Guidance on Open Science and Research Data Management in Horizon Europe proposals

Edited by Politecnico di Torino

Written by Shalini Kurapati, Federica Cappelluti

in collaboration with

Research Support Department (ARI) - Sara Rollino, Francesca Spinazzola, Silvia Tabotta, Chiara Biglia

Library - Open Access Group - Anna Benenti, Donella Cioffi, Paola Ferrero, Maria Girard, Monica Margara, Anna Perotti, Sandra Sacco, Lorenzo Torriani, Ivana Viglione

Acknowledgements for scientific support and revisions: Elena Giglia (University of Turin)

Last update – 24 September 2021

Contents

Introduction

'Open Science' (OS) stands for the transition to a new, more open and participatory way of conducting, publishing and evaluating scholarly research. Central to this concept is the goal of increasing cooperation and transparency in all research stages. This is achieved, among other ways, by sharing research data, publications, tools and results as early and open as possible. Open Science leads to more robust scientific results, to more efficient research and (faster) access to scientific results for everyone. This results in turn in greater societal and economic impact[1].

Being a different way of doing research, Horizon Europe will consider Open Science within the "Methodology" section and will evaluate the proposal on how it will adopt and adapt Open Science practices.

Horizon Europe calls have introduced a set of required and recommended practices that concern research outputs such as scientific publications, FAIR [=Findable, Accessible, Interoperable, Reusable] research data and possibly the entire research workflow:

- **Mandatory OS practices** refers to the open dissemination of the project, once it has been approved. It is important to address how they will properly implemented already at the proposal stage.
- **Recommended OS practices** are not mandatory but <u>are strongly encouraged, and they are an integral part of the criteria to evaluate the proposal.</u>

This document has been developed for PoliTo researchers, support staff and anyone who is getting started with OS practices, to guide them in writing data and open science paragraphs for Marie Curie actions and other Horizon Europe proposals.

This document contains

1. A recap on mandatory practices as set out by the Grant Agreement
2. Detailed guidance on sections of Horizon Europe proposals dealing with OS practices and research data management
3. A sample text for research data management section to inspire researchers to formulate their own
4. A non-exhaustive dictionary of open science terms

If you are new to OS, please read the dictionary of terms before you fill in the OS section of the proposal.

Disclaimer: This is a general-purpose guidance. OS practices vary from discipline to discipline. Please contact open.science@polito.it if you have more specific questions.

[1] Open science definition, Qeios, https://doi.org/10.32388/838962

1 Mandatory OS practices

Rules are set in the Grant Agreement, Annex 5, Art. 17 - Communication, Dissemination, Open Science And Visibility.

Concerning publications, you are mandated to **make OA all your peer-reviewed scientific publications** (including articles and long-text formats, such as monographs and other types of books), by taking the following actions:

- at the latest upon publication
    - depositing the Author Accepted Manuscript (AAM) or Version of Record (VoR) in a **trusted repository** (see box), AND
    - providing to the deposited publication **immediate open access via the repository under CC BY or equivalent license** (CC BY-NC/CC BY-ND are only allowed for long-text formats such as monographies)
- you must maintain the sufficient rights to comply with the immediate access mandate. A clause to be added to the publishers' agreement is provided in the Programme Guide, p. 49
- Deposition in a trusted repository is always requested, even if you publish in an Open Access journal, for preservation and text mining purposes
- Via the same repository you will provide information about any research output/tools/instruments needed to validate the conclusions of the scientific publication. For example, you will be able to link to the DOI of datasets, code, etc...
- Metadata of publications must be open under CC 0 or equivalent, in line with the FAIR principles and provide information about the licensing terms and persistent identifiers, amongst others.

Concerning **research outputs other than publications** generated in the action (referred to as '**data**'), you have to manage them in a responsible way, according to the FAIR principles, and then taking all of the following actions:

- establish a **data management plan ('DMP**') (normally within month 6, and to be regularly updated)
- as soon as possible and within the deadlines set out in the DMP, **deposit the data in a trusted repository**
- as soon as possible and within the deadlines set out in the DMP**, ensure open access — via the repository — to the deposited data**, under the latest available version of the Creative Commons Attribution International Public License (**CC BY**) or Creative Commons Public Domain Dedication (**CC 0**) or a licence with equivalent rights, following the principle '**as open as possible as closed as necessary**'. If you need to keep your data closed, you have to specify the reasons in a section of the DMP.
- provide information via the repository about any research output or any other tools and instruments needed to re-use or validate the data. Outputs and tools, too, are expected to be "as open as possible".

