CHALLENGE 6

ABSTRACT

Open Science is becoming a new paradigm in scientific research and complex changes are being done. This new way in knowledge development requires a great transformation that will allow science to adapt efficiently and effectively to the urgency of the problems to be solved while ensuring the reproducibility, transparency and reliability of scientific results. This chapter analyzes the impact of this change of model, the challenges to be addressed and the expected benefits.

KEYWORDS

open science | open knowledge | reproducibility | transparency | FAIR | big data | scholarly communication | outreach | scientific assessment

CHALLENGE 6

OPEN SCIENCE: REPRODUCIBILITY, TRANSPARENCY AND RELIABILITY

Coordinators
Agnès Ponsati
(URICI, CSIC, Coordinator)
Fernando Aguilar
(IFCA, CSIC-UC, ADJUNT Coordinator)

Participant researchers and centers
Reme Melero (IATA, CSIC)
Lourdes Verdes-Montenegro (IAA, CSIC)
Javier Moldón (IAA, CSIC)
Rafael Garrido (IAA, CSIC)
Susana Sánchez (IAA, CSIC)
Julián Garrido (IAA, CSIC)
Gemma Rius (IMB-CNM, CSIC)
Álvaro López (IFCA, CSIC – UC)
Pablo Orviz (IFCA, CSIC – UC)
Ismael Ràfols (INGENI, CSIC – UPV)
Isabel Bernal (URICI, CSIC)
Adrian Garcia (IFISC, CISC - UIB)

1. EXECUTIVE SUMMARY

Scientific research is a global, collective enterprise, today more than ever collaborative. Its main objective is the production of knowledge that allows society to advance in diverse areas with a strong social, economic, cultural and environmental impact. It is an activity that demands important resources and their long-term sustainability. A very high percentage of it is financed with public money, so society expects in return positive outcomes impacting on health and quality of citizens' life. The current model of scientific research is facing the unquestionable challenge of a great transformation that will allow it to adapt efficiently and effectively to the urgency of the problems it must solve while ensuring the reproducibility, transparency and reliability of scientific results. The CSIC is called upon to move towards this new paradigm, in which it must change the way science is done and, above all, how its results are managed, communicated, used and reused. It should seek ways for a more open, transparent, interdisciplinary, collaborative, sustainable global system of scholarly communication. This chapter analyzes the institutional potential impact of this change of model, the challenges to be addressed and the expected benefits.

2. INTRODUCTION AND GENERAL DESCRIPTION

Research institutions are under a hard pressure of scientific, health and socio-economic challenges. It is time to transform individual research efforts into a more collective effort. Sustainable and innovative solutions must be supported on time by an efficient, transparent and dynamic scientific endeavor, not only from the scientific community, but from other key stakeholders and the entire society. Open Science embodies the need to transform, open and democratize the entire knowledge generation to ensure that every scientific challenge is faced and really drives and allows the achievement of the United Nations Sustainable Development Goals [United Nations, 2020].

Driven by unprecedented advances in our digital world, the transition to Open Science allows scientific information, data and results to be more accessible (Open Access) and be more reliable (Open Data/Open Software) with the active participation of all relevant stakeholders (Open to society/Citizen science). However, in the fragmented scientific and political environment, there is not yet a global understanding of the meaning, opportunities and challenges of Open Science.

Furthermore, academia and scientific institutions need to regain and keep control of its research outputs (publications, data, methods, software, etc.). The control over all scientific results largely due to research funded with public money must be regained because the largest part of the scientific results are under the control of private commercial publishers and access to scholarly outputs is behind paywalls. Open Science is not only about Open Access advocacy and open and foster sharing and reuse of data and software. We must take another step forward and open the validation, use and reproducibility of processes, code, methods and protocols to open the doors. Change the way we do science, following the scientific method (a process established in the 17th century) in the digital era. Open Science is rooted in the tradition of established principles of good scientific practice, but the goal now is to critically reflect traditional scientific culture and to transfer it into the present era of linked-up Open Research.

Digital technologies today allow the setup of working environments for the development of research activity that drives a more trustable, shared, participatory, transparent, optimized and efficient science to meet the aforementioned challenges. In the era of the Big Data infrastructures, the use of artificial intelligence, deep learning technology applied to the analysis of big or long tail data will bring capabilities and results unimaginable a few years ago. But for this to happen, it is necessary a real time exchange of data, which must be organized, properly curated using open and interoperable standards for machines.

