Open Science for a Global Transformation

1. Key aspects of a transition to Open Science: Summary as input towards the UNESCO Recommendation

Importance of a UNESCO Recommendation
A UNESCO Recommendation on Open Science is timely, important and urgent. Open Science is the mode or paradigm for science in the 21st century. There are numerous examples of Open Science practice, but they are unevenly distributed. There is a need for governmental intervention to support open science, to reduce inequalities of adoption and to mitigate some negative consequences.

We support a UNESCO Recommendation as a mechanism both to form and to reflect global consensus. It should aim to play an important advocacy role, to win hearts and minds, in scientific communities and in governments of member states. Furthermore, and most practically, it will serve as a reference to encourage concrete action, policy interventions and investment.

Key messages for the UNESCO Recommendation.
In the following sections of this document, we have emphasized (in bold) certain key statements or recommendations which we think are important messages for inclusion in the UNESCO Recommendation. The most important of these are presented below.

Open Science is best characterised as the necessary transformation of scientific practice to adapt to the changes, challenges and opportunities of the 21st century digital era to advance knowledge and to improve our world. This requires changes in scientific culture, methodologies, institutions and infrastructures. These changes are already present in many research domains and institutions, where their transformative effects can be witnessed, but they are unevenly distributed. One of the purposes of Open Science viewed as a call for transformation, is to ensure that 'no-one is left behind'.

What is Open Science and why is it timely and important?

- Open Science is a key enabler for a post-COVID global transformation of reasoned discourse and informed decision-making.
- Open Science is a major advance and force to ensure full and equitable participation in the creation of knowledge, through approaches that are transparent, subject to scrutiny and critique, and verifiable.
- Open Science has become an important transformative advance, with implications for policy, the funding, the evaluation and practice of science systems. Because Open Science implies changes in practice, in policy, and in the priorities of scientific infrastructure, it is important and timely that UNESCO is preparing a recommendation based on global consultation.

What are the objectives and benefits of Open Science?

- Open Science aims to maintain and promote good practice and scientific reproducibility by maximising access to robustly described data, code and methods underpinning scientific conclusions.

- Open Science aims to maximise the reuse and (re-)combination of data and code, and to maximise the benefits of investment in science and scientific infrastructure.
- Open Science aims to maximise the benefit of science for society and the engagement of society with science.
- Open Science, through responsible governance, allows and requires necessary and proportionate protection of data, its sources, and derived information. *It categorically does not mean indiscriminate openness.*
- The UNESCO Recommendation should encourage research communities to agree on Open Science frameworks (which define community practices for data sharing (including ethics and Intelligent Openness), data formats, metadata standards, tools and infrastructure. The recommendation should also encourage governments, funders and other stakeholders to support this process.
- Open Science has the potential to be a major force for positive change in LMICs as well as in economically advanced countries. LMIC Open Science and open data initiatives are needed to ensure that data is gathered, stewarded and analysed by LMIC researchers according to their own priorities and research objectives.

What are the neglected aspects of Open Science that need to be addressed?

- The infrastructure, procedures and priorities established through Open Science measures have substantive political and socio-economic implications, which need to be considered whenever designing, developing and implementing Open Science infrastructure and projects.
- It is imperative that the broader impact of Open Science is regularly monitored and mechanisms are in place to adapt to shifts in social and scientific circumstances.

Open Science Infrastructures: the technical infrastructures required for Open Science

- Open Science requires investment in a federated and diversified information infrastructure that nevertheless follows certain core principles.
    - Every digital object of significance for science (whether a datum, a dataset, metadata, code, a publication etc) needs a persistent identifier (PID).
    - Data require rich metadata covering a broad range of attributes which allow the data to be assessed, accessed, understood, and reused.
    - Open Science depends on a global network of data repositories, the trusted organisations that take responsibility for the long-term stewardship of data for defined research communities.
- Proper application of 21st century technology will require consensus within the open science community on mechanisms to certify automated and artificial intelligence-generated research outputs in respect of trustworthiness.

Open Science Infrastructures: optimal governance, funding and ownership arrangements

- Scholarly publishing needs fundamentally reenvisioning such that it is unambiguously put to the service of science as a global public good.
- There is a need for a strategic vision and process to identify the services and 'core resources' for Open Science and to ensure that the funding streams and business models that maintain them are robust.

- Research funders need to recognise that to fund research and Open Science must mean to ensure also that the infrastructure to support it is dependable, reliable, and available for the long term.
- Open Science infrastructure needs to be resourced and staffed in ways that recognise and reward the high level of expertise required to develop and update it.
- Coordination of largely nationally-oriented funding is necessary to ensure that the global and international imperatives of science, and the infrastructure necessary for Open Science, are sufficiently addressed.
- Market dominance is cause for concern in a world where so many services, including the tools of Open Science, are dependent on the 'giants of the web'.

Capacity Building for Open Science

- To take advantage of the opportunities and to use FAIR data and Open Science for good, requires significant investment in capacity building, education and training.
- Priority investment is needed in the areas broadly defined as data science and data stewardship.
- A core set of data science and data stewardship skills should be regarded as part of the foundational expertise of all researchers and incorporated into the 'research skills' curriculum (from at least undergraduate level, if not before).
- Advanced graduate level training programmes are needed to develop the professions of data scientist and data steward, to support and enable 21st century Open Science.
- It is essential to provide training in the ethos and ethics of Open Science.

Negative impacts of Open Science and how to address them

- Research-performing organisations and research funders need to ensure that there is effective reward and incentives for good practice Open Science activities and outputs.
- UNESCO must promote the development of effective governance measures in order to address inequality and prevent related predatory behaviours.
- It is crucial that measures used for the governance of research materials be proportionate and reasonable, and that consensus-based criteria are identified and employed to define what *proportional* and *reasonable* mean in practice. UNESCO can play an important role in encouraging such conversations and enabling the identification and implementation of effective governance for Open Science.

What are the obstacles to reaching global consensus on Open Science and how can they be addressed?

- There is an opportunity by means of the UNESCO Recommendation to chart out a roadmap and a set of shared, global Open Science Goals.

Open Science and COVID-19

- The size and scale of the COVID-19 challenge, the urgency of effective science to inform evidence based decision-making, underline the need for Open Science and its components.

2. Introduction: why is Open Science important and timely?

This document has been prepared by an expert group, coordinated by CODATA, and including representatives from the ISC World Data System (WDS), GO FAIR and the International Council for Scientific and Technical Information (ICSTI). These global data and information organisations have responsibilities for policy advice and promoting collaborations in relation to data and information; they are advocates of Open Science and are convinced of its necessity and benefits. The organisations believe—fundamentally—that it is important to win hearts and minds, to significantly transform science policy and practice, in order to facilitate the application of the technologies and methodologies associated with Open Science to improve our world and address global challenges. Open Science requires policy interventions and investments to ensure equitable participation and access to its benefits and to mitigate some negative consequences. **The perspective brought by these organisations lays particular emphasis on maximising appropriate access to (re-)usable data.**

Open Science is a key enabler for a post-COVID global transformation of reasoned discourse and informed decision-making. The scientific approach, resting on the transparent, scrutinizable and verifiable use of data, evidence and methods, is essential to the creation of reliable knowledge and effective policy. The major challenges of the 21st century—including but not limited to pandemics, climate change, loss of biodiversity—require the massive use of data and analysis to comprehend and intervene in complex systems. Datasets are often too large and complex to analyse without daily assistance of computational methods, and therefore data must be machine actionable wherever possible, in ways that facilitate human scrutiny and judgement. Open Science is fundamental to efficient and responsible scientific collaboration, and thus provides the tools, methods and processes for the analysis of global and complex systems.

We argue, therefore, that Open Science is best characterised as the necessary transformation of scientific practice to adapt to the changes, challenges and opportunities of the 21st century digital era to advance knowledge and to improve our world.

As expressed in the vision of the International Science Council, **science[1] is a global public good**.[2] This means both that science must serve the global public good, and that it is a good to which all humanity must have access, regardless of geography, gender, ethnicity or socio-economic circumstances. Too often, the (gendered and racist) social structures that have surrounded science have excluded significant parts of humanity from participation in science and the use of its outcomes. Too often, the (false) findings of scientists have been deployed to reinforce (gendered and racist) systems of exclusion or domination. In that context, it is not surprising to encounter suspicion towards science, Open Science and the Eurocentric (or Atlanticist, Northern) character of much discourse around it. Nevertheless, inherent in the scientific approach is the fundamental notion of

[1] In this document **we use and endorse the International Science Council's definition of science as 'the systematic organisation of knowledge that can be rationally explained and reliably applied'. Essential and fundamental to this definition is that science is a process and an approach which requires that findings are transparent, scrutinizable and verifiable**. This definition of science is inclusive of the physical and natural sciences, the social sciences, the health sciences, the engineering and computational sciences and the empirical humanities. It rejects the anglophone distinction between the natural sciences and other branches of verifiable knowledge as an unnecessary accident of language and history.