Trusted repository as defined in the Annotated GA

– Certified repositories (e.g. CoreTrustSeal, nestor Seal DIN31644, ISO16363) or disciplinary and domain repositories commonly used and endorsed by the research communities (e.g. **Europe PMC** for life sciences including biomedicine and health or **arXiv** for physics, mathematics, computer science, quantitative biology, quantitative finance and statistics). Such repositories should be recognised internationally.

– General-purpose repositories (such as **Zenodo**) or institutional repositories (such as **PORTO@IRIS**) that present the essential characteristics of trusted repositories, i.e.:

* display specific characteristics of organisational, technical and procedural quality such as services, mechanisms and/or provisions that are intended to secure the integrity and authenticity of their contents, thus facilitating their use and re-use in the short- and long-term. Trusted repositories have specific provisions in place and offer explicit information online about their policies, which define their services (e.g. acquisition, access, security of content, longterm sustainability of service including funding etc.).
* provide broad, equitable and ideally open access to content free at the point of use, as appropriate, and respect applicable legal and ethical limitations. They assign persistent unique identifiers to contents (e.g. DOIs, handles, etc.), such that the contents (publications, data and other research outputs) are unequivocally referenced and thus citeable. They ensure that contents are accompanied by metadata sufficiently detailed and of sufficiently high quality to enable discovery, reuse and citation and contain information about provenance and licensing; metadata are machine- actionable and standardized (e.g. Dublin Core, Data Cite etc.) preferably using common non-proprietary formats and following the standards of the respective community the repository serves, where applicable.
* facilitate mid- and long-term preservation of the deposited material. They have mechanisms or provisions for expert curation and quality assurance for the accuracy and integrity of datasets and metadata, as well as procedures to liaise with depositors where issues are detected. They meet generally accepted international and national criteria for security to prevent unauthorized access and release of content and have different levels of security depending on the sensitivity of the data being deposited to maintain privacy and confidentiality.

NB: Personal websites and databases, publisher websites, as well as cloud storage services (Dropbox, Google drive, etc) are not considered repositories. Academia.edu, ResearchGate and similar platforms do not allow open access under the terms required and are **NOT** considered repositories.

2 Open Science in Horizon Europe Proposals

You will need to address Open Science matters at the following points of the proposal template:

- PART A – "Researchers involved in the proposal": you have to list up to 5 research outputs relevant to the project, such as publications, datasets, software… They are supposed to be "as open as possible, as closed as necessary", which for publications means they have to be available in a repository (e.g. PORTO@IRIS).
- PART B – Sec. 1.2 "Methodology": under the 'Excellence' criterion, you are expected to properly apply Open Science practices as a method to carry out you research project. Aspects that will be evaluated include: quality of open science practices, such as early sharing, FAIR management of research outputs, and engagement of citizens, civil society, and end users where appropriate.
- PART B – Sec. 2.2 "Measures to maximize impact - Dissemination, exploitation and communication": Open Science ("Fostering diffusion of knowledge and Open Science") is one of the Key Pathway Indicators for Impact[2]
- PART B – Sec. 3.2 "Capacity of participants and Consortium as a whole": under the 'Quality and Efficiency of Implementation' criterion, the expertise in Open Science, also in terms of scientific products such as papers and research outputs (described in Part A online), is evaluated.

A comprehensive overview is presented below in the [infographic curated by Ghent University]

[2] https://ec.europa.eu/info/research-and-innovation/strategy/support-policy-making/shaping-eu-research-and-innovation-policy/evaluation-impact-assessment-and-monitoring/horizon-europe_en

