Europa Media Trainings

# Open Access and Data Management

**Iasmina Cioroianu**
Project Manager, Europa Media Non-profit
EMG Group

Dissemination and Open Access in Horizon Europe
20 April 2023
Jožef Stefan Institute Ljubljana


TRAININGS

**01** The origins

**02** The practices

**03** Open Science & proposal writing

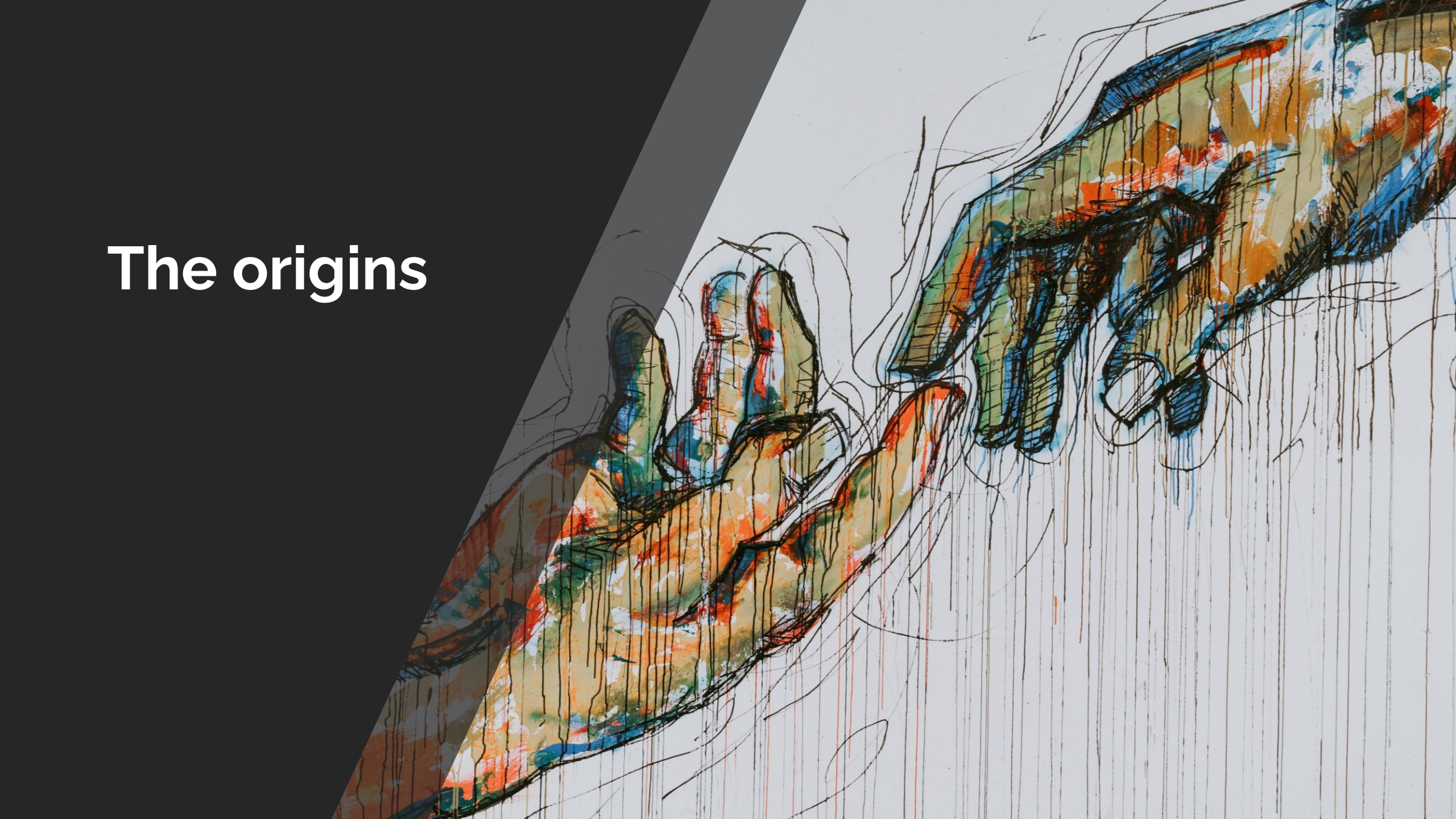
**04** Open Science & evaluation

**05** Open Science & implementation

**06** EC tools

**07** Other sources

# The origins



# Open Science

## The origins

### Open Science

= approach based on cooperative work and systematic sharing of knowledge and tools **as early and widely as possible**

### Responsible Research and Innovation (RRI)

= societal actors (researchers, citizens, policy makers, business, third sector organisations, etc.) work together during the whole R&I process to better align both the process and its outcomes with the values, needs and expectations of society.

- Public engagement
- Gender equality
- Ethics
- **Open Science**
- Science education
- Governance

# Open Science
## The 3 Os

### Open Innovation

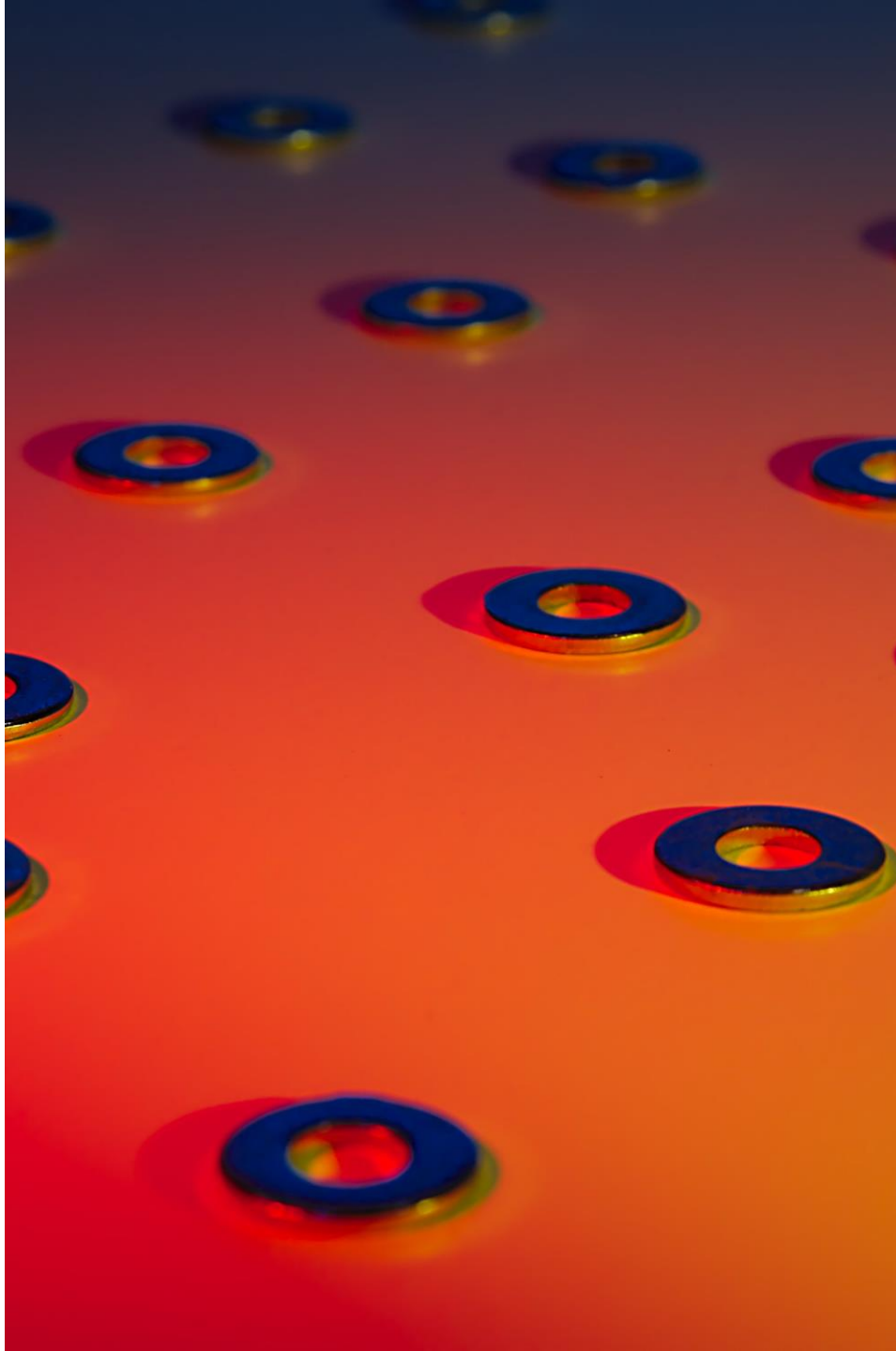
- ✓ in your methodology for collaboration with stakeholders you highlight how that leads to open innovation