[2] 'The vision of the Council is to advance science as a global public good. Scientific knowledge, data and expertise must be universally accessible and its benefits universally shared.' https://council.science/actionplan/isc-vision-and-mission/

scrutiny and critique. It is not just the findings of science that must be transparent, subject to scrutiny and verification. So must the practices, social structures and methods of science be transparent, subject to scrutiny and critique. Although neither of these are always the case in practice, they are fundamental to the principles and philosophy of science. **Open Science is a major advance and force to ensure full and equitable participation in the creation of knowledge, through approaches that are transparent, subject to scrutiny and critique, and verifiable.** As such, Open Science is an important tool to protect scientific reasoning and the validity of its findings against attacks from anti-science movements, religious fundamentalists, populists, ultra-nationalists and those that use groundless discourses around gender and race to maintain inequitable social structures and systems of power.

Open Science must aim to emancipate science from systems of gendered, racist and socio-economic exclusion. It is a mode of doing science that maximises the participation of all members of society in the creation of knowledge. For this reason, Open Science should embrace all epistemological approaches that can be scrutinised and verified, not just those that rely on Eurocentric (or Atlanticist, Northern) definition and social structure of doing science for their authority.

We endorse, therefore, the definition presented by the ISC. Open Science is:

> *Science that is open to scrutiny and challenge, and to the knowledge needs and interests of wider publics. Open science makes the record of science, its evolving stock of knowledge, ideas and possibilities accessible [...] to all, irrespective of geography, gender, ethnicity or [socio-economic] circumstance. It makes the data and evidence of science accessible and re-usable by all, subject to constraints of safety, security and privacy. It is open to engagement with other societal actors in the common pursuit of new knowledge, and to support humanity in achieving sustainable and equitable life on planet Earth.*[3]

Open Science is a means to an end. It aims to deploy the tools, methods and practices of the 21st century to increase understanding of society, of our world, of the universe. It is the hope of Open Science that this robust, evidence-based understanding will also be used for transparent, reasoned decisions and action to improve our world and address global challenges. The increased openness of scientific discourse, the engagement of the wider population with knowledge creation, will also—ultimately—have transformative and beneficial effects.

For these fundamental reasons, and others, **Open Science has also become an important transformative *advance*, with implications for policy, the funding, the evaluation and practice of science systems. Because Open Science implies changes in practice, in policy, and in the priorities of scientific infrastructure, it is important and timely that UNESCO is preparing a recommendation based on global consultation.**

This submission largely follows the structure of the UNESCO Open Science survey but in a number of places provides more detail and justification of the statements made. Some specific sections of the questionnaire are not formally included in order to avoid repetition, but we think we have addressed

[3] From the ISC submission to the UNESCO review of Open Science 'Open Science for the 21st Century: Draft ISC Working Paper'.

the key issues. Through the document we have put in bold some key statements which we hope will be useful in the process of gathering material for the Recommendation. Overall, we hope that this will be useful and constructive input to the global consultation.

3. Data Together Organisations and Open Science

The four major international data organisations (CODATA, RDA, WDS and GO FAIR) produced the Data Together Statement, published in March 2020. In it they outline their joint commitment to optimise the global research data ecosystem and outline their complementarity and joint collaboration in this effort. Driven by the core mission and expertise of each organisation - be it advancement of agreements on data policy and protocols (CODATA), building consensus and developing standards (RDA), data stewardship and access provision (WDS) or agreements and joint implementations for FAIR data (GO FAIR) - they join forces to contribute to promotion and advancement of the data aspects of Open Science.

4. What are the objectives and benefits of Open Science?

Open Science is sometimes viewed as an umbrella term, comprising a number of related issues, perspectives and initiatives. These include but are not limited to Open Access to scientific literature, Open Source software, Open Data including access to public data for research or to scientific data for scrutiny of results and for reuse. Although this compound understanding of Open Science serves some purpose it fails adequately to communicate its fundamental need, justification and benefits. Science has always been, by definition and often in practice, as open as possible. In many respects, Open Science has a long history, including centuries-old efforts to share data, protocols and instrumentation in astronomy, meteorology and natural history. Nevertheless, there are features of the current phenomenon of Open Science that are distinctive to our age.

We argue that Open Science is best characterised as the necessary transformation of scientific practice to adapt to the changes, challenges and opportunities of the 21st century digital era to advance knowledge and to improve our world. This requires changes in scientific culture, methodologies, institutions and infrastructures. These changes are already present in many research domains and institutions, where their transformative effects can be witnessed, but they are unevenly distributed. One of the purposes of Open Science viewed as a call for transformation, is to ensure that 'no-one is left behind'.

This box presents key points in conceptualising Open Science and its benefits:

- Science has always required transparent access to data, methods, evidence.
- The digital era offers significant technological and methodological opportunities for science.
- Most notably these opportunities include the ability to create data at far greater rates, to analyse data at greater scale and mechanisms for unprecedented engagement of the global population in science.
- Open Science can be characterised as the adaptation of scientific practice to take advantage of these opportunities.
- It requires changes in scientific practice, institutions and relationships with society.

- These changes are already present in many research domains and geographies but are unevenly distributed.
- The 21st century offers many challenges (characterised by digital transformations, the anthropocene and so on). The scientific approach is necessary to address these challenges and Open Science is the mode of doing science most adapted to this era.
- The ultimate goal of Open Science is to advance knowledge, to understand human and natural processes, to improve our world and to address major challenges.

Open Science aims to maintain and promote good practice and scientific reproducibility by maximising access to robustly described data, code and methods underpinning scientific conclusions. Transparency to allow the scrutiny of data, methods, analysis and reasoning is fundamental to the scientific approach, as is the principle of reproducibility. The scientific outputs required for independent verification need to be open. Increased data volumes, use of large scale analysis, algorithms and machine learning are some of the factors which imposed adaptations in scientific communication and infrastructure to allow transparency and scrutiny. Rich metadata, contextual information and provenance, and where possible machine actionability, are increasingly important.

Open Science aims to maximise the reuse and (re-)combination of data and code, and to maximise the benefits of investment in science and scientific infrastructure. Some scientific domains have benefitted immeasurably from large scale data creation and sharing. Some scientific fields require considerable investment in data creation and related infrastructure: both scientific and economic imperatives require maximal use and reuse of quality controlled data and software. Maximising reuse of data and associated workflows has not been an imperative in all domains of research, but almost all would benefit from it. This requires a larger degree of openness than that needed for reproducibility and verification.

Open Science aims to maximise the benefit of science for society and the engagement of society with science. Important in the objectives of Open Science is to encourage economic and societal benefit through promoting collaboration between research and innovation sectors. This implies 'Open Innovation' and maximising the creation of pre-competitive agreements and sharing. Often neglected in discussions of Open Science is engagement with society more broadly, through citizen science, engaged research and transdisciplinary research methods. Open Science should not just be viewed as of benefit to the science sector, providing the tools to do science more effectively at greater scale: it should be viewed as comprising the fundamental involvement of societal actors from all sectors with science and knowledge creation.

The Open in Open Science is often misunderstood and mischaracterized. **Open science, through responsible governance, allows and requires necessary and proportionate protection of data, its sources, and derived information.** ***It categorically does not mean indiscriminate openness.*** All serious policy statements on Open Science and Open Data include necessary qualifications and caveats that cover sensitive data, personal data, appropriate commercial interests, ethical issues, public good and security matters, protection of endangered species, indigenous cultures and so on. The point of emphasising Open is to make clear that the opportunity to access and use scientific outputs, including literature, data and code, should be the ***default*** unless one of these conditions apply. The maxim 'as Open as possible, and *only* as closed as necessary' communicates the need to respect protections and to do so proportionately.

