

Landscape study on (semi-)automatic publishing workflows/integration between RI and repository services

From the EOSC Future WP3 Working Group on the EOSC Interoperability Framework for Research Product Publishing

Edited by

Alessia Bardi, CNR-ISTI

Contributors

- Jose Benito Gonzalez Lopez, CERN, Zenodo.org technical manager
- Paolo Manghi, OpenAIRE, OpenAIRE Nexus coordinator, EOSC Future TCB member
- Chris Ariyo, EUDAT B2SHARE service owner
- Andreas Czerniak, Bielefeld University Library/OpenAIRE-Nexus
- Paul Gondim van Dongen, SURF
- Georgios Kakalettris, NEANIAS, Project Technical Manager
- Raul Palma, PSNC, RELIANCE coordinator
- Silvio Peroni, University of Bologna, Director of OpenCitations
- Hans van Piggelen, SURF
- Mark van de Sanden, Technical Coordinator EUDAT, EOSC Future TCB member
- Diego Scardaci, EGI, Technical Solution Team Lead, EOSC Future TCB member
- Jochen Schirrwagen, project coordinator at Bielefeld University Library, OpenAIRE-Nexus
- Debora Testi, CINECA, DICE project coordinator
- Raphaël Tournoy, CNRS, Episciences Project Manager
- Irena Vipavc, Social Science Data Archive, University of Ljubljana
- Deborah Grbac, Library of Università Cattolica del Sacro Cuore di Milano
- Carl-Fredrik Enell, EISCAT Scientific Association
- Guido Aben, CS3MESH4EOSC
- Ivan Heibi, OpenCitations
- Jorik van Kemenade, SURF, Manager of the SURF data repository

Summary

Open Science calls for researchers to publish as soon as possible any type of research product in such a way their research activity can be transparently assessed, reviewed, reproduced, and rewarded in all its aspects. However, the publishing process has become more and more a burden for scientists, who must, most of the time, spend time to publish their articles, data, software, and other products in the many institutional or thematic repositories of reference. Scenarios include first-time publishing of new resource products or double-publishing of research products, to satisfy institutional mandates and community practices. Such tedious work is often incomplete, with some products ending up unpublished and others showing incomplete or imprecise metadata.

As a solution to these problems, some communities investigated and realised the integration of their research performing services, from RIs and Clusters, with repositories. The integration ensures that outcomes of such services are deposited by the services, prior authorization of the users, into a given repository, giving life to an end-to-end scientific workflow, from experimentation to publishing.

This document reports the experiences of the WG members, describing solutions at different maturity levels (design, prototype, beta, production) and involving different types of services (repository, analysis/research tool, publisher, other scholarly service) for the (semi-)automatic deposition steps of research assets produced in a research infrastructure to a target service in the scholarly communication ecosystem. It also presents a list of scenarios that would benefit from an interoperability framework for research product deposition.

Methodology

The WG studied the status quo of (semi-)automatic publishing workflows by reviewing existing experiences about the design and implementation of (semi-)automatic deposition workflows of research assets produced in a research infrastructure to a repository service. The landscape study revealed 15 implementations, which were then classified according to the following axes:

- Maturity level (design, prototype, beta, production)
- Type of submitter/receiver (repository to repository, analysis/research tool to repository, scholarly Service to repository, publishing platform to repository)
- Implemented API/protocol (Zenodo API, B2Share API, SWORD)

The WG interacted with the team of the EOSC Future cross-infrastructure use-cases and went through a brainstorming session for the identification of scenarios, even beyond those implemented by the use cases reported for the landscape study, that would benefit from an interoperability framework for research product deposition.

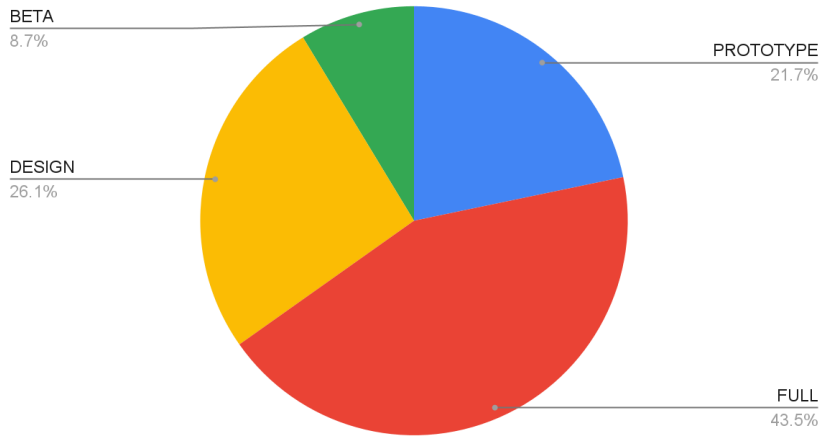
Existing solutions for (semi-)automatic publishing

The solutions identified by the WG are described below and summarised [in this online spreadsheet](#).