2.1 Guide to Template Section 1.2 - Excellence – Methodology

2.1.1 Open Science practices

Text from Research and Innovation Action (RIA) proposal template: it is under the **Excellence** Section (Section 1) in the **Methodology** Paragraph (Paragraph 1.2). A similar paragraph is usually requested in Horizon Europe calls and the suggested length may vary among actions (e.g. in RIA, IA, MSCA-DN is 1 page; in CSA is half page).
Open Science practices: Describe how appropriate **open science practices** are implemented as an **integral part of the proposed methodology**. Show how the choice of practices and their implementation are adapted to the nature of your work, in a way that will **increase the chances of the project delivering on its objectives**. If you believe that none of these practices are appropriate for your project, please provide a justification here. Open science is an approach based on open cooperative work and systematic sharing of knowledge and tools as early and widely as possible in the process. Open science practices include early and open sharing of research (for example through preregistration, registered reports, pre-prints, or crowd-sourcing); research output management; measures to ensure reproducibility of research outputs; providing open access to research outputs (such as publications, data, software, models, algorithms, and workflows); participation in open peer-review; and involving all relevant knowledge actors including citizens, civil society and end users in the co-creation of R&I agendas and contents (such as citizen science). ⚠ *Please note that this question does not refer to outreach actions that may be planned as part of communication, dissemination and exploitation activities. These aspects should instead be described below under 'Impact'*

This paragraph expects you to display a general awareness and knowledge of OS principles and to describe how you plan to put these principles into practice as an integral part in every step of your research. In this paragraph you must state the types of actions you wish to undertake to embed mandatory and recommended OS practices and tools in your research process. Although OS may seem all-encompassing, as a researcher you may focus on three or four key areas mentioned in this paragraph. Please bear in mind that some of these practices may not be applicable for all research disciplines. If you reckon none of the Open Science practices – see also a list and useful examples in the Programme Guide, p.42-58 – is suitable for your project, this is the section in which you have to justify your choices.

"Research outputs" shall be intended here not only as the very final results of the research work (like publications, deliverables, data, etc.) but pieces of work evolving during the research flow and exchanged among and/or elaborated by different people/partners. These can be open (to the public) or closed (reserved to the consortium) depending on the nature of the project or specific activity: the EU principle is "as open as possible, as closed as necessary".

Concerning mandatory OS practices, use one or two sentences to recap:

- Approach to open access to scientific publications and FAIR research data. Indicate which trusted repository(ies) you will use (you will provide more details in Sec. “Impact”, see Sec. 2.2).
- Setting and delivering of documentation for validation and re-use
- Management of research data in line with FAIR principles (you will provide more details in the following paragraph, see Sec. 2.1.2)

Among recommended OS practices associated with the research workflow, you might consider:

1. **Tools and practices to enhance Collaboration, Transparency and Research Integrity**: You may be aware that due to the pressure to publish and the perverse benefits of high impact factor journals, researchers are forced to publish only novel and positive research results. This has caused a significant crisis in terms of research reproducibility and research integrity issues in addition to the lack of transparency involved in the scientific publication process. There are several **tools and practices** that researchers can adopt **to uphold scientific integrity.** Consider some of these tools if they can be applied to your research and mention if and how you will use them during your research: publishing registered reports, opting for open peer review, or publishing a pre-print.

   Politecnico di Torino adopted a programmatic Manifesto and Regulations for Research Integrity (https://www.polito.it/ricerca/integrita/index.php?lang=en)

2. **Management of research flow, Collaboration, Reproducibility:** Think about research outputs that you can co-work on throughout the project life. Examples include research data, software (think GitHub/Gitlab projects and similar), lab protocols, methodologies, models, hardware, physical objects, and samples. **Collaborative Open Science tools** include shared reference libraries (e.g., Zotero), commenting openly (e.g., Hypotesis or PundIt), Pre-registering (e.g., OSF registries or As Predicted), platforms for sharing data and/or code, lab notebooks (e.g., Open Lab Notebook), protocols (e.g., Protocols.io). See here for a quite rich map of OS tools on research workflow and follow this link for an interactive and extensive list. You should mention that your best practices in Research data/output management will not only make your day-to-day research activities more efficient but will also guarantee long term preservation of your research outputs. If you are planning to make open any partial result during the project, you should highlight this and refer to **open licenses (e.g. CC0 or CC BY for data or free open-source licenses for software). Reproducibility** is one of Open Science’s cornerstone; a very useful tool is The Turing Way, a complete handbook to make your research workflow reproducible and you data FAIR.

3. **Co-creation with stakeholder and end users**: Initiatives like crowd sourcing and citizen science can help citizens actively contribute to one's research. If your research project is suitable for such initiatives, mention how you plan to execute them. See Co-Creation menu. Also mention collaboration and links with Research Infrastructures if any.

2.1.2 Research data management and management of other research outputs:

Concerning the management of research data, software and other outputs, the template calls for a specific paragraph (see table below) where you shall describe your operational steps on ensuring the open/FAIR research outputs in detail. This entails data management during project life.