Today, because of the huge data production of all kinds of scientific results, to guarantee quality and scientific veracity is paramount. The only way to face this issue and successfully achieve it is through openness, transparency, accessibility and interoperability to reproduce the whole scientific process. The different steps along the research life cycle must be linked and their relationships identified not only to create data, but also to transform into information and, eventually knowledge, technology advance and human progress.

Reproducibility, a cornerstone of the scientific method, is still a challenge. In the era of computational research, reproducing an experiment can be impossible due to the lack of access to data, configuration parameters, software environment, analysis tools, as well as to annotations and provenance information describing all those elements. Therefore, identifying these components is needed, which means that the research life cycle needs to improve its transparency. The reproducibility problem is currently increasing due to the information and data deluge generated by worldwide scientific infrastructures. Solutions are starting to emerge, involving e-science technologies to enhance scientific collaboration and knowledge sharing, ensuring transparency, opening data and methods, and encouraging Open Science methods.

Now more than ever before, due to the planetary challenges we face it is time to foster Open Science practices at CSIC. The institution will thus demonstrate its social accountability, leverage its role in the advancement of worldwide science in the search for solutions to the great problems that today's society is experiencing, as well as the unknown ones that it will have to face in the future to come. This chapter is about defining why, what and how CSIC should adapt to the new paradigm of Open Science already marked by all science agendas worldwide [Netherlands, 2017] [CNRS, 2019], how as one of the most relevant European scientific institutions plans to participate in building a global Open Research infrastructure that improves and adds value to the research workflow by promoting collaboration, sharing of scientific assets and knowledge, avoiding fragmentation and duplication. This is the frame of the European Open Science Cloud (EOSC), a new environment to perform Science and to promote the collaboration among disciplines and researchers in Europe.

The journey through this global cultural change, that means understanding science as a public good, will need efforts and investment in providing capacities (infrastructures), arranging the needed resources, both financial and human, endowed with the necessary qualities and capabilities. Last but not least, the adoption of a new framework for research evaluation and incentive systems that recognizes researchers' commitment in line with the Open Science values and fundamentals of reproducible science.

3. IMPACT IN BASIC SCIENCE PANORAMA AND POTENTIAL APPLICATIONS

The expected impact of developing Open Science practices is global and spans across all disciplines within the institution. Benefits of implementing openness along CSIC's community outweigh the potential costs the institution will need to face as takeoff. Among them we can highlight the following classified within main categories:

3.1. Institutional-Science-Society benefits

Promoting Open Science Publishing Publishing in Open Access journals, open publishing platforms or via institutional repositories means that scientific results will be accessible to a wider audience who can read without paywalls, accelerating information and knowledge transfer to society.

Science and social responsibility CSIC's image as a science generator and its social responsibility and commitment will be reinforced for its collaboration in the democratization of the scientific knowledge produced, by improving access to scientific resources; by enabling the participation of a wider community in the research process; and by making science more comprehensible to the society at large as well as more transparent.

Scientific efficiency Among benefits expected as a result of the implementation of an Open Research strategy, and specifically those that have to do with scientific efficiency, it is worth mentioning to ensure accountability: Open Access to scientific research outputs provides greater institutional accountability but also to research funders and taxpayers. Wider availability of knowledge resources will make high quality research faster, cheaper and research success more likely. More fluent collaboration among heterogeneous, multidisciplinary knowledge actors will amplify CSIC's collective intelligence and creativity.

Open Access infrastructures will also increase efficiency because it enables the use of common capacities and facilities that will interconnect knowledge production, reusing online available data to arrive at new findings (data driven intelligence). CSIC research community will be able to interact more fluently, collective intelligence will be amplified by the mere fact of being able to share, validate and quickly rule out different ideas, assumptions, hypotheses or new avenues of inquiry. Accelerating knowledge transfer and reducing delays in the sharing and reuse of CSIC's scientific outputs.

Foster a more open and collaborative scholarly environment @CSIC With the promotion of an Open Science strategy the institution encourages collaboration among all data dependent disciplines, provides infrastructures that facilitates data sharing, analysis to be taken to the next level in order to further scrutinize the data avoiding duplication. The role of the software will be particularly emphasized within the proposed strategy as a key enabler for the production, analysis and visualization of the research data in a scientific study.

