apply to open and the open solutions we invent also vary quite widely. There is common ground, however, and working together on this common ground can help us come up with open solution approaches and policies that are much less disjointed than now and ultimately more beneficial for the broader cause of open information.
- **The key recommendation (steps UNESCO should take)** from this report is that going forward, UNESCO should host and promote an open solutions approach that focuses on identifying and achieving our common goals from open—not just grand challenges like tackling climate change, but also more discrete common goals like improving science. This goal-based approach is based on a widely-used and well-developed model (Theory of Change), and built on the common ground that unites all open solutions and communities.
- **The key assessments (needed indicators and dimensions)** the global community needs with regard to open solutions will evolve from this goals-directed effort. These indicators and dimensions may end up being very narrow or discipline-specific, or may measure progress by institution, region, or even globally. We won't know until we first discuss our common goals and our framework of action for achieving these goals, from which we will then be able to develop the kind of assessments and tools we want and need. In the meantime, there are a number of assessments that aren't directly related to openness that need to be considered, such as forecasting what impact the fracturing of the open solutions space will have; getting a better grasp on the scope and impact of current inequities with regard to knowledge access; mapping out the stakeholder space to understand explicitly where more and better connections can be made; understanding the impact of SciHub and predatory publishing on open usage; and other detailed concerns. .

These are all very broad takeaways, of course, and for a complex topic like this, it's important to understand the full context from which they arise. The starting point is this: Increasingly, the entire suite of open solutions is merging into open policy discussions—open data, open source, open science, open government, open educational resources, and more. At the same time, aggressive open policies are paving the road ahead, and in the process, separating the policy landscape into winners, unintended consequences, and even losers.

What is the best way to move forward on the future of open policy? As separate camps, each advocating their own narrow visions (which only work for narrow communities)? Or with a broad, inclusive vision that is motivated by shared goals and that works for all communities and for every open solution? UNESCO is championing a broad, open solutions approach that may become the new normal. Growing from a deep understanding of the global open landscape, across regions, fields, and stakeholder groups, and rooted in deep knowledge and experience rather than ideology, this approach can lead the world to a future of open that is without compare.

What does this approach look like and how does it fit in with the needs and perspectives of researchers, publishers, scholarly societies, universities, libraries, funders, and other stakeholder groups? How can we transition from our current straight-jacket mindset on openness to an approach that embraces the diversity of the open landscape and works for an open future that is imaginative, integrative, collaborative, and bold?

We've attempted to answer these questions in this report. The issues are complicated—we thank our readers for their investment in engaging with these complexities. In summary, it's an understatement to note that the open solutions space is vast and diverse. This report has put forward recommendations that attempt to define this space, and also define the best policy framework for this space.

With regard to the former, there is significant overlap across open solutions—more than enough to merit thinking about these solutions as having a common orientation in terms of goals. Doing so will allow us to build on common ground, developing the philosophy, methods, action details and community bridges that will lead to the development over time of robust, global and sustainable open policies. Our goals should not be to supplant various open solutions with one single definition, but to embrace the diversity in this space as a vibrant community of committed experts and invested stakeholders addressing a variety of needs with a continuum of solutions that are infinitely stronger when woven together than torn apart.

Many scholars agree that the way we approach open solutions in the future cannot be ideological, one-size-fits-all answers imposed unilaterally on researchers around the world. Too many needs and perspectives exist for any one stakeholder group or institution, however well intentioned, to design solutions that have any chance of working globally and being adopted by researchers everywhere.

**Many scholars agree that the way we approach open solutions in the future cannot be ideological, one-size-fits-all answers imposed unilaterally on researchers around the world. Too many needs and perspectives exist for any one stakeholder group or institution, however well intentioned, to design solutions that have any chance of working globally and being adopted by researchers everywhere.**

As Martin Paul Eve and Jonathan Gray note in their newly-released tome on open access (Eve 2020), “understandings drawn from a diverse set of geographic locations and histories are important for policymakers, for publishers, for academics, and for funders. Without such understandings, we become

**A global open policy needs to be...more than just words. Without creating a framework for open solutions along with a framework for policy engagement through which diverse voices can be heard and community action can be pursued together, our open actions and policies will not have the truly transformative impact we need them to have.**

trapped in repetitive loops, reinventing wheels, and lacking that most fundamental of activities for scholarly communication: communication itself.”

Rather, the only way to truly move forward with open is to develop a broad and flexible policy framework that embraces this diversity and can help guide the global community toward common goals, find answers to missing parts of the puzzle, build needed tools, and further support the community through education, training, and the development of standards and best practices.

These kinds of common ground open solutions are not being sought today. It is true that some parts of our community are trying to develop broad solutions by working together in limited communities or with like-minded institutions. This constellation of community engagement is admirable and should be embraced and nurtured. At best, however, this engagement in its current form, without fully considering the

needs and concerns of all stakeholders and regions of the world, will not lead to global open solutions that are well designed or sustainable.<sup>172</sup> And at worst, some of these efforts, like Plan S, are leading to unilateral policy solutions being imposed by major regions or funders without broadly consulting the global stakeholder community or research community. This dynamic is diverting attention away from efforts that are more reflective of the global community, and creating tensions in this community that are going to be difficult to overcome.

The push for open solutions is not an existential challenge. The clock is ticking. We need to act together now on issues ranging from COVID research to climate change to food security, education, and developing economic opportunity and security for people everywhere. A neutral arbiter in the international space needs to champion this cause. UNESCO would be an outstanding choice for this task, provided it doesn't first damage its credibility and the cause it is seeking to champion by unveiling a forthcoming open science policy that is currently more ideological than practical, and provided that it works quickly to involve other interested multilaterals in this effort as well. This group of policy managers needs to speak with one voice if it's going to be effective in the policy arena. Who does this matter. Legitimacy is key (does this organization have the necessary expertise, reach and gravitas for this task?), reinforced by neutral and objective management, significant and sustained funding, broad inclusion, and robust participation from the open community, particularly from researchers and research institutions.

The future manager of our global open solutions policy needs to be all these things, just as this future open solutions policy needs to be more than just words. Creating a framework for open solutions, along with a framework for policy engagement through which diverse voices can be heard and community action can be pursued together, is necessary for our open actions and policies to truly embody the transformative impact we need them to have.

All of this activity in the open solutions space, of course, is taking place against the backdrop of broader and deeper societal change. As society drifts toward ever greater expectations of openness and our technology demonstrates what can be done with open information, we also expect more openness

---

172. Which is not to say that these engagements aren't attempting to be inclusive, just that there is so very much to understand.

from research and huge benefits from this openness. In response to these pressures, different fields, regions and institutions have been taking different approaches to openness, and globally, stakeholder groups have been buffeted to varying degrees by changes in open requirements and practices—funding agencies, researchers, publishers, universities, libraries and more.

These paths have been varied not just because of varying needs, but because openness itself is at once a philosophy, a practice and a goal, experiencing wide variation by field, region, and type of open. The challenge now is how to incorporate these philosophies into new information sharing norms. Devising an approach that embraces the diversity of opinions and definitions across research fields, regions, and different open solutions is critical for understanding open and arriving at recommendations that aren't ideologically or philosophically wedded to one open field, solution or perspective.

The caveat to all this chaos is that outside this mindset bubble, most researchers have never heard of SPARC, BOAI, FAIR, and every other acronym that is so common in this space. Likewise, most research, government policy and academic groups (outside of libraries) are decidedly uncommitted. They want solutions that will bring clarity and guidance, and are still waiting after 20 years of debate. UNESCO has an opportunity to provide this kind of guidance now, but it needs to be realistic and future-oriented, not just a distillation of the same ideologically-grounded guidance that has already been heard and rejected.

The first step in getting there from here is to try to see the big picture—as this report has discussed, not only the overlap and parallels between open access, open data, open source, open science, open government and open educational resources, but also the broader policy picture and our broad, common goals for open from which we can then develop workable action plans and metrics. Our common ground policy foundation will have multiple, irregularly-shaped points of intersection on multiple points of common interest, including but not limited to: working together to get all research materials somewhere onto the DARTS open spectrum; working together to improve all open outcomes; working together to immediately improve access where it's most needed; working together to improve our understanding of open and its associated standards; working together to identify and address urgent needs; building and piloting new open systems and policies; and looking beyond.

What goals do we share? What challenges will we face along the way? What tools and systems should we build? To what end (specifically)? What role will artificial intelligence have in being able to synthesize research? What forms of research communication might be more efficient than articles in today's research environment (the answer will differ from one field to the next)?

How is all this different than our current approach? It's different because at the moment there are no major policy instruments or initiatives in the open solutions space that incorporate a globally diverse set of views and perspectives across all major stakeholder groups and that are driven by shared and community-driven goals.<sup>173</sup> Instead of relying on one-size-fits-all approaches and solutions to open

**Transitioning to this environment of collective understanding and action will take time and will involve robust and dedicated efforts. Creating UNESCO's open solutions policy framework is only part of the solution. The first step will be for a critical mass of supporters in this community to stop thinking of the open solutions universe as something that belongs to any single institution, stakeholder or interest group.**

---

173. Of course, some agencies and institutions have indeed been successful at integrating several open solutions approaches

powered by ideology we can create an inclusive open movement that is informed and empowered by diversity and opportunity.

If we can work together, then in due course we can reach an Open Renaissance where open is clearly defined and supported; open solutions are robust, inclusive, broad, scalable and sustainable; the global access gap is nonexistent; connected issues like impact factors are resolved; and more. After a time, this will lead to many other kinds of improvement for research, the research ecosystem will grow exponentially more powerful, new fields and directions will emerge based on easier and more robust interdisciplinary work, funding efficiency will improve, and discovery will accelerate. The social impacts of research will surpass today's levels, including improved literacy, public engagement, and public policy impact, with knowledge becoming more of a global public good, and society reaping the benefits.

To begin the path forward, we should agree that inclusive, effective, sustainable open solutions policies intended to benefit all researchers everywhere must be researcher-focused, collaborative, connected, diverse and flexible, informed, ethical and accountable, common-goal oriented, equitable, sustainable, transparent, understandable and simple, and beneficial to research—as noted earlier in this report. Our tactical approach will include activities like education, support, and building needed open infrastructure. All of these activities are vital and will have their own ecosystem of actors and actions.

Transitioning to this environment of collective understanding and action will take time and will involve robust and dedicated efforts. Creating UNESCO's open solutions policy framework is only part of the solution. The first step will be for a critical mass of supporters in this community to stop thinking of the open solutions universe as something that belongs to any single institution, stakeholder or interest group. Instead, we need to see open solutions as a rich, brilliant, diverse universe of ideas, actors and actions that share a wealth of common ground. The sooner we embrace this diversity and stop arguing over definitions and ownership, the sooner we will be able to truly begin building the open future we need, and that is so nearly within our grasp.

## REFERENCES

Adams, J. 2015. Impact of Open Science methods and practices on the economics of research and science: Case Studies from Life, Mathematical and Social Sciences. European Commission, Directorate-General for Research and Innovation. doi: 10.2777/14

Anderson, R. 2019. OSI Issue Brief 3: Deceptive Publishing. Open Scholarship Initiative. doi:10.13021/osi2019.2419

Archambault, E. 2018. Universalisation of OA scientific dissemination. <https://www.slideshare.net/scielo/2-eric-archambault>

Atkins, DE, JS Brown, and AL Hammond. 2007 (Feb). A Review of the Open Educational Resources (OER) Movement: Achievements, Challenges, and New Opportunities. Report to The William and Flora Hewlett Foundation. <https://bit.ly/33cJSmi>

Baldwin, M. 2018. Scientific Autonomy, Public Accountability, and the Rise of “Peer Review” in the Cold War United States. *ISIS*, volume 109, number 3.

---

for years already (particularly leading research institutions), but the emphasis of this recommendation is to create a broad and inclusive approach involving all major stakeholder groups and institutions.