There is an important relationship between Open Science and the FAIR principles, which argue that in order to have maximum scientific utility, for analysis by humans and by machines, scientific data should be Findable, Accessible, Interoperable and Reusable.[4] On the one hand, FAIR is independent of Open Science. Even when data are necessarily closed and protected, or when they are embargoed while a research group or consortium's analysis continues, their utility is increased by being FAIR. On the other hand, all the benefits of Open Science cannot be achieved without FAIR. It is simply not enough to make data (or other research outputs) Open, on an arbitrary technical platform, without additional information that allows them to be discovered, that makes them intelligible. Data have significantly attenuated utility if they are not richly described, accessible by humans and machines, clearly reusable and repurposable. Put another way: **Open is not a necessary precondition to FAIR, but FAIR is a necessary precondition to Open Science.**

The FAIR principles cover the functional and technical principles to ensure that data and other scientific outputs have maximum utility. They do not aim to do more than this. **It is important to augment FAIR with other sets of principles, that have been developed for important (but distinct) service and ethical dimensions of Open Science.** The recent TRUST Principles (Transparency, Responsibility, User focus, Sustainability and Technology) are 'a set of guiding principles to demonstrate digital repository trustworthiness' and are important for one component of the Open Science ecosystem, those organisations that look after research data.[5] The CARE Principles for Indigenous Data Governance (Collective Benefits, Authority to Control, Responsibility, Ethics) add an ethical dimension which is essential to Open Science.[6]

Open Science and FAIR frameworks are needed that define community practices for data sharing (including ethics and Intelligent Openness[7]), data formats, metadata standards, tools and infrastructure.[8] Where most successful, these reflect a community agreement covering such things as 'this is how we communicate and debate our science, this is how we provide access to data, this is how we describe key concepts and key measurements and observations'. Sometimes such agreements are piecemeal or implicit: the more explicit and consensual they can be the better.[9] An important component of Open Science transformation is to encourage such agreements and to make them as explicit and reusable as possible. In doing so the facilitating role of policy and funding is important. **The UNESCO recommendation should encourage research communities to agree on Open Science frameworks. The recommendation should also encourage governments, funders and other stakeholders to support this process.**

---

[4] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[5] Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. Sci Data 7, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7

[6] CARE Principles for Indigenous Data Sovereignty https://www.gida-global.org/care

[7] The Royal Society's Science as an Open Enterprise report coined the useful term 'Intelligent Openness', which comprises practices closely aligned to the FAIR principles as well as the necessary and legitimate boundaries of openness.

[8] Turning FAIR into reality. Final report and action plan from the European Commission expert group on FAIR data. https://doi.org/10.2777/1524

[9] There are many examples of such agreements: the famous Bermuda and Fort Lauderdale agreements in genomics, the role of the IUCr in crystallography, the International Virtual Observatory Alliance, the Dryad Joint Data Archiving Policy in evolutionary biology and biodiversity. Each of these varies in scope and have different status in different domains, but there is a family resemblance of a particular research area making a consensus statement about 'how we do our science, how we share our data, what definitions and standards we use'.

Open Science has the potential to be a major force for positive change in lower and middle income countries (LMICs) as well as in economically advanced countries. In mobilising the processes and technologies of the 21st century, it provides the tools to help address complex and pressing challenges and to improve informed decision-making. There are potential benefits also of broader societal engagement and genuine global participation in Open Science and knowledge creation. None of these benefits will be realised without policy intervention, societal engagement and investment. LMICs have, in many cases, been slow to embrace Open Science. There are a variety of reasons for this, which include (but are not limited to) the expensive technological and infrastructural requirements, resistance to change in scientific communities, concerns about data and cultural exploitation and the use of Indigenous Knowledge without compensation, a fear of losing data sovereignty and Intellectual Property, and fear of damage to individuals, groups, or the environment.

Yet, the benefits of Open Science for LMICs are clear. A number of African countries are exploring strategies for Open Science, open data and open innovation as drivers of transparency and economic development.[10] There is a recognition of the transformative effects of technologies and of Open Science as a means of harnessing them for societal benefit. Mainstreaming Open Science, it is argued, 'open science will allow universities, research centres and society to take advantage of the benefits of collaborative, transdisciplinary approaches to knowledge development and sharing.'[11]

The more pressing risk is the converse: that the accelerating, global Open Science initiative will lead to further widening of the unequal scientific contribution and output between the 'global north' and the 'global south' and the concomitant control of scientific agendas. **LMIC Open Science and open data initiatives are needed to ensure that data is gathered, stewarded and analysed by LMIC researchers according to their own priorities and research objectives.** Promising examples are extensive and growing: they include the H3Africa research collaboration, the network of Health Demographic Surveillance System (HDSS) sites and collaborations like INDEPTH and ALPHA that use this data, and emerging initiatives such as the Malaysian Open Science Cloud and the African Open Science Platform.

5. Neglected aspects of Open Science

Political, social, ethical and economic factors underpin many of the decisions involved in implementing Open Science, and it is imperative that such factors are recognised and explicitly discussed among the relevant stakeholders. The collection and handling of most datasets is not a purely technical exercise, but rather involves socially sensitive dimensions which need to be taken into account at all times.[12] Similarly, the adoption of specific methods, procedures, publication routes and communication tools encompasses value judgments and considerations of interest that are

[10] See for example, Kenya Open Data http://icta.go.ke/open-data/; the Uganda Open Data Policy https://www.ict.go.ug/wp-content/uploads/2018/06/Open-Data-Policy-First-Draft-vX.pdf; Botswana Open Data Open Science White Paper https://zenodo.org/record/830626/files/POSITION-PAPER-ON-OPEN-DATA-AND-OPEN-SCIENCE-BOTSWANA.pdf

[11] South Africa, Department of Science and Technology White Paper on Science Technology and Innovation https://www.dst.gov.za/images/2018/white-pate-on-STI-7_09-FINAL.pdf; see also the SA-EU Open Science Dialogue Report https://doi.org/10.5281/zenodo.2559469

[12] See for instance British Academy & Royal Society Report on Data Governance (2017) https://www.thebritishacademy.ac.uk/publications/data-ai-management-use-governance-21st-century/ ; Cathy O'Neill (2016) *Weapons of Maths Destruction: How Big Data Increases Inequality and Threatens Democracy*. Penguin; European Approach to Artificial Intelligence (2020) https://ec.europa.eu/digital-single-market/en/artificial-intelligence .

grounded in the historical, geographical and social positioning of the researchers in charge.[13] Lack of acknowledgment of such positioning facilitates the exclusion of minority views from scientific discussions, and contributes to enhancing the racial, gendered and socio-economic divides that are often entrenched in the conduct of research.[14] If Open Science is to counter such tendencies, it needs to explicitly call attention to discriminatory practices and social dilemmas raised by research.

Too often, the need to reach agreement on how to implement Open Science measures is viewed as a technical or administrative issue, which can be resolved for instance through the development and implementation of standards by a small group of technical experts. By contrast, **the infrastructure, procedures and priorities established through Open Science measures have substantive political and socio-economic implications, which need to be considered whenever designing, developing and implementing Open Science infrastructure and projects**.

Facilitating relevant debates and agreements, and the regular review of these agreements, within and across disciplines and social domains are crucial. The circumstances of Open Science implementation may be changing all the time depending on the research domain and location at hand. Moreover, the effects and implications of Open Science measures may take some time to become apparent. **It is imperative that the broader impact of Open Science is regularly monitored and mechanisms are in place to adapt to shifts in societal and scientific circumstances.**

Relatedly, there is a tendency to neglect the importance of equitable, sustainable and responsible approaches to Open Science, to ensure maximal uptake among relevant stakeholders in all sectors and geographies and encourage use of such data for socially beneficial purposes. The ongoing discussion around CARE principles on indigenous data governance, among other initiatives, signals that reference to FAIR principles is not enough to address concerns with digital divides and inequality relating to Open Science and particularly Open Data.

The national scope of most funding sources and research infrastructures limits our ability to address major global research challenges. Open Science provides the tools to overcome those limits. To do this, the development of shared principles and practices for Open Science, frameworks and standards for interoperability are important. There is also a pragmatic intention to enable federation between numerous small scale infrastructure or even major national or regional Open Science Platforms. But this will not happen without concerted and major effort to build consensus and promote coordination. **If Open Science is to be a key enabler for a post-COVID global transformation, it needs to be accompanied by a bold vision for international cooperation, resourcing, practice and implementation.**

This broad vision has very concrete implications for the management of specific components of Open Science, and particularly data. **Maximising (re)use of data is needed and yet it remains neglected.** The FAIR principles have provided a framework for making data reusable, which centres on the need for data to be readable, and actionable where possible, at scale by machines to facilitate Machine Learning and Artificial Intelligence applications. This typically includes both generic and

[13] Douglas, H (2009) Science, Policy and the Value-Free Ideal. Pittsburgh University Press; Leonelli, S. (2017) Global Data Quality Assessment and the Situated Nature of "Best" Research Practices in Biology. *Data Science Journal* 16(32): 1-11. DOI: 10.5334/dsj-2017-032

[14] Nature editorial, 2020 Systemic racism: Science must listen, learn and change. https://www.nature.com/articles/d41586-020-01678-x

domain-specific standards and recommendations. It also requires systematic reflection on what makes data, algorithms and related services trustworthy. A good example for repositories are the recently published TRUST principles but **further consideration of the notion of trust for data themselves, for algorithms and the outcomes of machine learning is needed.**

6. Open Science Infrastructures

This box presents key considerations in relation to Open Science Infrastructures

- A persistent, trusted, and evolving infrastructure that will ensure that more research data and other digital research outputs and tools are easily used and more effectively reused is a precondition for Open Science.
- Most of the current Open Science infrastructure, beyond Open Access to publications, must be considered to be in its infancy. Stable and useful data infrastructure is lacking for many disciplines and there is almost no dedicated infrastructure for standards, research software, and methods in general.
- Such infrastructure will not be achieved by a single organization, and even not by many single nations. Key components may need to be provided by globally available commercial entities.
- Infrastructure should be funded as such (not as short-duration projects, as it is common today), and better be based on a sustainable business model, serving the needs of a sizeable user community.
- Some elements of a global infrastructure require collective funding, or at least collective governance
- Data-centred science initiatives will increasingly depend on regional and national or even global, commercial and non-commercial data clouds.
- It might be beneficial to achieve consensus on a minimal infrastructure (generic or otherwise; national, regional or global) which should be available to each scientist.
- There is a pressing need to come to terms with the role of commercial service providers and how to maintain openness when these become involved.