The collected list of solutions for (semi-)automatic publishing of research products reveals a high interest in the topic. The group identified 14 use cases involving 14 submitters (one use case may have several submitters, one submitter can participate in more than one use case) and 10 receivers or targets (one receiver can participate in more than one use case).

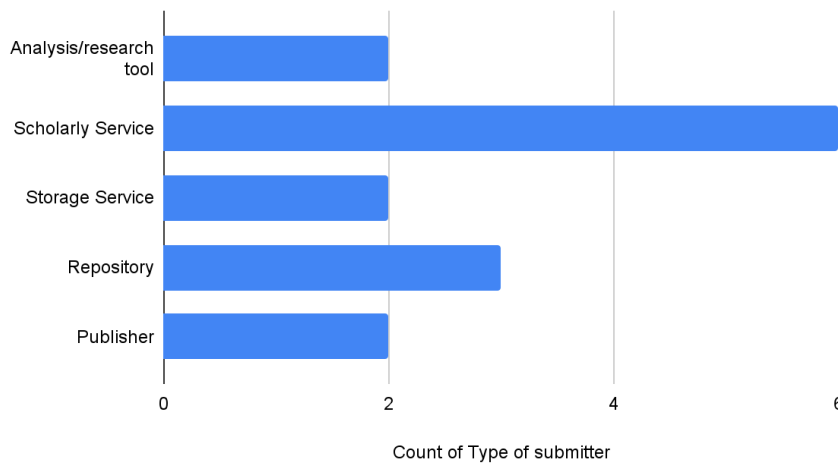
More than 40% of the solutions are fully implemented and operational in a production environment.

Count of Maturity



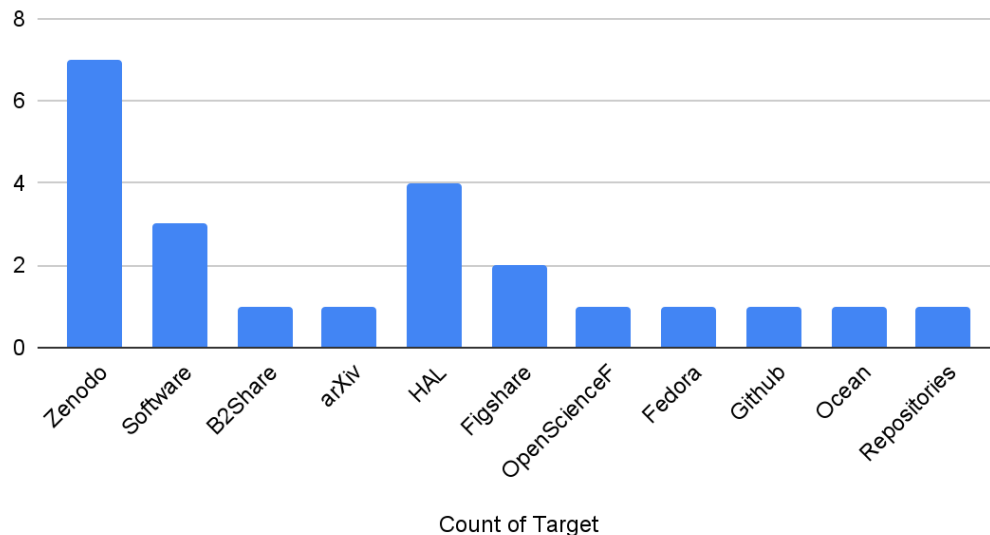
There are a variety of types of submitters. The most recurring is “Scholarly service”, not surprisingly due to its broad spectrum covering services related to scholarly communication in general (e.g. OpenCitations, PCI, Argos, ROHub). Repositories are also represented, followed by storage services, analysis/research tools, and one publishing platform (episciences.org).

Count of Type of submitter



Looking at the receivers (targets) is useful to understand which are the protocols, APIs and standards used for the implementation of the publishing workflows. The most recurring API is the API of Zenodo, followed by those of HAL, Figshare, and SWORD (used by SoftwareHeritage).