3.2. Methodological benefit for the researcher career

Be ready for Open Science Funding CSIC will meet the actual and future requirements established by funding agencies worldwide that promote the Open Science realisation. Thus, CSIC researchers will be better positioned for the upcoming challenges formulated at the international level.

Increase visibility By boosting open practices, CSIC's Open Science strategy practitioners will have the opportunity to get more visibility for their research outputs, through seamless access and reuse (citations), and impact beyond classical bibliometric, media attention, potential collaborators, job opportunities and more recognition.

Tackle the Research Reproducibility problem If researchers release their data, software, and materials into the appropriate repositories with the appropriate metadata curation, then this increases the accessibility, and reproducibility of the findings by other researchers. Ultimately, when researchers share their scientific outputs, they are exposed to further analysis by the scientific community, and thus this fact contributes to increase confidence in their own research. At the same time CSIC research results are preserved and accessible for the future.

Advancement of collaboration Success in research requires collaboration, but, often, it is difficult to identify and connect with the appropriate collaborators. Practicing Open Science within or beyond CSIC will make it easier for researchers to connect with one another because the discoverability and visibility of their work will increase. This in turn will facilitate share, reuse and easy access to novel data and software resources. Researchers will also be in a privileged position with respect to evaluation of their careers in terms of altmetrics.

3.3. Sample of CSIC's use cases

Open Science involves every single discipline represented within CSIC, so this new paradigm emphasizes the intrinsic multi- and inter-disciplinarity of the institution, enabling existing and new ways of collaboration. There are a number of potential new applications thanks to Open Science, both at general and at discipline level, ranging from applications in the fields of health and biotechnology, climate change, management of natural resources and ecosystems, astrophysics, demographic crisis or management of public policies… to mention some. During the last times, this interdisciplinary character of CSIC has been reflected in the creation of the Thematic Interdisciplinary Platforms, which aims to create collaborative groups with CSIC research groups to address complex scientific goals.

Reproducibility review "service" for bio sanitary crisis, the immediate case of COVID-19: During the worldwide challenge to face the bio sanitary crisis produced by the COVID-19 pandemic, an unprecedented volume of scientific information has been released to address the problem from different perspectives: clinical, biomedical, societal, etc. This means that researchers from diverse areas have started to collaborate supported by digital technologies and adopting an Open Science vision. This unexpected problem requires sharing of diverse research results as well as services. Sharing of computational resources and expertise in data science is already happening in CSIC in order to face the COVID-19 pandemic in an interdisciplinary way. Complementary to this, more than ever reproducibility and opening of the scientific methods is key to accelerate research, so that:

1. The results can be more rapidly compared
2. Reuse of methods by different teams in order to analyze different datasets/samples is facilitated
3. Reliability of the results and the exchange of knowledge is reinforced

If papers on COVID-19 can be reproduced, it will be possible to access the code that has been used, identify possible errors or improvements in the data sets, reuse them for new data that becomes available, etc. In practice this would translate in supporting end-to-end reproducibility through a research workflow. Key steps include creation of software environment, access to raw data, steps taken to process them, access to the data, simulations or models, as well as scripts (e.g. in the form of Jupyter notebooks) underlying the results, including making tables and figures in the paper reproducible (data, images, fits and their error bars, simulation models, etc.). Advantages include: faster implementation of peer review (re-running modified notebooks re-generate figures in "a click", cloud access to any other team, validation, re-use, re-purpose, etc.).

Nowadays it is very difficult for the researchers/scientists to do this without support of data stewardships or software engineers, and the investment of time required is often considered to be "not worth it." In the rush to find a solution to the COVID-19 crisis, teams need more than ever support to ensure that their digital research can be easily accessible, shared and reused. Several teams at CSIC have expertise in Open Science and reproducibility applied to different disciplines. In particular, CSIC has released the Global Health thematic interdisciplinary platform, started as a framework to establish relationships among CSIC research groups and also with external entities to develop vaccines and medicines, create new kinds of tests or to propose measurements of social distancing among others. Within this platform, some groups are oriented to facilitate the data management and sharing as well as provide computational resources. This is a clear example of how Science can be more efficient adopting an Open philosophy.