As noted above, our premise is that Open Science is best characterised as the necessary transformation of scientific practice to adapt to the changes, challenges and opportunities of the 21st century digital era. **Realising the potential of Open Science requires investment in technical and skilled human infrastructures.** Technical infrastructures are the subject of this section, while human capacity is considered in the next. Both are essential.

Following this argument, this section on technical infrastructures deals with two related questions:
1: What are the key technical infrastructures required for Open Science?
2: What are the optimal governance, funding and ownership arrangements to ensure that Open Science remains open and serves the global, public good?

The Technical Infrastructures Required for Open Science

On behalf of scientific domains, research funders invest in data creation and analysis exercises, ranging from small individual projects, to national data collection activities (including household surveys), to major investments in international initiatives, such as CERN, SKA and so on. Optimal return on investment generally requires that the data thus created is made available for reuse. The reuse statistics for the Hubble Space Telescope demonstrate the value of this.[15] The 21st century digital era has increased the value for science of the (re-)use of data and the imperative for infrastructures that facilitate access to well documented and usable data and other research outputs, including code. Mechanisms for this range from deposit in subject repositories, code hubs, or programmatic access to federated databases in multiple locations.

Historically, each scientific domain has determined the approaches taken for data creation, description and sharing, generally mediated through national funding regimes, but with some notable examples of international cooperation. It remains essential that scientific domains take a lead in determining agreements and frameworks (mentioned above) that define community practices for data sharing (including ethics and Intelligent Openness), data formats, metadata standards, tools and infrastructure. However, Open Science and FAIR provide general principles that are valid across all domains, although implementation will vary. The analogy of an ecosystem is frequently used to emphasise the interrelatedness of components, which nevertheless operate in particular (subject or geographical) niches.

Open Science requires investment in a federated and diversified information infrastructure that nevertheless follows certain core principles. Centralised planning is not feasible or appropriate and would not be effective. Yet without consensus and (a relative degree of) alignment, convergence and standardisation around certain technologies the objectives of Open Science cannot be realised and research will remain excessively fragmented and piecemeal, unable to apprehend and address major, global and holistic challenges.

It is possible to provide a catalogue of the key technical components necessary for FAIR and Open Science. This is not the place to provide a detailed roadmap for these infrastructures. But it is important to stress the importance of developing global consensus around these components. 21st century information technology cannot function without the common application of technologies for the internet (TCP/IP) and the web (HTTP). The technical enablers of FAIR and Open Science are more diverse, but the analogy holds.

Every digital object of significance for science (whether a datum, a dataset, metadata, code, a publication etc) needs a persistent identifier (PID). Ideally, such an identifier should be globally unique, persistent and resolvable. Without effective and global systems of PIDs (such as Handles and the DOI implementation of the Handle system) it is difficult for humans and machines to reference, cite, link, access digital objects without ambiguity. This is a key element in finding and accessing data and in enabling machine usability.

Data require rich metadata covering a broad range of attributes which allow the data to be assessed, accessed, understood and reused. This covers a large range of necessary information including how the data was created and transformed; the definition of what was measured, observed, recorded; the relationship that different classes of variables may have within a dataset and so on. Developing and maintaining clear and robust definitions (vocabularies, terminologies) and defining for particular fields the relationships between concepts (ontologies) is fundamental to the scientific enterprise. The availability, reuse and (ideally) machine actionability of these semantic tools is a fundamental enabler of FAIR and Open Science. Some domains have applied concerted efforts to

[15] Publication statistics of the Hubble Space Telescope https://archive.stsci.edu/hst/bibliography/pubstat.html

the development and maintenance of terminologies, ontologies or metadata schema. In others the work is piecemeal and poorly sustained.

Storage is not stewardship. **Open Science depends on a global network of data repositories, the trusted organisations that take responsibility for the long-term stewardship of data for defined research communities.** Sometimes these relate to specific data creation exercises. Often they are nationally funded and serve particular research domains. A number of catch-all repositories exist while some institutions have also taken responsibility to look after data outputs. Commercial publishers are increasingly showing an interest in providing a service for research data outputs, particularly those associated with publications. Building and maintaining trust, reinforced by the organisation's commitment to serve a research community and to apply good practice, is essential. As with metadata (and other semantics), the importance of repositories in the scientific process is often neglected and their operational sustainability insufficiently resourced and secured.

Open Science requires and exploits other features of 21st century technology (including but not limited to high speed networks, HPC, cloud computing and virtualisation, bespoke instruments and analysis software, libraries of software tools and developer platforms). It can be argued, however, that the three listed above: persistent and unique identifiers, metadata and other semantics, and a network of trusted repositories are the most distinctive and essential for FAIR and Open Science because it is they that enable the effective use, at scale, of the other technologies. In addition, **proper application of 21st century technology will require consensus within the Open Science community on mechanisms to certify automated and artificial intelligence-generated research outputs in respect of trustworthiness.**

The importance of technology, and of specific technologies, poses a major challenge for Open Science as a global endeavour and for international inclusion. **Full global and inclusive participation in the technologies and methodology of Open Science is essential if the 'digital divide' and the gap in scientific output is not to widen even further.** Technological infrastructure initiatives such as the African NRENs and HPC capacity is essential in this regard. LMIC research communities need a robust system of repositories so data can be discovered, reused and their project credited and participate in the global system of Open Science.[16] It is also vital that LMIC research communities participate in the technologies listed above. Researchers from around the world, including from LMICs, need to participate in the development of metadata schema, conceptual frameworks, terminologies and ontologies. Without that participation, such schema will remain dominated by 'north Atlantic' priorities and concerns. Defining the things that are measured and digitally encoded (whether social categories, soil types, species, phonemes etc) is fundamental to science and enables FAIR and Open Science. Almost all such definitions are subject to bias, conscious or not, and therefore this work needs to be a global, humanity-wide endeavour.

Optimal Governance, Funding and Ownership Arrangements

The international perspective of Open Science and the importance of Open Science as a global public good, means that **it is essential to consider issues of ethics, equity and governance**. Where Open Science receives public funds, it is vital to consider how such funds are disbursed most effectively for public benefit and maximum return on investment.

[16] Bezuidenhout, L., Leonelli, S., Kelly, A. and Rappert, B (2017) Beyond the Digital Divide: Towards a Situated Approach to Open Data. *Science and Public Policy* 44(4): 464-475 doi: 10.1093/scipol/scw036

This box presents key issues in relation to the governance and funding of Open Science infrastructures

- Open Science is intimately tied to an important critique of the current mechanisms for scientific communication and the related system of reward and recognition.
- Paywall publishing does not serve the interests of (publicly-funded) science and the profit margins of the traditional publishers, largely from public funding, are unacceptable.
- The private, profit-making sector is an important and necessary partner for innovation, for professional level services and for commodity services that are not core business of research institutions.
- One of the objectives of Open Science is open innovation.
- As with all private sector involvement in parts of the economy that are subsidised by public funds, it is essential that there is robust regulation.
- For Open Science, this needs to cover certification for quality and service, regulation to avoid lock-in, regulation of profits from public funds, and this should be applied to the science sector just as it is to commodity and utility providers in other sectors of the economy with public investment.
- The business models and sustainability of key social and technical components of Open Science infrastructure need to be considered and assured internationally as a global public good.