- Berman, F. 2020. The Research Data Alliance – The First Five Years. White paper posted on the RDA website at <https://bit.ly/3pZUw9O>
- Björk, BC. 2018. Evolution of the scholarly mega-journal, 2006–2017. *PeerJ* 2018; 6: e4357. doi:10.7717/peerj.4357
- Bol, T, M de Vaan, A van de Rijdt. 2018. The Matthew effect in science funding. *Proceedings of the National Academy of Sciences* May 2018, 115 (19) 4887–4890. doi: 10.1073/pnas.1719557115
- Bosch Stiftung. 2019. 14th Berlin Debate on Science and Science Policy Who Owns Science? Reshaping the Scientific Value Chain in the 21st Century. Conference report. <https://www.bosch-stiftung.de/en/project/berlin-science-debate/berlin-debate-2019>
- Bosman, J, and K Kramer. 2017. Defining Open Science Definitions. Blog post accessed 5/22/20.
- Budzinski, O, T Grebel, J Wolling, et al. 2020. Drivers of article processing charges in open access. *Scientometrics* 124, 2185–2206. doi: 10.1007/s11192-020-03578-3
- Cai, X, CV Fry, and C Wagner. 2020. International Collaboration During the COVID-19 Crisis: Autumn 2020 Developments. SSRN. doi: 10.2139/ssrn.3729672
- Caviggioli, F, A De Marco, F Montobbio, and E Ughetto. 2020. The licensing and selling of inventions by US universities. *Technological Forecasting and Social Change*. Doi: 10.1016/j.techfore.2020.120189
- Chen, X, S Dallmeier-Tiessen, R Dasler, et al. Open is not enough. *Nature Phys* 15, 113–119 (2019). doi: 10.1038/s41567-018-0342-2
- Clarivate. 2020. Highly Cited Researchers 2020. <https://bit.ly/39XI99U>
- Colavizza G, I Hrynaszkiewicz, I Staden, K Whitaker, and B McGillivray. 2020. The citation advantage of linking publications to research data. *PLoS ONE* 15(4): e0230416. doi: 10.1371/journal.pone.0230416
- College Board. 2019. Trends in College Pricing. <https://research.collegeboard.org/pdf/trends-college-pricing-2019-full-report.pdf>
- COPE (Committee on Publication Ethics). 2018. Principles of Transparency and Best Practice in Scholarly Publishing. Draft policy by COPE/DOAJ/OASPA/WAME. <https://bit.ly/3nY4NS5>
- COUNTER. 2018. The Code of Practice for Research Data Usage Metrics. [https://www.projectcounter.org/wp-content/uploads/2019/02/Research\\_Data\\_20190227.pdf](https://www.projectcounter.org/wp-content/uploads/2019/02/Research_Data_20190227.pdf)
- Cousijn, H, A Kenall, E Ganley, et al. 2018/ A data citation roadmap for scientific publishers. *Sci Data* 5, 180259. doi: org/10.1038/sdata.2018.259
- Dallmeier-Tiessen et al. 2017. Connecting data publication to the research workflow. doi: 10.2218/ijdc.v12i1.533
- Davies, T, SB Walker, M Rubinstein, and F Perini (eds). 2019. *The State of Open Data: Histories and Horizons*. African Minds, IDRC. ISBN 9781928331957. Book pdf from <https://www.idrc.ca/en/book/state-open-data-histories-and-horizons>. HTML rendition at <https://www.stateofopendata.od4d.net/>
- Digital Science. 2019. The State of Open Data. doi: 10.6084/m9.figshare.9980783

Dubinsky, E. 2014. A Current Snapshot of Institutional Repositories: Growth Rate, Disciplinary Content and Faculty Contributions. *Journal of Librarianship and Scholarly Communication*, 2(3), p.eP1167. doi: 10.7710/2162-

Elofsson A, B Hess, E Lindahl, A Onufriev, D van der Spoel, and A Wallqvist A. 2019. Ten simple rules on how to create open access and reproducible molecular simulations of biological systems. *PLoS Comput Biol* 15(1): e1006649. doi: 10.1371/journal.pcbi.1006649

European Commission (EC). 2017. Providing researchers with the skills and competencies they need to practice Open Science. [https://ec.europa.eu/research/openscience/pdf/os\\_skills\\_wgreport.pdf](https://ec.europa.eu/research/openscience/pdf/os_skills_wgreport.pdf)

European Union (EU). 2015. Realising the European Open Science Cloud. [https://ec.europa.eu/research/openscience/pdf/realising\\_the\\_european\\_open\\_science\\_cloud\\_2016.pdf](https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf)

European Union (EU). 2018. The Future of Scholarly Publishing and Scholarly Communication

Eve, MP, and J Gray (eds.). 2020. Reassembling Scholarly Communications: Histories, Infrastructures, and Global Politics of Open Access. MIT Press. doi: 10.7551/mitpress/11885.001.0001

Faniel, I. 2020. What researchers need when deciding to reuse data: Experiences from three disciplines. NIH Workshop, Session 3: Enabling Data Reuse. <https://bit.ly/3mendwQ>

Fecher, B, and S Friesike. 2013. Open Science: One Term, Five Schools of Thought. SSRN. doi: 10.2139/ssrn.2272036

Good, R. 2016. Cultural Alzheimer: 404s and the Unstoppable Linkrot. Blog post on medium.com. <https://bit.ly/35ns1Jw>

Graf, C, D Flanagan, L Wylie, et al. 2019. The open data challenge: An analysis of 124,000 data availability statements, and an ironic lesson about data management plans. *Authorea*. October 31, 2019. doi: 10.22541/au.157253515.58528497

Hampson, G, M DeSart, J Steinhauer, EA Gadd, LJ Hinchliffe, M Vandegrift, C Erdmann, and R Johnson. 2020b (June). OSI Policy Perspective 3: Open science roadmap recommendations to UNESCO. Open Scholarship Initiative. doi: 10.13021/osi2020.2735

Hampson, G. 2019 (2nd ed.). OSI Policy Perspective 1: Plan S & the quest for global open access. Open Scholarship Initiative. doi: 10.13021/osi2019.2450

Hampson, G. 2020. OSI Policy Perspective 2: Common ground in the global quest for open research. Open Scholarship Initiative. doi: 10.13021/osi2020.2725

Hariri, RH, EM Fredericks, and KM Bowers. 2019. Uncertainty in big data analytics: survey, opportunities, and challenges. *J Big Data* 6, 44. doi: 0.1186/s40537-019-0206-3

Hendler, J, and T Pardo. 2012. A Primer on Machine Readability for Online Documents and Data. Blog post on data.gov. <https://bit.ly/33gXM6K>

Higman, R, D Bangert, and S Jones. 2019. Three camps, one destination: the intersections of research data management, FAIR and Open. *Insights*, 32(1), p.18. DOI: doi: 10.1629/uksg.468

Hinchliffe, L. 2019 (Apr 23). Transformative Agreements: A Primer. Scholarly Kitchen blog post. <https://scholarlykitchen.sspnet.org/2019/04/23/transformative-agreements/>

- Holdren, J. 2013. Memorandum for the heads of executive departments and agencies. US Office of Science and Technology Policy (OSTP). <https://bit.ly/3m5NzAG>
- Holmberg, K, J Hedman, TD Bowman, et al. 2020. Do articles in open access journals have more frequent altmetric activity than articles in subscription-based journals? An investigation of the research output of Finnish universities. *Scientometrics* 122, 645–659. doi: 10.1007/s11192-019-03301-xl
- Hrynaszkiewicz, I, N Simons, A Hussain, R Grant, and S Goudie. 2020. Developing a Research Data Policy Framework for All Journals and Publishers. *Data Science Journal*, 19(1), p.5. doi: 10.5334/dsj-2020-005
- Huston P, VL Edge, and E Bernier. 2019 (Oct 3). Reaping the benefits of Open Data in public health. *Can Commun Dis Rep.* 2019;45(11):252-256. doi: 10.14745/ccdr.v45i10a01
- Institute of Medicine (IOM). 2015. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk.* Washington, DC: The National Academies Press. doi: 10.17226/18998.
- Jefferies, M, et al. 2018. Data2paper: Giving Researchers Credit for Their Data. doi: 10.3390/publications7020036
- Johns, A. 1998. *The Nature of the Book: Print and Knowledge in the Making.* University of Chicago Press
- Kamerlin, L. 2019. Reaction of Researchers to Plan S: Too Far, Too Risky. <https://sites.google.com/view/plansopenletter/open-letter>
- Khodiary, V. 2020. What Role Can Publishers Play in the Open Data Ecosystem?. NIH Workshop on the Role of Generalist and Institutional Repositories to Enhance Data Discoverability and Reuse
- Khoo, S YS. 2019. Article Processing Charge Hyperinflation and Price Insensitivity: An Open Access Sequel to the Serials Crisis. *LIBER Quarterly*, 29(1), 1–18. doi: 10.18352/lq.10280
- Klar, S, Y Krupnikov, JB Ryan, K Searles, Y Shmargad. 2020. Using social media to promote academic research: Identifying the benefits of twitter for sharing academic work. *PLoS ONE* 15(4): e0229446. doi: 10.1371/journal.pone.0229446
- Knoth, P, and N Pontika. 2015. Open Science Taxonomy. Figshare. doi: 10.6084/m9.figshare.1508606.v3
- Korn, N, and C Oppenheim, C. 2011 (June). *Licensing Open Data: A Practical Guide (v. 2.0).* JISC circular. [http://discovery.ac.uk/files/pdf/Licensing\\_Open\\_Data\\_A\\_Practical\\_Guide.pdf](http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf)
- Laakso, M, L Matthias, and N Jahn. 2020. Open is not forever: a study of vanished open access journals. Pre-print. <https://arxiv.org/ftp/arxiv/papers/2008/2008.11933.pdf>
- Larivière, V, S Haustein, and P Mongeon. 2015. The Oligopoly of Academic Publishers in the Digital Era. *PLoS ONE* 10(6): e0127502. doi: 10.1371/journal.pone.0127502
- Lee, R. 2020 (Nov 12). Open Access: An Asian Perspective. Council of Science Editors (CSE) webinar, “Understanding Open Access, Plan S, and Other Global Initiatives in the Current Publishing Ecosystem: An Overview and International Perspective During a Pandemic”
- Lewis, CL. 2018. The Open Access Citation Advantage: Does It Exist and What Does It Mean for Libraries?. *Information Technology and Libraries*, 37(3), 50-65. doi: 10.6017/ital.v37i3.10604

- Mangravite, L., A Sen, JT Wilbanks. 2020. Mechanisms to Govern Responsible Sharing of Open Data: A Progress Report <https://sage-bionetworks.github.io/governanceGreenPaper/v/7ef288619fb46e6d-9319433c64fbc5bef6250fe7/>
- McNutt, M. 2019 (Jan). Opinion: “Plan S” falls short for society publishers—and for the researchers they serve. *Proceedings of the National Academy of Sciences*, 201900359. doi: 10.1073/pnas.1900359116
- Molloy, J. 2011. The Open Knowledge Foundation: Open Data Means Better Science. *PLoS Biol* 9(12): e1001195. doi: 10.1371/journal.pbio.1001195
- Motter, K. 2020 (Nov 12). Untitled presentation. Council of Science Editors (CSE) webinar, “Understanding Open Access, Plan S, and Other Global Initiatives in the Current Publishing Ecosystem: An Overview and International Perspective During a Pandemic.”
- Murray, A, and A Ireland. 2018. Provosts’ Perceptions of Academic Library Value & Preferences for Communication: A National Study. *College & Research Libraries*, [S.l.], v. 79, n. 3, p. 336, apr. 2018. ISSN 2150-6701. doi: 10.5860/crl.79.3.336
- Murray-Rust, P, C Neylon, R Pollock and J Wilbanks. 2009. *Panton Principles*. <https://pantonprinciples.org/>
- NAS. 2018. *Open Science by Design*. doi: 10.17226/25116
- National Academies of Sciences, Engineering, and Medicine (NASEM). 2020. *Reflections on Sharing Clinical Trial Data: Challenges and a Way Forward*. Workshop proceedings. Washington, DC: The National Academies Press
- National Academies of Sciences, Engineering, and Medicine (NASEM). 2020. *Reflections on Sharing Clinical Trial Data: Challenges and a Way Forward: Proceedings of a Workshop*.
- Newton, C. 2020. Research and open access from low- and middle-income countries. doi: 10.1111/dmnc.14513
- Nicholaş, D., B Rodríguez-Bravo, A Watkinson, C Boukacem-Zeghmouri, E Herman, J Xu, A Abrizah, and M Świgoń. 2017. Early career researchers and their publishing and authorship practices. *Learned Publishing*, 30: 205-217. doi: 10.1002/leap.1102
- NIH. 2020 (Feb). NIH Workshop on the Role of Generalist and Institutional Repositories to Enhance Data Discoverability and Reuse. <https://datascience.nih.gov/data-ecosystem/NIH-data-repository-workshop>
- NSF. 2020. *The State of US Science & Engineering*. National Science Foundation. <https://ncses.nsf.gov/pubs/nsb20201/downloads>
- OECD. 2015. *OECD Science, Technology and Industry Scoreboard 2015: Innovation for growth and society*, OECD Publishing, Paris. doi: 10.1787/sti\_scoreboard-2015-en
- OECD. 2015. “Making Open Science a Reality”, *OECD Science, Technology and Industry Policy Papers*, No. 25, OECD Publishing, Paris. doi: 10.1787/5jrs2f963zs1-en
- OSI. 2016. Report from the “What is Open?” Workgroup. Open Scholarship Initiative 2016 conference. doi: 10.13021/G8XK5R