Count of Target



The landscape studies and the discussion among the WG members confirmed the need of creating a bridge towards repositories and that the bridges are not currently implemented with standards, but by implementing the proprietary API of each repository service they want to target. The only exceptions were the standards SWORD (mentioned in the study because of its implementation on SoftwareHeritage) COAR Notify and Signposting (mentioned in the use cases of integration between HAL, Episciences, Software Heritage and Peer Community In).

In the following subsections each use case is briefly described.

1. Publishing step for the EPOSAR data workflow

EPOS, the European Plate Observing System (<https://www.epos-eu.org/>), is a multidisciplinary, distributed research infrastructure that facilitates the integrated use of data, data products, and facilities from the solid Earth science community in Europe.

In the context of the EC OpenAIRE-Advance project, OpenAIRE and CNR-IREA, one of the Italian service providers of EPOS, collaborated to integrate its service EPOSAR in an open science publishing workflow with Zenodo.

EPOS scientists benefit from the on-demand EPOSAR service (accessible via the ESA's [Geohazards Exploitation Platform](#) (GEP)) that implements an advanced Differential Synthetic Aperture Radar (SAR) Interferometry (DInSAR) technique, referred to as Parallel Small Baseline Subset (P-SBAS) approach, to detect Earth surface displacements with sub-centimetre accuracy. The P-SBAS approach allows the generation of surface deformation time series and

velocity maps by processing in a parallel and efficient way SAR dataset consisting of tens or hundreds of images acquired over the same investigated area. EPOSAR outputs are extremely effective to investigate natural (earthquakes, volcanic unrests, landslides) and/or man-made (tunnelling excavations, aquifer exploitation, oil and gas storage and extraction, infrastructures monitoring) hazards.

However, they are often not shared as independent scientific products, provenance of datasets is not provided, and the experimental tools are not shared.

In order to provide scientists with an easy, integrated deposition option, CNR-IREA:

- Extended the EPOSAR data workflow with a step to let the user decide if to publish an experiment and relative results, thereby implementing a “on-demand publishing” workflow.
- Developed a new component for GEP capable of fetching experiment and dataset information from the local databases, package them as products as required by Zenodo and deposit them via APIs on behalf of the EPOSAR authorised service under the EPOS community. More specifically, for each experiment, two research products are deposited in Zenodo and linked to each other: one “research object” representing the experiment itself as a machine readable file and one “dataset” for the output displacement time series or the velocity maps.

The activity resulted in a software prototype (<https://github.com/dott109/EPOSAR>) and was presented at the RO2018 workshop at IEEE eScience Conference 2018 (<https://doi.org/10.5281/zenodo.1314671>).

User Interface: N/A
Protocol: Zenodo API
Metadata model: Datacite (see Zenodo API)

2. ARGOS DMP Publishing

Argos is an open and collaborative platform developed by OpenAIRE to facilitate Research Data Management (RDM) activities concerning the implementation of Data Management Plans. It uses OpenAIRE guides created by the RDM Task Force to familiarise users with basic RDM concepts and guide them throughout the process of writing DMPs. It also uses the OpenAIRE pool of services and other sources to make DMPs more dynamic in use and easier to be completed and published. Argos is based on the OpenDMP open source software and is available through the OpenAIRE Service catalogue and the EOSC.

Argos is integrated with several 3rd party services, one of which is an external AAI (OpenAIRE AAI currently), as well as vocabulary services.

In its Data Management Plan production workflow it includes one step where the Data Management Plan is deposited into a repository, optionally obtaining a PID from this deposition. This step currently supports the API of Zenodo. It supports SSO using the account of the user

on Zenodo for publishing the DMP, however a “service” account may be also used for publishing datasets, if a case requires so, however this is not a preferred strategy. To publish a dataset, it may be formed in either a human readable form (e.g. PDF) or in a machine readable form (e.g. the maDMP specification issued by the DMP working group of RDA).

Currently Argos is offered as a free service from OpenAIRE and is adopted by several projects to produce their Data Management Plans.

User Interface: Web user interface as dedicated web app;
Protocol: **[Consuming]** Zenodo API (for pushing DMPs to repositories); OpenID Connect (SSO); Simple REST for vocabulary harvesting; [TBC: OAI PMH]
[Serving] Proprietary REST API for DMP manipulation; RDA maDMP json, proprietary XML and OpenXML for DMP exchange;
Metadata model: Zenodo metadata; RDA maDMP (for DMP description);

3. NEANIAS Zenodo Publishing Bridge

NEANIAS (Novel EOSC Services for Emerging Atmosphere, Underwater & Space Challenges) is an ambitious project that comprehensively addresses the challenges set out in the ‘Roadmap for EOSC’ foreseen actions. NEANIAS promotes Open Science practices and plays an active role in the materialisation of the EOSC ecosystem by efficiently engaging large scientific and professional communities; actively contributing to the technological, procedural, strategic and business development of EOSC. NEANIAS drives the co-design and delivery of innovative thematic services, derived from state-of-the-art research assets and practices in three major sectors: Underwater research, Atmospheric research and Space research, each engaging multitudinous academic and business groups, numerous researchers, professionals and governmental entities.