An important driver in the Open Science advance, though not the only one, is a critique of the current mechanisms for scientific publication and the reward system that has grown on top of it. Paywall publishing does not serve the interests of (publicly-funded) science and the profit margins of the traditional publishers, largely from public funding, are unacceptable. The Open Access movement (in a number of forms), the arrival of major Open Access publishers, both for profit and not-for-profit, the success of 'pre-print servers' and open peer review are significant developments, a much needed corrective and have had transformative effects in some domains. The landscape remains complicated. Open Access business models are perceived as a threat to the financial models of learned societies, which themselves have served the interests of science. In some instances, it has been hard to disambiguate for-profit Open Access publishers from predatory publishing, which has contributed to the (false) perception of lowering publishing standards. Journals with hybrid models have allowed publishers to charge for subscriptions and for Open Access article processing charges. In many respects, **scholarly publishing needs fundamentally reenvisioning such that it is unambiguously put to the service of science as a global public good.**

With a few exceptions, academic publishing still deals very badly with data and despite marketing around 'the article of the future' has failed to adapt to the necessary realisation that scientific communication now needs to comprise a package of things, including the human readable explanation, the data (and its associated metadata) and sources, the code and methods that allow reproducibility, the machine readable 'assertions' and results that can be automatically combined into systematic reviews and surveys. As it became impractical or impossible to include the factual basis (data) for findings in journal articles, electronic publishing platforms provided for their attachment as "supplementary material", alongside articles, often only supplying derived information, and not the data itself. Furthermore, many, if not most, of these supplements are not independently findable, let alone reusable, and are hardly ever machine readable (whether data tables are provided as PDF, or in proprietary formats, and are lacking standardized descriptions). **The**

principle that scientific conclusions should be transparent, scrutinizable and verifiable is central to the philosophy of science, but current (publication) practices are such that it is really questionable whether this principle is actually honoured in practice.

All data, sources, methods, code and so on which allow the findings presented to be scrutinised and verified should be accessible and intelligible. Trusted data repositories, not supplementary materials, are the right place to look after the data (and code) underpinning and associated with publications. Some publishers, including the established houses and newer SMEs, are refocusing their strategies with the aim to become companies for “data, and information processing and reuse”. There is a long way to go and a lot of distrust to overcome.

There is a need for a strategic vision and process to identify the services and ‘core resources’ for Open Science and to ensure that the funding streams and business models that maintain them are robust. Even more fundamentally, research funders need to recognise that to fund research and Open Science must mean to ensure also that the infrastructure to support this exists: to pretend otherwise is a failure of mission and responsibility. Research infrastructures need to serve the needs of research domains The ESFRI (European Strategy Forum on Research Infrastructures)[17] processes and the South African Research Infrastructures Roadmap[18] are good examples of consultative approaches to identify strategic priorities. The approach of the Elixir research infrastructure in the life sciences to identify ‘core data resources’[19] is also worthy of mention. Such processes need to be holistic and take into account less well resourced research areas and interdisciplinary research. Research infrastructures also need to be designed flexibly in order to be fit for the pressing cross-domain research challenges of our age.

Research infrastructure has to be dependable, reliable, and available for the long term. Where there is an understandable caution among research funders to long term commitments it must be tempered by the realisation that adequate funding of research requires adequate funding of research infrastructure. Many of the research infrastructures upon which Open Science relies are operated on a financially insecure basis, too reliant on soft money, on time limited project funding.[20] This is even the case with many fundamentally important services, the ‘core resources’, such as those providing persistent identifiers, reference vocabularies or data standards, as well as data repositories and services that extract information and data of various sorts to present them in highly curated, reference databases. Too many of such components, the nuts and bolts of the global research data ecosystem, have tiny funds and rely on uncosted, voluntary effort. This results in vulnerability and an inability to scale and to provide robust, user-friendly and service-level infrastructure. Once a resource, initially developed in a scientific domain, becomes widely used and in part routine, the university, institute or consortium that developed it usually cannot maintain it effectively. Often such infrastructures end up being financed from piecemeal project and grant funding or run by volunteers or as a side activity. It is essential therefore to have mechanisms for transition to professional service models, where needed, for key services and infrastructure. Dividing resources between infrastructure and research funding agencies is not an effective solution either, as it widens the gap

[17] https://www.esfri.eu/
[18] https://www.dst.gov.za/images/Attachments/Department_of_Science_and_Technology_SARIR_2016.pdf
[19] See https://elixir-europe.org/platforms/data/core-data-resources and Durinx C, McEntyre J, Appel R et al. Identifying ELIXIR Core Data Resources [version 2; peer review: 2 approved]. F1000Research 2017, 5(ELIXIR):2422: https://doi.org/10.12688/f1000research.9656.2
[20] OECD and CODATA Report: Business models for sustainable research data repositories (2017) https://doi.org/10.1787/302b12bb-en

between the ‘service providers’ and the research communities, condemning the former to resort to endless, duplicative surveys of ‘user needs’. **Open Science infrastructure needs to be resourced and staffed in ways that recognise and reward the high level of expertise required to develop and update it.**

Most funding is only available on a national basis, which means that the global and international imperatives of science, and the infrastructure necessary for Open Science, are insufficiently addressed. Technical and organizational approaches to “federate” national infrastructures—using relatively meager international funds—have so far produced more vision and hope than concrete success, and will be further complicated by recent trends which see international collaboration as antithetical to national interest. There is concern that the rate of progress is slow compared with globally operating, commercial (cloud) infrastructure providers. Considering the growing importance and usefulness of “big data”, and their interdisciplinary combination (think of biodiversity data, satellite data, and climate predictions in the context of disease vectors), there is a technical and economical need to “move the code to the data”, and thus a growing value in providing analytical processing capacity as part of data infrastructures. For reasons of cost, convenience and technical facility, researchers and institutions increasingly look to the major commercial cloud providers, while federations of institutional or national clouds do not provide the same technical or economic advantages.

Cloud computing is a 21st century commodity, so the private sector will have a role. But few people not employed by the GAFAM[21] would pretend that this is a well-operating market. Nothing comes for free. Research infrastructure and projects lured by free or low cost services have then been stung (sometimes in their naivety) by the cost of high-volume downloads (“egress”). The importance of open research software is increasingly recognised. Github, a commercial entity, has become the primary venue and platform for collective development and maintenance of research software, with few alternatives. The freemium model has allowed considerable innovation with an extremely useful tool. But the dependency is a source of potential vulnerability (and one that also masks the continued absence of organisational solutions to the challenge of keeping software operational over more than a few years). **Market dominance is cause for concern in a world where so many services, including the tools of Open Science, are dependent on the ‘giants of the web’ and this should be addressed in the UNESCO recommendation.**

The “public good” dimension of Open Science should guide the funding and provision of research infrastructure. There is clearly a need for robust and professional services as well as for innovation. Private companies, not-for-profit organisations and state funded entities should all have a role in a vibrant economy around Open Science. Private sector service provision and innovation is valuable. It is not acceptable, however, that the science budget should feed excessive profit, nor that the such profit should be reinforced by the privatisation of publicly-funded knowledge and data, uncosted publicly funded effort and a reward system that does not serve the public good. Open Science will require investment in research infrastructure. The current market dominance of traditional publishers and major cloud providers presents a risk which will need to be countered by robust policy and regulation.

[21] Google, Amazon, Facebook, Apple, Microsoft.

Open Science infrastructures need to be technically sound, based on scientifically robust core resources, with a sustainable business model and certification that ensures optimal openness and trustworthiness. There is clearly a role for innovative private sector companies, public-private partnerships and providers of robust, industry standard commodity services, provided there is appropriate certification and regulation to prevent vendor lock-in and predatory behaviour. The role of commercial providers of services and of data renders the call for Open availability of data and transparency about their quality and provenance even more urgent.[22]

Although these discussions are important for Open Science in many parts of the world, they should not blind us to dramatic global inequalities of access to the resources, tools and infrastructure that are fundamental to 21st century Open Science. Many researchers and institutions all over the globe lack even the 'basics', such as reliable electric power and high-speed Internet access. **Open Science can and must be global: therefore the fundamental, essential technological enablers must be addressed as well as the cutting edge technical challenges. There should be a growing concern over the openness of efficient access to research data and the tools of Open Science, for all. No-one and no region should be left behind.** This applies most urgently to the data and Open Science tools needed to address major, global grand challenges. There should be a constant focus and continuing efforts to empower researchers and other users from all countries to participate in global Open Science on an equitable basis. Capacity building, investing in human potential, is essential.

7. Capacity Building for Open Science

Even in the most developed countries and research institutions, capabilities and capacities in data management and research software engineering are scarce, standing in the way of achieving the goal of Open Science and FAIR data. So-called "data scientists", who are adept at actually making use of (big) data in a scientifically meaningful and responsible manner, are also scarce. The shortage of these skills and capacity is even more notable in LMICs. There are a very few cases, globally, where well-funded data infrastructures provide capacity building and development of the skills to support Open Science and FAIR as a substantial part of their mission and budget. A number of universities have started offering Masters programmes in data science. By and large, these target the commercial marketplace, rather than eventual open science applications. There exist a number of initiatives to develop software and data skills, to train data stewards. Nevertheless, these activities are relatively few and in their infancy. There remains a recognised and well documented skills shortage of data scientists and data stewardships in the research sector. Graduate programmes are still behind the curve in incorporating fundamental data skills, research data management and stewardship and responsible data governance into their curricula for foundational research skills.