Biodiversity: sharing data, publications or any kind of development will enable the possibility to improve the information we have about not only the distribution of certain species but also how they interact among them. Furthermore, due to the transverse character of Open Science, the interactions between the species and any other process on Earth will be potentially understood thanks to the integration of data sources and knowledge from different disciplines. CSIC coordinates the Spanish node of GBIF (Global Biodiversity Information Facility) and it is actively involved in a few Biodiversity-related ESFRIs like LifeWatch ERIC or DiSSCo.

Earth and Marine Sciences: building a new and innovative infrastructure for mapping seafloor features that may represent potential geological hazard. The initiative will allow defragmenting the geoscience community working at sea through the incorporation and integration of collections, data-archives, services that nowadays remains dispersed through the scientific groups, institutes and services. In fact, today there is a huge volume of data that were initially recovered for specific scientific objectives and that now remain dispersed throughout and stored in research institutes, universities, and exploration and survey companies. These data should be findable, reusable and accessible and thus to ensure that they keep having great value of scientific continuity and also for other end users (industry, public administration). This infrastructure should also ensure their sustainability and offer new methods based on machine learning and artificial intelligence to increase accurate mapping of marine geohazard features, and to build a link with risk assessment; also, it would be transferable to other disciplines.

Astronomy and Astrophysics: Astronomy has a long tradition in Open Science, being a reference for the establishment of standards: the International Virtual Observatory Alliance (IVOA) was created in 2002 to provide standards for data and software interoperability, as well as Open Access to data in Astronomy, with a special emphasis on optical Astronomy. CSIC is leading the Spanish participation in the Square Kilometre Array (SKA), which will constitute the largest scientific infrastructure on Earth and the largest generator of open data.

This brings the challenge of providing access to SKA (Big) data products, tools and processing power to international distributed teams. Creation of new standards for radio astronomy is crucial to enable bringing the new facilities in construction into the Virtual Observatory ecosystem, facilitating multimessenger science to a wider community. The construction of the new mega science projects and infrastructures are moving the field into a new era in which large international alliances of scientists will be required to analyse the data deluge. We are hence in a race to exploit ever larger datasets, and in our quest for "efficiency" we risk forgetting about reproducibility. In this challenging new era in which Big Data will be analysed by teams largely distributed along the globe, only if we are ready to change the way in which we work, we will be able to increase the degree of reproducibility of our research, and consequently improve the quality of Science. A new kind of infrastructure and platform to share knowledge, big/complex data, and methods, will be required to follow the FAIR principles [Wilkinson, 2016] (Findable, Accessible, Interoperable, Reusable) and achieve reproducibility.

4. KEY CHALLENGING POINTS

4.1. Ensure reproducibility: standards and interoperability building capabilities

Ensuring full openness and reproducibility will take time and major efforts to operate with interoperable and reproducible data. A questionnaire on reproducibility published in Nature (2016 [Baker and Penny, 2016]) found that 70% of researchers cannot reproduce another's experiment while > 50% their own ones. Furthermore, the academic paper represents just a small part of the research if it doesn't come together with input data, configuration parameters, software environment and analysis tools, as well as annotations and provenance information describing all those elements. The actual standard formats for publications come with a loss of at least 40% of the knowledge [Bechhofe et al., 2010].

Open and reproducible science often consumes time, mostly in the early phase of implementation of a new system. Archiving, documenting, and quality controlling of source code and data takes time, but the resultant benefits are manifold for the creators, their institutions and the broader user community. Just as the quality practices in the software development life-cycle shall be enforced from the very early stages, curation activities must start as close to data creation as possible in order to ensure reproducibility and replicability. Different types of additional practices will be needed, for instance: enough time to learn and implement emerging tools, methods and standards; time and incentives for researchers to change habits and an experienced dedicated staff should be engaged in the process of data management and curation according to such standards and best practices. Further, CSIC multidisciplinary poses an extra challenge in this regard, as the adoption of Open Science tools, standards and good practices should meet specific discipline requirements and foster at the same time high degrees of interoperability across all institutional infrastructures and associated contents. DIGITAL.CSIC compliance with several standard metadata schemes and various data management best practices move towards greater interoperability within the institution and in the global Open Science ecosystem.