As a result NEANIAS produces several thematic services (more than 9 in total) that produce or visualise scientific datasets. For those that produce data, the approach of publishing those onto a well known repository, compliant with Open Science and FAIR principles has been promoted by the project technical management.

In order to accomplish this in a uniform manner across different services, an integration component, currently labelled “NEANIAS Zenodo Publishing Bridge” has been designed and implemented, currently in its prototypical form, that mediates among NEANIAS technologies (storage and metadata descriptors) and Zenodo for simplifying the publication of datasets onto Zenodo. It creates the information package required to publish onto Zenodo API, presents the user options for curating the metadata, acquires data from NEANIAS storage and finally invokes Zenodo API for the depositing of the dataset and its metadata.

User Interface: HTTP redirect calls for services to invoke. Small web UI for metadata filling;

Protocol: **[Consuming]** Zenodo API (for pushing datasets and metadata to repositories); OpenID Connect (SSO); WebDAV or simple HTTP to obtain data files; Proprietary XML for research product description;

Metadata model: Zenodo metadata;

4. Dryad-Zenodo Integration

The Dryad Digital Repository is a curated resource that makes research data discoverable, freely reusable, and citable. Dryad provides a general-purpose home for a wide diversity of data types. Dryad originated from an initiative among a group of leading journals and scientific societies to adopt a joint data archiving policy (JDAP) for their publications, and the recognition that open, easy-to-use, not-for-profit, community-governed data infrastructure was needed to support such a policy.

Dryad is well integrated with editorial workflows (publishers) but it is covering only the data objects. Other research outputs need to be archived somewhere else. That’s why an integration between Dryad and Zenodo made it easy for authors to quickly upload their data, software, and other supplementary material, through Dryad to Zenodo, all while in the same process of uploading the publication to the publisher.



Upload Your Files

You may upload data via two mechanisms: directly from your computer, or from a URL on an external server (e.g., Box, Dropbox, AWS, lab server). We do not recommend using Google Drive.

We require that you include a README file to provide key information for understanding and using your data.

Software and Supplemental Information can be uploaded for publication at [Zenodo](#). You will have the opportunity to choose a separate license for your software on the review page.

Files

Filename	Status	URL	Type	Size	Actions
2021-03-29_report.csv	New		data	17.01 MB	Remove
dataset-analysis-output_reports.Rmd	New		software	4.35 KB	Remove
MakeDataCount_8.jpg	New		supp	138.10 KB	Remove

Researchers submitting data for curation and publication at Dryad now have the option to upload code, scripts, and software packages on a new tab “Upload Software”. Anything uploaded here will be sent directly to Zenodo. Researchers will also have the opportunity to select the proper licence for their software, as opposed to Dryad’s CC0 licence. The Dryad upload form includes an option to upload code files that will be triaged and sent to Zenodo.

Protocol: Zenodo API
Metadata model: Datacite (see Zenodo API)

5. Software Heritage-Zenodo integration

Software Heritage (SWH) is the universal software archive that intends to collect and preserve all existing software in source code form. Through a SLOAN grant, Software Heritage funded the development, by Cottage Labs of an integration between Zenodo (via InvenioRDM) and Software Heritage archive. CottageLabs has extensive expertise in InvenioRDM and also in designing [SWORD](#) protocol. This protocol will be the base connector that will allow all instances of InvenioRDM to safely and efficiently archive in Software Heritage the source code of all research projects that will be deposited in them.

The usual workflow will be: researchers will deposit software in Zenodo or any other InvenioRDM-based repositories, and those depositions will be sent to SWH for archival. A bidirectional linking will be in place between the repository and SWH.

Protocol: SWORD

6. ROHub-[Zenodo/B2Share] integration for publication of ro-crate snapshots

ROHub is a holistic solution for the storage, lifecycle management, publishing and preservation of scientific investigations and campaigns via research objects, exportable as RO-crates, and in line with FAIR principles and EOSC Semantic Interoperability Framework, A research object aggregates and describes all the resources (internal/external) and context entities related to the investigation. ROHub makes these resources available to others, allows publishing and releasing them through a DOI, and allows the discovery and reuse of pre-existing scientific knowledge. ROHub serves as a backbone to a wealth of RO-centric applications and interfaces across different scientific communities.

