D8.1 EOSC-Life Methodology framework to enhance reproducibility within EOSC-Life – revised version

Florence Bietrix, José-Maria Carazo, Salvador Capella-Gutierrez, Frederik Coppens, Maria-Luisa Chiusano, Romain David, Maria Fernandez Jose, Maddalena Fratelli, Jean-Karim Heriche, Carole Goble, et al.

EOSC-Life

EOSC-Life: Building a digital space for the life sciences

D8.1 – Methodology framework to enhance reproducibility within EOSC-Life – revised version

WP8– International Impact, Innovation and Sustainability
Lead Beneficiary: INFRAFRONTIER and HU
WP leader: Michael Raess and Jing Tang
Contributing partner(s): CSIC, BSC, VIB, EMBRC, ERINHA, IRFMN, EMBL-HD, UNIMAN, Fraunhofer, BBMRI-ERIC, NKI, EMBL-EBI, FZJ, CSR4, IMTM, INFRAFRONTIER, HU

Authors of this deliverable: **Florence Bietrix**, **José Maria Carazo, Salvador Capella-Gutierrez, Frederik Coppens, Maria Luisa Chiusano, Romain David, Jose Maria Fernandez, Maddalena Fratelli, Jean-Karim Heriche , Carole Goble, Philip Gribbon, Petr Holub, Robbie P. Joosten, Simone Leo, Stuart Owen, Helen Parkinson, Roland Pieruschka, Luca Pireddu, Luca Porcu, Michael Raess, Laura Rodriguez-Navas, Gary Saunders, Andreas Scherer, Stian Soiland-Reyes, Jing Tang**

Contractual delivery date: **30 September 2022**
Actual delivery date: **30 September 2022**
H2020-INFRAEOSC-2018-2

Grant agreement no. 824087
Horizon 2020
Type of action: RIA

Table of Contents

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Executive Summary

The original scope of task 8.3 is to develop a framework to assess the impact on reproducibility of the availability of life-science open data and workflows in the cloud.

Such a framework should include the understanding of the tools and services produced and made available, the gaps they are filling up and their expected impact.

A great part of the activities within EOSC-Life is actually related to reproducibility and provides in several ways tools that will have an impact. We describe here such activities and explain why and how they will have an impact on reproducibility in life sciences. In addition, for each action, we provide an indication of how we could measure its impact in practical ways, using adequate proxies. For these reasons, we changed the title of the deliverable, from “framework to assess ...” to “framework to enhance reproducibility”.

First of all, we reasoned on what can be the contribution of open science to improve the reproducibility of research. Publicly sharing data, protocols, tools and computational workflows makes it possible to compare or combine the data and outcomes from different studies within a discipline as well as integrate data across scientific domains. It allows conclusions to be validated and possibly corrected as well as being reinforced by meta-analyses. Replication data and test/training data can also be used in many applications to contribute to reproducible research. Moreover, new hypotheses, different from the original aims of the study, can be explored. Datasets can be re-used to develop and test new methods, to conduct scientific and technical benchmarking activities and to support training activities. Therefore, in addition to generating more value from research investments, data sharing has the potential to increase confidence in research outcomes and increase knowledge dissemination. These benefits of open sharing have long been recognized in some fields such as bioinformatics, which has a long history of publicly sharing data with, for example, public repositories for nucleotide sequences going back 30 years and the Protein Data Bank (PDB), a repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies, that celebrates its 50th birthday this year.

In this notion, any improvement in sharing of data, tools and workflows among scientists and across disciplines, which is the aim of EOSC-Life and the wider EOSC, will contribute to reproducible science. In addition to this general scope, several specific actions to frame transparency in the reporting of experimental protocols, data and analytical workflows warrant the reproducibility of every single object (experimental results, data, or workflows) that is made available on the cloud.