- OSI. 2016a. Report from the “What is Publishing?” Workgroup 2. Open Scholarship Initiative 2016 conference. doi: 10.13021/G8CS33
- OSI. 2017a. OSI2017 Conference Report. <http://osiglobal.org/wp-content/uploads/2017/11/OSI2017-summary-report-v2final.pdf>
- OSI. 2017b. Report from the Standards, Norms, and Best Practices Workgroup. Open Scholarship Initiative 2017 conference. doi: 10.13021/G8osi.1.2017.1932
- Parsons, J. 2016 (Mar 4). Who Pays for Open Access? Library Journal
- Pinfield, S, S Wakeling, D Bawden, and L Robinson. 2020. Open Access in Theory and Practice: The Theory-Practice Relationship and Openness. Routledge.
- Piwowar, H, J Priem, V Larivière, JP Alperin, L Matthias, B Norlander, A Farley, J West, and S Haustein. 2018. The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. PeerJ 6:e4375. doi: 10.7717/peerj.4375
- Plutchak, TS. 2018. OSI Issue Brief 1: What do we mean by open? Open Scholarship Initiative. doi: 10.13021/osi.v3i0.2367
- Poynder, R. 2019. Open access: Could defeat be snatched from the jaws of victory? <https://richardpoynder.co.uk/Jaws.pdf>
- Poynder, R. 2020. Open access: “Information wants to be free”? <https://poynder.blogspot.com/2020/12/open-access-information-wants-to-be-free.html>
- Priyadarshini, S. 2018 (Aug 28). India targets universities in predatory-journal crackdown. Nature. doi: 10.1038/d41586-018-06048-2
- RDA (Research Data Alliance). 2014. The Data Harvest: How sharing research data can yield knowledge, jobs and growth. <https://bit.ly/33h0SYG>
- Reller, T, E Torie and W Gunn. 2020. Access for patients – and other ways to get Elsevier articles without a subscription. Elsevier blog post. <https://bit.ly/2UXSw3x>
- Research England. 2018. Monitoring sector progress towards compliance with funder open access policies. UKRI (RE-P-2018-05). <http://re.ukri.org/documents/2018/research-england-open-access-report-pdf/>
- Rodrigues RS, E Abadal, and BHK de Araújo. 2020 Open access publishers: The new players. PLoS ONE 15(6): e0233432. doi: 10.1371/journal.pone.0233432
- Ruijter E, S Grimmelikhuijsen, J van den Berg, and A Meijer. Open data work: understanding open data usage from a practice lens. International Review of Administrative Sciences. 2020;86(1):3-19. doi: 10.1177/0020852317753068
- Scaria, AG, and R Shreyashi. 2018. Open Science India Report. OSF Preprints. doi: 10.31219/osf.io/aj9gw
- Schonfeld, R. 2020. Global Science and the China Split. Ithaka S+R blog post. doi: 10.18665/sr.314295
- Science-Metrix. 2017. Imagining a Gold Open Access Future: Attitudes, Behaviors, and Funding access availability of scientific publications. [https://www.science-metrix.com/sites/default/files/science-metrix/publications/science-metrix\\_open\\_access\\_availability\\_scientific\\_publications\\_report.pdf](https://www.science-metrix.com/sites/default/files/science-metrix/publications/science-metrix_open_access_availability_scientific_publications_report.pdf)

Science-Metrix. 2018. Analytical support for bibliometrics indicators: Open access availability of scientific publications

Siler K, S Haustein, E Smith, V Larivière, and JP Alperin. 2018. Authorial and institutional stratification in open access publishing: the case of global health research. *PeerJ* 6:e4269. doi: 10.7717/peerj.4269

Smith, M, and KMA Reilly (eds). 2013. *Open development: Networked innovations in international development*. IDRC (MIT Press). ISBN: 9780262525411

Smith, M, and RK Seward (eds). 2020. *Making open development inclusive: lessons from IDRC research*. IDRC (MIT Press). ISBN 9780262539111

Springer Nature. 2020. *Open for all: Exploring the reach of open access content to non-academic audiences*. White paper. doi: 10.5281/zenodo.4143313

Staunton, C., S Slokenberga, and D Mascalzoni. 2019. The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks. *Eur J Hum Genet* 27, 1159–1167. doi: 10.1038/s41431-019-0386-5

STM. 2020. *Introducing the STM2020 Research Data Year*. <https://www.stm-researchdata.org/>

Stuart, D, G Baynes, I Hrynaszkiewicz, K Allin, D Penny, M Lucraft, and M Astell. 2018. *Practical challenges for researchers in data sharing*. Springer Nature white paper. doi: 10.6084/m9.figshare.5975011

Tan, SC, D Flanagan, E Morris, et al. 2020 (Jul 8). *Research data repositories chosen by researchers across broad range of disciplines, from an analysis of 145,000 data availability statements*. Authorea. doi: 10.22541/au.159422974.49069472

Taylor & Francis. 2019. *Taylor & Francis Researcher Survey*. <https://bit.ly/3koHgrX>

Tennant, J, A Ritwik, K Baždarić, D Brassard, T Crick, DJ Dunleavy, TR Evans, et al. 2020. “A Tale of Two ‘opens’: Intersections Between Free and Open Source Software and Open Scholarship.” *SocArXiv*. March 6. doi: 10.31235/osf.io/2kxq8

Tennant, J, JE Beamer, J Bosman, B Brembs, NC Chung, G Clement, T Crick, et al. 2019 (January). *Foundations for Open Scholarship Strategy Development*. doi: 10.31222/osf.io/b4v8p

Tennant, J, N Bielczyk, BG Tzovaras, P Masuzzo, and T Steiner. 2020a. *Introducing Massively Open Online Papers (MOOPs)*. *KULA: knowledge creation, dissemination, and preservation studies*, 4(1), p.1. doi: 10.5334/kula.63

Tenopir, C, E Dalton, L Christian, M Jones, M McCabe, M Smith, and A Fish. 2017. *Scenarios among Authors of Academic Scholarship College & Research Libraries*, 78(6), 824. doi: 10.5860/crl.78.6.824

UNESCO. 2018. *World Trends in Freedom of Expression and Media Development*. ISBN 978-92-3-100242-7

VI BRISPE (6th Brazilian Meeting on Research Integrity, Science and Publication Ethics). 2020 (Aug). *Research Integrity, Open Science and Authorship: Challenges and Opportunities for the Research Systems*. Proceedings (unpublished as of December 2020)

Wellcome Trust. 2018. *Wellcome and COAF open access spend 2016/17*

Wells, HG. 1937. World Brain

White, N. 2020. Open Access in Action: The Last 12 Months Reviewed. Wiley Publishing internal report. <https://bit.ly/3lIQAHm>

Wiley. 2018b. Interview of Fiona Murphy. <https://www.wiley.com/network/archive/open-research-and-data-sharing-are-we-hearing-what-researchers-are-telling-us>

Wiley. 2019a. Wiley Open Research Survey. <https://www.wiley.com/network/archive/researchers-tell-us-about-their-open-research-practices>

Wiley. 2019b. Wiley Open Data Survey <https://www.wiley.com/network/researchers/licensing-and-open-access/what-do-researchers-think-about-open-data>

Wilkinson, MD. et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18.

Woodward, G. 2017. The Top 5 Training Needs of Early Career Researchers. Wiley blog post. <https://bit.ly/37VOcl0>

Wootton, D. 2015. *The Invention of Science: A New History of the Scientific Revolution*. HarperCollins

Worlock, D. 2020 (Dec 7). Scholarly Communication Re-Invented. Keynote presentation to the 2020 Frankfurt Academic Conference, "A New Era for Academic Publishing in Europe and the US." <https://bit.ly/39RWYcl>

## ANNEX 1: DRAFT UNESCO PROCLAMATION ON OPEN SOLUTIONS<sup>174</sup>

### Preamble

Recognizing the urgency of addressing complex and interconnected environmental, social, health and economic challenges for the people and the planet;

Acknowledging the vital importance of factual information to respond to these challenges;

Committed to leaving no one behind with regard to access to factual information;

Recalling that one of the key functions of UNESCO is to maintain, increase and diffuse knowledge by encouraging cooperation among the nations in all branches of intellectual activity;

Recognizing the potential of open solutions to reduce existing global inequalities, accelerate progress toward needed solutions, and achieve the United Nations' Sustainable Development Goals;

Further recognizing that open solutions have a wide variety of definitions, motivations, goals, and adaptations that vary widely by field, region, and institution;

And taking fully into account, in the adoption and application of this recommendation, the great diversity of laws, regulations and customs which will determine how this recommendation will ultimately be adopted, be it hereby resolved that Member States:

1. Adopt the following recommendation on open solutions;
2. Take appropriate steps to give effect within their jurisdictions to the principles of this recommendation; and
3. Engage with UNESCO in the further development of the open solutions roadmap and action items at such dates and manner to be determined, in pursuance of this recommendation.

### I. OBJECTIVE OF RECOMMENDATION

The objective of this Recommendation is to provide an international framework for open solutions policy and practice that recognizes the broad global diversity of open solutions actions and perspectives, and that also sets forward a roadmap for continued international engagement on open solutions and a growth in open solutions best practices, standards, and accomplishments.

### II. DEFINITION OF OPEN SOLUTIONS

Open solutions is an umbrella concept that is being used by UNESCO to describe various movements and practices variously aimed at making scientific knowledge, methods, data and evidence more available and accessible (particularly to researchers from lower resourced regions and institutions); increasing scientific collaboration and the potential from this collaboration; increasing the sharing of

---

174. Much of the proposed language here is distilled and adapted from UNESCO's draft recommendation on open science.

information of all kinds for the benefit of science and society; opening knowledge to societal actors beyond institutionalized communities; improving the reliability and factualness of information through increased transparency and replicability; and other similar motives. Not all open solutions practices share the same motives, nor the same goals, methods, actions or stakeholders. Additionally, a complex of intersecting and overlapping ‘open’ elements are generally involved in the conduct of open solutions, including but not limited to open access (generally meaning users being able to gain free access to research reports published in science journals), open data (generally meaning that research data is licensed in such a way that it can be reused without permission), open source/code, open government, open educational resources, and more.

### III. ROADMAP FOR ACTION

The roadmap of global action on open solutions adopted by Member States should itself be open and transparent, developed by Member States and the full international community of stakeholders, and respecting the wide diversity of needs and perspectives surrounding open solutions. At its core, UNESCO, the United Nations, and UN Member States should avoid regulating what we don’t yet fully understand, or adopting one-size-fits-all solutions that may make open solutions adoption impossible for some, or making open solutions dynamics worse for developing countries. In general, our approach must be:

1. **USER-FOCUSED.** Open solutions tools, services and options must be developed with heavy input from the research community, with solutions and approaches driven by user needs and concerns;
2. **COLLABORATIVE.** Successful and sustainable solutions will require broad collaboration, not just to ensure that all perspectives are considered, but also to ensure there is broad ownership of ideas;
3. **CONNECTED.** There are a great many interconnected issues in the open solutions space. Developing an effective future for open solutions will require a systemic approach;
4. **DIVERSE AND FLEXIBLE.** There are no one-size-fits-all solutions to open solutions reform. Instead, there are many different pathways to reform, likely including many that have not yet been conceived or deployed. Diversity, creativity and flexibility in this undertaking should be paramount, at the same time noting that common ground actions will be critical;
5. **INFORMED.** We need a better understanding of key issues in open solutions before moving forward. The more accurate and honest our assessments, the more accurate and honest our reform efforts can be, the easier these efforts will be to promote, and the more successful they will be;
6. **ETHICAL AND ACCOUNTABLE.** We need enforceable, community-developed standards to ensure the integrity of publishing, archiving, and other related activities and products, and to ensure that unethical approaches are not embraced;
7. **COMMON GOAL ORIENTED.** We must discuss and plan for what the future of open solutions means, beyond just having easier access to information, including defining what we plan to do with open information, where we need data interoperability, what tools and procedures we need to achieve this interoperability, and more. By doing this, we can better focus on and strive for our community’s common goals;

8. **EQUITABLE.** People everywhere need to be able to access and contribute content to the global body of research information with minimal barriers. To the extent practicable, information—particularly information central to life and health—should not be unreasonably constrained by issues such as high access costs, poor journal indexing, and a lack of capacity-building programs;
9. **SUSTAINABLE.** Open solutions reform approaches need to be sustainable, which flows from all the other elements in this list. The reform solutions we design need to be achievable, affordable, popular, effective, and otherwise maintainable over the long term;
10. **TRANSPARENT.** The global community needs to maintain as much transparency as possible in this effort (with regard to pricing, usage, ownership, and so on) in order to maintain trust in this effort;
11. **UNDERSTANDABLE AND SIMPLE.** The global community needs to agree on high-level, common-ground goals for open solutions reform—a general set of goals that are understandable, achievable, and adaptable. By setting out general goals that can be easily achieved, participation can be made simple and easy, with low barriers to entry.
12. **BENEFICIAL.** In the end, these reforms need to benefit research first and foremost. While the argument to improve benefits to society is central, these benefits need to be matured carefully, deliberately, and realistically in order to ensure that societal benefits are indeed being conveyed as intended, and that research is not being harmed in the process.