Because the cost of not FAIR data for research is high, strict adherence to FAIR principles to ensure sustainability (research data and software) alike data preservation are other major challenges. Applying FAIR criteria will guarantee discovery, access and reuse of research outputs by humans and machines, while long term sustainability and preservation will depend on the adoption of Core and Trust Seal principles [CoreTrustSeal, 2020] by CSIC repositories. Also a robust institutional preservation policy is needed behind, such as repositories, with a clear mission, budget, priorities and strategy and based on the OAIS preservation model. This institutional infrastructure should cover the maximum level of compliance of NSDA recommendations. Ensuring digital preservation will contribute to eliminating science's blackout tomorrow.

FAIR principles do not just apply to data, but also to other research objects. In particular, research software sustainability shall be given a high level of importance and visibility in order to develop Open Science, since it is an undeniable key piece. It will be necessary to establish comprehensive quality criteria for software production and delivery, including documentation, code management and accessibility, as well as validation and verification processes. The challenge is not only to manage new research outputs from an Open Science perspective but also to fully identify, document and safeguard software heritage produced by CSIC in past years. Documenting, managing and making workflows accessible will also be a requirement to enable computational reproducibility.

Good data management is critical for ensuring validation, transparency of research findings, as well as to maximize impact and value of publicly-funded research through data reuse. CSIC repositories and other Open Science infrastructures would need to provide crucial services and tools that manage and enable full description, date and provenance registration, access, versioning of data and software, publications, software quality assessment and a wide array of other types of scholarly content, as well as value added tools for the community to ensure the good management of the research results.

CSIC must contemplate a strong investment in training and development of knowledge and skills. Groups exist in CSIC that are a reference at international level in these aspects applied to their domain, but this should be expanded, and not only must the scientific community be trained, but also the technical staff, and that will need a major skills upgrade in order to assist it in adopting practices and guidelines for the development of an Open Science environment ("Metadata curators", "Knowledge Commons Specialists" "Research Software Engineer", "User support staff", "Data scientists").

4.2. Providing capacities by implementing Open Research infrastructures

CSIC would need to be able to deploy a set of reliable and open infrastructures (European Open Science Cloud driven) that covers the needs of a multidisciplinary institution with a huge diversity of requirements. To cover all scientific and technical communities, assuring none of them is left behind. The infrastructures must cover the complete research workflow, tailored for big and long-tail data environments. These infrastructures shall provide an open environment with sufficient computing capacity for managing the research process, including added-value services and tools that allow the scientific collaboration, FAIR data and software quality assessment, setting the path for the reusability and exploitation of the scientific results and digital research objects (Text Data Mining, Deep and Machine Learning, Extracting, Visualization, etc.) The challenge of setting up a distributed, powerful, granular and scalable infrastructure must also be considered.

Best practices akin to FAIR principles can be adopted by Open Research infrastructures aiming at properly providing the required functionalities. Some of the resources already available for the researchers are not always well-known or they cannot easily be found (Findability). Furthermore, the set of services provided need to be seen as a common daily basis tool, just like a teleconference room or an E-mail client. In that case, the learning curve needs to be reduced and some training needs to be provided, making the use of research infrastructures and e-infrastructures more accessible. Some of the required components to compose the Open Research infrastructure environment are already available. For example, repositories providing Open Publications/Data and metadata, computing infrastructures enabling technologies for data analytic or data preservation services storing long term data. The future infrastructures will not need to reinvent the wheel but adopt an interoperable practice of the already available facilities. This can be done thanks to the federation of resources, implementing machine-actionable features to allow services to interact among them or promoting the sharing of research products, not only papers but open-source software or data.

One of the main critical challenges to be addressed by new research infrastructures is the capability to support the growing volume of scientific data being gathered. The high data rate and volume as well as the complexity of data produced by disciplines like those above mentioned (see Sample of CSIC's use cases) will necessarily transform how scientists access, share and analyse information. Furthermore, to address complex and global challenges and problems to be faced, an interdisciplinary way to do science needs to be implemented. The future research infrastructure needs to be able to manage Big Data from different perspectives (Volume, Variety, Velocity, Veracity, and Value [Ishwarappa and Anuradha, 2015]) in a FAIR manner and it needs to facilitate the data sharing and collaboration among researchers.

