ROSiE

DELIVERABLE 6.1. – SEPTEMBER 2021

Preliminary analysis and mapping of existing European and national Open Science infrastructures with regard to promoting responsible Open Science

- o 0 o -

Summary

Introduction

Part of the basic objectives of ROSiE are to guide and equip Open Science (OS) users after exploring the ethical, integrity, social, and legal challenges of OS; mapping existing technologies that safeguard responsible OS, and engaging stakeholders in an open dialogue. Science relies largely and increasingly, especially for its opening process, on research infrastructures which are either (i) generic in their scope but dedicated to OS such as OpenAIRE, D4Science, etc., or (ii) thematically focused but with a commitment to opening scientific processes and especially the release of data and publications such as DARIAH-EU, BBMRI-ERIC and other ERICs (European Research Infrastructure Consortia). During its first 18 months, ROSiE will map and analyze existing research infrastructures and their relation to, or their role in, promoting responsible OS. Most EC-supported research infrastructures have a solid commitment and experience in addressing responsible OS issues. ROSiE sees the analysis of these infrastructures' current practices as a robust first step towards its goal. For this purpose, a workshop was organized online on July 2021.

It is understood in the present document that "open science" is not exactly a product or an object, but a process: the opening of the entire scientific activity. This systemic transition does take some effort, and the European research community is committed towards the creation of the European Open Science Cloud (EOSC), i.e., "an environment for hosting and processing research data to support EU science" in order "to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and re-use data, tools and services for research, innovation and educational purposes" (quoting the EOSC Portal https://eosc-portal.eu/). It was therefore natural to co-organize and co-brand the workshop with EOSC.

The workshop was advertised during the EOSC symposium (June 2021) of which it was actually considered an extension, and through the EOSC liaison platform. It was however the organizers' choice to target a specific audience for this first workshop so that the focus was on research infrastructures and more specifically ERICs. It was decided to not combine practitioners with policy makers for instance, keeping in mind that a future workshop with other stakeholders will be organized in the coming months. The workshop gathered a large panel of ERICs: of the twenty-two currently established ERICs, seven were represented, covering a diversity of scientific fields such as humanities, life sciences, environment sciences, etc. The attendance was also broadened to include other parties of interest as listed in the Annex. It was, however, considered important to have a relatively small audience (15 participants, see annexed list) so that they could actively take part in the exchanges. The workshop took place online due to sanitary restrictions. The chat box attached to the videoconference service was very useful for the attendees to provide complements (comments, links, etc.).

Given this productive but limited audience, the ambition of the workshop was not to produce a comprehensive mapping of the situation, but rather to identify the main describers of the situation, to be explored through additional focused workshops in the coming months. Therefore, the participants were specifically invited to witness and comment about (1) the RE/RI/legal challenges and gaps they identified in their respective fields, (2) the (types of) solutions they have implemented so far and (3) their expectations of ROSiE and the form ROSiE's outputs are to be funneled through the Knowledge Hub, the guidelines, the trainings, etc. Beside this general organization, the chairs of each session were careful to leave the discussion flow without constraining it: the purpose was to listen.

General considerations and limitations

As expected, the overall organization of the discussions according to the status of each issue (challenges, current implementation, expectations) did not favor a clear distinction to be made between research ethics, research integrity, and the social and legal aspects of responsible science. Instead, most of the attention quickly became mainly focused on two matters, (1) compliance to the GDPR (General Data Protection Regulation, EU 2016/679) and (2) informed consent.

→ *The mapping and analysis of the practices of OS platforms will therefore require some specific questions to be examined, such as for instance "As an end-(re)user, how can I trust that the online data set which I am about to re-use is not the result of falsification or fabrication? How can I trust that it has not been produced against ethical rules such as informed consent of questionable validity or breaches to animal welfare, among others?"*

It was also observed that much of the focus was on data during the workshop: this is only one of the products of science, but an illustrative one.