#### IV. AREAS OF ACTION

With this roadmap for action in mind, the four general areas of action that should be supported by Member States are to:

1. **DISCOVER** critical missing pieces of the open solutions puzzle so we can design our reforms more effectively;
2. **DESIGN**, build and deploy an array of much needed open infrastructure tools to help accelerate the spread and adoption of open solutions practices;
3. **WORK TOGETHER** on finding common ground perspective solutions that address key issues and concerns; and
4. **EDUCATE** and listen to the research community about open solutions, and in doing so design solutions that better meet the needs of research.

To the extent possible and at a more detailed level, Member States are also recommended to pursue these 10 specific areas of action, taking into account their individual political, administrative and legal contexts:

1. **PROMOTE** a common understanding of open solutions as defined in this recommendation within the scientific community and among the different open solutions actors at the institutional, national and regional levels;
2. **ENSURE** that public research funders require open solutions practices and that all information outputs from publicly funded efforts are as open as possible, and only as closed as necessary;

3. **EMBRACE** and combine the efforts of the many different actors in the open solutions space, including research funders, universities, journals, and scientific journals;
4. **ENGAGE** the private sector in discussion about the ways in which the scope of open solutions principles and priorities can be enlarged and mutually shared;
5. **DEVELOP** or encourage policies, including those at the institutional and national levels, that are supportive of a transition to open solutions. This includes but is not limited to helping establish regional and international funding mechanisms for promoting and strengthening open solutions; supporting the creation and maintenance of effective collaborative networks to exchange best open solutions practices and policies; promoting cooperation among countries in capacity building for data management and stewardship; and investing in open solutions infrastructure and services;
6. **COMBAT** the practice of predatory publishing, wherein ‘fake’ publishers publish anything for a fee regardless of merit and without adequate gatekeeping mechanisms in place, and in doing so corrupt the global body of factual information;
7. **REVIEW** research assessment and career evaluation systems in order to align them with the principles of open solutions;
8. **LEARN** more about the open solutions space by helping fund additional studies and fact-finding efforts as needed to ensure that open solutions efforts are fully informed and optimally effective and efficient;
9. **COLLABORATE** on finding solutions to urgent science-based challenges such as climate change, medical research and food security. Demonstrating the value of open solutions collaboration efforts will advance the cause of open solutions while at the same time providing an urgently needed service to humankind;
10. **ENTRUST** UNESCO with the mission to coordinate, in consultation with stakeholders and Member States, the development and adoption of an evolving and detailed global framework for action on open solutions, which will guide and stimulate international cooperation to advance open solutions for the benefit of humankind and planetary sustainability.

## V. MONITORING

Member States should, according to their specific conditions, governing structures and constitutional provisions, monitor policies and mechanisms related to open solutions using a combination of quantitative and qualitative approaches, as appropriate. The UNESCO-established coordinating body, in a mechanism to be determined, will collect these statistics and share them with Member States.

## ANNEX 2: THE INTERNATIONAL OPEN DATA CHARTER

The Open Data Charter has been signed by 22 countries, 55 local and regional governments, and 68 non-state organizations around the world. The charter is non-binding. While the focus of this charter is on open government, it specifically tackles the data that governments generate. The charter, at present, doesn't make specific mention of open access or open science. See <https://opendatacharter.net/principles/> for more details.

### PREAMBLE

Open data is digital data that is made available with the technical and legal characteristics necessary for it to be freely used, reused, and redistributed by anyone, anytime, anywhere.

1. The world is witnessing a significant global transformation, facilitated by technology and digital media, and fueled by data and information. This transformation has enormous potential to foster more transparent, accountable, efficient, responsive, and effective governments and civil society and private sector organizations, and to support the design, delivery, and assessment of sustainable development goals at a global scale.

Open data is at the center of this global shift.

2. Building a more prosperous, equitable, and just society requires that governments are transparent and accountable, and that they engage regularly and meaningfully with citizens. Accordingly, there is an ongoing global data revolution that seeks to advance collaboration around key social challenges, provide effective public oversight of government activities, and support innovation, sustainable economic development, and the creation and expansion of effective, efficient public policies and programs.

Open data is crucial to meeting these objectives.

3. Open data enables governments, citizens, and civil society and private sector organizations to make better informed decisions. Effective and timely access to data helps individuals and organizations develop new insights and innovative ideas that can generate social and economic benefits, improving the lives of people around the world.

Open data presents an opportunity that must be seized.

4. Open data allows user to compare, combine, and follow the connections among different datasets, tracing data across a number of programs and sectors. When data can be effectively combined and compared, it can help highlight trends, identify social and economic challenges and inequities, and benchmark progress in public programs and services.
5. Open data can empower governments, citizens, and civil society and private sector organizations to work toward better outcomes for public services in areas such as health, education, public safety, environmental protection, human rights, and natural disasters.
6. Open data can contribute to the generation of inclusive economic growth by supporting the creation and strengthening of new markets, enterprises, and jobs. These benefits can multiply

as more civil society and private sector organizations adopt good open data practices and share their own data with the public.

7. Open data can help improve the flow of information within and among governments, and make government decisions and processes more transparent. Increased transparency promotes accountability and good governance, enhances public debate, and helps combat corruption.
8. Open data presents opportunities to provide innovative, evidence-based policy solutions and support economic benefits and social development for all members of society. Open data can do this by, for example:
  - Supporting evidence-based policy making: Encouraging governments' use of data in policy development and evidence-based decision-making, which enables improved public policy outcomes and underpins sustainable economic and social development;
  - Enabling cross-sector collaboration: Supporting collaboration among governments, citizens, and civil society and private sector organizations on the design of policies and the delivery of better public services;
  - Following the money: Showing how and where public money is spent, which incentivizes governments to demonstrate that they are using public money effectively;
  - Improving governance of natural resources: Increasing awareness about how countries' natural resources are used, how extractives revenues are spent, and how land is transacted and managed;
  - Monitoring impact: Supporting assessments of the impact of public programs, which in turn allows governments and civil society and private sector organizations to respond more effectively to the particular needs of local communities.
  - Promoting equitable growth: Supporting sustainable and inclusive growth through the creation and strengthening of markets, enterprises, and jobs;
  - Geolocating data: Providing geospatial and earth observation references, which support comparability and interoperability and effective analysis by allowing data to be layered geographically; and
  - Improved decision-making: Enabling citizens to make better informed choices about the services they receive and the service standards they should expect.

When used in these ways, open data is a key public good which people can use to generate value, insights, ideas, and services to create a better world for all.

9. We, the adherents to the International Open Data Charter, recognize that governments and other public sector organizations hold vast amounts of data that may be of interest to citizens, and that this data is an underused resource. Opening up government data can encourage the building of more interconnected societies that better meet the needs of our citizens and allow innovation, justice, transparency, and prosperity to flourish, all while ensuring civic participation in public decisions and accountability for governments.

10. We therefore agree to follow a set of six principles that will be the foundation for access to data and for the release and use of data. These principles mandate that data should be:
  1. Open by Default
  2. Timely and Comprehensive
  3. Accessible and Usable
  4. Comparable and Interoperable
  5. For Improved Governance and Citizen Engagement
  6. For Inclusive Development and Innovation
11. We will develop action plans or identify existing mechanisms or policies in support of the implementation of the Charter principles and associated resources. We agree to commit the necessary resources to work within our political and legal frameworks to implement these principles in accordance with the technical best practices and time frames set out in our action plans.
12. This Charter has been developed with a view to adoption by governments of all levels and by multilateral institutions. While the focus of the Charter is on open government data, other organizations, such as those from civil society or the private sector, are also welcome to adopt these principles.

#### **PRINCIPLE 1: OPEN BY DEFAULT**

13. We recognize that the term “government data” includes, but is not limited to, data held by national, regional, local, and city governments, international governmental bodies, and other types of institutions in the wider public sector. The term government data could also apply to data created for governments by external organizations, and data of significant benefit to the public that is held by external organizations and related to government programs and services (e.g. data on extractives entities, data on transportation infrastructure, etc.).
14. We recognize that free access to, and subsequent use of, government data is of significant value to society and the economy, and that government data should, therefore, be open by default.
15. We acknowledge the need to promote the global development and adoption of resources, standards, and policies for the creation, use, exchange, and harmonization of open data.
16. We recognize that open data can only be unlocked when citizens are confident that open data will not compromise their right to privacy, and that citizens have the right to influence the collection and use of their own personal data or of data generated as a result of their interactions with governments.
17. We will
  - a. Develop and adopt policies and practices to ensure that all government data is made open by default, as outlined in this Charter, while recognizing that there are legitimate reasons why some data cannot be released;
  - b. Provide clear justifications as to why certain data cannot be released;

- c. Establish a culture of openness, not only through legislative and policy measures, but also with the help of training and awareness programs, tools, guidelines, and communication strategies designed to make government, civil society, and private sector representatives aware of the benefits of open data;
- d. Develop the leadership, management, oversight, performance incentives, and internal communication policies necessary to enable this transition to a culture of openness in all government departments and agencies, including official statistics organizations;
- e. Observe domestic laws and internationally recognized standards, in particular those pertaining to security, privacy, confidentiality, and intellectual property. Where relevant legislation or regulations do not exist or are out of date, they will be created and/or updated; and
- f. In accordance with privacy legislation and standards, anonymize data prior to its publication, ensuring that sensitive, personally-identifiable data is removed.

## **PRINCIPLE 2: TIMELY AND COMPREHENSIVE**

- 18. We recognize that it may require time and human and technical resources to identify data for release or publication.
- 19. We recognize the importance of consulting with data users, including citizens, other governments, and civil society and private sector organizations to identify which data to prioritize for release and/or improvement.
- 20. We recognize that in order to be valuable to governments, citizens, and civil society and private sector organizations, data must be comprehensive, accurate, and of high quality.
- 21. We will
  - a. Create, maintain, and share public, comprehensive lists of data holdings to support meaningful consultations around data prioritization, publication, and release dates;
  - b. Release high-quality open data in a timely manner, without undue delay. Data will be comprehensive and accurate, and released in accordance with prioritization that is informed by consultations with open data users, including citizens, other governments, and civil society and private sector organizations;
  - c. To the extent possible, release data in its original, unmodified form, and link data to any relevant guidance, documentation, visualizations, or analyses;
  - d. To the extent possible, release data that is disaggregated to the lowest levels of administration, including disaggregation by gender, age, income, and other categories;
  - e. Allow users to provide feedback, and continue to make revisions to ensure data quality is improved as necessary;
  - f. Apply consistent information lifecycle management practices, and ensure historical copies of datasets are preserved, archived, and kept accessible as long as they retain value;

- g. Consult data users on significant changes to the structure or supply of data in order to minimize the impact to users that have created tools based on open data; and
- h. Be transparent about our own data collection, standards, and publishing processes by documenting these processes online.

### **PRINCIPLE 3: ACCESSIBLE AND USABLE**

- 22. We recognize that opening up data enables governments, citizens, and civil society and private sector organizations to make better informed decisions.
- 23. We recognize that when open data is released, it should be easily discoverable and accessible, and made available without bureaucratic or administrative barriers, which can deter people from accessing the data.
- 24. We will:
  - a. Publish data on a central portal, so that open data is easily discoverable and accessible in one place;
  - b. Release data in open formats to ensure that the data is available to the widest range of users to find, access, and use. In many cases, this will include providing data in multiple, standardized formats, so that it can be processed by computers and used by people;
  - c. Release data free of charge, under an open and unrestrictive licence;
  - d. Release data without mandatory registration, allowing users to choose to download data without being required to identify themselves; and
  - e. Ensure data can be accessed and used effectively by the widest range of users. This may require the creation of initiatives to raise awareness of open data, promote data literacy, build capacity for effective use of open data, and ensure citizen, community, and civil society and private sector representatives have the tools and resources they need to effectively understand how public resources are used.