We describe here the initiatives in EOSC-Life to implement existing tools for reproducibility as well as to develop new tools for its enhancement. As the final goal of EOSC-Life is to make data resources available to the wider community of life scientists, although necessarily technical in several points, this document aims at a general readership, including experimental in addition to data scientists.

This project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 824087.

Project Objectives

With this deliverable, the project has reached/contributed to the following objectives:

a. Develop metrics to assess impact of life-science open data in the cloud on data reproducibility.

Detailed Report on the Deliverable

1. Reproducibility definition(s)

It is important to note that reproducibility is not an absolute, all or none, concept, but a spectrum. In fact, there is a wide range of types of actions that can be applied in different combinations and with different degrees of accuracy in different situations. In addition, different types of experiments may need to aim at different levels of reproducibility. Indeed, the purpose of research may be roughly dissected into two fields: an exploratory mode of research, which aims at generating scientific hypotheses that can tolerate a certain level of uncertainty, and confirmatory mode of research, aiming at demonstrating such hypotheses, generating evidence that enables decisions and/or builds confidence. Although the request of reproducibility standards should be the same in both cases, the uncertainty and imprecision are usually greater in exploratory studies. Therefore, a complete documentation may not be requested for exploratory studies.

Experts in different fields use different terminologies to define reproducibility, sometimes in conflicting ways. For instance, computational disciplines and social/life science domains have sometimes different understanding of the same terms [1].

We will adopt here the definition that has recently been suggested by the Committee on Reproducibility and Replicability in Science [2].

- "Reproducibility is obtaining consistent results using the same input data, computational steps, methods, and code, and conditions of analysis. This definition is synonymous with computational reproducibility"
- "Replicability is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data"

These definitions reflect the notions more often used in computational disciplines and were adopted also by the scoping report on reproducibility of scientific results in the EU, recently released by the European Commission [3]. However, it is not always easy to apply these definitions in the experimental setting. As reported in Appendix-1, experimental reproducibility is defined based on four different features: methods, results, inferential and external reproducibility.

Moreover, it is important to note that other authors, in the area of life sciences, use quite the opposite interpretation from the one adopted here and this may be a cause of confusion.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

For instance, Plesser et al [4] report the following:

- "Repeatability (Same team, same experimental setup): The measurement can be obtained with stated precision by the same team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same location on multiple trials. For computational experiments, this means that a researcher can reliably repeat her own computation."
- "Replicability (Different team, same experimental setup): The measurement can be obtained with stated precision by a different team using the same measurement procedure, the same measuring system, under the same operating conditions, in the same or a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using the author's own artifacts."
- "Reproducibility (Different team, different experimental setup): The measurement can be obtained with stated precision by a different team, a different measuring system, in a different location on multiple trials. For computational experiments, this means that an independent group can obtain the same result using artifacts which they develop completely independently."

A similar interpretation is used by Drummond [5], where replicability, intended as the reproducibility concept adopted by the present document, is considered not necessarily a primary goal in science.

2. Reproducibility requirements

The requirements for the results of an experiment or an analysis to withstand scrutiny are the subject of extensive studies and debates [6], which have led to many position documents and recommendations.

As a very general outline, prerequisites of reproducibility of life science experiments are availability of raw data, and detailed description of methodology and metadata. Transparency and access to data allow evaluation of the results through re-analyses and tests of reproducibility. Metadata and protocols are essential for the process of understanding the experiment design and performance during the pre-analytical (e.g. animal caretaking, cell culture, sample storage, sample processing, reagents, patient selection criteria, personnel) and the experiment phase (e.g. design, handling of samples and reagents during an experiment, data collection, laboratory personnel). Experiments are life cycles of individual steps along which quality control checks are essential.

For a detailed account of the guidelines for designing and reporting experiments, which is not the scope of the present document, we refer to the EQUATOR network [7] and to FAIRsharing, a comprehensive data and metadata standards resource [8,9]. In the following chapter we will report on tools favouring compliance to FAIR principles for experiments (in particular for COVID-19 drug repurposing assays and chemosensitivity assays) and on the development of provenance standards for biological materials, data generation and data processing within EOSC-Life.