Text from Research and Innovation Action (RIA) proposal template: it is under the **Excellence** Section (Section 1) in the **Methodology** Paragraph (Paragraph 1.2). A similar paragraph is usually requested in Horizon Europe calls and the suggested length may vary among actions (e.g. in RIA, IA, MSCA-DN is 1 page; in CSA is half page).
Research data management and management of other research outputs: Applicants generating/collecting data and/or other research outputs (except for publications) during the project must provide maximum 1 page on **how the data will be managed in line with the FAIR principles (Findable, Accessible, Interoperable, Reusable)**, addressing the following (the description should be specific to your project): Ø Types of data/research outputs/research outputs (e.g. experimental, observational, images, text, numerical) and their estimated size; if applicable, combination with, and provenance of, existing data. Ø Findability of data/research outputs: Types of persistent and unique identifiers (e.g. digital object identifiers) and trusted repositories that will be used. Ø Accessibility of data/research outputs: IPR considerations and timeline for open access (if open access not provided, explain why); provisions for access to restricted data for verification purposes. Ø Interoperability of data/research outputs: Standards, formats and vocabularies for data and metadata. Ø Reusability of data/research outputs: Licenses for data sharing and re-use (e.g. Creative Commons, Open Data Commons); availability of tools/software/models for data generation and validation/interpretation /re-use. Curation and storage/preservation costs; person/team responsible for data management and quality assurance. *Proposals selected for funding under Horizon Europe will need to develop a detailed data management plan (DMP) for making their data findable, accessible, interoperable and reusable (FAIR) as a deliverable at mid-term and revised towards the end of a project's lifetime.* *For guidance on open science practices and research data management, please refer to the relevant section of the HE Programme Guide on the Funding & Tenders Portal.*

The section can be dubbed as a data paragraph, which is a precursor for a detailed Data Management Plan (DMP, deliverable at month 6) should the application be successful. In general, the European Commission expects the data paragraph to briefly answer: the type of data that you produce/reuse, the size of the data, how you will describe the data (metadata standards), in which trusted repository you will safely store the data, how do you plan to make it FAIR and what are the costs roles, and responsibilities involved in this process. If you reckon that some of data or outputs generated during the project will have to be closed, this is the section in which you have to state and justify the reasons why. More details will be included in the full DMP.

You can briefly describe this section in terms of: how you will manage your data during the research according to the FAIR principles and what you will do after your research project ends in terms of long-term preservation. A short recap about FAIR:

- **Findable:** You will describe your research output with rich metadata standards followed by your research community, see for example. You have file naming conventions and easy to understand folder organization structure to make your files both human and machine readable. You can include these information in a ReadMe file (template). When you plan to share your research outputs you will choose a suitable trusted repository, for example Zenodo or Dryad, which mints a DOI [Digital Object Identifier] for your output making it findable in a persistent manner.
- **Accessible:** "Accessible" does not mean "Open", but stored in a way that can be accessible both by humans and machines respecting reuse conditions. Sometimes data cannot be open due to privacy concerns or commercial/IPR constraints. You can provide accessibility conditions also through a dedicated license if applicable.
- **Interoperable:** the data and research outputs should be structured in such a way that they can interact with other datasets, models and/or software. This includes metadata standards and ontologies or controlled vocabularies used in the research communities. Preservation friendly, non-proprietary data formats (for example csv, pdf etc. instead of excel and word) are recommended, in case of proprietary data and software mention the version and release information.
- **Reusable:** The important consideration here is that these outputs will be useful for the scientific community and public to not only validate your research but also for reuse according to the license and terms of reuse set by you. Make data "reusable" means to provide them with a license or terms of use that describe how the data can be used and under what conditions.. Please note that licenses can vary between data and software. For more info on license see for example. Reuse also requires sufficient documentation on how to properly use the dataset.

To check if your plans meet the FAIR principles see e.g. https://fairaware.dans.knaw.nl/

A good tool with technical suggestions to make your data FAIR is the FAIR Cookbook.

Responsibilities and costs associated with data management can vary between research projects. It can vary from data collection, documentation, storage, security, quality control and long-term preservation. To get an idea of the type of costs and responsibilities involved, refer to this OpenAIRE guide.

Costs for data management are eligible for reimbursement (Grant Agreement 6.2.C3).