### Open Science

- ✓ new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools
- ✓ practices like data management

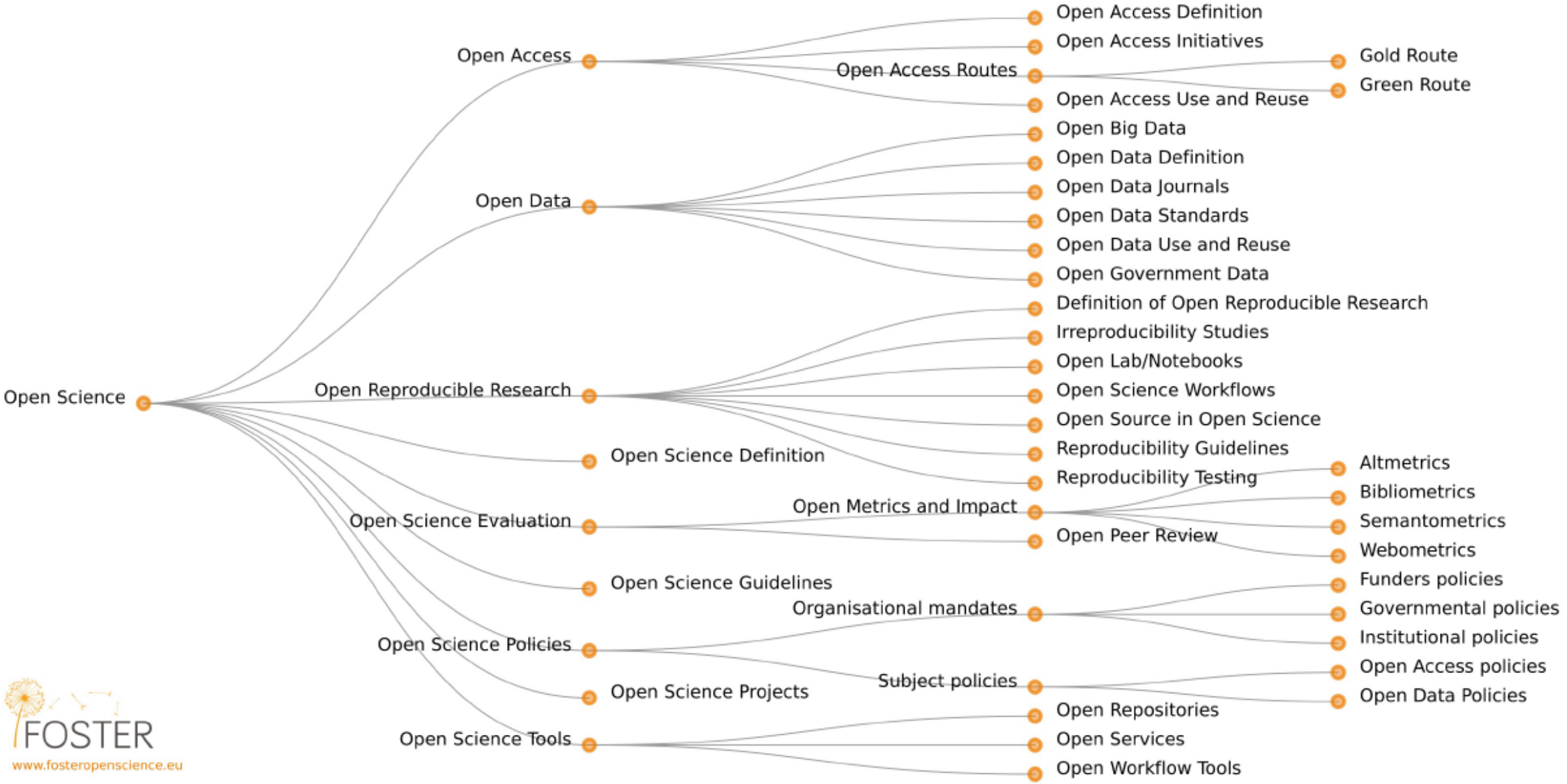
### Open to the World

- ✓ consider the existing international collaborations, the EU's regional and bilateral agreements



# Open Science Taxonomy

TRAININGS




FOSTER
www.fosteropenscience.eu

FOSTER Open Science taxonomy

Zenodo http://doi.org/10.5281/zenodo.2595951

# The practices (1) Open Access & Data Management

# Legal obligations
## GA Art 17 + Annex 5

### Open Access to scientific publications

The beneficiaries must ensure open access to peer-reviewed scientific publications relating to their results. In particular, they must ensure that:

- **at the latest at the time of publication, a copy of the published version or the final peer-reviewed manuscript accepted for publication**, is deposited in a trusted repository for scientific publications
- **immediate open access is provided to the deposited publication via the repository**, under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a licence with equivalent rights
- **information is given via the repository about any research output or any other tools and instruments** needed to validate the conclusions of the scientific publication.

**Metadata** of deposited publications must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles.

# Open access (1)

= **online access at no cost for the end user** of research outputs (e.g. scientific publications, data, software, algorithms, electronic notebooks etc.)

**Research data** = data used/produced in the research process or is the result of the research process

- raw and processed data
- simulation data
- observational data
- survey transcript data
- measuring sequences, code, algorithms, scripts
- audio-visual data
- physical objects (samples, artefacts)

# Open access (2)

**To consider:**

- Scientific peer-reviewed publications must be **open access under open licenses** (e.g. Creative Commons)
- **Retain sufficient IPR**
  - Retain the copyright on the work and grant, nonexclusive licenses to publishers
  - Put in place institutional policies to ensure copyright retention and compliance with the open access requirements
- Publication in venue of choosing but **publication fees are reimbursable only if publishing venue is FULL open access**
- In the proposal**: highlight the venues** that would qualify as providing open peer review
- 'As open as possible as closed as necessary'

# Open access (3)

## How to… Creative Commons

| Creative Commons license type | Share (copy + redistribute) | Use for commercial purposes | Adapt (transform, build upon) | Give appropriate credit | Maintain copyright, database rights |
|---|---|---|---|---|---|
| **CC BY** | Yes | Yes | Yes | Yes | Yes |
| **CC BY NC** | Yes | No | Yes | Yes | Yes |
| **CC BY ND** | Yes | Yes | No | Yes | Yes |
| **CC BY NC ND** | Yes | No | No | Yes | Yes |
| **CCo** | Yes | Yes | Yes | No | No |

# Legal obligations
## GA Art 17 + Annex 5

### Research Data Management

The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:

- **establish a data management plan ('DMP')** (and regularly update it)
- 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 (CC0) or a licence with equivalent rights
- **provide information via the repository about any research output or any other tools and instruments** needed to re-use or validate the data.

**Metadata** of deposited data must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles.

# Research Data Management
## The concept

### Research Data Management (RDM)

= the **process** within the research lifecycle that includes the data collection or acquisition, organisation, curation, storage, (long-term) preservation, security, quality assurance, allocation of persistent identifiers (PIDs), provision of metadata in line with disciplinary requirements, licencing, and rules and procedures for sharing of data.