As previously noted, we characterise Open Science as the necessary transformation of scientific practice to adapt to the changes, challenges and opportunities of the 21st century digital era to advance knowledge and to improve our world. **To take advantage of the opportunities and to use FAIR data and Open Science for good, requires significant investment in capacity building, education and training.** For example, the first EOSC High Level Expert Group estimated that over half a million 'Core Data Experts' would be needed within a decade if the initiative were to achieve its

[22] See the recent case regarding data on treatments with Hydroxichloroquine, "An open letter to Mehra et al and The Lancet", DOI: 10.5281/zenodo.3862789

potential.[23] **Priority investment is needed in the areas broadly defined as data science and data stewardship.** These categories and skillsets are indicative, and are helpful in designing curricula, but should not be viewed as hermetically distinct: the roles overlap and individuals will likely need to draw on skills from each of the categories.[24]

Data Science can be broadly defined as the theory and practice of handling, processing and analysing data to draw insights from it in a scientifically valid and responsible manner. Data science comprises knowledge and skills drawn from computer science, software development, statistics, visualisation and machine learning. It also covers computational infrastructures and knowledge of information modelling and algorithms and the use of technologies relating to 'big' data including 'clouds' and 'data cubes'.

Data Stewardship consists of the skills and knowledge needed to ensure data are properly managed, shared, curated and preserved throughout the research lifecycle and in subsequent archiving. Important in this are information management skills including the ability to complete and oversee the execution of data management plans, the application of good practice in terms of documentation, use of metadata and semantics. It is essential also to address the skills required research software engineering and for support roles in research performing organisations.

All researchers need at least a foundational set of data skills (comprising those drawn from data science and data stewardship) in order to make adequate use of the data, technologies and methods of the 21st century. Many of the data skills described should be and are recognised as intrinsic to research. The success of initiatives such as the Carpentries,[25] R Ladies[26] and the CODATA-RDA Schools of Research Data Science,[27] demonstrate the need and demand for these skills. **A core set of data science and data stewardship skills should be regarded as part of the foundational expertise of all researchers and incorporated into the 'research skills' curriculum (from at least undergraduate level, if not before).**

Nevertheless, the degree of specialisation, advanced knowledge and the effort required means that many research groups increasingly need to draw on the expertise of specialised individuals in data science or data stewardship (in 'x'informatics or in the development of algorithms or other code, for example). **Advanced graduate level training programmes are needed to develop the professions of data scientist and data steward, to support and enable 21st century Open Science.**

Cutting across each of these general categories are the competencies and knowledge of the approaches that distinguish Open Science. **It is essential to provide training in the ethos of Open Science**, in empirical scepticism, in the fundamental importance of scrutinising evidence, methods and data - and having one's evidence, methods and data scrutinised. Thorough understanding of research ethics and the appropriate limits of Openness are also essential. **Most of the important research issues to address societies' pressing challenges can only be tackled with interdisciplinary**

[23] Realising the European Open Science Cloud (European Commission, 2016): https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf
[24] See the analysis and recommendations in Turning FAIR into reality. Final report and action plan from the European Commission expert group on FAIR data. https://doi.org/10.2777/1524
[25] The Carpentries https://carpentries.org/
[26] R Ladies https://rladies.org/
[27] CODATA-RDA Schools of Research Data Science https://codata.org/initiatives/strategic-programme/research-data-science-summer-schools/

and transdisciplinary methods, by Open Science that engages with society and through international research collaboration. It is important then, to develop the skills and competencies that facilitate these approaches.

It cannot be sufficiently emphasised that **Open Science is a broad societal issue and advance that aims to maximise societies' engagement with science and maximise the role of science in informed decision making.** Generally speaking, among citizens and decision-makers, there is insufficient understanding of how to evaluate and make use of scientific facts and findings, and in particular how to examine their trustworthiness, limits, and level of (un-)certainty. There is also a need to improve understanding of the distinction between observed facts and models, or the concept of verifying (or falsifying) a hypothesis or theory. Such insights are indispensable for informed decision making, based on evolving scientific knowledge. Education in these fundamental aspects of Open Science (scrutinising and verifying assertions, understanding and critiquing data, and so on), should infuse the basic curriculum from as early an age as possible.

The currently most visible societal challenges of the COVID-19 pandemic, climate change and loss of biodiversity have also uncovered a lack of fully understanding such effects as exponential developments, feedbacks and tipping points on short and long timescales, forces of almost instantaneous global interconnectedness on the one hand and huge inertia of global systems on the other. Improved science communication is important but it is not the only issue to be addressed in the science-policy interface. Equally important is to understand what science cannot do. Science provides knowledge necessary to inform decisions, but cannot determine policy decisions by itself: that requires political and social deliberation, which needs also to be made as transparent as possible (as called for by the Open Government movement, which parallels and shares some of the motivations of Open Science). **It is necessary to recognize and close, as far as possible, the gaps between science, society and policy, in order to enable fair and fruitful communication (or even negotiation) between scientists and citizens, and scientists and decision makers. Open Science provides the means to do this.**

Maximising the benefits of Open Science requires investment. Some of the tools and infrastructures of Open Science are cutting edge and expensive. The most important investment, however, as always, is in education and developing the capacity for data science, data stewardship and Open Science. This investment is important for the conduct of Open Science, but it is also essential for the broader engagement of society with science and to maximise the effectiveness of the science-policy interface.

8. Negative Impacts of Open Science and How to Address Them

Many of the negative impacts of Open Science relate to a lack of appropriate governance and stewardship of research components and procedures, and particularly of data. Lack of investment in long-term strategies for data management, curation and credit systems to reward Open Science contributions makes it hard for researchers to follow the FAIR, TRUST and CARE principles mentioned above. **The concerns and potentially negative impacts that follow can all be addressed through coordinated efforts to acknowledge and reward the work involved in supporting and implementing Open Science.**

Among the most prominent concerns are the following:

- **Problems with the premature sharing of results**, and especially concerns around the quality and reliability of the data, models, methods and analysis that are thus disseminated. For instance, data creators may not take the time to properly annotate data with relevant metadata, thus making them unusable or unwittingly encouraging misuse/mishandling by others; and algorithms and models may be shared that contain flaws and mistakes, which if undetected may become shaky ground for future research. There are also strong concerns among researchers in both public and private institutions around "scooping" (publication by a third party of research using given data before the creators of that data have published) or the lack of proper attribution and recognition (re-use of research outputs (particularly data) without citation). The latter is particularly problematic, because citation practices for data, methods and models lag behind those used to cite papers.
  Possible mitigation: **research-performing organisations and research funders need to provide: (1) incentives towards the constructive evaluation of the resources being shared**, which would be aimed at improving their quality and elucidating more or less reliable ways to use those resources; **and (2) clear and widely recognised procedures to track the provenance of the research components being shared and ensure that those who contribute to Open Science activities are appropriately recognised and rewarded**.

- **Open Science may contribute to an increase in the digital divide between researchers in high-income and low-income countries**. It has been demonstrated that research communities with unreliable access to relevant infrastructure, support and/or training have a diminished ability to take advantage of Open Science measures than research communities who have better resources. Without mitigation, this will lead to a deepening of inequality between researchers who are able to exploit Open Science to improve their work and researchers who are not. There is a risk of **predatory behaviour** by researchers based in well-resourced institutions or companies, who may be able to exploit Open Science measures in order to appropriate valuable data from low-resource sites (a situation sometimes referred to as "digital feudalism").
  Possible mitigation: **UNESCO must promote the development of effective governance measures** (such as those proposed via the CARE principles on indigenous data governance and the Nagoya protocol for the exchange of biological materials) **in order to address inequality and prevent related predatory behaviours**. Equal opportunities for the uptake of Open Science also involves a political commitment to investments in the digital infrastructure, computing facilities and reliable bandwidth for use by researchers around the world.

- **Widespread confusion between "fast" and "open" science**: too often the former is assumed automatically to accompany the latter, thanks to the emphasis placed on the extent to which Open Science can help to speed up research and discovery (as for instance in the case of global health emergencies, such as Ebola, Zika and COVID-19). There is a risk that if the speed of Open Science is prioritised above other considerations, it will encourage sloppy, irresponsible research practices.