4.3. Open Science as the norm in research practice: meet the agendas of the funding agencies

Managing an Open Science environment in the CSIC will involve facing a change in the scientific culture of the organization and its members. Conceive how to generate science and produce knowledge based on the values of Open Science. Understand knowledge as a common good, defend a cognitive justice that promotes plurality and diversity, be inclusive and foster equity to the fullest. Respect the right to make research, fulfill collaboration in terms of equality between partners and advocate for sustainable and inclusive infrastructures for all. It is relevant to take into account that the idea of Open Science started as a bottom-up initiative, directly from the community (Declaration on Research Assessment (DORA), Metric Tide, Leiden Manifesto, Altmetrics, etc.) and just recently became top-down with initiatives like "cOALition S" or through the implementation of different working groups at the European Commission.

Open Science is one of the European framework pillars, and will become the modus operandi of Horizon Europe [Horizon Europe, 2020], requiring Open Access to publications and data. CSIC, as a research institution that manages numerous national and international projects and especially those under the umbrella of the European Commission, has to harmonize the protocols necessary for researchers to comply with the requirements of regulations and guidelines that affect open publication and data management. It must provide the tools and infrastructure that allow access and interoperability between the different systems of the ecology of scientific communication, and foster the implementation of monitoring processes to measure compliance, compatible with similar procedures in other countries, regarding variables and processes for its calculation.

A challenge that CSIC must face in the near future will be how to adapt our infrastructure and protocols to comply in the next future with the Directive (EU) 2019/1024 of the European Parliament and of the Council on Open Data and the re-use of public sector information [European Parliament and Council, 2019]. The Directive extends its coverage to research data resulting from scientific research activities subsidized by public funding or co-funded by public and private-sector entities, according to the principle “as open as possible, as closed as necessary”. This will require that the ethics committees take into account all those cases in which the investigation deals with individuals. CSIC should provide tools, resources, and training for the good management especially of those personal and sensitive data to comply with the data protection regulations, as well as the implementation of protocols and software applicable to data anonymization.

On the other hand, Plan S [Plan S, 2020] endorsed by the European Commission and the ERC, and supported by more than 20 European research funders, considers both the golden route and the green route to Open Access scholarly outputs. To comply with the requirements for Open Access repositories (H2020, ERC, CSIC), technological and standard requirements must be implemented along with the mandatory criteria. In this regard, Confederation of Open Access Repositories [COAR, 2020] (COAR), of which CSIC is a member, is working on the recommendations to adapt the repositories to those requirements, based on among others in the FAIR principles (Findable, Accessible, Interoperable, Reusable). Regarding publications, the European Commission has recently awarded the development of a platform [European Commission, 2020] for open scientific publications that meets the requirements of its calls and that will be launched early in 2021.

4.4. Research Evaluation

Current CSIC indicators of performance (Strategic plan 2018-2021) are based on global figures as number of papers (I1, I2), books (I3), PhDs (I4), protected technologies (I5), created companies/spin-offs (I6), or economical value in terms of budget of signed research contracts (I7) or return of licensed technologies (I8). There is wide bibliography (including e.g. a Nature special on Open Science in 2010 [Braun, 2010]), stating that science is killed by numerical ranking. Open Science denotes new ways of generating and using scientific knowledge that facilitate collaborative modes of research. Such collaboration connects individuals from many different contexts thus dissolving the sharp institutional boundaries that had so far segmented the processes of knowledge generation and use. Technicians, scientists, and citizens can get involved, through large teams and informational platforms for sharing in a seam-less manner in complex processes of knowledge generation and use.

The consequences of this shift for research evaluation are nothing short of momentous. Current evaluation practices typically assume a strict division of tasks. Separate organisations are assumed to produce specific types of outputs (publications, patents, etc.), which are used in incentives schemes that revolve around the individual as the central evaluand. Given the focus in Open Science on the facilitation of knowledge flows across diverse stakeholders, research evaluation has to shift from the current focus on journal outputs and visibility, to a focus on communication and collaboration processes. This means that the traditional evaluation approaches that reward individuals based on performance, and assess such performance through the analysis of narrow sets of indicators (like publications and citations) are not fit for use in a context of Open Science.