### **PRINCIPLE 4: COMPARABLE AND INTEROPERABLE**

- 25. We recognize that in order to be most effective and useful, data should be easy to compare within and between sectors, across geographic locations, and over time.
- 26. We recognize that data should be presented in structured and standardized formats to support interoperability, traceability, and effective reuse.
- 27. We will:
  - a. Implement consistent, open standards related to data formats, interoperability, structure, and common identifiers when collecting and publishing data;
  - b. Ensure that open datasets include consistent core metadata and are made available in human- and machine-readable formats;

- c. Ensure that data is fully described, that all documentation accompanying data understand the source, strengths, weaknesses, and analytical limitations of the data;
- d. Engage with domestic and international standards bodies and other standard- setting initiatives to encourage increased interoperability between existing international standards, support the creation of common, global data standards where they do not already exist, and ensure that any new data standards we create are, to the greatest extent possible, interoperable with existing standards; and
- e. Map local standards and identifiers to emerging globally agreed standards and share the results publicly.

## **PRINCIPLE 5: FOR IMPROVED GOVERNANCE AND CITIZEN ENGAGEMENT**

- 28. We recognize that the release of open data strengthens the governance of and trust in our public institutions, reinforces governments' obligation to respect the rule of law, and provides a transparent and accountable foundation to improve decision-making and enhance the provision of public services.
- 29. We recognize that open data encourages better development, implementation, and assessment of programs and policies to meet the needs of our citizens, and enables civic participation and better informed engagement between governments and citizens.
- 30. We recognize that engagement and consultation with citizens and civil society and private sector organizations can help governments understand which types of data are in high demand, and, in turn, can lead to improved data prioritization, release, and standardization practices.
- 31. We recognize that city or local governments are often the first point of interaction between citizens and government, and that these governments therefore have a crucial role in supporting citizen engagement on open data.
- 32. We will:
  - a. Implement oversight and review processes to report regularly to the public on the progress and impact of our open data initiatives;
  - b. Ensure that information published as a result of transparency or anticorruption laws is released as open data;
  - c. Provide training programs, tools, and guidelines designed to ensure government employees are capable of using open data effectively in policy development processes;
  - d. Engage with the Freedom of Information / Access to Information / Right to Information community to align the proactive release of open data with governments' obligation to release information on request;
  - e. Engage proactively with citizens and civil society and private sector representatives to determine what data they need to effectively hold governments accountable;

- f. Respect citizens' right to freedom of expression by protecting those who use open data to identify corruption or criticize governments; and
- g. Encourage the use of open data to develop innovative, evidence-based policy solutions that benefit all members of society, as well as empower marginalized communities.

## **PRINCIPLE 6: FOR INCLUSIVE DEVELOPMENT AND INNOVATION**

- 33. We recognize the importance of openness in stimulating creativity and innovation. The more governments, citizens, and civil society and private sector organizations use open data, the greater the social and economic benefits that will be generated. This is true for government, commercial, and non-commercial uses.
- 34. We recognize that open data can help to identify social and economic challenges, and monitor and deliver sustainable development programs. Open data can also help meet global challenges such as poverty, hunger, climate change, and inequality.
- 35. We recognize that open data is, by its nature, an equitable resource that empowers all people by allowing them to access data regardless of who they are or where they live. However, we also recognize the existence of a global digital divide in regard to technological tools and expertise; this divide limits the ability of socially and economically marginalized people to access and use open data.
- 36. We recognize the role of governments in promoting innovation and sustainable development does not end with the release of open data. Governments must also play an active role in supporting the effective and innovative reuse of open data, and ensuring government employees, citizens, and civil society and private sector organizations have the data they need and the tools and resources to understand and use that data effectively.
- 37. We will:
  - a. Encourage citizens, civil society and private sector organizations, and multilateral institutions to open up data created and collected by them in order to move toward a richer open data ecosystem with multiple sources of open data;
  - b. Create or explore potential partnerships between governments and with civil society and private sector organizations and multilateral institutions to support the release of open data and maximize the impact of data through effective use;
  - c. Create or support programs and initiatives that foster the development or co-creation of datasets, visualizations, applications, and other tools based on open data;
  - d. Engage with schools and post-secondary education institutions to support increased open data research and to incorporate data literacy into educational curricula;
  - e. Conduct or support research on the social and economic impacts of open data;
  - f. Build capacity and share technical expertise and experience with other governments and international organizations around the world, ensuring that everyone can reap the benefits of open data; and

- g. Empower a future generation of data innovators inside and outside government by building capacity and encouraging developers, entrepreneurs, civil society and private sector organizations, academics, media representatives, government employees, and other users to unlock the value of open data.

### ANNEX 3: THE ALL-SCHOLARSHIP REPOSITORY

As noted in the text, most broad approaches to the future of open solutions are still in the realm of science fiction in this community’s policy debates. Still, at the risk of detracting from the credibility of this report, it’s important to note that these broad approaches have purpose and appeal (see, for example, the sentiments expressed in NASEM 2020 and NIH 2020), even if the exact solutions are still elusive. In principle, one such approach would be to build something akin to an “All-Scholarship Repository” (ASR). An ASR approach would replace the thousands of government and institutional repositories currently in use with one ultra-high functioning repository—a single Amazon for all things research—and would also simplify the flow of research information. All information—not necessarily in research paper format but in all kinds of reporting formats—would simply be deposited into ASR and picked up from there by publishers and other researchers. This information dump would be analogous to simply filing information on a computer, but in an organized and standardized way. Identifying the most salient information and developing it for audiences would be a separate process left mostly to publishers, researchers, and research communication teams, who would troll this massive database for content.



In this sort of information ecosystem, new opportunities and challenges would emerge for publishers as more and more innovation and competition developed around identifying and developing the best content from the ASR for audiences and also developing the most compelling interfaces that provide the most value for audiences. Some public interfaces would certainly be free, but competitive new interfaces supported by ad and subscription revenue would also likely emerge. In addition to providing premium content (e.g., peer reviewed, formatted and edited, supplemented by interactive databases and research connection services, etc.) these publishers would have a strong, valuable and viable presence in the research publishing world, but they would not be gateways to the raw information itself. Much like Internet news today, anyone would be able to find out anything, but the most consumed and trusted information would still be curated, fact-checked, and come from reliable sources.

The core database for ASR would be structured like the Digital Public Library of America, where every institution, publisher, discipline, subject and author would be able to design their own user interface drawing from the same common information base. Every institution would be able to adjust content permissions as needed for now, but the long-term aim would be to figure out how to release more and more existing repository content to the public over time, and to enhance existing content with data and other value (such as commentary, “impact” ratings, research connections, researcher databases and more).

In order for ASR to reach its full potential, it would need to be surrounded by a new communications capacity in research institutions. Otherwise, the task of publishing data to the ASR and maintaining this resource would be untenable. These research communication teams would bear the primary responsibility for maintaining ASR, and over time they might also eventually perform functions such as standardizing data prior to publishing, connecting research, and liaising with publishers, policymakers, educators and the public, thereby relieving researchers of these communication responsibilities (many of which are new) and also improving participation in and sustainability of these important functions (over time, the ASR team itself will also develop these capacities, especially to assist communication teams with their work).

Once ASR becomes fully functioning, it would be a deep resource for researchers and publishers. With a massive, connected database of research information, new disciplines and industries would emerge to try to connect the dots. “Peering” experts would be needed—a new class of scientists would look for and develop connections within and between research—as well as experts who would take on the heady task of standardizing data within and between fields.

ASR would eliminate a host of bottlenecks in the current scholarly publishing system, and simultaneously address a number of contentious and even seemingly intractable cultural issues, all while keeping existing stakeholders at the table, and also spurring discovery and innovation as researchers from institutions, business, and the public begin digging through and connecting research in ways that have never been possible before.

## ANNEX 4: DRAFT MARKETING PLAN FOR UNESCO'S OPEN SOLUTIONS POLICY

One of the deliverables for this policy paper is to offer a draft marketing plan for UNESCO's open solutions policy. This paper is going to fall short in this regard in a useful, meaningful sense because a marketing plan presumes that the agency in question is fully behind the content. It's not clear whether this is the case here at the moment. Once this commitment is clearer, the marketing plan can be clearer as well—the author team would be happy to revisit this question. As it currently stands, a marketing plan for this proposal would be rather convoluted, looking something like this:

- 1. Compete publicly with UNESCO's Open Science policy:** Unfortunately, unless UNESCO's nascent Open Science policy becomes more aligned with its Open Solutions vision, there's going to be internal conflict and division, as well as external messaging problems. Something will have to give. Ideally, that something will be that UNESCO's open science policy comes more in line with UNESCO's open access and open solutions thinking, which has been 20 years in the making and has benefited from many years of input, conversation and reflection. If there is no alignment, then depending on the internal will and resources of UNESCO, different divisions will be promoting different, competing visions for the future of open. This "competition" creates uncertain messaging dynamics. That is, should the open solutions proposal be aimed directly at researchers and bypass policy audiences? Should it be critical of the open science plan? Should it try to carve out a separate policy space altogether (that is, let the open science plan stand as is and offer the open solutions approach as a next generation policy geared more toward global involvement and action)?
- 2. Advertise the Open Solutions policy:** Do this through a broad array of activities and mediums, including:
  - a. A new open solutions website
  - b. A second, more authoritative version of this paper, with forwards written by prominent scholars and endorsements from major governments and agencies
  - c. An outreach campaign directed at researchers, universities and funders, describing this vision (based on widely circulated short videos and infographics)
  - d. A new series of fact gathering efforts, not geared toward deciding what to do but toward understanding the specific challenges that need to be addressed and finding areas of common ground and action
- 3. Take action:** This policy will not stand on its own without also aggressively working on improving how we do open. UNESCO should take the lead in helping the world develop "easy open" solutions and "bold open" actions:
  - a. New guidelines to streamline open publishing and data deposits
  - b. More flexible approaches to open

- c. Leadership on using open the attack urgent crisis such as COVID and climate change

The dynamics of this plan are also going to be affected by:

- 1. Buy-in:** If UNESCO can get a slew of high profile organizations and governments to “endorse” its open solutions approach, marketing it will be easier. Boilerplate policies and marketing/out-reach materials can be developed for signatories so they can spread the word through their own networks.
- 2. Turnkey success stories:** Flowing from buy-in, adoption will beget more adoption as this approach develops more momentum and adoption examples and best practices.
- 3. Budget:** If UNESCO can commit a significant budget toward advertising and action, then dynamics 1 and 2 will be accelerated. If the budget commitment isn’t vigorous, the only “weight” in this conversation would come from the fact that UNESCO is behind this effort—not insignificant, but also not sufficient in itself to draw enough attention to guarantee success.

These factors considered, then, the recommended marketing approach for this policy is as follows, in this order of priority:

- 1. Figure out where UNESCO stands:** Which course of action does UNESCO want to pursue? A global policy alignment course? Adopting OSI’s Plan A? Creating an open science policy plus an open solutions policy? It isn’t possible to even begin the conversation about how to market a policy without knowing what it is and in what context it will be pursued.
- 2. Figure out what UNESCO is willing to do to push the ball forward:** Will this be a build it and move on effort, or does UNESCO plan to stay at the center? The former approach just means marketing a concept. The latter means marketing both marketing a concept and leading and developing a vision, rallying everyone under the UNESCO banner. The first task is simple and low-budget; the second is complex and long-term and requires strategically thinking about UNESCO’s long-term budget commitment to this effort, not just for marketing but for product development, convening conversations and overseeing policy development efforts, and general thought leadership in this space.
- 3. Mobilize resources both internally and globally:** Depending on the direction and scale of the ambition here, the lift could be enormous and require a significant and long-term investment of resources. UNESCO would be wise, therefore, to engage as many governments as possible to help support this effort through whatever mechanism can be long-term sustainable. For example, IFC’s Technical Assistance and Trust Fund vehicle allows for different countries to pool resources together for financing set objectives.
- 4. Act quickly, realistically and collaboratively:** The open space is evolving quickly. A worst-case outcome would be for UNESCO to take years to finalize a course of action and years more to get started working in an arrangement that was bureaucratic and unresponsive to market conditions. Whatever mechanism is created needs to be light, nimble, inclusive and accountable—able to work quickly and effectively with all stakeholders and develop solutions that work well and quickly with maximum flexibility and minimal burden. Anything overly rigid or burdensome would be viewed as nothing more than another Plan S, and would not be likely to attract participants and attention.