In computational research, the same general concepts of transparency and access described above, apply, but with several specificities. Complete reporting is the first stage necessary to ensure reproducibility in computational research [10]. Checklists are available to verify that the

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

reporting process is exhaustive [11–13]. The availability of a complete software development environment, processed data and computational scripts is the second stage necessary to ensure reproducibility of the results. The two-stage procedure previously described needs facilities and standardization [14]:

a. Common standards for sharing tools and documentation (guidelines and software support, product configuration and customization, installation and update procedures, maintenance pack information, archived versions for previous releases). At the moment the landscape is complex, with a number of individual solutions as well as a number of community-specific tools. We will describe in the next chapter several contributions of EOSC-Life in using, improving and developing bio.tools, bioschemas tools specifications, package managers and container registries.
b. Availability of data with proper metadata and curation. It is important to note that not all datasets can be shared publicly for reasons of privacy and copyright. In some particular cases, for instance, the data sets used in the process are of sensitive nature and, therefore, need access-control. If appropriate infrastructures are used, e.g. European Genome-phenome Archive (EGA) for human sensitive data, repetition would be possible after gaining access to data sets of interest. A guide to the sharing of genomic and health-related data is provided by the Global Alliance for Genomics and Health (GA4GH) 15. We will describe in the next chapter an example of a workflow specification for analysis of undisclosable data developed in EOSC-Life.

3. Reproducibility-related activities within EOSC-Life

As anticipated above, several activities are ongoing within EOSC-Life to support the scientific community in improving reproducibility and replicability. Many of the described efforts are necessarily applied in a specific field of interest, and they are used here to provide concrete examples. However specific, they set the way for addressing similar needs in other disciplines.

3.1. Replicability in COVID-19 drug repurposing (WP1, D1)

The Demonstrator project D1, supported by WP1, covered integration of Chemical and structural biology data with a focus on deployment of FAIRified fragment and small molecule screening data sets into public repositories. During 2020, workflows established in D1 were adapted to support COVID-19 drug repurposing screening studies originating from EOSC RI partners and the wider scientific community. A key issue facing users of COVID-19 repurposing data sets is the limited degree of replicability between phenotypic assay readouts (Figure 1). Although screening efforts to identify anti-viral phenotypes typically profile the same finite collections of marketed clinical stage compounds, a wide variety of assay conditions (endpoints, readouts, cell models, virus MOI's, time of exposure etc.,) were used in each screening protocol. This lack of community-level standardisation in assay prosecution has contributed to non-overlapping populations of hits being reported (Figure 1). The availability of complete and accurate assay metadata is therefore essential to allow for the correct analysis and interpretation of results from across multiple COVID-19 studies. Appropriate alignment of results by cell model (eg VERO-E6 or Calu-3) or

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

readout (eg; qPCR or cell viability marker), is necessary to interpret the data and drive decisions related to progressions of compounds towards clinical studies.