Elements to consider in your project's RDM:

- **Persistent identifiers** (PIDs) to ensure findability of research outputs and data
- **Standardised metadata frameworks** for the findability of research outputs and their potential reuse
- **Trusted repositories** for the deposition of and access to publications and research data

# Research Data Management
## Access options

**Access levels**

- open to everyone (public use file)
- embargo (accessible after a certain date/period)
- limited user group (only registered users, campus use files etc.)
- on request
- closed (on-site-use/access only; data centres)

**Important for**

- Proposals
- Consortium Agreement
- Joint ownership agreements

# Research Data Management

## Ethical issues

- Ethical and privacy issues regarding sharing data
- Organisation's ethical and data protection policies
- Confidential or sensitive information (written consent)
- Data need to be anonymized?
- Secure storage and transfer of sensitive data

# Legal obligations
## GA Art 17 + Annex 5

### Additional practices

- **Where the call conditions impose additional obligations** regarding open science practices
- **Where the call conditions impose additional obligations regarding the validation of scientific publications**, the beneficiaries must provide (digital or physical) access to data or other results needed for validation of the conclusions of scientific publications
- **Where the call conditions impose additional open science obligations in case of a public emergency**, the beneficiaries must (if requested by the granting authority) immediately deposit any research output in a repository and provide open access to it under a CC BY licence, a Public Domain Dedication (CC 0) or equivalent. As an exception, if the access would be against the beneficiaries' legitimate interests, the beneficiaries must grant non-exclusive licenses —under fair and reasonable condition. **This provision applies up to 4 years after the end of the action.**

# The practices (2)
# Other Open Science practices

# Early and open sharing

= make available research, methodologies, outputs, and findings as soon as possible in the research process.

**HOW?**

- **Preregistration** in a public repository
- **Registered reports:** research articles that are peer-reviewed and published in 2 stages
- **Preprints:** scientific manuscripts that are publicly shared prior to peer-review and journal application via preprint platforms

**Resources**

- ORION
- The Centre for Open Science
- Sherpa Romeo
- Preregistration repositories: OSF, AsPredicted, etc.
- Preprint servers: Zenodo, Preprints, ArXiv, SocArXiv, etc.

# Open peer-review

= like peer-review but more transparent and accountable

- Authors and reviewers are aware of each other's identity
- Review reports are published alongside the relevant article
- The wider community is able to contribute to the review process
- Manuscripts are made immediately available in advance of the formal peer-review procedure
- Review or commenting on the final 'version of record' is made possible
- Direct, reciprocal discussion between authors and reviewers and/or between reviewers is allowed and encouraged
- Review can be decoupled from publishing when facilitated by a different organisational entity than the venue of publication (e.g. publishing platforms)

# Reproducibility of results

= possibility for the scientific community to obtain the same results as the originators of specific findings.

**Practices to increase reproducibility**

- Specify the research design and methodologies applied
- Specify how you deal with negative results
- Make prior searches and checks on existing results and data to avoid duplication
- Specify how you are making use or preprints, preregistration
- Detail steps you will take to make your research process and tools transparent
- Mention what steps you will take to ensure validity and quality of the project process and results (e.g. peer review)
- Plan to use the DMP and make sure your data are FAIR

# Citizen, civil society and end-user engagement

= refers to **opening of R&I processes to society** to develop better, more innovative and more relevant outcomes and to increase societal trust in the processes and outcomes of R&I

**Activities to consider**

- **Co-design** = workshops, focus groups to develop R&I agendas, roadmaps, policies
- **Co-creation** = involve citizens or end users directly in the development of new knowledge or innovations
- **Co-assessment** = assisting in monitoring and evaluation of the project progress and ensure interaction with citizens, civil society and end users on quality, utilization and impact of project outputs

# Open science & Proposal writing

# Section 1.2
## Methodology – Open Science example

***Open Science practices, research data and output management***

partners are organisations committed to open science. One Health-related science will be further boosted by through open sharing of research data, methodologies, protocols and research collaborations with other projects and initiatives to prevent and combat potential new pandemics.

partners have implemented good practices in terms of open access, data management (e.g. in MOOD and COMBAT Projects) that will also be applied in . Key Research Data Alliance recommendations will be followed.[46] The ethical framework will be fully in line with data management and open science practices ensuring interoperability, easy re-use and replicability.

| OS Practice | Adoption and implementation in |
|---|---|
| *Open access* | Green or preferably gold open access will be provided to scientific publications with deposited metadata. Open access will be ensured to research data, genomic sequences, the ABM model, integrated risk assessment strategies, algorithms, and participatory workflows. |
| *Early and open sharing of research* | Regarding scientific publications and open datasets, early and open sharing will be achieved via the (i) pre-registration at OSF, sharing time-stamped, read-only versions of publications, and (ii) upload of open datasets to trusted repositories (see below). Publishers' policies will have to be considered - Journals that will target include: *Emerging Infectious Diseases*, *Eco Health*, *One Health*, *Futures*, *Journal of Environmental Management*, etc. |
| *Open peer review* | will consider using Open Research Europe (ORE) for some publications. |
| *Co-creation* | The participatory approach is a key aspect in overall methodology. Actions will be implemented together with all relevant knowledge actors including citizens, civil society and end users not only in implementation and validation, but also in the co-creation of R&I agendas and |

# Section 1.2
## Methodology – Data Management example

**Research Data Management** will fully comply with the FAIR principles and will be monitored following the data management plan with strong collaboration of data harmonisation principles (WP2). data will be:

| Pre-existing datasets | New data |
|---|---|
| • Human and veterinary public-health (PH/VH) agencies' data, national statistical offices' data;<br>• MOOD platform data;<br>• Eurostat, Worldclim, Earthdata, ESA's Copernicus, European Environment Agency's (EEA) data repositories, GBIF and Worldpop bibliographic datasets collected from PubMed, WoS, Google Scholar;<br>• Epidemiological data on past epidemics and outbreaks sourced from available literature and gathered through text mining of social media, Internet queries, and news media by using relevant key words identified and validated for each pathogen; and<br>• Genomic sequence data related to the relevant pathogens that will be retrieved from publicly available databases based on availability of metadata and its precision - INSDC (GenBank) database. | **Sampling data:** epidemiological data and genomic sequence will be created and collected;<br>Open format: GenBank, FASTA sample collection date (ISO 8601) and sample geo-localisation (ISO 19115/32018);<br>**CF covariates list:** GIS/spatial/satellite data – GeoTIFF and GPKG formats;<br>**Metadata:** use of thematic and general controlled vocabularies such as Mesh or UMLS (Unified Medical Language System) for health data, GCMD (Global Change Master Directory) or CF (Climate Forecast) for co-variates;<br>**Statistical data:** Environmental, agricultural, climatic, demographic, and socio-economic data, as contextual data to be integrated with epidemiological/disease datasets; and<br>**Observational data** from participatory workshops. |

# Section 1.2
## Methodology – Data Management example

**Accessibility of data/research outputs:** The data created by the partner institutions will be accessible following their open access/science policies. also has the Dataverse, an institutional platform for data resulting from its research. All partner policies generally follow the principle of providing access to their data on a royalty-free basis, but access is provided upon request. In case the datasets would be used for commercial purposes, partner institutions would sign an agreement with the third party based on fair and reasonable conditions. Access modalities will follow the Horizon Europe's Programme Guide's recommendations related to open science.

**Interoperability of data/research outputs:** will follow pre-existing standards recommended by scientific communities (such as: norm ISO 8601 for the dates, ISO 19115/32018 for geolocalisation, EML for ecological data, Norm DDI for survey and social data (this norm is available in the Dataverse), those of the ECDC (TESSy) for epidemiological data, as well as WHO, OIE and RDA recommendations). In principle, data and metadata will be requested, stored and transferred (across partners) in a comma-separated values (CSV) format. MS Excel-compatible files, including .xls(x) format will be also accepted to facilitate exchange. Other data formats that will be used in include .fata, .png, .gml, and .nexus. For statistical purposes, other formats include .sas7bdat (SAS), .RData (R), .SAV (SPSS) and .mat (matlab). As far as possible, the consortium will reuse existing controlled vocabularies for providing metadata to resources.