2.1.3 EXAMPLE: A sample research data management paragraph to inspire you to formulate your own

This project will generate microscopic image data. Measurements and quantification of the images will then be recorded in spreadsheets. Micrograph data is expected to total between 100GB and 300GB over the course of the project. The data will be stored in password protected laptop computers and regularly backed up on institutional cloud storage provided by the Politecnico di Torino. Files will be named according to a pre-agreed convention. The dataset will be accompanied by a README file which will describe the directory hierarchy and file naming convention. Each directory will contain an INFO.txt file describing the experimental protocol used in that experiment. It will also record any deviations from the protocol and other useful contextual information. Microscope images capture and record a range of metadata (field size, magnification, lens phase, zoom, gain, pinhole diameter etc.) with each image. This

metadata, which is standard in the community, will help other researchers to understand and reuse the data if necessary.

It is not anticipated that this study will generate any patentable data or proprietary data so the data can be openly shared. Datasets together with the associated metadata from this work will be deposited in the certified Zenodo repository.* Zenodo provides a Digital Object Identifier (DOI) to the dataset as well as a suitable license to attribute to the dataset, in our case a CC-by license. Images will be stored as .tif. Data in spreadsheets will be stored as .csv. Data in free text documents will be stored as .txt. These formats are platform agnostic and should support future access and reuse. Any data which must be stored in a proprietary format will have the necessary software (including version number) noted in the associated INFO.txt file as part of the metadata and documentation information associated with the dataset.

The DOI issued to datasets in Zenodo can be included as part of a data citation in publications, allowing the datasets underpinning a publication to be identified and accessed. DOIs will also be linked with appropriate publications, to enhance visibility of datasets. Metadata about datasets and the license attribution will make the datasets publicly searchable and discoverable and will indicate how and on what terms the dataset can be accessed and reused.

[Researcher1] will be responsible for ● data capture ● data management ● metadata production ● storage and backup ● data archiving ● data sharing. We have already accounted and allocated costs in time and effort within the grant, as well as the institutional overhead in terms of IT and computing services to prepare the data for sharing / preservation. We have sufficient storage and equipment to undertake these tasks.

***Further Practical tip:** you can create a **Zenodo Community** with the project acronym and use it to deposit all the material related to the project (preprints, open access publications, datasets, code, images etc., project deliverables). For each of them Zenodo assigns a DOI, first step for FAIR management, and you will be able to select a different type of access (closed or public and license).

This example has been curated from actual research projects.

2.2 Guide to Template Section 2.2 Impact - Measures to maximise impact - Dissemination, exploitation, and communication

Over the course of Horizon 2020 (2014-2020) the Research Support Department (ARI) produced several handbooks for PoliTO staff to support them through cross-cutting themes - such as Impact, Communication & Dissemination, Ethics and Open Access (the former edition of this Guide) - which are available at this link on the intranet. Please refer to these manuals for a general overview on how to approach these aspects.

The Horizon Europe is an "impact-driven" programme[3] and therefore it has introduced nine Key Impact Pathways indicators, grouped in Scientific, Economic and Societal Impact: your proposal should take into account all of them, among which

- KIP1 "Creating high-quality new knowledge"
- **KIP3 "Fostering diffusion of knowledge and Open Science"**

As such, in this paragraph you shall

- demonstrate a responsible approach to dissemination and exploitation of project results. Horizon Europe sets openness as the default, but always according to the "as open as possible as closed as necessary" principle. A strategy on IPR management shall also be devised in this Section, if relevant.
- show your awareness about open access to publications/FAIR research outputs. You can briefly outline your publication plan and mention how you plan to disseminate your publications in open access (see detailed rules below), bearing in mind that this is one of the mandated Open Science practices. You can also mention that PoliTo has a dedicated OA policy, which encourages and supports OA by mandating authors to deposit their publications in the institutional open access repository PORTO@IRIS. Concerning the other research outputs (data, software, methods, etc.) describe the strategy for depositing them in a trusted repository and opening them as soon as possible, according to the approach and timeline set in the DMP (see "Excellence" section).
- emphasize the positive implications of the OS practices adopted in Sec. "Excellence" in terms of impact, e.g. ensuring reproducibility to and accessibility of FAIR research outputs associated to your publications will help to make them highly influential (KP1) and/or facilitate the uptake by stakeholders (see also KIP on Societal and Economic / Technological Impact).
- highlight initiatives like publication of pre-prints, contribution to open platforms for dissemination/uptake of results or diffusion of knowledge towards a specific target audience (opening conference presentations, videos, demonstrators, contributing to open learning initiatives such as Wikipedia, etc.)
- bear in mind that dissemination initiatives should be always adapted and commensurate to the intended audience (e.g. publication, conferences for academic audience; briefing paper for policy makers; leaflets for citizens…)
- as Horizon Europe is strongly linked to the UN Sustainable Development Goals, and Open Science has been recognized as an accelerator to achieve them, a mention of how the project will have impact on the issues tackled by the SDGs will be appreciated.