Mitigation: While there is no doubt that Open Science plays an important role in speeding up research in many cases, **acceleration in the pace of research does not necessarily involve disregarding quality evaluations and/or responsible behaviour.** Rather, there is some evidence that the expectations that others will be scrutinizing one's work makes researchers feel more accountable for the quality, reliability and longer-term implications of their results. Furthermore, **Open Science encourages engaged reflection on the widening of the user base for scientific research and its implications.** For example, citizen science projects are often time-consuming to organise and manage, and yet yield large benefits in terms of being able to engage and consult a wide variety of stakeholders in the development of new findings.

- **Negative social implications of Open Science,** which need to be considered and evaluated especially for sensitive or policy-relevant work. In some areas of research, secrecy or embargos on the sharing of results may be justified by the destructive impact that such results could have on the public sphere. A recent example is the moratorium imposed by preprint archives in biology and biomedicine over non-peer-reviewed projections on COVID-19 contagion and transmission rates, after some such reports were picked up by the media and fed into public discourse without any critical scrutiny of their validity.
  Possible mitigation: **UNESCO should support venues or platforms where the implications of sharing specific research components can be discussed and evaluated, and decisions can be taken on how best Open Science can serve human development.** Such governance is crucial to the implementation of OS, as we signalled in our definition of OS above when emphasising that OS does not involve "indiscriminate openness".

- **Confidentiality is not antithetical to openness**. Again as stressed in our definition of Open Science above, "open" does not need to mean "accessible to all, for any reason, at all times". Instead, depending on the data types or situations, "Intelligent Openness", as defined by the Royal Society in 2012,[28] may need to be facilitated by an ensemble of measures, institutions and infrastructures responsible for overseeing access and reuse of research components.
  Mitigation: There is a need for a portfolio of standardised, machine readable, widely applicable protocols, management systems and licenses for access to (and use of) sensitive data. Additionally, a slew of promising systems are being developed for searching through datasets without necessarily viewing them, thus permitting data (re)use without requiring data access. **It is crucial that measures used for the governance of research materials be *proportionate* and *reasonable*, and that consensus-based criteria are identified and employed to define what proportional and reasonable mean in practice**. **UNESCO can play an important role in encouraging such conversations and enabling the identification and implementation of effective governance for Open Science.**

9. A Global Consensus on Open Science: is it important and urgent?

A notable global consensus has been achieved with the United Nations Sustainable Development Goals. The UNESCO recommendation on Open Science should aim to have comparable ambition in its

[28] Royal Society 2012, Science as an Open Enterprise
https://royalsociety.org/topics-policy/projects/science-public-enterprise/

sphere. **The Sustainable Development Goals are a set of shared objectives that enunciate the necessary components for peace and prosperity, for human flourishing on this planet. In the same way, it is possible to define a set of shared goals for Open Science, the technical and societal components which are necessary to make science open and of global, public benefit. An important part of the UNESCO Recommendation on Open Science should be the development of a set of shared, global Open Science Goals.**

Ultimately, the importance of a UNESCO recommendation is to develop and obtain global consensus that to address the major challenges of our age, we need shared agreement to adopt the shared practices of Open Science. Achieving this consensus and UNESCO's role in this is important to enable equitable access to and participation in the benefits of Open Science and to help ensure that no one is left behind. The consensus is urgent, because—viewed as the necessary transformation of scientific practice to adapt to the changes, challenges and opportunities of the 21st century digital era—Open Science provides the tools to help identify solutions to the major challenges of our age, including the COVID-19 pandemic and the effective realisation, monitoring and evaluation of the SDGs.

Specifically in relation to the SDGs, we argue that:

- the SDGs can't be achieved without the input of Open Science;
- the SDGs can't be effectively monitored without the processes, methods and tools of open science (most notably, FAIR data and the large scale analysis and modelling it enables);
- Open Science, and the skills that support it, should be included in SDG4 on education and elsewhere as appropriate.

Historically, there are a number of important examples of global consensus on the need to gather data for global monitoring and scientific research in relation to common human concerns. The network of data centres that emerged as a result of the International Geophysical Year of 1958 (part of the forerunners to the ISC World Data System) is a notable example: this is one of the many data endeavours that prefigure Open Science. Ranging across 17 goals and 231 unique indicators, touching every aspect of human and planetary concern, the SDGs require a data collection and scientific endeavour of a different order again. In sum, they require Open Science.

Monitoring the SDGs is an enormous challenge in itself. Addressing them requires the mobilisation of participatory Open Science and mechanisms to turn its findings into informed decisions and action. The SDGs require us to be able to apprehend enormous complexity, in trends, patterns and processes, in the context of a continuum of urgencies (from basic needs to issues of global interdependence) and at manifold scales (local to regional to national to global). Comprising initiatives for vast data gathering and the FAIR principles that enable analysis at scale, Open Science is necessary for us to apprehend and understand this complexity. Comprising transdisciplinary research, participatory and transparent science that is engaged with stakeholders, Open Science is necessary to provide solutions and help narrow the science-policy interface.

10. What are the obstacles to reaching global consensus on Open Science and how can they be addressed?

There are a number of obstacles to reaching a consensus on Open Science and to implementing the changes necessary to support and promote the new practices in a global and inclusive way. A number of these have already been discussed above in sections on . The most obvious include:

- the resistance, particularly in relation to Open Access of the paywall publishers, concerned to retain their profit margins;
- the caution of parts of the scientific community to endorse changes in the familiar structure of recognition and reward;
- concerns whether the necessary limits of Open, and the protection of sensitive data, will be adequately respected;
- national rivalries and a narrow interpretation of self-interest which emphasises scientific competition rather than the need for cooperation on major global issues;
- understandable concerns about the potential for 'data extractivism' or 'digital feudalism' if Open Science is not accompanied by equitable collaboration and rules of good practice;
- concerns about the unequal distribution of the technologies, skills and practices of Open Science and the costs of implementing these less well resourced countries, science systems and research performing organisations.

As we argue above, all these obstacles have mitigations. The most important point we would like to communicate and emphasise here is that **the global science community and its stakeholders have a profound shared interest in transitioning fully to the practices and ethos of Open Science.**

On the basis of consensus and the recognition of shared interest, there is an opportunity by means of the UNESCO Recommendation to chart out a roadmap and a set of shared, global Open Science Goals.

The authors of this submission, and the Data Together organisations, would be happy to contribute towards a consultation on such a roadmap. The outlines can be discerned from the recommendations presented through this document and gathered in Section 2. As a minimum they should include steps in the following areas:

Shared principles. Steps to build consensus and convergence around core principles of Open Science, including the recognition of Open Science as a global public good; the need for transparent, scrutinizable and verifiable use of data, evidence and methods; the FAIR, TRUST and CARE principles; and other necessary ethical protections and good practice.

Policy instruments. Encouraging the development and implementation of policy instruments (where possible at institutional, national, regional and global levels) that promote the good practices of Open Science, maximise access to all the outputs of publicly funded research, while affording necessary protections.

Open Science and FAIR frameworks. Open Science and FAIR frameworks are needed that define community practices for data sharing (including ethics and Intelligent Openness), data formats, metadata standards, tools and infrastructure. The UNESCO recommendation should encourage research communities to agree on Open Science frameworks. The recommendation should also encourage governments, funders and other stakeholders to support this process. UNESCO should support and use its convening power to assist such processes that are underway involving the ISC, scientific unions and national or regional research infrastructures.

Investments in the necessary enabling infrastructures. With particular attention to the components of the Open Science ecosystem described in Section 6, and to ensuring maximum global access to this ecosystem. This should include developing and advancing initiatives, such as the African Open Science Platform, that will ensure scientists in LMICs have access to infrastructure and technologies for data collection, stewardship and analysis.

Investments in the necessary enabling skills and education. Including the formalisation and professionalisation of data science and data stewardship skill sets as well training in the ethical and society-oriented approach that is integral to Open Science.

Major steps to review, reform and renew the processes for scientific communication. As argued above, the principle unit of scientific communication (the article) is outdated, the level of access to publicly funded scientific research insufficient, the extraction of public funds by traditional publishers excessive and unacceptable, the systems of recognition and reward based so heavily on the journal impact factor not fit-for-purpose. There is a need for a profound review of the mechanisms of scientific communication that are needed for Open Science and for the global public good. We need to find new ways to measure and reward scientific impact, and to find sustainable alternative models on which to build the publishing enterprise. With its members, programmes and the broader scientific community, the International Science Council is initiating such a debate. UNESCO can add its own convening power to encourage a global process for real change.

11. Open Science and COVID-19

We offer here some brief observations on the implications of the COVID-19 pandemic for Open Science. We intend to follow up on this submission with a summary of the way in which the COVID-19 pandemic has further demonstrated the need for Open Science and FAIR data. We recommend that there should be a concerted process to examine the scientific response to COVID-19 from the perspective of Open Science.