Therefore, the challenge goes beyond the definition of new sets of "Open Science" indicators, but requires the design of a different research governance system. It needs to shift to forms of evaluation of labs, teams or institutes that value not only the knowledge but crucially also how this knowledge is made available and the collaboration with a variety of social stakeholders. Given the great diversity of communication and collaborative practices across disciplines, Open Science evaluation cannot rely on a fixed new set of indicators, which replaces the 'late 20th century' bibliometric indicators. Instead, the evaluation of Open Science requires contextualized evaluation methods in accordance with principles put forward in the DORA or the Leiden Manifesto. In some disciplines and contexts sharing 'software', well-curated 'data' or training local experts may be as or more important than publishing. Evaluation of Open Science needs the flexibility to capture the varying values of these activities in different research environments. This is why they cannot rely on universal indicators, as reported by the EC Expert Group on Open Science evaluation [Wouters et al., 2019].

While Open Science is a challenge for research evaluation it is also a great opportunity to put the value of scientific learning, collaboration and communication at the center of the evaluation effort – and to move away from top-down, one size fits all, evaluation practices that may have harmed research in the last two decades.

4.5. Sustainable scholarly communication system (Authors rights management, OA licenses and GDPR)

A system based on Open Science by design contemplates that all products of the research process will be available as soon as possible at no charge. This view conflicts with the current CSIC scientific communication system that relies on the *modus operandi* imposed by the third-party content subscription system. It is thus necessary to look for alternative models (e.g. Publish and Read agreements). All this, complemented with other options of promoting Open Access publication (Gold route), through repositories (Green route) or through an Open Access institutional publishing platform (Red way).

In the current scholarly publication system there is another doubtful component, the actual peer-review system. Its opacity can jeopardize the truth or scientific integrity, thus opening the door to scientific fraud, data manipulation, poor quality peer review and conflicts of interests. Alternative models supporting wider transparency and accountability have been put in practice over the last years by several publishers and repositories. Opening up identities of reviewers and/or associated reviews is contributing to create a new scientific validation environment that is more participatory, transparent and informative. However, major challenges and reticence remain within the scientific community. For instance, DIGITAL.CSIC survey [Bernal and Román-Molina, 2018] to CSIC researchers about peer review practices and opinion about the repository's Open Peer Review Module shed light about current lack of incentives to embrace new peer review practices, and other factors such as fear of retaliation and time consuming-ness.

There is an urgent need to establish an institutional system whereby CSIC centrally manages the intellectual property rights derived from the works produced by institutional researchers. According to national legislation, exploitation rights of intellectual outputs by employers in the public sector belong to their institutions however there is a high degree of confusion amongst researchers as to the proper management of such rights. An institutional copyright management that avoids the transfer of exploitation rights of CSIC intellectual works to third parties is a must in a system based on Open Science. By signing copyright transfer agreements and exclusive licenses to publishers use and reuse of CSIC outputs (even by CSIC itself, for instance when it comes to self-archiving into DIGITAL.CSIC) are highly compromised and this widespread practice contributes to the permanence of the current system of scholarly communication. On the contrary, a CSIC policy is needed so as to enable Open Access and reuse of its research outputs through the appropriate Open Access licenses ("as open as possible, as close as necessary"). This policy should include considerations for CSIC publications, research data and software as all these types of research outputs fall under copyright legislation.

There are more legal concerns around. CSIC must monitor the current landscape around the national transposition [FESABID] of the European Copyright Directive in order to make sure that it does not harm the interests of research, does not prevent certain types of Text and Data Mining actions or does not impose additional management workload on content sites. Last but not the least, the correct application of the new EU General Data Protection Regulation (GDPR) requires an extra effort of awareness raising, training and allocation of resources. From May 2018 collecting and processing personal data is regulated by the GDPR for all EU citizens. Personal data, and in particular sensitive data, are at the core of some research fields. However, there are research exemptions in the Regulation; it is crucial that CSIC includes GDPR considerations in the technical design and operational policies of its research repositories and infrastructures to avoid major legal breaches. At the same, a full understanding of GDPR amongst CSIC researchers and technical staff is needed considering the implications on the planning of data management plans and recommended data curation throughout the whole research cycle. In this regard, clear institutional guidance on best practices to safely store, process, anonymize and pseudononymize and share personal data will need to go accompanied with resources that enable the lawful management of this type of data while supporting Open Science.