You have 3 ways to be compliant with the Open Access mandate:

1) **Publish in ORE** – Open Research Europe. ORE is a platform the EU Commission provides for free. In ORE you deposit your preprint, which undergoes open peer review (securing as a plus one of the recommended Open Science practices) and then is published, deposited and indexed in

[3] Programme Guide, p.9: "The impact-driven design of Horizon Europe1 aims at maximising the effects of Research and Innovation investments, ensuring their contribution to the Commission's policy priorities. It marks a paradigm change in the design of the EU R&I Framework Programmes from an activity-driven to an impact-driven programme".

databases like Scopus and Web of Science. Data are published alongside the article, complying to another Open Science practice. It is for free, so you don't have to put any cost in the budget. ORE Articles are published under a CC-by license. Publishing in ORE you are straightforwardly compliant with Horizon Europe rules.

2) **Publish in an Open Access Journal**. You get immediate access but then you must provide to deposit the paper and the underlying data in a trusted repository. Reliable Open Access journals can be found in the Directory of Open Access Journals (DOAJ). 27% of them asks for APC, Article Processing charges, varying from 250 to 2900 euros. You need to check in advance and put these costs into the project budget in order to be reimbursed. Please notice that only APCs in full Open Access journals are eligible. Fees in hybrid journals (i.e. traditional subscription based journals offering an "Open Option" for a single article or journals under transformative agreements) are not eligible for reimboursement.
3) **Publish in a traditional commercial journal**. In this case you have not only to deposit the allowed version of the paper (usually, the "accepted manuscript") and the underlying data in a trusted repository, but you also have to check for a potential embargo period in the SHERPA RoMEO copyright policies database (learn how to read the policy here) . If the journal does not allow for immediate access, you must add upon submission a clause by which you notify the publisher your "prior obligation" to the funder funding your research and your need to maintain the rights on your paper in order to give immediate access from the repository.

As for ways 2) and 3) bear in mind that you must always retain sufficient rights to ensure **immediate open access via the repository under CC BY or equivalent license.**

Whatever is the approach, you shall always archive your publications in "Porto@IRIS" in order to fulfill not only the Horizon Europe mandate but also the Politecnico OA Policy's requirements.

Useful resources for **journals and books**

- SHERPA ROMEO to check publishers' open access policy
- How to read SHERPA RomEO pages, in Italian and English
- ORE – Open Research Europe, the publishing platform provided by the Commission
- DOAJ directory of peer-reviewed and trusted open access journals
- DOAB directory of peer-reviewed and trusted open access books and publishers of open access books
- OA BOOKS TOOLKIT a toolkit to help authors in publishing open access books
- OLH The Open Library of Humanities – funded by an international consortium of libraries - publishes open access scholarship with no author-facing article processing charges (APCs)

2.3 Guide to Template Section 3.2 Implementation - Capacity of participants and Consortium as a whole

Text from template – part B
Describe the **consortium**. How does it match the project's objectives and bring together the necessary disciplinary and inter-disciplinary knowledge? Show how this includes **expertise** in social sciences and humanities, **open science practices**, and gender aspects of R&I, as appropriate.

You shall highlight here the ability and commitment of the consortium to do Open Science, e.g. mentioning previous experience in Open practices, citizen science projects; infrastructures/services available at the institutions that enable OS practice such as collaborative platforms, repositories, etc..