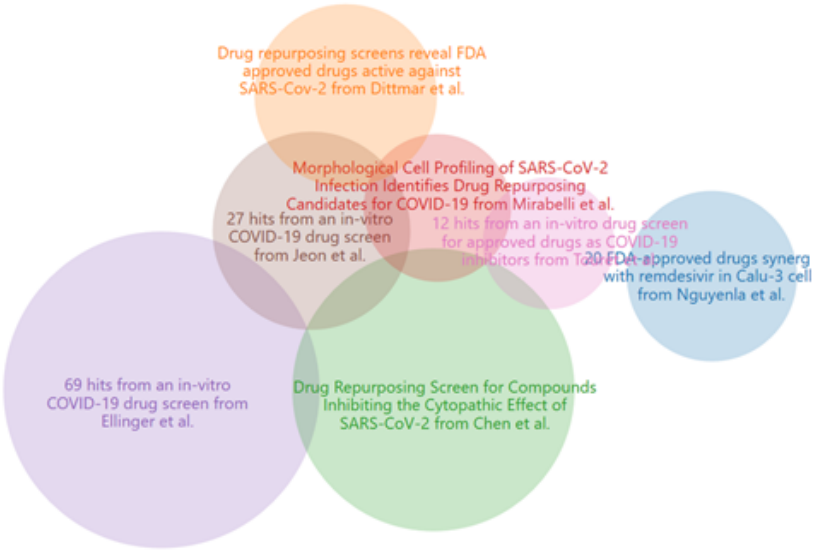

Figure 1: Venn diagram showing limited overlaps in the identity of putative anti-viral compound hits from seven SARS-CoV-2 anti-viral phenotypic screens[1].

In D1, replicability issues were addressed by elevating the FAIR status of general chemical biology analysis workflows and specific SARS-CoV-2 compound screening data within the ECBD, and chEMBL. In the European Chemical Biology Database (ECBD) compounds/targets descriptions were established based on MIABE [16] principles including a strong emphasis on the description of the process (methodology and instrumentation). The ontologies and vocabularies adopted included: BioAssay ontology; BRENDA; Cellosaurus; NCBI Taxonomy; Reactome Pathway Ontology (biological pathways) and Units of Measurement Ontology. The InChI/InChIkey was used for compounds and UniProt IDs/ChEMBL IDs for targets whilst NCBI Tax IDs was used for organisms. ECBD data access is facilitated Web UI for data upload, database dump via PostgreSQL and a REST API. SARS-CoV2 datasets related to the repurposing screens against the key viral protease 3CL-Pro and PL-Pro are now in the process of being curated for deposition into the ECBD. These will be used to confirm the validity of these targets in future anti-viral projects. In ChEMBL the aims of the D1 associated SARS-CoV2 curation and FAIRification were to compile a list of drugs that might be possible candidates for repurposing in COVID-19. The sources of data were drugs targeting proteins identified as important in SARS-CoV-2 infection; drugs in current clinical trials for COVID-19; drugs active in cell-based assays for SARS-CoV-2 inhibitory activity; and other approved anti-inflammatory/immunomodulatory and anti-coagulant drugs. FAIR data curation also provided access to parallel cell viability data which is essential in differentiating between cytotoxic and anti-viral mediated response, thereby enhancing confidence in the replicability of observed effects. At present 9 phenotypic data sets have undergone the ChEMBL SARS-CoV-2 drug repurposing FAIRification workflow. Large scale analyses are now underway with these data to elicit how

[1] https://maayanlab.cloud/covid19/#nav-drugs-table

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

differences in assays (cell lines, assay type, SARS-CoV-2 strain etc) affect the overall replicability of drug repurposing in COVID-19.

3.2. FAIR chemosensitivity assays (WP1)

Furthermore, to facilitate the open data in translational medicine as well as to have a standardized protocol based on minimal information principles for the annotation of chemosensitivity experiments, EATRIS has initiated MICHA (Minimal Information for CHemosensitivity Assays) [17,18]. MICHA is an integrative pipeline to annotate chemosensitivity assays based on four major components, including 1) compounds, 2) samples, 3) reagents, and 4) data processing references as outlined in Figure 2.