4.6. Outreach and dissemination

Access to culture (including scientific culture) is a right enshrined in the Spanish Constitution [Fundación Acción Pro-Derechos Humanos] in which scientific centers are actively working. The number of initiatives to promote scientific culture is very wide in formats: from initiatives outside the centers such as the international “Pint of Science” festival in which scientists give outreach talks in bars, to science fairs promoted by universities, to television programs whose central theme is science. But just as new technologies have revolutionized all aspects of society, the distribution of scientific culture is no exception.

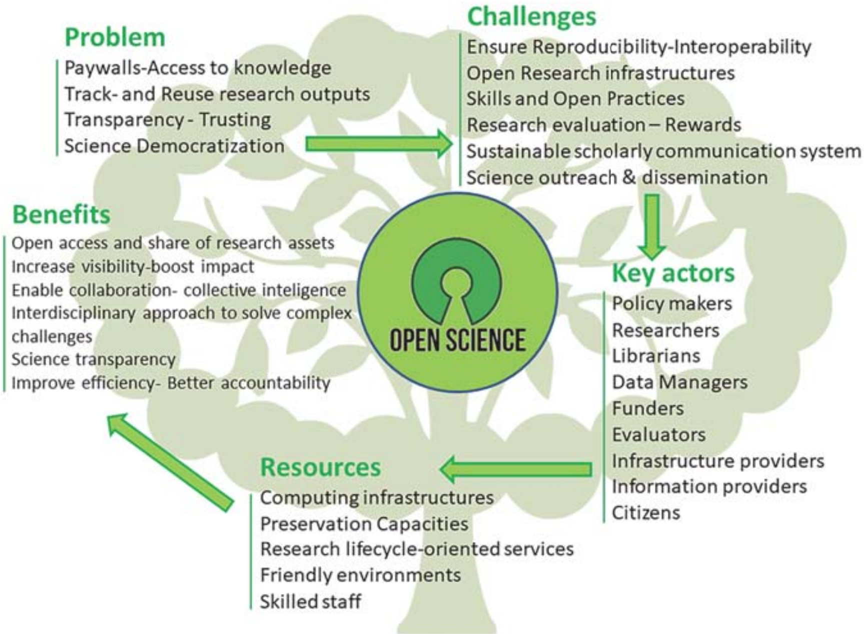
The digitalization of everyday life has led to a change in the consumption of information. The latest report on the social perception of science and technology in Spain [FECYT, 2018] by the Spanish Foundation for Science and Technology (FECyT) highlights the fact that 63.4 % of those surveyed report on science and technology issues through the Internet, the digital press, social networks and other websites. In the 15-44 age range, the preference for this medium is always above 70%. In this context, it is clear that an effective strategy to digitize scientific culture is necessary.

Since scientific communication is a field with so many possibilities, there are many digital formats in which to carry it out. However, there are common strategies among most CSIC centers, such as the management of their own self-managed dissemination websites to collect these materials, which are listed in the Dissemination section of the CSIC website [CSIC Divulgación]. On these platforms, digital material such as videos or blogs produced by the center’s own research and/or technical staff can be found. These materials not only have value in themselves, but are often used by the traditional media as support when they echo related research. Having the capacity to generate this type of audiovisual format has gained importance in the digital era.

However, in current times, the amount of information that digital communication channels are bombarded with, much of it fake or malicious, can make efforts to digitize scientific culture go to waste and become diluted in the midst of so much noise. It is key to design strategies that not only tell the story of science in a rigorous way, but also do so in an aesthetically attractive way. In this scenario, one of the biggest challenges in the field of scientific communication is to adapt to these new realities; this requires dedicated staff and training, as well as providing centers with sufficient funding to be able to subcontract professionals from the sector who are capable of generating and distributing digital content for scientific communication.

However, this does not mean that face-to-face dissemination activities can be neglected. Close and face-to-face scientific communication remains very important, especially for those groups that may be more distant from the digital world, such as the elderly or children.

ANNEX: ONE SLIDE SUMMARY FOR EXPERTS



ANNEX: ONE SLIDE SUMMARY FOR THE GENERAL PUBLIC