NB. Open products relevant to the proposal shall be listed in Part A (online) - Researchers involved in the proposal - *List of up to 5 publications, widely used datasets, software, goods, services, or any other achievements relevant to the call.*

- For digital objects such as dataset, software, etc.., you shall include their digital object identifier (DOI) or other type of persistent identifier (PID). Remember that this is the first characteristic to be compliant with FAIR principles!
- Listed publications should be deposited (AAM or VoR version depending on the editor policy) in an open access repository, such as PORTO@IRIS, Zenodo, etc.
- For each item you shall include a short qualitative assessment of its impact. Do not confuse this with the journal Impact Factor, which is not to be mentioned (Programme Guide, p. 40)! You can quote some metrics, but the relevant point is to explain why that publication/product/achievement has been relevant in the scientific community

3 Dictionary of Open Science terms to get you started

Citizen Science refers to the practice of science by people who are not affiliated with a research organization as professional researchers, but who cooperate with or under the supervision of professional researchers. Examples of citizen science projects are here.

Co-creation refers to the practice of involving societal actors in the making of the proposal. There are different ways, see examples here.

Crowd sourcing has emerged as a rapidly growing field in research facilitated by the development of recent technologies, a greater incentive for outreach among researchers, a growing public interest in applied science and the desire to have a positive impact on the world. Here is an example of a crowdsourcing research project.

DOI (Digital object identifier) is a persistent identifier that serves as a permanent link to a publication, or a research output such a dataset. A persistent identifier assigned to an article or dataset remains the same even if the location and address of the publication or dataset change.

Electronic lab notebooks are a software system designed for scientists who conduct lab research to document their research, maintain reproducibility and share their findings easily. Electronic lab notebooks provide functionalities such as a text editor to replicate paper notebooks, spreadsheets for calculations, protocol templates and lab inventories and collaborative tools to share information. Check here to see how some researchers use these tools and some guidance on how to choose the right lab notebook tool.

FAIR principles according to FORCE11 principles and published in Nature Scientific Data are Findable, Accessible, Interoperable, and Re-usable, to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of appropriate scientific data and their associated algorithms and workflows.

HARKing means to hypothesize after the results are known, also a questionable practice fueled by the pressure of publishing.

Meta Data and Documentation Metadata is data providing information about data that makes it findable, trackable and (re)usable. It can include information such as contact information, geographic locations, details about units of measure, abbreviations or codes used in the dataset, instrument and protocol information, survey tool details, provenance, and version information and much more. Metadata can take many different forms, from free text to standardized, structured, machine-readable, extensible content. It is recommended to use a standard metadata format used in your field. Some useful resources include

- DCC Metadata Standards
- RDA Metadata Directory
- Fairsharing

Open access online access to research outputs provided free of charge to the end-user. Here is a handy guide written by Open Access experts at POLITO.

Open peer review is an umbrella term for various alternative review methods that seek to make classical peer review more transparent and accountable. Certain aspects of the review process (identity of the reviewer, the review report, or the platform itself, etc.) are open to the research community or the public. Here is an example of process from a publisher and of process in a publicly funded platform OCTOPUS.

Open-Source Software/Code: Availability of source code for a piece of software, along with an open-source license permitting reuse, adaptation, and further distribution. Follow the link to know more about open-source software.

Pre-print: The version of a scientific publication prior to formal peer review. Check here for more info.

P-hacking, also known as data dredging, is a questionable scientific practice that occurs when researchers collect or select data or statistical analyses until non-significant results become significant. See here for more information.

Post-print/Accepted Version is the version of the paper after peer-review, with revisions having been made (not the definitive version of record).

Registered Report/Pre-registration is a form of journal article in which methods and proposed analyses are pre-registered and peer-reviewed prior to research being conducted. This process is also known as 'pre-registration' or 'pre-reg.' Registered reports are considered an important innovation to promote research reproducibility since they aim at reducing publication bias of selective reporting, as well as discouraging questionable scientific practices like p hacking and Harking. An example of a stage 1 registered report is here. Publishers have specific submission guidelines on registered reports, to see an example follow the link for nature communications guidelines.

Repository is used to safely archive and preserve publications, datasets, and any other research output in the long term. There are several types of repositories for different research outputs. In general, a repository stores data safety, makes sure that the data has a persistent identifier such as a DOI, that the data is accessible and reusable, accepts metadata standards and allows the depositor to specify a license governing the access and reuse conditions. Zenodo and Dryad are widely used certified general-purpose repositories. For a larger overview of general and discipline specific repositories refer to re3data.org and EUDAT CDI - EUDAT.