ROHub support scientists throughout the research lifecycle to, among others, (i) create and maintain high-quality ROs that can be interpreted and reproduced ; (ii) to collaborate along this process; (iii) to publish and search these objects and their associated metadata; (iv) to semantically enrich the RO metadata with natural language processing techniques enhancing RO visibility and discoverability; (v) to manage their evolution; (vi) to keep account of its social impact; and (vii) to monitor and preserve them supporting their accessibility and reusability.

ROHub adopts RO-crate as the default standard format for exporting/exchanging research objects, treats data cubes as first class entities to facilitate access and discovery of Earth Observational (EO) data, and is integrated with other EOSC services, including EOSC AAI (EGI

check-in), to enable SSO across different EOSC services, publication of RO-crate snapshots in other EOSC repositories, particularly Zenodo and B2Share, and is currently being integrated with B2Drop as an alternative storage of aggregated resources. ROHub includes a backend service, exposing a RESTful API, a reference Web portal and python library facilitating its use under EOSC notebooks.

ROHub enables the publication of RO-crate snapshots in Zenodo and B2Share, for a wider dissemination of these objects. This publication process can request and assign a DOI to the research object directly to these repository services, or it can request ROHub to provide such DOI and include it in the Zenodo/B2Share entry.

Research Object editing

ABOUT ACTIVITY EXPLORE MY ROS PEOPLE SUPPORT

Basic information People & organizations Tags Sketches gallery Resources Related locations License & Funding Advanced metadata

Basic information

Title

Analysis of MOD_Aqua mass concentration chlorophyll concentration in sea water

Description

This Research Object aggregates the resources associated with the analysis of MOD_Aqua mass concentration chlorophyll concentration in sea water

Publication services

No Needed B2SHARE ZENODO

Create DOI

Snapshot & Exit Reset form Snapshot & Go to overview

User interface: Web portal, Python library (e.g., for Jupyter Notebooks)

Protocol: http (REST) ROHub API, Zenodo API, B2Share API, Research Objects (<https://www.researchobject.org/>)

Metadata model: ro-crate <https://www.researchobject.org/ro-crate/> reliance ro-crate profile:<https://reliance-eosc.github.io/reliance-ro-crate/> (and mappings to Zenodo metadata and b2share metadata)

7. HAL to Software Heritage publishing

Part of the Ministry's research infrastructure roadmap and of the National Plan for Open Science, HAL is the multi-disciplinary open archive chosen by the whole French scientific and university community for the dissemination of knowledge. It is both a unique repository and an application, a shared platform for institutional archives, disciplinary repository and electronic theses.

Created by Inria in 2015, Software Heritage aims to collect, preserve and share all softwares publicly available in source code form, with the objective of building a common and shared infrastructure at the service of industry, research, culture and society as a whole.

The opening of HAL repository to softwares is the result of a CCSD-Inria-SWH partnership. As of 2018, HAL and Software Heritage have been interconnected: when the researcher deposits the source code of his software in HAL, it is automatically transferred to Software Heritage; the HAL deposit is then updated with the archive identifier (SWHid).

Protocol: SWORD

8. Publishing Articles from Episciences overlay journals to open repositories (arXiv, HAL, Zenodo)

Episciences.org is a hosting platform for open access journals. It is developed and managed by the CCSD (Centre pour la Communication Scientifique Directe). The originality of the Episciences.org platform is that it is based on contents deposited in an open archive for creating overlay-journals. The Episciences.org platform allows the editorial process: submission of manuscripts, peer reviewing management, publication. The submission of a manuscript to a journal hosted by Episciences.org is made through a deposit in an open archive. It is freely available and therefore accessible by all, whether accepted or not by the journal.

Once the article is accepted by the overlay-journal, the open archive used for manuscript submission is updated with the publication references.

The first step is always to submit the preprint to a compatible repository. When the preprint is online it can be submitted to a journal. The submission process involves retrieving metadata from the repository. Articles Metadata are then imported from the open archives to the journal, with the OAI-PMH protocol and several APIs offered by the repositories. The files are always hosted by the repositories. When a file is consulted on a journal, the PDF file is downloaded from the repository, but it is not noticeable by the end user.