# Section 1.2
## Methodology – Data Management example

**Reusability of data/research outputs:** We expect core project partners to openly deposit their data using a Creative Commons version BY 4.0 licence, or equivalent. Data will be deposited in GenBank,[50] Dataverse and/or GBIF for biodiversity data. As noted, data will be visually/searchable available through the Ebo-Sursy Project system.

Data: The Data Management Plan will include a Data Transfer Form, guidance to prepare data for exchange, guidance to obtain access to data, and guidance for establishing a processing agreement for pseudonymised data. Genomic sequences will be fully accessible. Survey results or socio-economic data will be available upon request and only shared in anonymised format. All documentation concerning the data (protocols, survey questionnaires, data dictionary, etc.) will be accessible with the data (in the data repository) or published in open access.

Samples and related data: Prior to the exchange of samples, a material and associated data transfer agreement will have to be completed and signed between the data controller and the recipient (i.e. laboratory performing the analyses of the samples (data processor)). Transfer of all materials and their associated data shall be documented through the completion of a Material and Associated Data Transfer Record Form. will collaborate with the **Knowledge Centre for Biodiversity** as well as support the objectives of the **Biodiversity Partnership**.

**Curation and storage/preservation costs:** Storage and data sharing will be discussed in detail in the DMP. Data storage can be provided by the partner institutions. (https://data.inrae.fr), all have proper IT systems for high security data storage. Hot data storage can be managed by AVIA-GIS, and the use of a cold data archive (OpenGeoHub) will be also considered. Data sharing with external parties within or after the project will be addressed by . Such decisions will take existing collaboration agreements on data use and sharing into consideration, collaboration with MOOD and VEO Projects, as well as contributions to European or global data platform options. Data security is of major importance in the Project. Special attention will be given to the security of sensitive data. The protection of personal data will be ensured through procedures and appropriate technologies such as the use of HTTPS protocol for the encryption of all Internet transactions, and appropriate European and Internet security standards from ISO, ITU, W3C, IETF and ETSI. If data will be kept in a certified repository, then the security standards of that repository will apply.

# Section 3.2
## Capacity of participants and consortium as a whole - example

previously mentioned tasks. Given their expertise in development and maintenance of co-variate databases, will also play an important role in data management tasks.

has a 27-year history of market-focused agricultural livelihood programs in Cambodia, strong expertise in ethnographic/anthropological/qualitative research, and in designing innovative solutions using Human Centered Design/Design Thinking methodologies for complex problems. are also active in the facilitation of participatory activities that include a variety of stakeholders. They will be involved in WP3 –, specifically in the assessment of socio-economic risk factors, including wildlife trade, and WP5, WP7. is an expert on GESI and established a *Global Gender Equality and Social Inclusion Policy* - recognizes that biases and social norms prevent women, men, and socially excluded groups from exercising free choice and from taking full and equal advantage of opportunities for individual development, contribution and reward. By outlining 's GESI guiding principles, this policy aims to challenge discrimination in our global programs and our organization by improving our knowledge of the way our interventions have impacted women, men, and socially excluded or marginalized groups and to better understand historical imbalances among these populations so that we can ensure all clients and staff enjoy equitable opportunities to participate in and benefit from our work.

# Open science & evaluation

# Open Science
## Evaluation

**Excellence**

- Methodology: how open science practices are implemented

**Capacity of participants and consortium as a whole**

- How the consortium brings together the necessary disciplinary and interdisciplinary knowledge

**Part A**

- List up to 5 relevant publications, widely used datasets or other achievements
- Open access expected for publications
- Datasets are expected to be FAIR and ‘as open as possible, as closed as necessary’

# Open Science
## Evaluation

**Criterion 1 - Excellence**

Score:

**The following aspects will be taken into account, to the extent that the proposed work corresponds to the description in the work programme:**

**- Clarity and pertinence of the project's objectives, and the extent to which the proposed work is ambitious and goes beyond the state of the art.**

**- Soundness of the proposed methodology, including the underlying concepts, models, assumptions, inter-disciplinary approaches, appropriate consideration of the gender dimension in research and innovation content, and the quality of open science practices, including sharing and management of research outputs and engagement of citizens, civil society and end users where appropriate.**

*An extensive programme of Open Science initiatives is presented which is fully in line with EU recommendations. The initiatives for sharing of research outputs are sound and credible. Data management aspects are well-addressed, including basic ethical considerations, data findability, accessibility and interoperability.*

*The proposal properly endorses open science practices, with a plan to adopt and promote Open Access wherever possible. However, the proposed claims about data management, sharing of data and research outputs based on AI tools are not well founded.*

*Open Science practices are poorly addressed without adequate procedures for early and open sharing of research, management & reproducibility of research outputs. FAIR is mentioned without specifying the types of data and how to ensure FAIR.*

# Open Science
## Evaluation

**Criterion 3 - Quality and efficiency of the implementation**

Score: **5.00** (Threshold: 3/5.00 , Weight: -)

**The following aspects will be taken into account, to the extent that the proposed work corresponds to the description in the work programme:**

**- Quality and effectiveness of the work plan, assessment of risks, and appropriateness of the effort assigned to work packages, and the resources overall.**

**- Capacity and role of each participant, and the extent to which the consortium as a whole brings together the necessary expertise.**

*The consortium expertise is very well considered, representing the required fields of expertise needed for successful implementation, the partners' competencies building essential synergies. There is a good geographical coverage and an appropriate distribution of academic organisations and NGOs. Leadership roles are well considered and include both EU and LMICs partners. Expertise in SSH, Open Science and communication is well represented.*