→ *During the upcoming mapping and analysis, the focus will therefore have to be enlarged to diverse aspects of responsibility in the context of creating and using different types of products of open science such as data, open access research tools, resources.*

One participant suggested that Gaia-X be considered close to ROSiE in its objectives. Gaia-X aims at creating "the next generation of data infrastructure for Europe, its states, its companies and its citizens", that is an ecosystem ensuring digital sovereignty and fostering innovation based on trust in the digital environment. Indeed, ROSiE expects its outcomes on the determinants and practical tools ensuring responsible open science in Europe to be useful for such a strategic endeavor as Gaia-X's.

- o 0 o -

Design and enforcement of policies, principles and norms

Challenges and gaps

One of the main challenges mentioned by the participants during the ROSiE-EOSC workshop is that ERICs in particular, and other research infrastructures as well, feel that they have little or no power to force users to open up their collected data, and even more so in a responsible manner. Rather, this power lies in the hands of the institutions that employ researchers, and/or can only be enforced by national policies in each country.

However, the participants agreed that research infrastructures are in a position to raise awareness, to facilitate and encourage data sharing, to offer training, and to elaborate guidelines and practical tools for this purpose. Because they are by construction a meeting point for coherent communities from all origins, research infrastructures have the ability and the legitimacy to structure their respective communities and their sets of practices. This is an especially useful external resource for researchers in smaller universities and structures which do not have enough in-house resources such as the legal and/or ethics specialists employed in dedicated offices of larger universities.

Guidelines and training were believed to be essential, but they must be discipline-specific and/or area-specific. Research infrastructures, often being discipline-specific themselves, are therefore in a good position to implement or disseminate them, to stimulate or even participate in their design, to raise awareness among users and all stakeholders, etc.

Altogether, thus, the layers of (1) setting up a policy and general principles, (2) translating these into workable guidelines, norms and tools and (3) applying these in daily practice are operated by different types of stakeholders. These stakeholders interact and co-construct at all layers but the national and institutional levels are preeminent in (1), individual researchers and end-users in (3), and disciplinary or otherwise field-specific interfaces such as research infrastructures are in an excellent position for (2).

→ *This clarification will have to be made in ROSiE D6.2, in order to facilitate appropriation by all stakeholders.*

Examples of current implementation

Some ERICs maintain legal information platforms for their communities. Examples are:

- BBMRI in the field of biology/biobanking through its ELSI – Ethical, Legal and Societal Issues services: see https://www.bbmri-eric.eu/elsi/knowledge-base/
- CLARIN in the field of social sciences and humanities: see the legal information platform at https://www.clarin.eu/content/legal-information-platform

However, the research infrastructures reported difficulties to keep such platforms updated, especially with the diversity of member states' legal regulations of intellectual property. The ELSI services of BBMRI have published their return on experience at https://dx.doi.org/10.1089%2Fbio.2018.0018.

An example of how institutions deliver guidance on their policies is provided by Delft University of Technology (NL): https://www.tudelft.nl/library/tu-delft-open-science/os/open-publishing/about/policies. The general principles guiding the staff in all processes of their academic practice are listed, however without much details since those are better established on a smaller scale (i.e., disciplinary).

Funding agencies can also provide such information services, like for instance the platform on research integrity maintained by DFG, the main German research funding agency https://wissenschaftliche-integritaet.de/en/.

Depending on the universities and institutions, the support for the implementation of responsible science is provided by one or another office. All aspects of responsible science may not be addressed by a unique office in a given institution. For instance, at Stockholm University, while patent and intellectual property support services are often provided by the legal office, research ethics and research integrity services are provided by REIS, the office that provides services on research funding matters and the office which researchers are more familiar with.

Remark: maintaining specialized services is costly

When research infrastructures function as providers of tools to guide users through legal aspects, especially those linked with GDPR, they stress the elevated maintenance cost of such services. Moreover, the challenges faced for the maintenance of such tools are numerous, as exemplified by ELDAH CFW (see below).