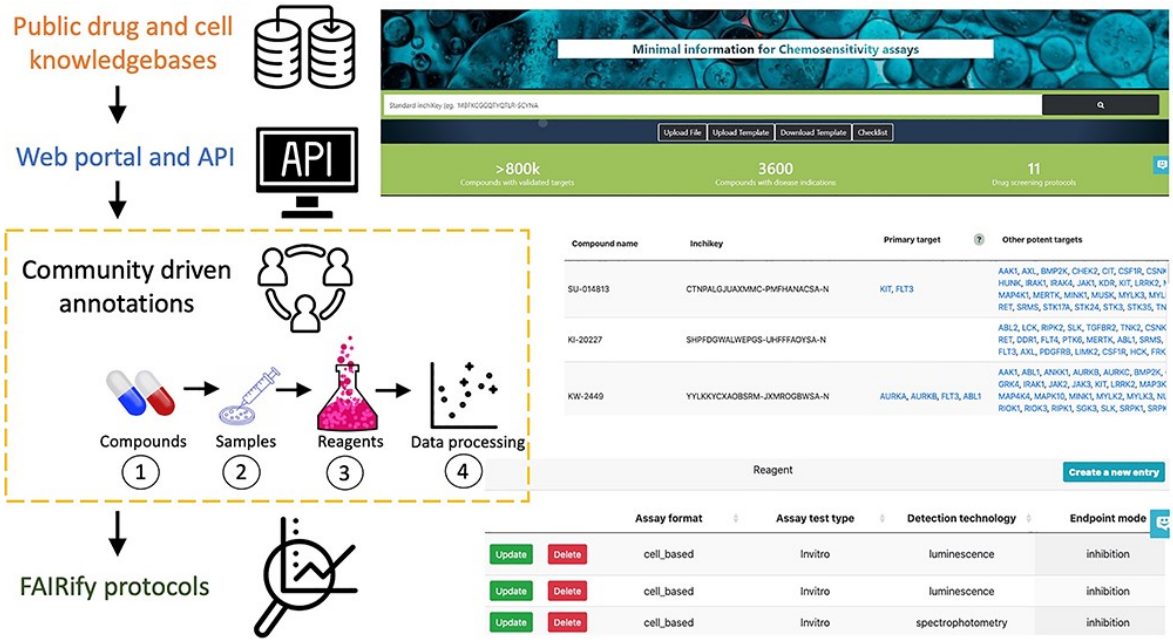

Figure 2: Graphical overview of MICHA.

Using the platform of MICHA, we aim to increase acceptance and adoption of the principles of FAIR (Findable, Accessible, Interoperable and Reusable), by making the assay annotation as smooth as possible with the help of data integration tools and databases. To consolidate the utility of MICHA, we provide FAIRified protocols from several major cancer drug screening studies (including CCLE, CTRP, GDSC), as well as recently conducted COVID-19 drug screening studies. With the integrative webserver and database, we envisage a wider adoption of the MICHA strategy to foster a community-driven effort to improve the open access of chemosensitivity assays as well as consensus on annotation protocol for chemosensitivity experiments.

3.3. FAIR protein structures (WP1, OC2020)

In computational structural biology studies, the starting point is a set of experimentally derived macromolecular structure models. If these experimental structure models were generated in the

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

course of the study, community-supported publication requirements prescribe deposition of said models and the underlying experimental data in the Protein Data Bank (PDB) [19,20], if the models were pre-existing they are referred to by their databank identifier. This databank can be either the PDB or the PDB-REDO databank [21,22] which provides alternative interpretations of the experimental data used to construct PDB entries. Both databanks use the same identifiers.

This approach of data reporting seems robust but nevertheless has several weaknesses that hamper methods and computational reproducibility as the input data is poorly defined:

1. When a large set of structure models is used, a complete set of databank identifiers is in many cases not reported in favour of a more general description of structure model selection. Apart from the risk of this description being rather vague, it should be noted that new entries are added to the PDB and PDB-REDO in the order of 10 thousand entries per year. A much smaller number of entries is obsoleted every year. A date when the data was selected should be included to clarify the structure models that were considered.
2. The PDB is often seen as a (historical) archive of structure models as this was its original purpose [23], however PDB entries are actually updated in terms of metadata (i.e. model annotation) and a recent policy change now also allows changing the actual structure model (i.e. the atomic coordinates). As such, reference to a PDB entry by just its identifier is no longer enough, a version number should be included.
3. The PDB-REDO databank is a “living” resource in which entries are updated regularly to incorporate changes made to the underlying PDB entries, but more importantly to apply new methodological advances in structure model generation. This means that each PDB entry can lead to several PDB-REDO entry versions. However, practical limitations meant that older PDB-REDO entry versions were not stored.