# Open science & implementation

# Your first deliverables…

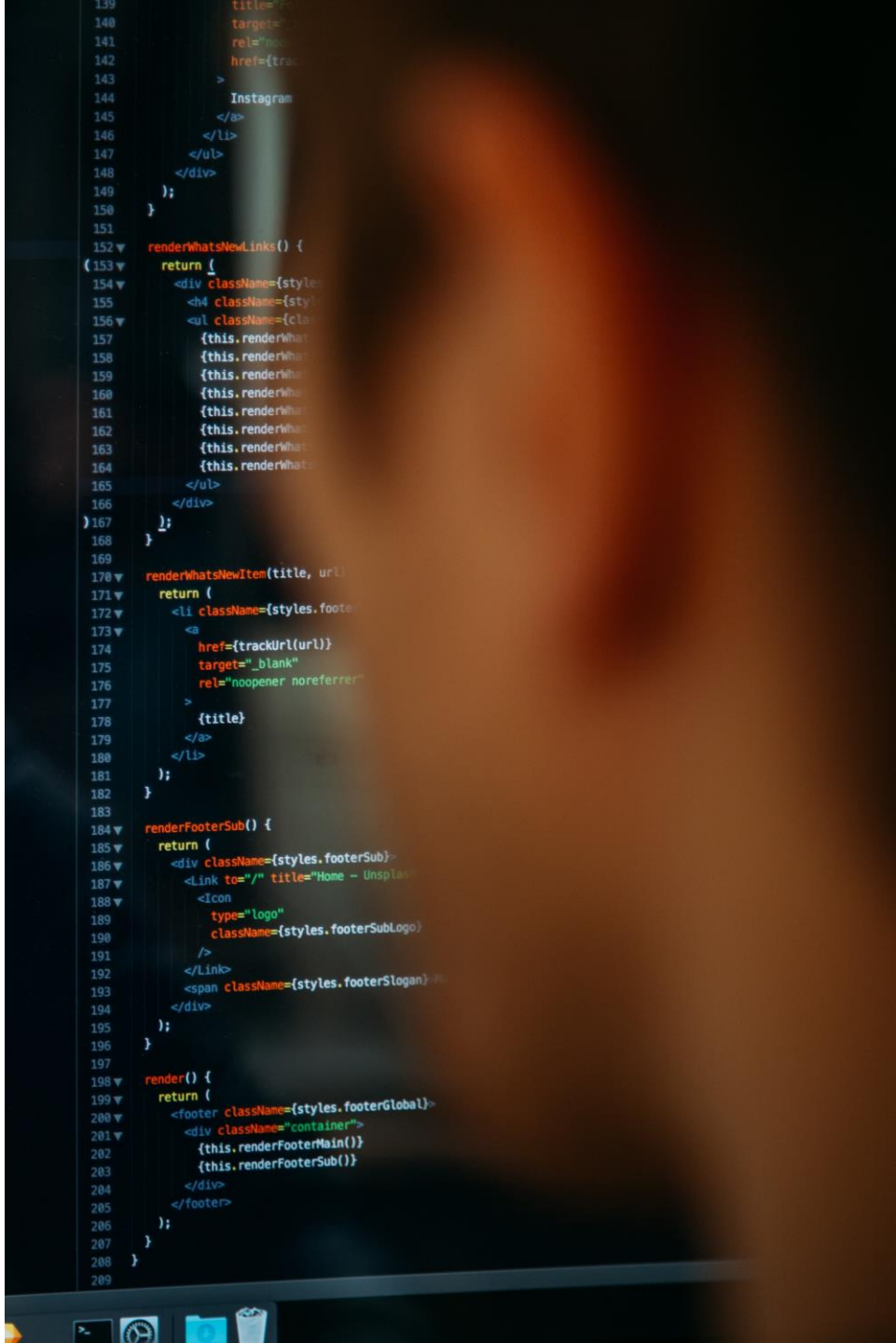
## Data Management Plan

**What is a Data Management Plan (DMP)?**

= your key to good data management

- Describes the data management life cycle
- The template = set of questions
- Living document **(!)**
- Deliverable – 1st version by M6

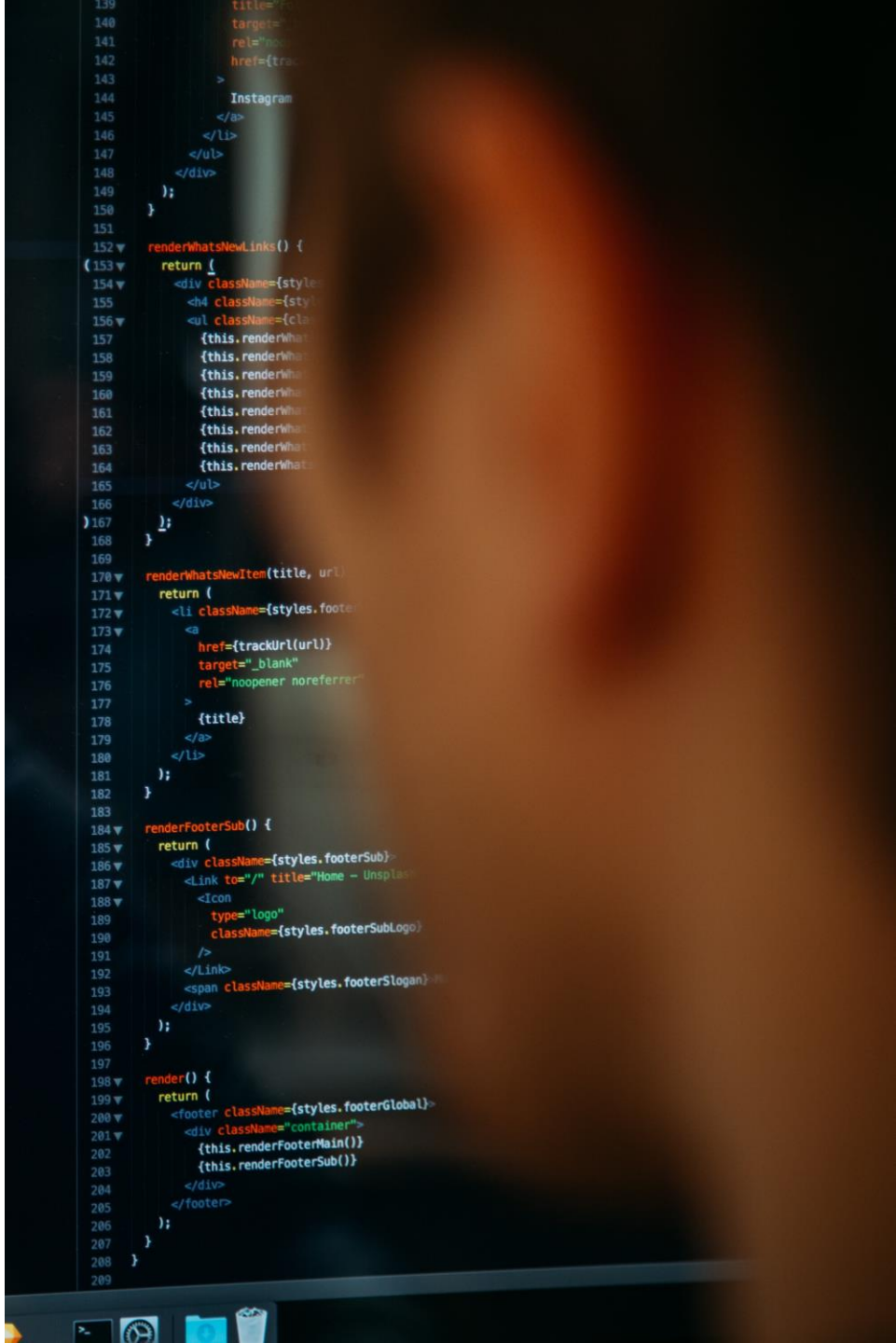
💡 Register your DMP as non-restricted public deliverable + publish it in journals, platforms or repositories (e.g. RIO, DMPOnline).