The world is currently experiencing an unprecedented worldwide epidemic caused by a new coronavirus. The WHO has recently declared that this virus will be endemic for the foreseeable future. Globally, the race is on to develop common sense strategies, repurposed interventions and—later in the roadmap—vaccines, novel therapeutics, and improved diagnostics to halt the COVID-19 pandemic, which threatens to infect a significant portion of the planet, and potentially will kill millions of the world's most vulnerable populations. Such interventions need to be based on the best and most timely and complete data possible, and Open Science will play a crucial role. However, the expert community faces dual challenges: on the one hand, a lack of access to relevant and protected real world observations (RWO); and on the other hand, (published) information overload and misinformation.

Under the pressure of the first wave of COVID-19 patients, which overwhelmed the standard health systems in many countries, many initiatives have been created at unprecedented speed.[29] Many articles have been published in pre-print servers, without, as of yet, proper peer review, but even in high-profile journals like the Lancet and the NEJM, papers have been published under high pressure that have rapidly had to be retracted due to problems with the data on which they were based.[30] Meanwhile, the selected CORD-19 Open Research Dataset[31] currently contains more than 60,000 articles deemed directly relevant for COVID-19. It is therefore beyond any researcher, institute or country, to fully comprehend, let alone judge the validity of even a subset of the available information on COVID. **The size and scale of the COVID-19 challenge, the urgency of effective science to inform evidence based decision-making, underline the need for Open Science and its components (including FAIR data, with appropriate, responsible and proportionate access control, the capacity to deploy large scale analysis and models, rich semantics and description, to maximise the development of conceptual linking and the effectiveness of Machine Learning).** The GO FAIR VODAN initiative provides an important example of a community exploring these approaches in the context of the COVID-19 pandemic.[32]

It became clear during the first wave of the COVID-19 pandemic that the disease progression is highly variable in different people. For some patients an infection results in minor symptoms, while others suffer from severe to critical symptoms that require hospitalisation and intensive care. It is increasingly recognised that the severe disease is in fact a complex, multifactorial interplay of virus induced aberrant immune responses, endothelial and vascular problems, as well as thrombotic events. The manifestations differ widely based on age, prior conditions and localisation of the major viral concentrations, leading to multiple forms of organ failure (lungs, heart, brain, kidney, etc.). Thus, multiple disease models will be needed to understand the phase and trajectory in which a particular patient is, and to rationalise personalised interventions.

Among all the measures taken to prevent the worst-case scenario of the COVID-19 pandemic, data and verifiable evidence should be at the core of our decisions. However, evidence-based interventions and its translation into policies in the case of COVID-19 so far was erratic at best and too slow. We could have discerned the pattern of vascular and thrombotic problems underlying much of the severe pathology much earlier if we had been able to access the RWO from the early outbreak locations on a near-real time basis. Because of the political battlefields around the national responses to the virus and the role of WHO, the situation in the case of COVID-19 is further complicated and fraught. This underlines the pressing need for ‘safe and custodian controlled data visiting’ and near real time community annotation by experts. Sticking to the old publication systems, and to old fashioned and non-interoperable data collection, effectively keeps objective information and the underlying data from researchers, health workers and the broader public, nationally and from other nations. This severely frustrates their ability to help mount adequate responses. As a result, we see deeply concerning panic reactions, widely variable and sometimes contradicting policy

[29] A Google scholar search lists over 125,000 results for COVID-19 since the start of the outbreak.
[30] Lancet: Retraction—Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis, https://doi.org/10.1016/S0140-6736(20)31324-6; NEJM: Retraction: Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. N Engl J Med. DOI: 10.1056/NEJMoa2007621
[31] https://www.semanticscholar.org/cord19
[32] VODAN (Virus Outbreak Data Network): https://www.nature.com/articles/s41431-020-0635-7

interventions, not always based on solid evidence, and sometimes more likely to aggravate the situation than to contribute to a solution.

The epidemic has also painfully exposed the potential risks of overly rapid publication and the lack of adequate peer review before scientific ‘findings’ are propelled into mainstream knowledge and discourse. We would argue, however, that the problems of hasty publication are not necessarily and irrevocably associated with Open Science. Indeed, Open Science practices contain a number of mitigations to the risks associated with hasty publication. **The most important lesson of the papers retracted from the Lancet and NEJM is the paucity of comprehensive, quality controlled data which resulted in the use of unverified and flawed data obtained from a private company. Open Science is not antithetical to thorough peer-review and insists on maximal appropriate access to the data underpinning publications.** What is more, open peer review through extensive pre-print discussions (for example arXiv or bioRxiv) or more controlled open peer review (Copernicus, F1000) are a feature of Open Science initiatives. For circumstances where there may be more sensitivity, Open Science platforms are exploring controlled, but large scale, expert peer-review (Frontiers). **Open Science requires thorough examination of innovative publication models, which make peer review as transparent and rigorous as possible.** This can include reimagining the purpose of peer review; deploying maximum community-wide expertise; understanding when caution may be necessary in relation to the release of preliminary findings; considering how data and code can also be included in the review process, when this is practicable and necessary and when it can be left to wider debate. All this requires new processes, reward systems and education for Open Science and its new way of working. The classical ‘article’ will not suffice any longer as the main carrier and unit of scholarly communication. In many fields—including those most urgent for the future of humanity and the planet, including the COVID-19 pandemic and the SDGs—Open Science will introduce a ‘continuum’ of gradually improving knowledge and ‘living reviews’ as well as machine readable models and knowledge representations, as opposed to carefully drafted and repeatedly peer reviewed articles that take so long to publish that they are outdated by the time they are finally ‘accepted and published’.

Appendix 1: the Data Together Organizations

CODATA's mission is to connect data and people to advance science and improve our world. As the 'Committee on Data of the International Science Council (ISC)', CODATA supports the ISC's mission of 'advancing science as a global public good' by promoting Open Science and FAIR data. CODATA convenes a global expert community and provides a forum for international consensus building and agreements around a range of data science and data policy issues, from the fundamental physical constants to cross-domain data specifications. CODATA's membership includes national data committees, scientific academies, International Scientific Unions and other organisations.

Global Open (GO) FAIR is a global initiative of community-driven implementation networks that enable data-related agreements, data services and protocols that support FAIR (Findable, Accessible, Interoperable and Reusable) sharing and reuse of data and related tools in an open manner. Established in 2017, GO FAIR draws its membership from domain-specific communities of practice that are invested in the development of best practices for data management fostering the coherent development of the global Internet of FAIR Data and Services.

The **Research Data Alliance** is a global, consensus-based, community-driven organisation of over 10,000 individual and institutional members from 144 countries whose mission is to provide a platform to drive innovation surrounding open data sharing and interoperability. RDA enables data to be shared across geographical, technological and disciplinary boundaries through outputs developed by focused Working Groups and Interest Groups of volunteer experts from around the world and drawn from academia, the private sector and government. Established in 2013, RDA draws its membership from individuals and organizations across the data management ecosystem.

The **World Data System** is a research programme of the ISC that enables data repositories to use infrastructure and protocols that ensure long term custody of data, and to define services and resources that achieve greater interoperability between repositories and applications that consume data. Established in 2008, WDS draws its membership from data repositories and service providers who implement agreements that strengthen these organisations, and promote best practices, including the open data sharing principles.

Appendix 2: Members of the Expert Group

This document has been prepared by an expert group, coordinated by CODATA, and including representatives from the ISC World Data System (WDS), GO FAIR and ICSTI, the International Council for Scientific and Technical Information.

The Expert Group comprised:

- Paul Arthur **Berkman**, Founding Director, Science Diplomacy Center, Professor of Practice in Science Diplomacy, Fletcher School of Law and Diplomacy, Tufts University; Associated Fellow, United Nations Institute for Training and Research (UNITAR); Member of the CODATA Data Policy Committee.
- Jan **Brase**, head of the Research and Development at the State and University Library, Göttingen; President of ICSTI.
- Richard **Hartshorn**, Professor, School of Physical and Chemical Sciences, University of Canterbury; Secretary General of the International Union of Pure and Applied Chemistry; member of the CODATA Executive Committee.
- Simon **Hodson**, Executive Director, CODATA.
- Wim **Hugo**, Director of Strategy, ISC World Data System.
- Sabina **Leonelli**, Turing Fellow and Professor in Philosophy and History of Science, University of Exeter.
- Barend **Mons**, Professor in Biosemantics, Leiden University Medical Center; President of CODATA; Director GO FAIR International Support and Coordination Office.
- Hana **Pergl**, Operations Manager, CODATA.
- Hans **Pfeiffenberger**, Consultant, scientific data infrastructures and policies, Member of the CODATA Data Policy Committee, founder and member of the advisory board of Earth Systems Science Data.