WP1 Open Call 2020 project “PDB-REDO Cloud: FAIR protein structures with deep versioning for scientific reproducibility and data provenance tracking” will provide remedies to the weaknesses described. The aim is to provide the PDB-REDO databank on EOSC with a stack of all previous versions. Versions will have provenance records describing the underlying PDB data versions and the versions of the software used to create the model (the PDB-REDO software pipeline consists of over 50 programs, some of which are updated frequently).

A query interface will allow users to select a dataset based on the metadata that describes the PDB-REDO entries. The dataset will have a full description of which entries and versions were selected in a machine-readable format that can be enclosed with a computational study so that the underlying data can be recovered for a very long time. This also takes away the need for researchers to archive the structure models used in their study locally.

3.4. Reproducible workflows (WP2, D7)

The Collaboratory brings together key components essential for reproducible workflows: workflow Management Systems for their reproducible design and execution; a workflow registry that spans their native repositories for documentation-driven reproducibility; registries for tools, containers and workflows; and platforms for testing and monitoring for execution-driven reproducibility (Figure 3).

This project has received funding from the *European Union’s Horizon 2020 research and innovation programme* under grant agreement No 824087.

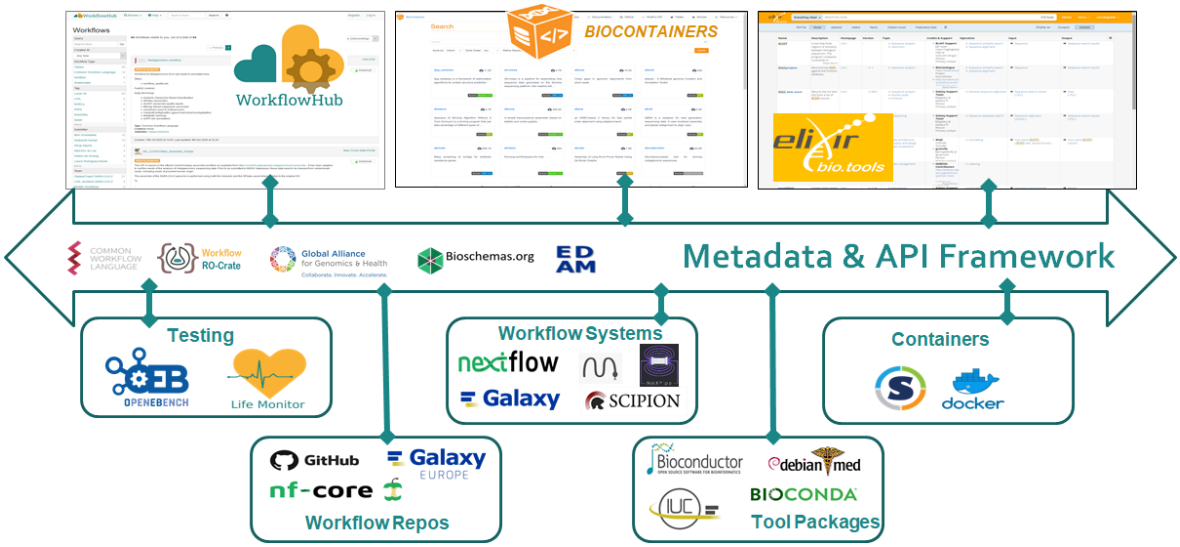

Figure 3: The Workflow and Tools Collaboratory framework

Within this framework, WP2 addresses reproducibility of computational workflows in four ways, detailed below:

1. installation and execution of software and workflows;
2. packaging of workflows;
3. specification of workflows;
4. FAIR+Reproducible (FAIR + R) workflows - description, documentation and registration.