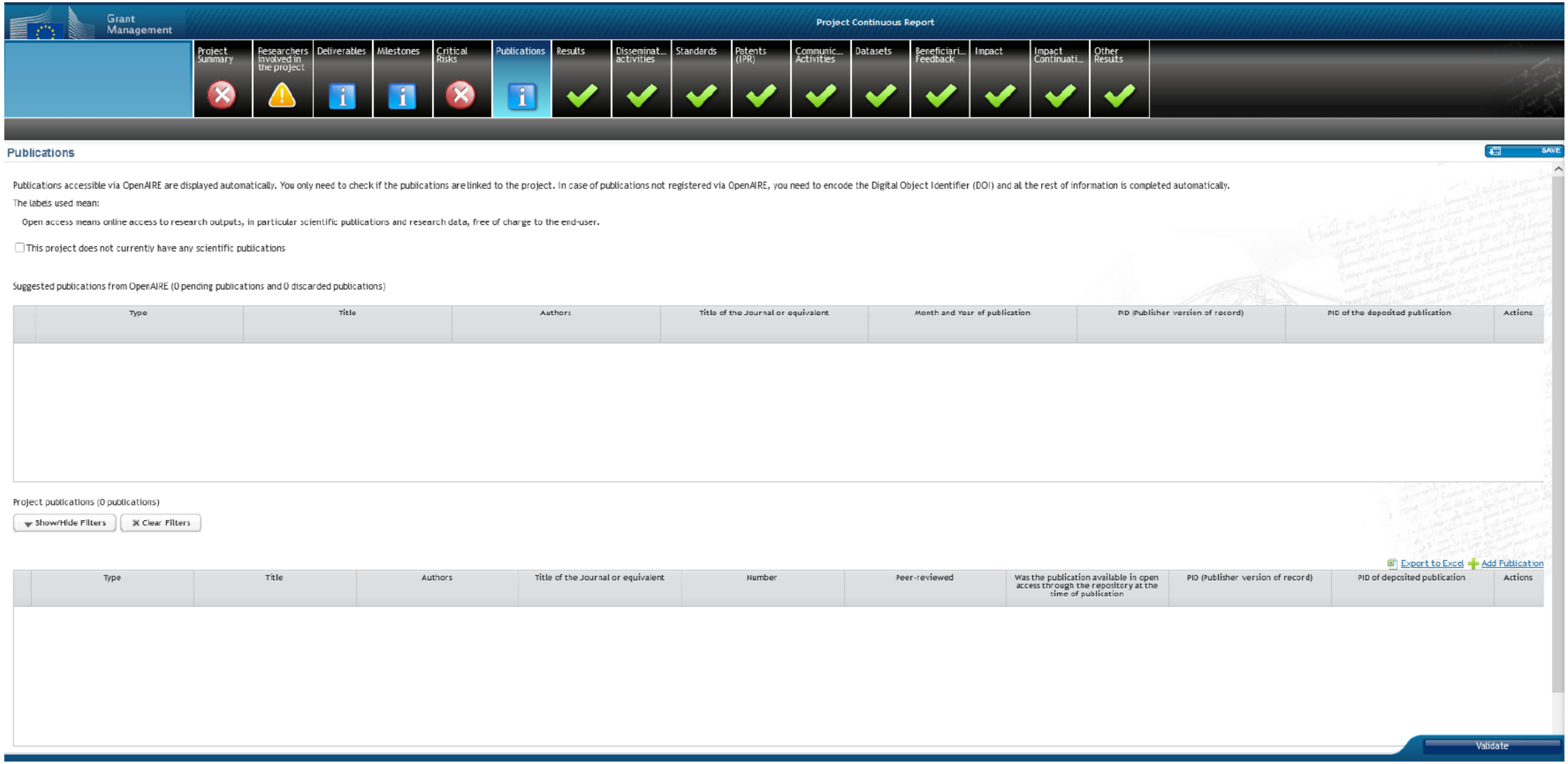
# Data Management Plan

## What to include?

- **Data set description:** what kind of data is your project generating or reusing? Estimate the size of the data set
- **Standards and metadata:** how do you structure your data and what protocols are you using?
- **Name and persistent identifier for data sets:** unique and persistent identifier and a stable link to directly access the data
- **Curation and preservation methodology:** how will you ensure the integrity of the data sets and for how much time? How will it be preserved and kept?
- **Data sharing methodology:** how can the data sets be accessed? Terms of use and license
- **Research output management** other than data and publications
- **Related costs and personnel:** data collection, documentation, storage, preservation, availability and reuse, person/team in charge