## ANNEX 5: COMMON LICENSE TYPES

There isn't a lot of overlap in license types between open data, open source and open access. The following licenses are generally the most commonly used. Not all are "optimally" open, or even open at all (like copyright); in these cases, however, information isn't necessarily hidden, just restricted in some sense (in terms of reuse, for example). Open government and OER are not listed here because they use these licenses below, depending on whether they are making text or data available.

LICENSE TYPE	OPEN SOURCE	OPEN DATA	OPEN ACCESS
<a href="#">Apache License 2.0 (Apache-2.0)</a>	X		
<a href="#">3-clause BSD license (BSD-3-Clause)</a>	X		
<a href="#">2-clause BSD license (BSD-2-Clause)</a>	X		
<a href="#">GNU General Public License (GPL)</a>	X		
<a href="#">GNU Lesser General Public License (LGPL)</a>	X		
<a href="#">GNU Free Documentation License (GFDL)</a>	X	X	
<a href="#">MIT license (MIT)</a>	X		
<a href="#">Mozilla Public License 2.0 (MPL-2.0)</a>	X		
<a href="#">Common Development and Distribution License 1.0 (CDDL-1.0)</a>	X		
<a href="#">Eclipse Public License 2.0 (EPL-2.0)</a>	X		
<a href="#">Public domain</a>		X	
<a href="#">Creative Commons Zero (CC-0)</a>		X	
<a href="#">Creative Commons BY (CC-BY)</a>		X	X
<a href="#">Creative Commons BY-SA (CC-BY-SA)</a>		X	X
<a href="#">Creative Commons BY-NC</a>		X	X
<a href="#">Creative Commons BY-NC-ND</a>			X
<a href="#">Creative Commons BY-NC-SA</a>		X	X
Copyright held by author			X
Copyright held by publisher			X
<a href="#">Open Data Commons Public Domain Dedication and License (PPDL)</a>		X	
<a href="#">Community Data License Agreement – Permissive (CDLA), Version 1.0</a>		X	
<a href="#">Community Data License Agreement (CDLA)– Sharing, Version 1.0</a>		X	
<a href="#">Open Data Commons Open Database (ODC-ODbL) License</a>		X	
<a href="#">Open Data Commons Attribution License (ODC-BY)</a>		X	
No license specified	X	X	X

Sources:

<https://opensource.org/licenses/category>

<https://help.data.world/hc/en-us/articles/115006114287-Common-license-types-for-datasets>

<https://resources.data.gov/open-licenses/>

<https://opendatacommons.org/licenses/>

<https://creativecommons.org/about/cclicenses/>

## ANNEX 6: OPEN DATA WEBSITES BY SECTOR/TOPIC

### OPEN GOVERNMENT DATA (SMALL SAMPLE)

SECTOR	WEBSITE
Agriculture	<a href="#">The USDA National Farmers Market Directory</a>
Agriculture	<a href="#">U.K. Department of Agriculture and Rural Development</a>
Budgets & Public Finance	<a href="#">WB Open Budgets</a>
Budgets & Public Finance	<a href="#">OpenSpending</a>
Budgets & Public Finance	<a href="#">International Budget Partnership</a>
Budgets & Public Finance	<a href="#">The International Aid Transparency Initiative (IATI)</a>
Budgets & Public Finance	<a href="#">U.S. IRS Tax Statistics</a>
Education	<a href="#">Ed Data Inventory</a>
Education	<a href="#">MyData Office of Educational Technology</a>
Education	<a href="#">CheckMySchool</a>
Energy & Extractive Industries	<a href="#">Extractive Industries Transparency Initiative</a>
Energy & Extractive Industries	<a href="#">U.S. Department of Energy</a>
Energy & Extractive Industries	<a href="#">Enel Open Data - Largest power company in Italy</a>
Environment	<a href="#">Open Climate Data</a>
Environment	<a href="#">Fuel Economy Data</a> , U.S. Environmental Protection Agency
Environment	<a href="#">New York City Environment Open Data</a>
Geospatial	<a href="#">OpenStreetMap</a>
Geospatial	<a href="#">Haiti Data geospatial information</a>
Health	<a href="#">The U.S. Department of Health &amp; Human Services</a>
Health	<a href="#">Agency for Healthcare Research &amp; Quality (AHRQ) Databases on healthcare cost &amp; utilization in the U.S.</a>
Health	<a href="#">WB Health Data</a>
Information & Communication Technologies (ICT)	<a href="#">Australian ICT Open Datasets</a>
Transport	<a href="#">OpenPlans</a>
Transport	<a href="#">European Public Sector Information Platform: Transport</a>
Water	<a href="#">Global water database</a>

Source: <http://opendatatoolkit.worldbank.org/en/essentials.html#uses>

### OPEN SCIENCE DATA (SMALL SAMPLE)

FIELD	OPEN ARRANGEMENT
<b>BIOLOGY: NUCLEIC ACID SEQUENCE</b>	
<a href="#">DNA DataBank of Japan (DDBJ)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">European Nucleotide Archive (ENA)</a>	<a href="#">view FAIRsharing entry</a>

<a href="#">GenBank</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">dbSNP</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">European Variation Archive (EVA)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">dbVar</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">MGnify</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">NCBI Trace Archive</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">NCBI Sequence Read Archive (SRA)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">NCBI Assembly</a>	
<b>BIولوجY: PROTEIN SEQUENCE</b>	
<a href="#">UniProtKB</a>	<a href="#">view FAIRsharing entry</a>
<b>BIولوجY: MOLECULAR &amp; SUPRAMOLECULAR STRUCTURE</b>	
<a href="#">Protein Circular Dichroism Data Bank (PCDDDB)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Crystallography Open Database (COD)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Coherent X-ray Imaging Data Bank (CXIDB)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Biological Magnetic Resonance Data Bank (BMRB)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Electron Microscopy Data Bank (EMDB)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Worldwide Protein Data Bank (wwPDB)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Structural Biology Data Grid</a>	<a href="#">view FAIRsharing entry</a>
<b>NEUROSCIENCE</b>	
<a href="#">NeuroMorpho.org</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">OpenNeuro (formerly OpenfMRI)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">G-Node</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Neuroimaging Informatics Tools and Resources Collaboratory (NITRC)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">EBRAINS</a>	<a href="#">view FAIRsharing entry</a>
<b>OMICS: FUNCTIONAL GENOMICS</b>	
<a href="#">ArrayExpress</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Gene Expression Omnibus (GEO)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">GenomeRNAi</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">dbGAP</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">The European Genome-phenome Archive (EGA)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Database of Interacting Proteins (DIP)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">IntAct</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Japanese Genotype-phenotype Archive (JGA)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Biological General Repository for Interaction Datasets *</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">NCBI PubChem BioAssay</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Genomic Expression Archive (GEA)</a>	<a href="#">view FAIRsharing entry</a>
<b>OMICS: METABOLOMICS &amp; PROTEOMICS</b>	
<a href="#">MassIVE</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">MetaboLights</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">PeptideAtlas</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">PRIDE</a>	<a href="#">view FAIRsharing entry</a>
<b>TAXONOMY &amp; SPECIES DIVERSITY</b>	
<a href="#">Environmental Data Initiative (formerly LTER Network Information System Data Portal)</a>	<a href="#">view re3data entry</a>
<a href="#">Global Biodiversity Information Facility (GBIF)</a>	<a href="#">view FAIRsharing entry</a>

<a href="#">Integrated Taxonomic Information System (ITIS)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">KNB: The Knowledge Network for Biocomplexity</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Morphobank.org</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Movebank Data Repository</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">NCBI Taxonomy*</a>	<a href="#">view FAIRsharing entry</a>
<b>MATHEMATICAL &amp; MODELING RESOURCES</b>	
<a href="#">BioModels Database</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Kinetic Models of Biological Systems (KiMoSys)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">The Network Data Exchange (NDEX)</a>	<a href="#">view FAIRsharing entry</a>
<b>CYTOMETRY &amp; IMMUNOLOGY</b>	
<a href="#">FlowRepository</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">ImmPort</a>	<a href="#">view FAIRsharing entry</a>
<b>IMAGING</b>	
<a href="#">Image Data Resource</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">The Cancer Imaging Archive</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">SICAS Medical Image Repository</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Coherent X-ray Imaging Data Bank (CXIDB)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Cell Image Library</a>	<a href="#">view FAIRsharing entry</a>
<b>ORGANISM-FOCUSED RESOURCES</b>	
<a href="#">Eukaryotic Pathogen Database Resources (EuPathDB)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">FlyBase</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Influenza Research Database</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Mouse Genome Informatics (MGI)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Rat Genome Database (RGD)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">VectorBase</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Xenbase</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Zebrafish Model Organism Database (ZFIN)</a>	<a href="#">view FAIRsharing entry</a>
<b>HEALTH SCIENCES</b>	
<a href="#">National Addiction &amp; HIV Data Archive Program (NAHDAP)</a>	<a href="#">view FAIRshaing entry</a>
<a href="#">National Database for Autism Research (NDAR)</a>	<a href="#">view FAIRshaing entry</a>
<a href="#">The Cancer Imaging Archive</a>	<a href="#">view FAIRshaing entry</a>
<a href="#">ClinicalTrials.gov</a>	<a href="#">view FAIRshaing entry</a>
<a href="#">SICAS Medical Image Repository (formally Virtual Skeleton Database)</a>	<a href="#">view FAIRshaing entry</a>
<a href="#">PhysioNet</a>	<a href="#">view FAIRshaing entry</a>
<a href="#">National Database for Clinical Trials related to Mental Illness (NDCT)</a>	<a href="#">view FAIRshaing entry</a>
<a href="#">Research Domain Criteria Database (RDoCdb)</a>	<a href="#">view FAIRshaing entry</a>
<a href="#">Synapse</a>	<a href="#">view FAIRshaing entry</a>
<a href="#">UK Data Service</a>	<a href="#">view FAIRshaing entry</a>
<b>CHEMISTRY &amp; CHEMICAL BIOLOGY</b>	
<a href="#">caNanoLab *</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">ChEMBL *</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">ioChem-BD Computational Chemistry Datasets</a>	<a href="#">view re3data entry</a>
<a href="#">NCBI PubChem BioAssay</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">NCBI PubChem Substance</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Beilstein-Institut, STRENDA</a>	<a href="#">view FAIRsharing entry</a>