The challenges for maintaining specialized services are not only in terms of costs, but also in terms of human resources, i.e., legal or other specialized competencies are challenging to hire. Their occasional turn-over, for instance when a specialist leaves, is also difficult and comes along with their successor not feeling comfortable updating and amending the work done earlier by someone else. Moreover, legal offices in universities and research institutions are understaffed, taking into account the growing

importance of GDPR in research. Finally, such platforms target an audience that is not accustomed to legal questions. The tool developed by legal specialists has to be understood by a non-specialized audience, making set-up and maintenance even more complex.

→ *Although this type of challenge is not the main focus of ROSiE, it does matter since it is a possible hurdle for implementation, so it should be documented during the preparation of D6.2.*

- o 0 o -

Informed consent: some clarification is needed

Challenges and gaps

In terms of informed consent, it was pointed out by the participants that it is sometimes difficult for researchers to distinguish between (1) the legal basis for the processing of personal data and (2) the fundamental ethical requirement for research involving human subjects or participants. The need for education and user-friendly instruments is evident, or the collection of informed consent may be seen as undue bureaucracy by researchers, with the risk that it is eventually overlooked or addressed *a-minima*.

It was also recognized rather unanimously that signing a paper or ticking a box on an online form is not a sufficiently robust expression of informed consent (or consent expression at large), despite the sometimes-different traditions in different EU member states (see below). It has to be the result of a more dynamic interaction with the community. Some institutions, especially Scandinavian ones, such as, e.g., the University of Stockholm (Sweden), officially advise researchers not to use informed consent as a legal basis for personal data processing even if, of course, they still need to obtain consent for participation in a given project or cohort: https://www.su.se/staff/researchers/research-ethics/how-to-inform-research-subjects-and-ask-for-their-consent-1.501340.

Two additional specific challenges were identified:

- Childrens' digital rights.
- In the Humanities, where primary data is often attached to actual persons, intellectual property is often managed as copyrights by scientific publishers who only rarely approve of delivering them to re-users.

Examples of current implementation

An online consent form wizard (CFW) has been co-developed by DARIAH (ELDAH – Ethics and Legality in Digital Arts and Humanities) and CLARIN: https://consent.dariah.eu/startpage in order to support researchers in the field of Humanities in the process of collecting valid consent for data processing in the context of their specific professional activity.

The project PANELFIT (EU Horizon 2020 grant agreement #788039) has recently delivered an analysis of the issues and gaps regarding informed consent in the domain of ICTs (information and communication technologies): https://www.panelfit.eu/wp-content/uploads/2020/11/D21-Issues-and-gaps-analysis-on-informed-consent-in-the-context-in-ICT-research-and-Innovation.pdf.

→ *The question of informed consent is at the crossroads of all aspects of responsible science: ethics, integrity, social and legal. It is therefore expected that D6.2 (and other deliverables of ROSiE) covers it thoroughly.*

- o 0 o -

A geographic heterogeneity of practices

Challenges and gaps

The different national traditions and practices between EU member states sometimes prevent ERICs and European infrastructures from establishing common guidelines at the European level.

For instance, while the University of Stockholm advises against using informed consent as a legal basis for further data processing and/or re-use (see above), which fits with, e.g., Scandinavian traditions and practices, for historical and cultural reasons Germanic countries such as Austria and Germany put emphasis on actually doing so. One participant insisted that in his experience, "*the distinction between consent as legal basis and consent for participation is actually one of the things that is the most challenging to get researchers and research participants to understand*".

In general, it may be difficult to apply common guidelines uniformly: rather, in some instances they must be adapted to national specificities. ERICs and other research infrastructures, therefore, must aim for flexibility and compatibility between countries.

Practical implementation of GDPR was also reported to differ between member states.