3.4.1 Installation and execution of software and workflows

The reproducibility of a computational analysis depends on the ability to reproduce the computational environment in which it was executed. This is a complex task because of the need to replicate the tree of dependencies of the various top-level software, down to the specific software package versions and low-level libraries. Moreover, software installation in scientific computing environments can be further complicated by permission issues and the need to involve system administrators, not to mention the often cumbersome installation procedures of scientific software.

Software package managers have been developed to help deal with these issues. Among these, the Conda[2] package manager has become a popular way of building and installing scientific software across many different scientific fields, computing environments and systems, and in particular for life sciences the BioConda[3] repository of domain-specific software, which is supported by WP2, ELIXIR and the Galaxy project.

For a complete solution, EOSC-Life WP2 also promotes the use of containerisation as a complementary technology, which enables computational reproducibility by encapsulating software tools with their dependencies in ready-to-execute and easily distributed virtual environments, independent from the context in which they are deployed. Some containerization

[2] https://docs.conda.io/
[3] https://bioconda.github.io/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

runtimes, such as Singularity, can also be used without special system privileges, thus removing the need for the intervention of system administrators; important in HPC environments. Because different workflows use multiple software tools, and in different versions, reproducibility of computational workflows is improved significantly by workflow management systems that make use of reproducible software deployment systems such as conda and containers. Therefore EOSC-Life promotes the use of workflow management systems such as CWL, Galaxy, Snakemake and Nextflow which integrate Conda and containers; as well as packaging individual tools to facilitate such use. Through continuous integration, updates to tools in BioConda automatically trigger the build of containers, ensuring these are available for use by these management systems.

3.4.2 Packaging workflows

A workflow is defined and run within a particular context, typically to address a scientific question using specific data. To capture the contexts of workflows, WP2 is contributing to the community standard for research output packaging RO-Crate[4] [24], as well as developing its specialization Workflow RO-Crate, aimed at packaging workflow definitions with their documentation and support data. Workflow RO-Crate[5] is also being extended to include test specifications, which further support reproducibility by helping ensure the workflow is operating as expected within a given computing environment.

For this, WP2 is developing the LifeMonitor workflow testing service, which uses the testing metadata to monitor the correct functioning of workflows over the longer term to continually detect problems which could jeopardize their reuse both to reproduce previous results and to perform new analyses.

In addition, WP2 and the RO-Crate community is collaborating with ELIXIR Cloud and Authentication & Authorisation Infrastructure and GA4GH to formalize recording of workflow execution provenance in a Workflow Run RO-Crate; collaborating with developers of CWL, Nextflow. Snakemake and Galaxy workflow engines. In this way we are facilitating RO-Crate both for prospective provenance (a workflow definition that is ready to be executed) and retrospective provenance (a particular execution of that workflow).

3.4.3 Workflow specification

Workflow management systems are diverse and as a result executable workflows are expressed in a diversity of formats. However, this diversity hinders re-use and therefore reproducibility. A workflow management system agnostic description of workflows is therefore needed. In this context, EOSC-Life WP2 has selected the Common Workflow Language[6] (CWL) [25] as its standard for describing workflows across different workflow engines.

The WP also leads a community effort to define the ComputationalWorkflow Bioschemas profile[7] to describe the metadata about a workflow, which have been integrated into RO-Crate and WorkflowHub. This Bioschemas markup enables FAIR search and discovery, not just by the Hub, but also by search engines and other aggregators such as Google and OpenAIRE. WP2 have also

[4] https://w3id.org/ro/crate
[5] https://about.workflowhub.eu/Workflow-RO-Crate/
[6] https://www.commonwl.org/
[7] https://bioschemas.org/profiles/ComputationalWorkflow/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

been assisting Bioschemas in aligning with and maturing the related profile Computational Tool[8] used by bio.tools[9] - it is worth pointing out that both of these profiles are generic and not specific to bioinformatics.