<a href="#">Crystallography Open Database (COD)</a>	<a href="#">view FAIRsharing entry</a>
<b>EARTH, ENVIRONMENTAL &amp; SPACE SCIENCES: BROAD-SCOPE EARTH &amp; ENVIRONMENTAL SCIENCES</b>	
<a href="#">NASA Goddard Earth Sciences Data and Information Services Center</a>	<a href="#">view re3data entry</a>
<a href="#">NERC Data Centres</a>	<a href="#">view re3data entry</a>
<a href="#">PANGAEA</a>	<a href="#">view re3data entry</a>
<a href="#">National Tibetan Plateau/Third Pole Environment Data Center</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">NOAA National Centers for Environmental Information (DOIs only assigned to deposited data on request)</a>	<a href="#">view re3data entry</a>
<a href="#">HydroShare (CUAHSI)</a>	<a href="#">view FAIRsharing entry</a>
<b>EARTH, ENVIRONMENTAL &amp; SPACE SCIENCES: ASSTRONOMY &amp; PLANETARY SCIENCES</b>	
<a href="#">SIMBAD Astronomical Database</a>	<a href="#">view re3data entry</a>
<a href="#">UK Solar System Data Centre</a>	<a href="#">view re3data entry</a>
<b>EARTH, ENVIRONMENTAL &amp; SPACE SCIENCES: BIOGEOCHEMISTRY &amp; GEOCHEMISTRY</b>	
<a href="#">EarthChem</a>	<a href="#">view re3data entry</a>
<a href="#">Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC)</a>	<a href="#">view re3data entry</a>
<b>EARTH, ENVIRONMENTAL &amp; SPACE SCIENCES: CLIMATE SCIENCES</b>	
<a href="#">World Data Center for Climate at DRKZ (WDCC)</a>	<a href="#">view re3data entry</a>
<b>EARTH, ENVIRONMENTAL &amp; SPACE SCIENCES: ECOLOGY</b>	
<a href="#">AEKOS - TERN Ecoinformatics</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Environmental Data Initiative (formerly LTER Network Information System Data Portal)</a>	<a href="#">view re3data entry</a>
<a href="#">Global Biodiversity Information Facility (GBIF)</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">KNB: The Knowledge Network for Biocomplexity</a>	<a href="#">view FAIRsharing entry</a>
<b>EARTH, ENVIRONMENTAL &amp; SPACE SCIENCES: Geomagnetism &amp; Palaeomagnetism</b>	
<a href="#">Magnetics Information Consortium (MagIC)</a>	<a href="#">view re3data entry</a>
<b>EARTH, ENVIRONMENTAL &amp; SPACE SCIENCES: OCEAN SCIENCES</b>	
<a href="#">Australian Antarctic Data Centre (AADC)</a>	<a href="#">view re3data entry</a>
<a href="#">Australian Ocean Data Network</a>	<a href="#">view re3data entry</a>
<a href="#">Marine Data Archive</a>	
<a href="#">Marine Geosciences Data System</a>	<a href="#">view re3data entry</a>
<a href="#">SEANOE</a>	<a href="#">view FAIRsharing entry</a>
<b>EARTH, ENVIRONMENTAL &amp; SPACE SCIENCES: SOLID EARTH SCIENCES</b>	
<a href="#">British Geological Survey</a>	<a href="#">view re3data entry</a>
<a href="#">EarthChem</a>	<a href="#">view re3data entry</a>
<a href="#">Magnetics Information Consortium (MagIC)</a>	<a href="#">view re3data entry</a>
<a href="#">Marine Geosciences Data System</a>	<a href="#">view re3data entry</a>
<a href="#">UNAVCO, Inc.</a>	<a href="#">view re3data entry</a>
<a href="#">Incorporated Research Institutions for Seismology (IRIS)</a>	<a href="#">view re3data entry</a>
<a href="#">OpenTopography</a>	<a href="#">view FAIRsharing entry</a>
<b>PHYSICS</b>	
<a href="#">HEPData</a>	<a href="#">view re3data entry</a>
<b>MATERIAL SCIENCE</b>	
<a href="#">NoMaD Repository</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Materials Cloud</a>	<a href="#">view FAIRsharing entry</a>

<b>SOCIAL SCIENCES</b>	
<a href="#">Archaeology Data Service</a>	<a href="#">view re3data entry</a>
<a href="#">Harvard Dataverse</a>	<a href="#">view re3data entry</a>
<a href="#">openCPSR</a>	<a href="#">view re3data entry</a>
<a href="#">Open Science Framework</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Qualitative Data Repository</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">UK Data Service</a>	<a href="#">view re3data entry</a>
<b>GENERALIST REPOSITORIES</b>	
<a href="#">Dryad Digital Repository</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">figshare</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Harvard Dataverse</a>	<a href="#">view re3data entry</a>
<a href="#">Open Science Framework</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Zenodo</a>	<a href="#">view re3data entry</a>
<a href="#">Mendeley Data</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Science Data Bank</a>	<a href="#">view FAIRsharing entry</a>

Source: <https://www.nature.com/sdata/policies/repositories>

Other data repository lists are available from:

- PLOS: <http://journals.plos.org/plosbiology/s/data-availability#loc-recommended-repositories>
- SpringerNature: <http://www.springernature.com/gp/group/data-policy/repositories>
- EMBO Press: <http://emboj.embopress.org/authorguide#datadeposition>
- Elsevier: <https://www.elsevier.com/authors/author-services/research-data/data-base-linking/supported-data-repositories>
- COPDESS: <https://copdessdirectory.osf.io>

## ANNEX 7: KEY OPEN DEFINITIONS, GROUPS & POLICIES

### DEFINITIONS

**Accepted author manuscript (AAM):** The version of a manuscript that has been accepted by a publisher for publication.

**Altmetrics:** Alternative ways of recording and measuring the use and impact of scholarship. Rather than solely counting the number of times a work is cited in scholarly literature, alternative metrics also measure and analyze social media (e.g., Facebook, Twitter, blogs, wikis, etc.), document downloads, links to publishing and unpublished research, and other uses of research literature, in order to provide a more comprehensive measurement of scholarships reach and impact.

**Article-level metrics:** All types of article-level metrics including download and usage statistics, citations, and article-level altmetrics.

**Article processing charge (APC):** A fee charged to the author, creator, or institution to cover the cost of an article, rather than charging the potential reader of the article. APCs may apply to both commercial and open access publications. APCs are sometimes charged to authors in order to cover the cost of publishing and disseminating an article in an open access scholarly journal.

**Bibliometrics:** The branch of library and information science concerned with the application of mathematical and statistical analysis to bibliography. Bibliometrics involves the statistical analysis of books, articles, or other publications.

**CC-BY license:** A Creative Commons license compatible with the most stringent stipulations of open access, and which allows the reuse, sharing, and remixing of materials providing the original author is appropriately attributed.

**Copyright:** The aspect of Intellectual property that gives creators the right to permit (or not permit) what happens to their creations, as opposed to trademark rights or moral rights.

**Creative Commons licenses:** A suite of licenses that set out the rights of authors and users, providing more “open” alternatives to the standard copyright.

**Current Research Information Systems (CRISs):** CRISs are an institution’s internal tools and systems that collect a wide array of research information in order to be able to describe the institutional research activity for reporting purposes, either at institutional, funder or governmental level.

**Data availability statement (DAS):** Statements accompanying published journal articles that detail where the data supporting research results can be found.

**Digital Object Identifier (DOI):** A unique text string that is used to identify digital objects such as journal articles or open source software releases.

**Double blind peer review:** When the reviewers don’t know who the authors are, and vice versa.

**Embargo period:** A length of time imposed on a research output for users who have not paid for access, or do not have institutional access, before it is made freely available.

**GNU GPL (General Public License):** A free license for software and other kinds of works.

**Gold OA:** Making the final version of record of a research paper freely available immediately upon publication by the publisher. In order to facilitate this, the author (or her institution) pays a fee to the publisher to cover publishing costs.

**Green OA:** Making a version of a research paper (typically not the final version of record) freely available in a repository.

**H-index:** A personal metric that relates the number of citations to the number of published papers for an academic.

**Hybrid journal:** A type of journal in which certain articles are made open access while others remain subscription access.

**Impact factor:** A numerical (and controversial) measure that indicates the average number of citations to articles published over the previous two years in a journal, and frequently used as a proxy for a journal's relative importance.

**Institutional repository:** An online database designed to collect the intellectual output of a particular institution or university, including digital collections such as electronic theses and dissertations, pre-prints, or faculty scholarship, and presents associated metadata regarding these items.

**Intellectual property (IP):** A legal term that refers to creations of the mind. Examples of intellectual property include music, literature, and other artistic works; discoveries and inventions; and phrases, symbols, and designs.

**Intellectual property rights:** The rights given to the owners of intellectual property, protected either automatically (e.g., copyright, design rights) or by registering or applying for it (e.g., trademarks, patents). Protecting intellectual property makes it easier to take legal action against anyone who steals or copies it. These rights can be legally sold, assigned or licensed by the creator to other parties, or joint-owned.

**Journal:** An aggregation of published research articles. Historically divided into volumes and issues.

**Journal level metrics:** Metrics that apply to all papers published within a journal. A common example is Thomson Reuters' journal impact factor.

**Mandate:** Authority to carry out a policy—in this context, largely to conform to open access policies.

**Megajournal:** A journal with editorial criteria based on scientific soundness instead of a priori estimated newsworthiness or impact.

**Open:** At its most open, information is available without cost, immediately upon publishing, and includes the right to repurpose without attribution. Other types of open can be more restrictive, including open information that carries limited reuse conditions, limited embargo periods, and/or has less than ideal discoverability. The DARTS Framework, developed by OSI participants, proposes that the openness of information exists along five dimensions: discoverability, accessibility, reusability, transparency, and sustainability. The result is a broad spectrum of open states, not binary open-closed values.

**Open access (OA):** Information (generally peer reviewed scholarly works) which is freely available, generally (although interpretations vary) without restriction, permitting anyone to use this for any lawful purpose, without barriers other than those inseparable from gaining access to the internet itself. Several different states of OA exist, including green, gold, hybrid, etc. (as defined in this list).

**Open data:** Data which is open. At its most open, data can be downloaded, copied, analyzed, re-processed, or used for any other purpose without barriers other than those inseparable from gaining access to the internet itself.

**Open educational resources (OER):** Openly licensed, online educational materials for sharing, use, and reuse.

**Open peer review:** When reviews are made openly available, typically alongside the article.

**Open source software (OSS):** Source code for a piece of software, along with an open source license permitting reuse, adaptation, and further distribution.

**Open spectrum:** The range of different types of open, from public access information (see "public access") to open access information and everything in-between. The DARTS Framework, developed by OSI participants, proposes that the openness of information exists along five dimensions: discoverability, accessibility, reusability, transparency, and sustainability. The result is a broad spectrum of open states, not binary open-closed values.

**Overlay journal:** An open access, electronic journal that does not produce its own content, but selects and curates groups of articles that are already freely available online.

**Paywall:** Restriction via a financial barrier to research. Can be removed by personal or institutional subscription.

**Peer review:** A process by which a research article is vetted by experts in community before publication. Several variations of peer review are used (as defined in this list.)

**Persistent identifiers (PID):** Any long-lasting reference to a digital resource.

**Post-print:** A manuscript draft after it has been peer reviewed.

**Post publication peer review:** Standard peer review, but after a research article has been formally published.

**Pre-print:** A manuscript draft that has not yet been subject to formal peer review, distributed to receive early feedback on research from peers.

**Public access:** A type of open, used predominately by US government agencies, in which information is made freely available following a brief embargo period and to which typical copyright restrictions (e.g., requiring author permission and/or source citation) normally apply.

**Publisher:** An entity (including corporation, university, research institution, society, trade group, etc.) who makes the outputs of research publicly available.

**Publishing:** The act of making research output available to the public.

**Repository (article):** An archive to deposit manuscripts. Legally, these can be personal or institutional, but aggregator sites such as ResearchGate also function as defacto repositories.

**Repository (data):** According to the US National Library of Medicine, “a data repository can be defined as a place that holds data, makes data available to use, and organizes data in a logical manner. A data repository may also be defined as an appropriate, subject-specific location where researchers can submit their data.”

**Repository (software):** A collection of files managed with version control software (e.g., bsr, hg, git, csv, svn, etc.). Can be hosted by third-party (e.g., github, bitbucket, sourceforge), by an institution, or self-hosted locally.

**Scholarly communication:** The creation, transformation, dissemination, and preservation of knowledge related to teaching, research, and scholarly endeavors; the process of academics, scholars and researchers sharing and publishing their research findings so that they are available to the wider academic community.

**Version of record (VOR):** The final version of a manuscript, after peer review and processing by publishers.

## GROUPS

**1Science:** A Montreal-based company specializing in open access research and OA solutions analysis and development.

**American Association for the Advancement of Science (AAAS):** An international non-profit organization, AAAS is the world’s largest multidisciplinary scientific society and a leading publisher, dedicated to advancing science for the benefit of all people.

**arXiv:** The world’s largest and most successful pre-print server, containing a large collection of open articles from physics, astronomy, and other disciplines. Operated by Cornell University.

**Association of Research Libraries (ARL):** A network of American university and research libraries, and a long-time global leader in OA advocacy and open solutions development.

**Australasian Open Access Strategy Group (AOASG):** A pioneer and leading voice in open science, working to advocate, collaborate, raise awareness, and build capacity for OA in Australia and New Zealand.

**bioRxiv:** A recent addition to the pre-print server universe and building off the success of arXiv, bioRxiv primarily serves the life sciences community. Operated by Cold Springs Harbor Laboratory.

**California Digital Library (CDL):** Serving the University of California system, CDL is at the forefront of innovative and collaborative approaches to improve open for the UC system and beyond.

**Canadian Association of Research Libraries (CARL):** CARL serves Canada's research ecosystem and has been a leader in the development and implementation of forward-thinking open policies.

**Center for Open Science (COS):** A US-based nonprofit organization whose mission to increase openness, integrity, and reproducibility of research. Toward this end, COS manages several projects that are helping build capacity in the open community.

**CERN:** CERN, the European Organization for Nuclear Research, is a world leader in high energy physics research, and also a world leader in open access research publishing through its partnership with APS on SCOAP3.

**Chan Zuckerberg Initiative (CZI):** A venture funded by Facebook founder Mark Zuckerberg, the priorities of CZI are still coming into focus. It's clear that they're interested in data, and they have a lot of it at their disposal.