9. Publishing and reviewing preprint hosted on HAL

The CCSD is implementing the COAR notify project to allow Peer Community In (PCI, a free recommendation process of scientific preprints based on peer reviews) and Episciences to notify the HAL repository that an article has been published by an Episciences journal or recommended by a PCI. Implementing this protocol will also allow researchers to submit their preprint form HAL to PCI and an overlay journal hosted by Episciences. In the future other preprint reviewing platforms and publishers will be able to communicate with HAL using the

same Notify Protocol. The mechanism also leverages the pattern or “protocol” [Signposting the Scholarly Web](#). It is used to advertise the location of several resources or metadata formats in order to automate the processes.

This implementation is funded by a grant from the French National Fund for Open Science for a project called ([HAL Open Workflows and INteroperability](#)). The software is in a [production phase](#) with HAL and Episciences. HAL is also ready to receive notifications from PCI, as soon as the protocol is deployed by PCI in production (2023). Signposting the Scholarly Web has also been deployed in production for HAL and Episciences.

10. OpenCitations Figshare uploader

OpenCitations stores all its citation data in Figshare (https://figshare.com/authors/OpenCitations_/3068259). The process developed for the production of new data creates zipped archives of gigabytes of data, that were then uploaded in Figshare by hand. Since the amount of uploads increased during the previous years, a script has been developed to automate the creation and upload of new Figshare items containing OpenCitations data.

Such an upload is performed by using the Figshare HTTP API (<https://docs.figshare.com/>) within a Python code. The necessary input that should be provided is a Figshare token (<https://help.figshare.com/article/how-to-get-a-personal-token>) that allows the use of their API programmatically. Currently, the script is only adopted for internal use, but it will be updated, extended, and then published for future consumption by the community.

User interface: Python script

Protocol: HTTP(S) (REST) Figshare HTTP API (<https://docs.figshare.com/>)

Metadata model: OpenCitations Data Model (OCDM, <http://opencitations.net/model>)

11. The CS3mesh4EOSC dataset selection, metadata annotation & repository deposit mechanism

The ScienceMesh (i.e., the platform being built by the [CS3mesh4EOSC](#) project) is built by meshing together existing national, uni cluster, or domain specific synch/share sites; e.g., [SURFdrive](#), [SwitchDrive](#), [CERNbox](#) – many of these have been in operation for close to a decade. As a rule, these nodes were built as low barrier of entry facilities, meant to provide a catchment area for data that might otherwise leak away to storage facilities outside of policy control (e.g., commodity commercial cloud services).

The consequence of several years of providing a low-barrier-of-entry store for research data is that these nodes have amassed Petabytes of data of various degrees of structure. The spectrum ranges from some volumes looking like research scratch spaces, to others that contain domain level reference data sets of excellent metadata quality. This leaves a broad middle ground of many PB of valuable research data, amassed by researchers who are

non-experts in the field of data structure, or FAIR in general – a truly excellent opportunity to engage in FAIR advocacy with these users. The task at hand for the CS3mesh4EOSC project, then, is to deliver tooling that exposes the maximum amount of these researchers to FAIR practices, and does it in a way that helps them annotate / metadata-enrich the maximum amount of their data and deposit it in appropriate repositories. Our first iteration of such tooling is based on the [PARADISEC](#) project's [Describo](#) data annotation and packaging tool, which packages data as ROcrate, and the [ScieboRDS](#) research data management tool, which allows researchers to define projects, list PI/CIs, and finally deposit project ROcrates into nominated repositories. It is ScieboRDS which in turn invokes Describo to take care of bundling the ROcrate'ed data into the project. The deposit mechanism currently supports Zenodo and OpenScienceFramework, with others under development.

For a video of ScieboRDS in action, see <https://cds.cern.ch/record/2750522>

For a video of Describo and its architecture, see <https://cds.cern.ch/record/2750519>

User Interface:	Describo web user interface (presented inside ownCloud/NextCloud)
Protocol:	Zenodo API (for pushing ROcrate to Zenodo); OSF API (for pushing ROcrate to OpenScienceFramework) SAML2 or OpenID Connect (Edugain-based SSO); cs3api for intra-service signalling
Metadata model:	ROcrate (with custom user-selectable schema, defaulting to schema.org)

12. Overleaf

Overleaf (<https://www.overleaf.com>) is a service for collaborative editing of LaTeX documents. It thus works best for scientific users familiar with LaTeX editing, although it does provide a rich text editor style interface in addition to the LaTeX code mode.

A large selection of LaTeX templates are available to choose from, including the style files of many scientific journals.

Documents can be shared among users in both read-only and editable mode, either by assigning registered users or sharing an access URL.