Reproducible Research is a broad term to establish scientific validity and verification of research results. Some of the commonly used terms are reproducibility, replicability, reliability, robustness, and generalizability. The method of verification varies from discipline to discipline, while the main terms mentioned in the Open Science domain are reproducibility and replication.

- **Reproducibility** refers to the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis (or more in general, methodology) to yield the same results
- **Replication** is the ability of a researcher to duplicate the results of a prior study if the same procedures are followed but new data are collected. It is also known as repeatability

Documenting reproducibility requires, at minimum, the sharing of analytical data sets (original raw or processed data), relevant metadata, analytical code, and related software.

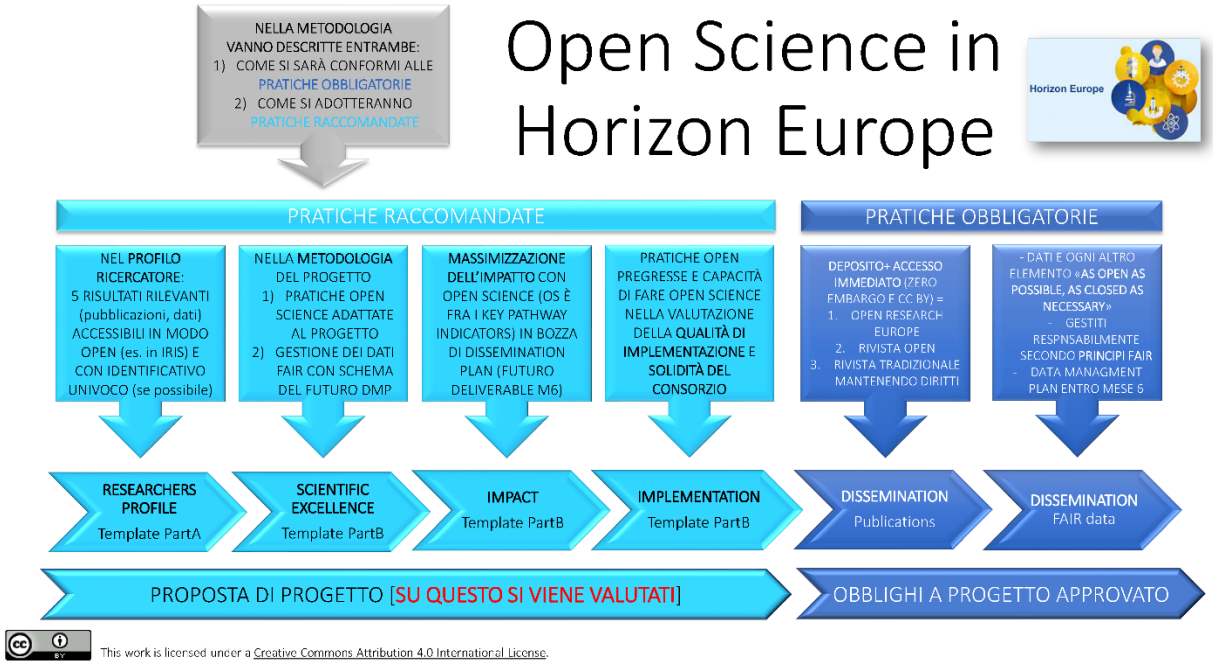
The Turing way handbook is an easy and complete tool to make your research reproducible and FAIR.

Please note that for qualitative research, research reproducibility does not focus on duplication of results, but it requires transparency, documentation, and the lineage of research processes. For more information on Open Science practices for qualitative research please refer to the Qualitative Data Repository.

Version Control is the management of changes to documents, computer programs, large web sites, and other collections of information in a logical and persistent manner, allowing for both track changes and the ability to revert a piece of information to a previous revision. Some tools that can help with version control are Git and Subversion.

4 Other resources

A summary (in Italian) of many useful recommendations can be found in the "Guida all'Open Science in Horizon Europe" written by Elena Giglia, University of Turin. Below an infographic to recap the main actions at the proposal and implementation stages.



……

Please note that Open Science is a dynamic and evolving topic. While we tried to cover most of the important aspects related to your proposal this guide will be updated based on new developments and requirements. We will keep you well informed of the changes through newsletter and website announcements. We have a dedicated slack channel for Open Science discussions if you wish to join here is the link:

https://join.slack.com/t/oscommunity-polito/shared_invite/zt-vwujvvpd-9zNspvNyTiLgVK3op3bUBw