→ *The mapping and analysis performed to prepare D6.2 will have to cover not only disciplinary- of thematic area-specific heterogeneity but also the differences observed in different EU member states and/or cultural areas.*

Examples of current implementation

The section (see above) about the successive layers of policy/implementation/practice applies to this kind of geographic heterogeneity.

Earlier in 2021, EOSC-Pillar (Horizon2020 Grant Agreement #857650) has delivered a documentation of the legal and organizational aspects of service delivery for open science in Europe: https://repository.eosc-pillar.eu/index.php/s/tbqe6B7rDycdFCJ. Much of the attention regarding responsible science as defined here is given to legal and ethical aspects (especially but not only GDPR) while research integrity and social aspects remain to be explored, although the 2019 brief published by SPARC Europe on "RI through open science and FAIR data" is referred to.

→ *The policies and practices will have to be further mapped.*

- o 0 o -

Not all science products can be responsibly opened

Challenges and gaps

Another challenge that arose in the discussion was linked to business confidentiality and the way it can be conciliated with openness and data sharing. For infrastructures involved in partnerships with industrial or private bodies, this requires discussions and agreement signing on this topic before the project starts. At this point, it was reminded that FAIR data does not necessarily mean open data. It is possible, according to the involvement of academic or industrial partners on a topic, to adapt the data sharing rules of the consortium.

Therefore, the guidelines must include conditions and possible restrictions for openness.

Examples of current implementation

The FAIR principles, listed at https://www.go-fair.org/fair-principles/ and published by Holub et al. (2018) (https://doi.org/10.1089/bio.2017.0110), provide a shared guidance on what products of science to open, and how.

- o 0 o -

In the digital era, a diversity of publications needs to be qualified

Challenges and gaps

Information and communication technologies have a profound impact in the way all steps of research are performed. Anyone can publish anything on social media or on more formalized or specialized platforms. While opening science is definitely a positive endeavor, there is a need to qualify these publications and to audit their results, following processes that will probably be as diverse as the form they take themselves.

Examples of current implementation

None was specifically mentioned.

- o 0 o -

Remark: relying on data management plans (DMPs)?

It was proposed by some participants that DMPs could be considered as a basis to implement some guidelines (non-discipline specific). A DMP could be used as a way of mapping to provide procedures and guidelines and translated into a roadmap and/or a tool of self-training for researchers. Such early mapping of the project data and their attached challenges in terms of responsible science would help structure the projects at an early stage, anticipation being considered a key to handling legal and other responsible science issues.

→ *This possible approach will have to be explored in the preparation of D6.2, and of the ROSiE Knowledge Hub.*

- o 0 o -

Next steps for the ROSiE consortium

In the coming 12 months, ROSiE will produce and deliver a full report with a more extensive and detailed mapping and analysis of how the current practices of research infrastructures facilitate responsible open science, based on the observations made in this preliminary report. The full report will also identify the gaps that the ROSiE knowledge hub and other dissemination activities will be committed to fill.

For this purpose, concerted action will be maintained with the following ROSiE WPs:

- WP1 regarding the ethical and research integrity challenges, while producing its report on the relationship between RI, the wider RE perspective and OS (D1.1);
- WP2 regarding the legal and social challenges, while producing its report on social and legal implications and challenges related to OS (D2.4);

- WP5 regarding how policies and guidelines address research infrastructures or how they rely on them, while producing its report on the state-of-the-art in policies and guidelines in the European Research Area (D5.1);
- WP3 and WP4 to valorize their active interactions with the communities of stakeholders and of SwafS projects respectively.

Targeted workshops will be organized from December 2021 to April 2022 to explore the specific questions raised above and/or specific communities.

During the preparation of the EOSC-ROSiE workshop, it was identified that several workshops dedicated to specific target groups needed to be organized separately and sequentially. Beside the ERICs and research infrastructures at large, which were considered a priority and were therefore gathered in July 2021, it was estimated that national policy-makers should also be consulted to explore the current challenges, existing solutions and possible gaps in implementing responsible open science. A workshop will be dedicated to this consultation of research policy-makers.