Like many systems of this kind, Overleaf is a freemium service with several levels of licensing from free to “professional” and institutional subscriptions. More advanced sharing features and version tracking are available only to paying subscribers. The added features in the paid subscriptions also include Dropbox, Git, Mendeley and Zotero integration. A document in these versions works as a Git repository (except for branching) that can be synced with Github and with users' local repositories.

In use cases where the only requirement is to be able to write a scientific paper collaboratively, Overleaf is an interesting option to consider if the users are familiar with LaTeX, especially if their institutes have institutional subscriptions.

User interface: Web, git

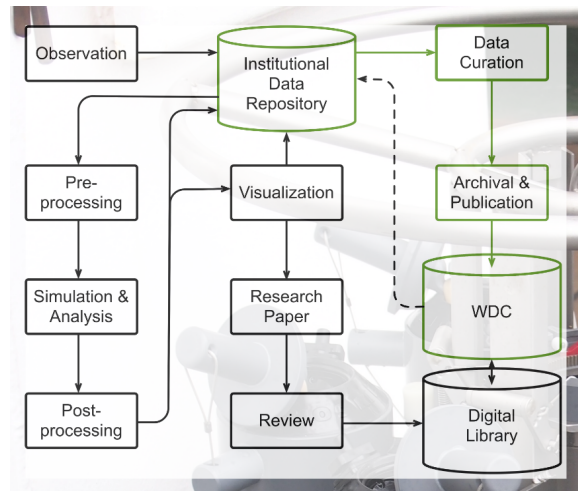
Protocols: https, OAuth2 (login with Google and OrcID accounts), git

13. PubFlow

PubFlow invents new approaches and a pilot application to work with scientific data in scientific workflows to increase the productivity in scientific work. PubFlow provides a data publication framework for scientific data, built on top of proven business workflow technologies such as BPEL and BPMN.

More information available at

- <https://www.pubflow.uni-kiel.de/en>
- <https://www.usenix.org/conference/tapp2014/agenda/presentation/brauer>
- <https://oceanrep.geomar.de/id/eprint/21112/7/Poster.pdf>



14. “Data literacy”, the Pierre Auger Observatory Open Data release example

“Data literacy” could help to fill the gap of non-existing automated system to harvest and transmit metadata explaining to users how to fill in metadata on requests. An instrument could be a pre-compiled Jupyter notebooks, containing both instructions and codes, to enable users to fill in data and to perform command to further exchange or deposit them. This could be useful as users in some cases, as humanities and social sciences, do not have the habit to deal with metadata, they do not even know what metadata are and most of the time they lack local support at university for using tools already at disposal (for instance: all the EUDAT connected services

<https://eudat.eu/events/webinar/webinar-introducing-dice-and-eudat-services-for-the-research-data-lifecycle>).

The use of Jupyter notebook is recently spreading in different platforms for enabling users to make autonomously some analyses on dataset extracted, an example could be the “Pierre Auger Observatory Open Data release”, a tool enabling users not only to download data, but by offering open education Jupyter notebooks to help users to do some autonomous analyses with

those data. This is possible by a dedicated website hosting a framework to help people to work with data with all explanations and a set of simple analysis on Python notebook (a couple of tutorials on how to use the data and five notebooks as an example reproducing results to enable people to work with data).

A similar support could be produced to help researchers to complete the filling, or formatting of metadata and to proceed with further metadata sharing and deposit tasks.

References:

A video with a brief presentation of the project “Lightning Talk: the Pierre Auger Observatory Open Data release: not only data”, recorded on 10th September 2021 for the OAI12 – The Geneva Workshop on Innovations in Scholarly Communication held virtually from 6-10 September 2021.

The introduction to the author and the talk is at disposal at: <https://oai.events/etn-speaker/dr-xavier-bertou/>

The video recorded on 10th September 2021, starting at the time 1 hour and 24 titled “The Pierre Auger Observatory Open Data release: not only data”, at disposal at: <https://oai.events/oai12/replay/>.

15. Publishing articles with relations to Software artefacts (Episciences + Software Heritage + HAL)

In the scope of the FAIRCORE4EOSC project, Episciences will accept SWHID references on its platform to import the software metadata hosted by Software Heritage and HAL (end of 2023).

This will enable the submission of code along the submitted preprints.

If a preprint is accepted and published by a journal, the software metadata will be displayed on the landing page of the article. At the same time, a notification using COAR Notify will be sent to HAL and/or Software Heritage to notify the repository that a published article references a software artefact.