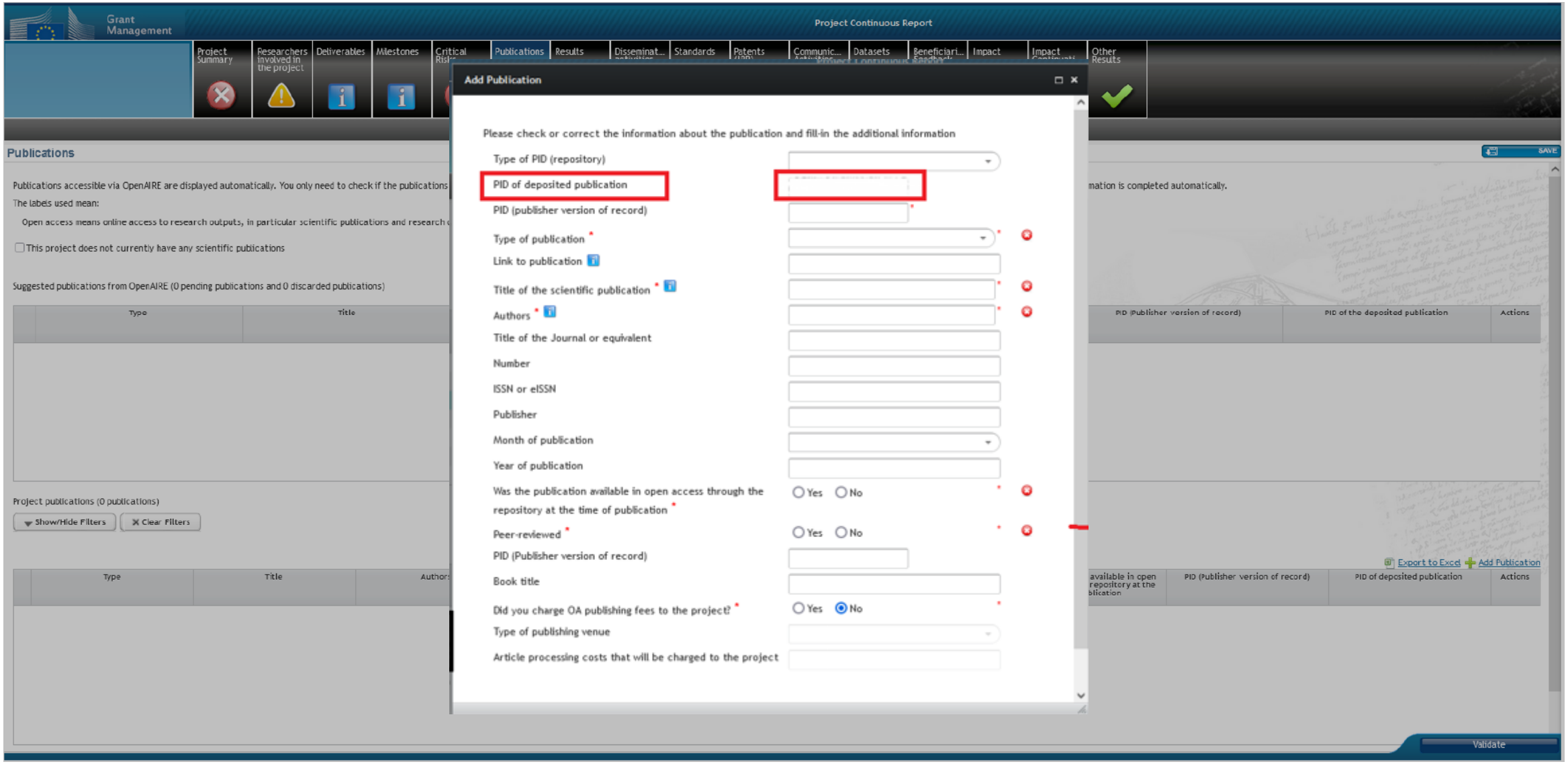


# Continuous reporting
## Publications

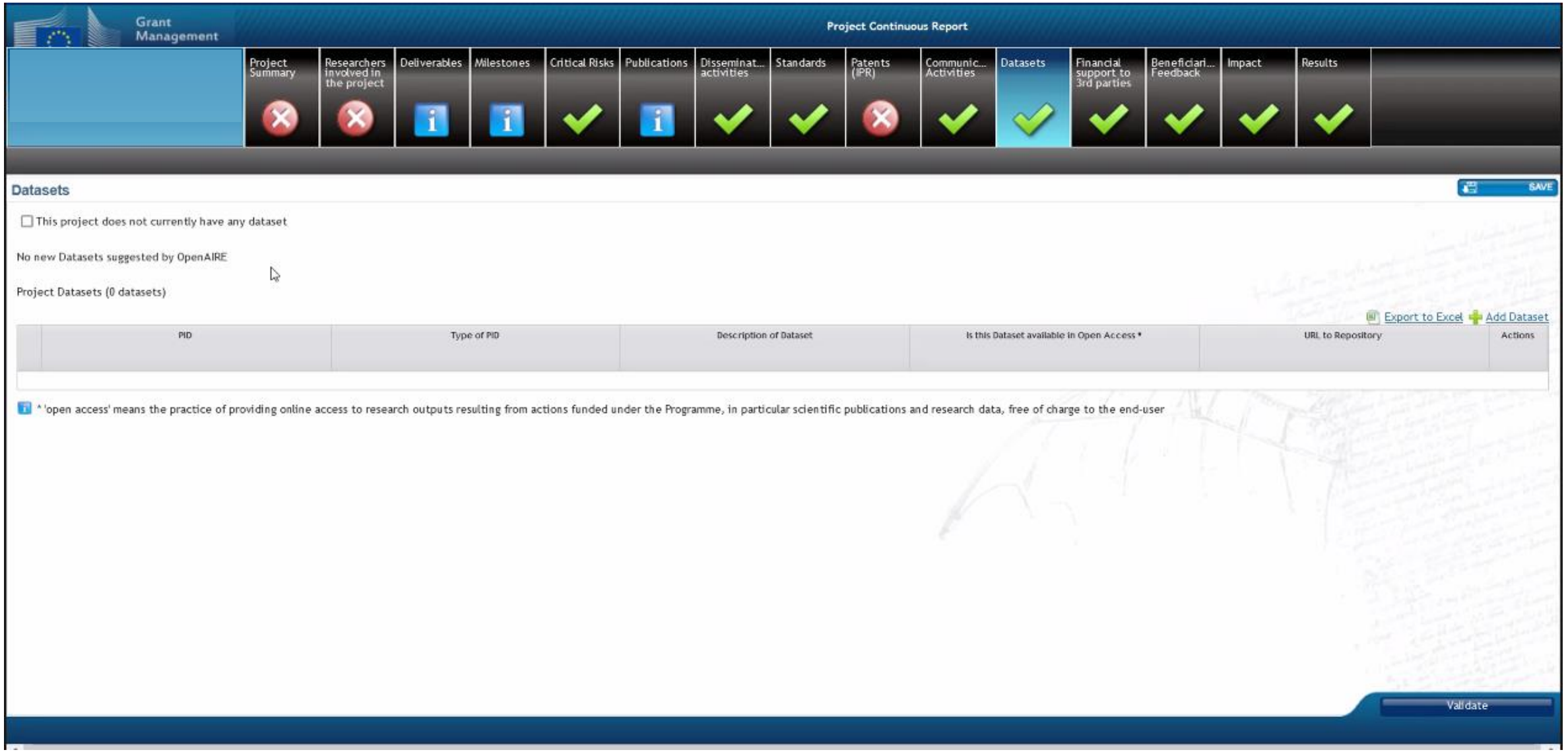


# Continuous reporting

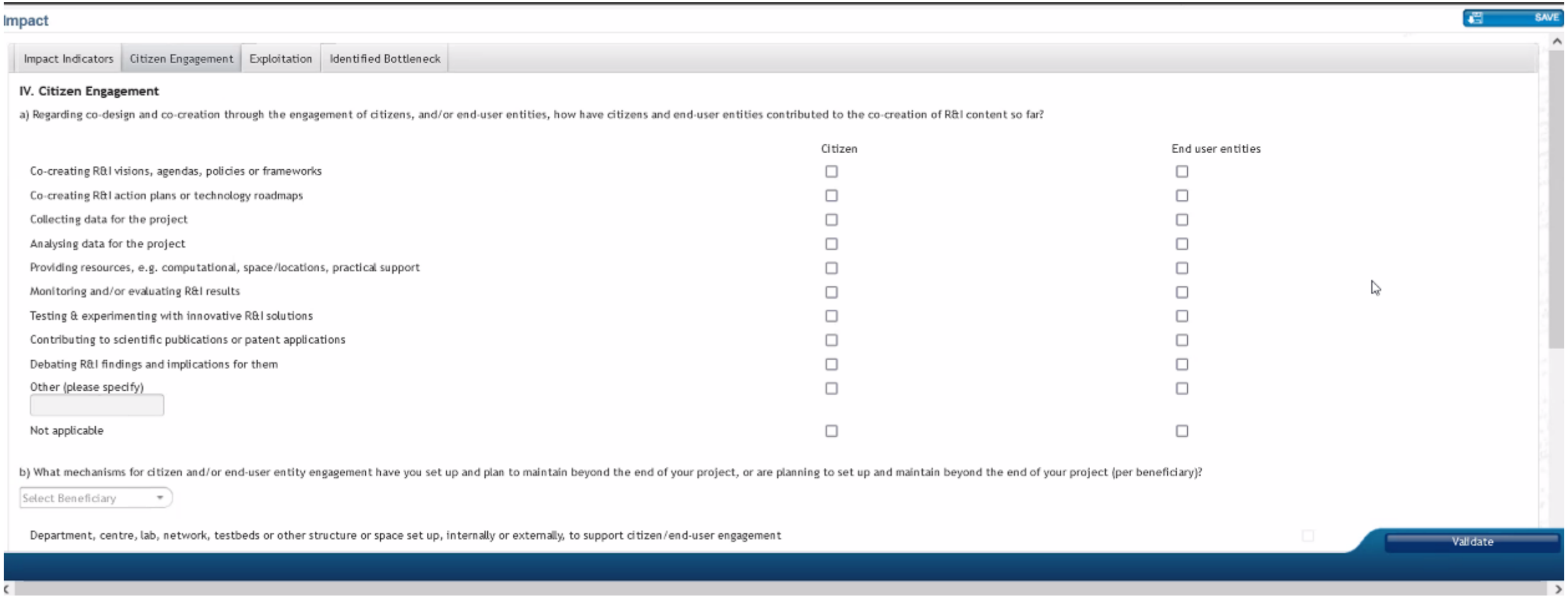
## Publications



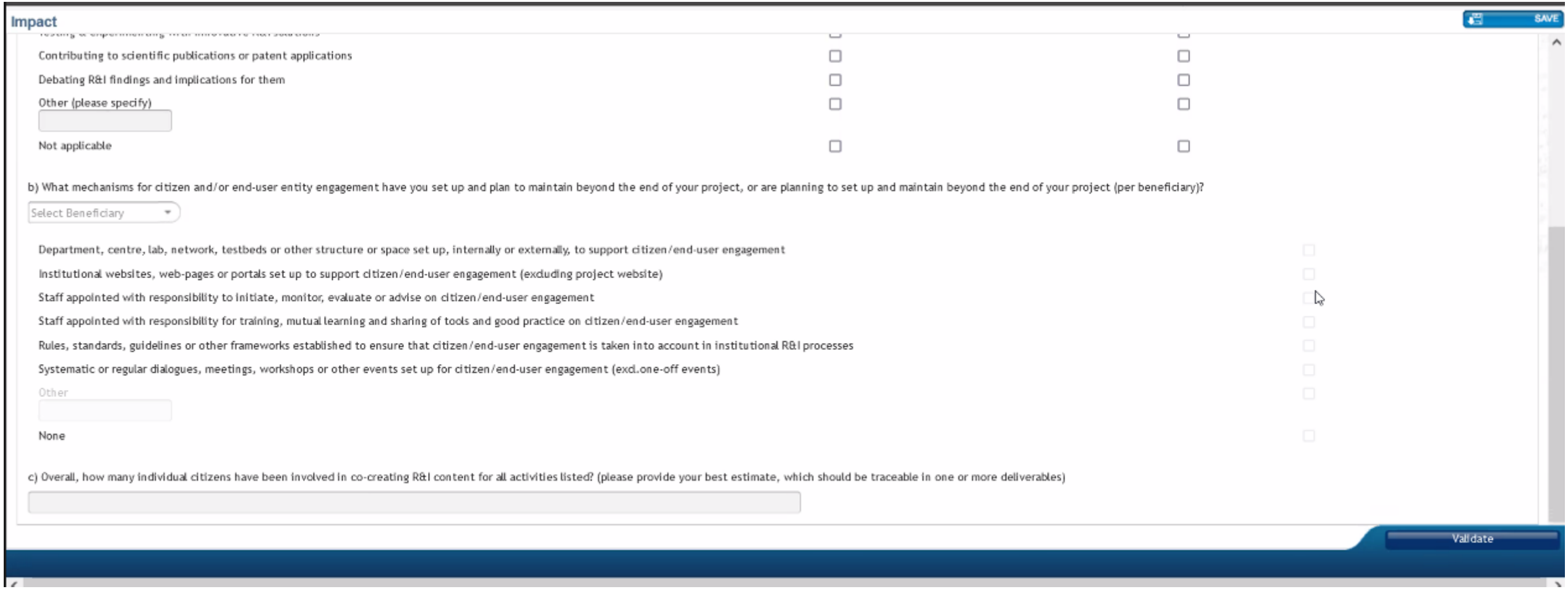
# Continuous reporting
## Datasets



# Continuous reporting
## Impact – citizen engagement



# Continuous reporting
## Impact – citizen engagement



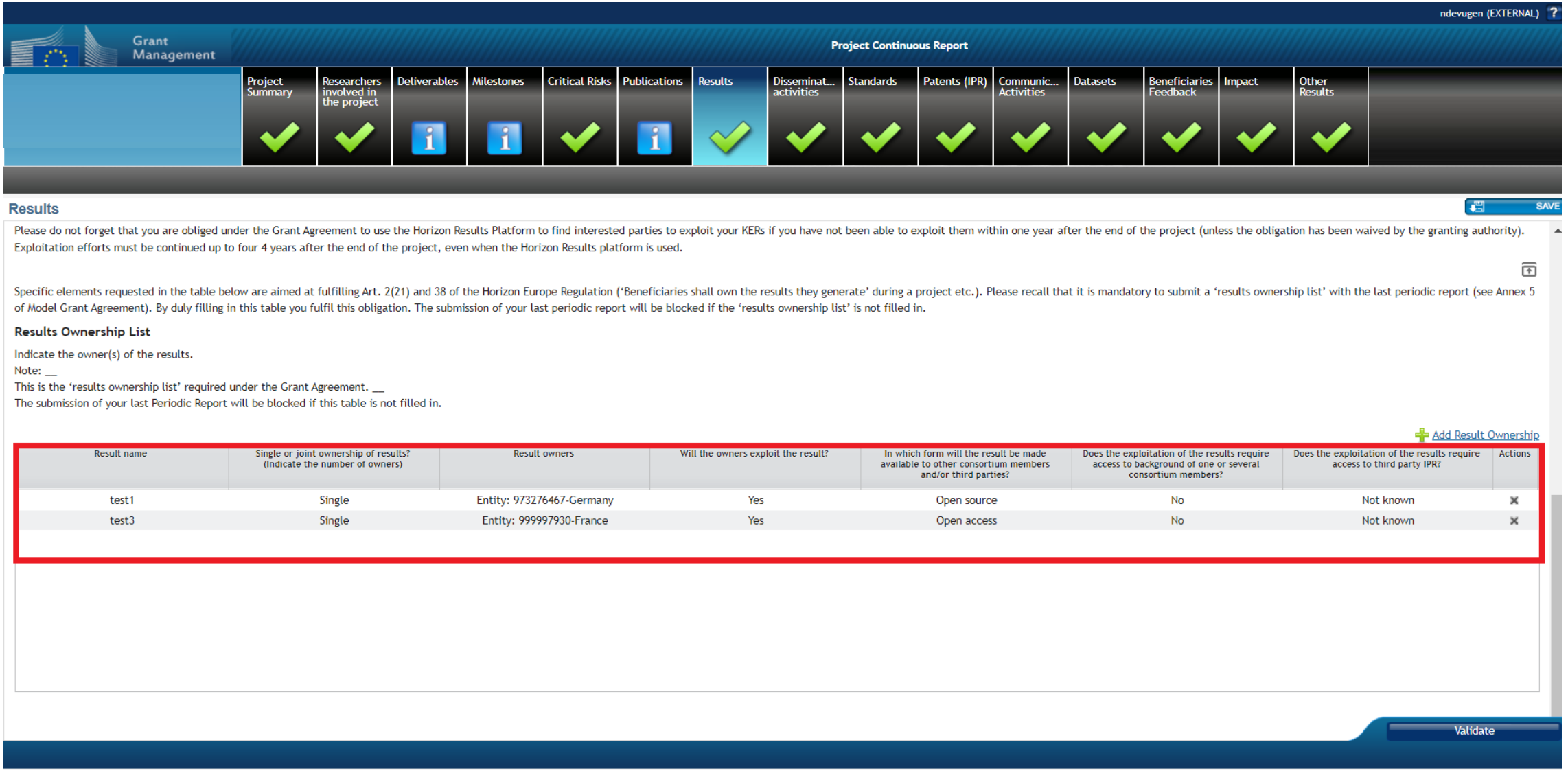
# Periodic reporting
## Part B, Section 4

Describe the Open Science practices related to **early and open sharing of research** (e.g. through pre-registration, registered reports, pre-prints or crowd-sourcing of solutions to a specific problem).

Describe the concrete measures that ensure the **reproducibility** of the results obtained during the action i.e., measures to ensure that the same results can be obtained by using the same data and/or methods, etc.

# Final reporting period

## Results Ownership List



# The EC tools

# European Open Science Cloud (EOSC)

= open trusted virtual cloud to enable researcher to store, share, process, analyze, and reuse research data, publications, and software across disciplines and borders.

- Use cases on EOSC in Practice
- EOSC community and stakeholders on events and news sections
- EOSC-Hub
- Catalogue & Marketplace for services and resources for researchers
- Training

💡 Work programmes may require the use of trusted repositories federated in EOSC for depositing research data

# OpenAIRE

= network of dedicated Open Science experts + training

= infrastructure harvesting research output from connected data providers

**Services and perks:**

- Integrated scientific information (links publications, project info, datasets in one place)
- Training sessions on Open Access/Open Science
- Zenodo = all-purpose open research repository (publications, datasets, code, posters, presentations…)
- Open Science helpdesk
- 34 National Open Access Desks (NOADs)
  - UNIT in Norway

💡 All OpenAIRE services can be used for free, however some of them may require logging in.