**CHORUS:** A cross-repository portal, designed to identify all public access and open-access materials, and improve identification of these materials, plus improve discovery, access, compliance, and preservation.

**Coalition for Networked Information (CNI):** CNI represents the digital information interests of a wide range of member organizations from higher education, publishing, information technology, government agencies, and libraries, and beyond, and fosters connections and collaborations between these communities.

**Coalition for Publishing Data in the Earth and Space Sciences (COPDESS):** COPDESS connects Earth and space science publishers and data facilities to help translate the aspirations of open, available, and useful data from policy into practice.

**Coalition of Open Access Policy Institutions (COAPI):** An organization that helps North American universities develop open access policies (membership levels are based on the degree of development of a university's open access policies).

**Committee on Data of the International Council for Science (CODATA):** CODATA exists to promote global collaboration to improve the availability and usability of data for all areas of research. CODATA supports the principle that data produced by research and susceptible to be used for research should be as open as possible and as closed as necessary. CODATA works also to advance the interoperability and the usability of such data: research data should be intelligently open or FAIR.

**Committee on Publication Ethics (COPE):** An international organization dedicated to educating and supporting editors and publishers and developing best practices that preserve and promote the transparency and integrity of the scholarly record.

**Creative Commons (CC):** Through its innovative and widely-used copyright licenses (such as CC-BY), Creative Commons is the recognized international leader in developing, supporting, and stewarding legal and technical infrastructure that maximizes digital creativity, sharing, and innovation.

**CrossRef:** An association of scholarly publishers that develops shared infrastructure to support more effective scholarly communication.

**Data.gov:** The central public-facing repository for all US government agency data.

**DataCite:** Global non-profit organisation that provides persistent identifiers (DOIs) for research data and other research outputs.

**Directory of Open Access Journals (DOAJ):** The world's leading directory of open access peer-reviewed journals.

**Dryad:** Community-owned multi-field data warehouse that strives to make more research data discoverable, freely reusable, and citable.

**European Open Science Cloud (EOSC):** An ambitious group funded by the European Commission to dramatically improve research and science interoperability in signatory states by around 2020 (the EOSC initiative also includes funding for data infrastructures, Horizon/OA 2020 and other programs).

**Force 11:** A global volunteer community of scholars, librarians, archivists, publishers and research funders working together to facilitate change toward improved knowledge creation and sharing.

**GitHub:** A web-based service that provides a source code repository.

**Global open data index (GODI):** The annual global benchmark for publication of open government data, run by the Open Knowledge Network.

**Google Scholar:** Google's popular and widely-used search engine for indexing scholarly literature.

**International Data Week (IDW):** An annual event is organised by CODATA, RDA, and the International Science Council (ISC), bringing together experts and stakeholders from around the world to explore how best to exploit the data revolution to improve science and society.

**International DOI Foundation (IDF):** Nonprofit membership organization responsible for issuing and managing digital object identifiers (DOI's).

**International Open Data Conferences (IODCs):** Annual global conferences designed to bring the global open data community together in order to learn, share, plan and collaborate on the future of open data and data for development.

**Joint Information Systems Committee (JISC):** A government and member-supported organization providing UK universities and colleges with shared digital infrastructure and services.

**Knowledge Unlatched:** A leading online platform offering free access to scholarly content, and providing libraries with a central place to support open access models.

**Max Planck Society:** One of the world's largest and most successful non-university research institutions, and also a leader in open access reform (see "global flip" in definitions).

**National Information Standards Organization (NISO):** International organization whose mission is to identify, develop, disseminate, and maintain voluntary, consensus-based technical standards for managing information in a changing environment.

**National Institute of Health (NIH):** US agency which provides the majority of US public funding for medical research, and has a leading role in the implementation of the US public access program. Also manages PubMed and PubMed Central.

**National Science Foundation (NSF):** US agency which provides significant funding for US natural and social science research.

**OCLC:** Nonprofit global library cooperative providing shared technology services, original research and community programs so that libraries can better fuel learning, research and innovation.

**Open Access Infrastructure for Research in Europe (OpenAIRE):** A multi-country EU effort whose goal is to promote open scholarship and substantially improve the discoverability and reusability of research publications and data.

**Open Access Scholarly Publishers Association (OASPA):** A trade association representing the interests of open access publishers globally in all scientific, technical and scholarly disciplines.

**Open Data 4 Development (OD4D):** IDRC funded organization committed to improving the use of data for sustainable development.

**Open Data Barometer (ODB):** A global measure of how governments are publishing and using open data for accountability, innovation and social impact.

**Open Scholarship Initiative (OSI):** UNESCO-backed global organization working with all stakeholder groups in scholarly communication to develop and promote globally fair and effective open policies.

**ORCID:** An organization which supplies and manages persistent digital identifiers to distinguish individual researchers, and also supports integration in research workflows.

**PLOS:** The world's largest open access publisher and a driving force in the open access movement.

**Public Knowledge Project (PKP):** A multi-university initiative developing (free) open source software and conducting research to improve the quality and reach of scholarly publishing. **SCOAP3:** A unique global partnership of libraries, funding agencies and research centers which converts key journals in high-energy physics to open access by paying publishers for OA costs.

**Public Knowledge Project (PKP):** A multi-university initiative developing free, open source software and conducting research to improve the quality and reach of scholarly publishing.

**Publishers:** This is an overly broad category of stakeholders, including every entity from large commercial publishers to university presses, society publishers, non-university research institutions and small startups. In the "large commercial publisher" subcategory, Elsevier, SpringerNature (Palgrave MacMillan), Wiley, and other major commercial scholarly publishers are highly influential; Elsevier alone accounted for 20% of all journal articles published in 2017 and 5% of all open access articles, and manages foundational publishing and discovery tools and services such as SCOPUS, ScienceDirect, and Mendeley. All of these companies are working to accommodate the wishes of their clients and build out open options that work in the marketplace, especially for open data. Similarly, many university presses, society publishers, and non-university research institution publishers are also experimenting with and rolling out open options and reforms.

**PubMed:** A repository consisting of more than 26 million citations for the biomedical literature.

**PubMed Central (PMC):** A free full-text archive of biomedical and life sciences journal literature at the US National Institutes of Health's Library of Medicine.

**Registry of Research Data Repositories (Re3Data):** A global registry of research data repositories that covers research data repositories from different academic disciplines.

**Research4Life:** A public-private partnership (including WHO, FAO, UNEP, WIPO, ILO, Cornell and Yale Universities, the International Association of Scientific, Technical & Medical Publishers and up to 175 international publishers) providing developing countries with free or low cost access to academic and professional peer-reviewed content online.

**Research Data Alliance (RDA):** An international research community organization started by the European Commission, the US National Science Foundation, NISO, and the Australian Department of Innovation, whose mission is to build the social and technical bridges to enable open sharing of data.

**Research Data Alliance and Preservation Association (RDAP):** Organization that brings together information professionals committed to creating, maintaining, advancing, and teaching best practices for research data, access, and preservation.

**Scholarly Publishing and Academic Resources Coalition (SPARC):** An international alliance of academic and research libraries working to create a more open system of scholarly communication, and long-time leader in open advocacy.

**SCHOLIX:** A community-wide collaborative initiative to establish a high level interoperability framework for exchanging information about the links between scholarly literature and data.

**Scientific Electronic Library Online (SciELO):** A major Latin American cooperative working across borders and institutions to develop a robust, common methodology for the preparation, storage and dissemination of scientific literature.

**SCOPUS:** Scopus, an Elsevier product, is the world's largest abstract and citation database of peer-reviewed literature: scientific journals, books and conference proceedings.

**SHERPA/Juliet:** A complement to Romeo, SHERPA Juliet is a searchable database and single focal point of up-to-date information concerning funders' policies and their requirements on open access, publication and data archiving.

**SHERPA/Romeo:** Global database of publisher copyright policies. Managed by JISC.

**Society for Scholarly Publishing (SSP):** A nonprofit organization promoting and advancing communication among all sectors of the scholarly publication community through networking, information dissemination, and facilitation of new developments in the field.

**UK Research and Innovation:** The primary research management body in the UK, operating with significant government funding and authority and combining the resources of seven different UK Research Councils.

**United Nations Educational, Scientific and Cultural Organization (UNESCO):** The UN body charged with developing and advocating global approaches and solutions to issues regarding education, science and culture.

**Unpaywall:** A searchable database of nearly 20 million legally free scholarly articles from 50,000 publishers, repositories, library systems and information products worldwide.

**Web of Science (WoS):** A Clarivate product, WoS is one of the most widely used citation indexes in science.

**Wellcome Trust:** One of the world's private philanthropies, focused largely on funding biomedical research. Wellcome has created significant open policies for its funded researchers, and these policies are being followed by many other organizations.

**World Bank:** Long-time leader in international development with highly influential open data repositories and data reporting policies.

**World Intellectual Property Organization (WIPO):** The United Nations agency charged with managing global intellectual property services, policy, information and cooperation.

## **POLICIES (SAMPLING)**

**DORA:** The San Francisco Declaration on Research Assessment is a set of recommendations for improving scholarly publishing, mostly centered around reducing the use of impact factors. As of mid-2018, around 500 organizations and 12,000 individuals have signed this declaration.

**FAIR:** FAIR is a set of four foundation guiding principles for data management and stewardship in scholarly research: Findability, Accessibility, Interoperability, and Reusability. These four principles are intended to guide data producers and publishers in order to help maximize the value from digital scholarly research publishing.

**Gates Foundation open policy:** The Gates Foundation's open access policy requires its researchers to publish in fully open formats—CC-BY in open access journals with no embargo period. Whether this model spreads is too early to tell—the Wellcome Foundation created a similar model as well.

**Harvard open policy:** In 2008, Harvard became the first US university to roll out an open policy for faculty, and perhaps the first in the world where this policy was adopted by faculty. In the years since, Harvard's model has been considered by many other universities in the development of their own open policies.

**Horizon2020:** The largest EU Research and Innovation program ever (covering 2014 to 2020), including new requirements for open publishing in research. (Note: Horizon2020 is the 8<sup>th</sup> framework program for this work; FP9 is the working name for the next framework program, which will run from 2021-2027.)

**Joint Declaration of Data Citation Principles (JDDCP):** A synopsis and harmonization of the recommendations of major science policy bodies, developed by the Publishers Early Adopters Expert Group as part of the Data Citation Implementation Pilot (DCIP) project, an initiative of FORCE11.org and the NIH BioCADDIE program.

**OA2020:** The so-called "global flip" effort, this is a project being led by the Max Planck Institute with the goal of converting more subscription journals to open access journals.

**Open, Public, Electronic and Necessary Government Data Act (OPEN):** The US government law that powers data.gov, providing a sweeping, government-wide mandate for US federal agencies to publish all their information as open data using standardized, non-proprietary formats.

**Open access button:** A bookmarklet for browsers that helps readers get legal versions of research papers quickly. By clicking the button, when an article isn't freely available the OA button staff asks authors to share it (if this can be done) by putting it into a repository so requesting readers others can get access.

**Open Data Charter (ODC):** A collaboration between over 100 governments and organisations working to open up data based on a shared set of principles (see annex section for ODC charter). ODC is an extension of the G8 Open Data Charter.

**Open Science Framework:** A suite of open research and collaboration tools built by the Center for Open Science.

**Pathways to Open Access toolkit:** Produced by the University of California system, this toolkit analyzes the many approaches and strategies for advancing the large-scale transition to OA, and identifies possible next action steps for UC system-wide investment and experimentation. (Note: As a follow-on, the UC system will be hosting a pathways to open conference in October 2018.)

**Registry of Open Access Repository Mandates and Policies (ROARMAP):** A searchable international registry charting the growth of open access mandates and policies adopted by universities, research institutions and research funders.

**Research Excellence Framework (REF):** The system for assessing the quality of research in UK higher education institutions.

**US Public Access program:** Open access is widely seen as the benchmark for open. However, the US public access program is much larger in terms of the volume of materials being made available. Under this program, US federal agencies are required to make research publicly available, but not "open access," meaning that research can still carry traditional copyright, for instance, as long as it is made freely available to the public within a reasonable period of time.

**Wellcome open access policy:** Similar to the impact of the Gates Foundation's open access policy, Wellcome's policy also generated a lot of discussion in the open community.

Source: The initial version of the groups and definitions list was drawn from the work of Jon Tennant and Ross Mounce, "Open Research Glossary," May 2015, Figshare. <http://dx.doi.org/10.6084/m9.figshare.1482094>, Additions were made from the OSI website ([osiglobal.org](http://osiglobal.org)) and various other resources.



# OPEN SCHOLARSHIP INITIATIVE