Details of the activities can be followed at

<https://github.com/CCSDForge/episciences/issues/332>

<https://github.com/CCSDForge/episciences/issues/326>

<https://github.com/CCSDForge/episciences/issues/327>

Scenarios for the interoperability guidelines for research product publishing

The WG went through a brainstorming session for the identification of scenarios, even beyond those implemented by the use cases reported for the landscape study, that would benefit from an interoperability framework for research product deposition. The session, plus the collaboration for the EOSC Future cross-infrastructure use cases on the Data Transfer Service, resulted in the five scenarios depicted in figure 1.

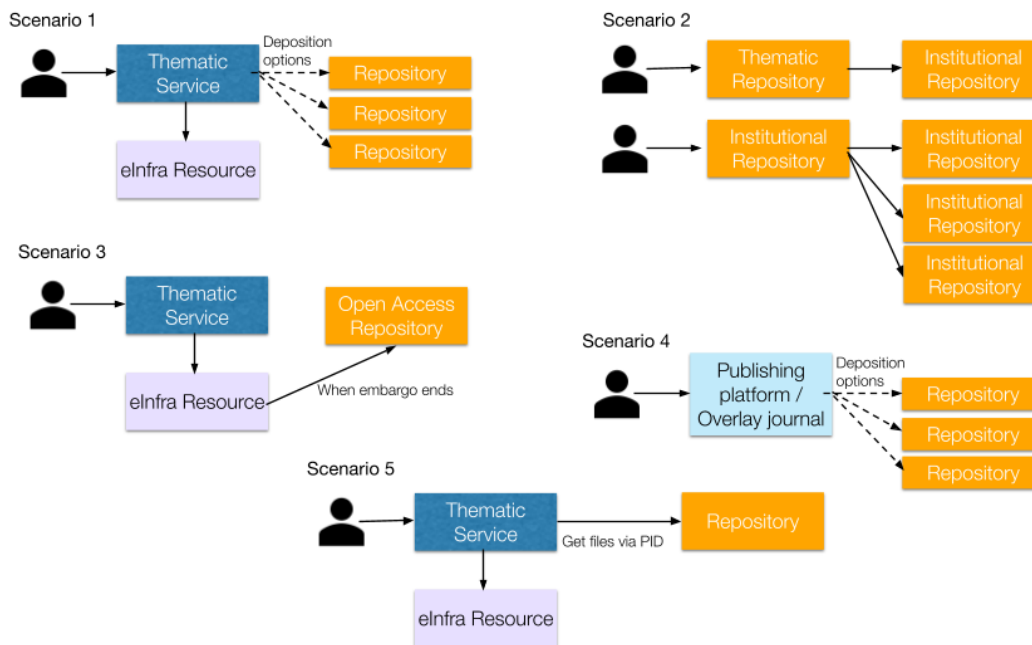


Figure 1. The five scenarios for the EOSC IF on research product publishing

1. **Scenario 1 - thematic service and deposition options**

A researcher uses a thematic service and produces a research product (e.g. a dataset) that is stored in a storage resource of the underlying e-infrastructure. The thematic service can propose to the researcher several deposition options, targeting repositories that are known to be compatible with the EOSC Interoperability Framework for research product publishing.

2. **Scenario 2 - Automatic deposition into repositories of reference**

A researcher deposits a research product (e.g. a publication) on a thematic repository. By knowing the affiliation of the researcher, the thematic repository can automatically deposit the publication also in the institutional repository of reference. In a similar way, if the researcher deposits the publication in the institutional repository and provides information about the co-authors, the repository can automatically deposit the publication in the institutional repositories of the co-authors.

3. **Scenario 3 - Automatic Open Access deposition from e-infra service**

A researcher uses a thematic service and produces a research product (e.g. a dataset) that is stored in a storage resource of the underlying e-infrastructures. The researcher wants to deposit the product on an Open Access repository, but has to wait for an embargo period. The storage resource, based on a configuration provided by the researcher, can automatically deposit the product on an Open Access repository when the embargo period ends (so the researcher does not need to remember that the asset prepared months before can now be made available in Open Access).

4. **Scenario 4 - Seamless submission process via publishing platform and overlay journals**

Overlay publishing platforms and overlay journals typically suggest their users to deposit their research assets in repositories and get back with the identifier/PID/URI of the deposition so as to proceed with the publication workflow. With the EOSC IF for research product publishing, the platforms and overlay journals could offer their users an integrated experience.

5. **Scenario 5 - Automated access to files of a deposition via its PID**

A researcher uses a thematic service to run analysis on a dataset available on a repository. Instead of downloading the dataset files from the repository and uploading them to a storage resource of the e-infrastructure, the researcher gives the dataset's PID as input to the thematic service, which can get the files and store them where they can be analysed.