# Open Research Europe

= open access publishing platform for the publication of research coming from H2020 and HE funding

**Some characteristics:**

- Helps beneficiaries comply with the open access terms of their funding
- Publishing venue to share results and insights rapidly
- Facilitate open, constructive research
- Author-driven model = authors make sure the article is peer-reviewed by independent experts
- All articles are published open access under a CC-BY license

**Process:**

1. Article submission
2. Publication & data deposition
3. Open peer review & article revision
4. Send to indexers & repositories

# Other sources

# Tools and platforms

Digital profile

- ORCID (for researchers)
- CRIS (for organizations)
- ImpactStory
- Publons
- Open Science Framework (OSF)

OpenAIRE – Zenodo – Argos

- Putting your work into OpenAIRE-compliant repositories ensures that
  - you comply with H2020 mandate on Open Access
  - saves you time as you can import your project publications into the F&T Portal in one click

Open Research Europe

European Open Science Cloud (EOSC)

# Open Science
## Sources and guides

- HE Programme Guide
- OpenAIRE guide: *Guiding you in Open Science* https://www.openaire.eu/guides
- Research Data Alliance
  - https://www.rd-alliance.org/
  - http://rd-alliance.github.io/metadata-directory/
- Open Science Framework https://osf.io/
- Re3Data https://www.re3data.org/
- GitHub https://guides.github.com/
- Choosing a License https://choosealicense.com/
- FOSTER Open Science https://www.fosteropenscience.eu
- FIT4RRI project https://fit4rri.eu/guidelines/

# Research Data Management
## Sources and guides

- Research data management (RDM) open training materials (Zenodo)
- FOSTER Open Science e-learning
- Data Management Plans
  - DMPonline
  - OneHealth EJP DMP Guide
  - Webinar (video: DOI: 10.5281/zenodo.2564974; slides: DOI: 10.5281/zenodo.2565750)
- EC Guide for FAIR data management in H2020


QUESTIONS?

# thank you

## *for your attention*

Iasmina Cioroianu
iasmina.cioroianu@europamedia.org

**FOLLOW US!**

@EuropaMedia
@jesuisiasmi