Following the EOSC-ROSiE workshop, another need emerged: to conduct a series of thematic workshops since some of the challenges, solutions and gaps appeared to be specific while others were generic.

As stated above, it is difficult for many research communities to distinguish between research ethics, research integrity, and the social and legal aspects of responsible science. GDPR and informed consent, which focused much of the attendees' attention, seemed to currently be the aspects of responsible science which are the most readily recognized by actors[1]: this needs to be confirmed and possibly complemented. Further workshops will be committed to "unbias" this perception, and consciously focus on other aspects of responsibility in (open) science.

Specific questions were identified during the July preliminary workshop and highlighted in the present report. They will be addressed in subsequent workshops:

- An end-user perspective on the provision of "responsibility" certification services by open science infrastructures;
- A consideration of the cost of development and of maintenance of such specialized services for OS infrastructures.
- A documentation of the question of informed consent, in various disciplinary or thematic contexts.
- A mapping of the challenges, current solutions, gaps and expectations across disciplines, thematic fields and also geographical situations.
- The possibility that data management plans can be efficient vectors of responsible science principles in practice, to be valued by the ROSiE knowledge hub.

[1] See also Buljan, I., Pina, D. G., and Marušić, A. (2021). Ethics issues identified by applicants and ethics experts in Horizon 2020 grant proposals. F1000Research, 471. doi:10.12688/f1000research.52965.2.

Annex: the ROSiE/EOSC online workshop, 16 July 2021

Participants

- EOSC:
  - Karel Luyben (president of EOSC)
- ROSiE:
  - Heidi Beate Bentzen (WP2)
  - Rose de la Cruz Bernabe (WP1, WP9)
  - Lisa Häberlein (WP3)
  - Arild Jansen (WP6)
  - Panagiotis Kavouras (WP6, WP8)
  - Teodora Konach (WP5)
  - Olivier Le Gall (WP6)
  - Tom Lindemann (WP3)
  - Maud Medves (WP6; organizer)
  - Signe Mežinska (WP7)
  - Kadri Simm (WP1)
  - Lisa Tambornino (WP3)
  - Mariana Vidal Merino (WP4)
- ERICs:
  - CLARIN-ERIC: Paweł Kamocki
  - DARIAH-ERIC: Erszébet Toth Czifra
  - ECCSEL-ERIC: Sébastien Dupraz
  - EMBRC-ERIC: Gemma Gimenez Papiol
  - Euro Bioimaging-ERIC: Aastha Mathur
  - Lifewatch-ERIC: Christos Arvanitidis
  - BBMRI-ERIC: Michaela Theresia Mayrhofer, Mónica Cano Abadia, Ilaria Anna Colussi
- Other participants
  - Jonas Åkerman (Univ. Stockholm & EARMA-ERION)
  - Claude Kirchner (French National Pilot Committee for Digital Ethics)
  - Antonia Schrader (Helmholtz Association)
  - Mojca Kotar (Univ. Ljubjana & OpenAIRE)
  - Iñigo de Miguel Beriain (University of the Basque Country & Project PANELFIT)
  - Denise Amram (Sant'Anna School of Advanced Studies & PANELFIT project)

Program

- Welcome address and presentation of ROSiE (Rose de la Cruz Bernabe)
- Introduction of the workshop (Maud Medves)
- Session 1 - RE/RI/legal issues: experience, challenges & gaps (chair: Olivier Le Gall)
- Session 2 - RE/RI/legal issues: current technical/legal solutions (chair: Rose de la Cruz Bernabe)
- Session 3 - RE/RI/legal issues: Training, knowledge hub & other desirable environment (chair: Panagiotis Kavouras and Signe Mežinska)
- Open exchange
- Conclusion and final words (Rose de la Cruz Bernabe)