3.4.3.1 A particular workflow specification for analysis of undisclosable data

The Demonstrator Project D7, which is strongly supported by WP2, has been focused on a scenario involving the analysis of sensitive human data, e.g. samples from rare diseases cases. D7 has been developed by BSC and CRG, including teams from EGA and CNAG, for scenarios where it is quite common to analyse experimental data which cannot be disclosed. The pipeline used for the demonstration purposes, provided by CNAG-CRG, is a variant calling pipeline that requires paired-end sequenced raw genomic data (in FASTQ format) and reference genome data. It runs a mapping and variant calling pipeline and in turn produces unannotated GVCF files, which can be further submitted to the RD-Connect GPAP portal or analysed on their own.

BSC has developed a high-level workflow execution service (WfExS[10]) backend which fulfils all the requirements of human data analysis scenarios. The first implementation iteration is supporting both CWL and Nextflow workflows. The workflow to be executed has to be available either in a Git repository or be findable at a GA4GH TRSv2 compatible service, which supports describing the workflows through RO-Crate. This fits nicely with the ongoing implementation of WP2 EOSC-Life WorkflowHub. In this particular scenario, the selected workflow is available at the GitHub repository[11] as well as in WorkflowHub, both CWL[12] and Nextflow[13] implementations of the same pipeline.

As illustrated in Figure 4, WfExS receives a high-level description of what has to be done. As inputs, it can receive both inline values and file-like parameters. The outputs are either proposed filenames for those outputs or wildcard patterns to match local filenames. Input file-like parameters are described through URIs, currently standard URLs, and in a future iteration some CURIE namespaces from identifiers.org / n2t.net will be supported, for instance EGA files/datasets. These inputs are downloaded and cached, whenever it is allowed, in order to avoid downloading the very same copy of the workflows or reference genomes, for instance.

If the workflow is fetched from a TRS endpoint, e.g. EOSC-Life WorkflowHub, RO-Crate semantic annotation of the workflow provides the repository of the workflow, as many workflows depend on additional resources, like profiles or subworkflows. If the workflow corresponds to one of the supported engines, the workflow is analysed, and an appropriate version of the workflow execution engine is also downloaded and installed, if it is needed.

Another pre-condition to be able to execute the workflow are the containers with the software used for the workflow steps. For replicability and reproducibility matters, all the workflow steps have to be based on public docker or singularity containers. Considering the kind of environments where scientific workflows are usually run, we have decided to use Singularity[14] runtime, which is

[8] https://bioschemas.org/profiles/ComputationalTool/0.5-DRAFT/
[9] http://bio.tools/
[10] https://github.com/inab/WfExS-backend
[11] https://github.com/inab/Wetlab2Variations/tree/eosc-life/
[12] https://workflowhub.eu/workflows/107
[13] https://workflowhub.eu/workflows/106
[14] https://sylabs.io/docs/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

HPC-friendly. Another reason is that any Docker container image can be pulled and re-assembled as one usable by Singularity. For the very same reasons, the preconditions materialization phase, which is where containers are fetched, is detached from the execution one, as HPC execution environments usually have restricted internet access.

Some features of the minimum viable product milestone we are pursuing: trusted, secure and reproducible workflow execution using EGA files and datasets. In order to provide all the needed hints to have reproducible executions, output metadata from the WfExS backend is an RO-Crate with all the execution provenance: concrete repository checkout hashes, the concrete engine used, the complete list of inputs (even implicit ones). But, at the same time, a secure execution is going to be achieved using FUSE encfs encrypted directories for intermediate results, and final results are encrypted using crypt4gh GA4GH standard[15] and the public keys of the researchers, so the results can be safely moved outside the execution environment through insecure networks and storages.

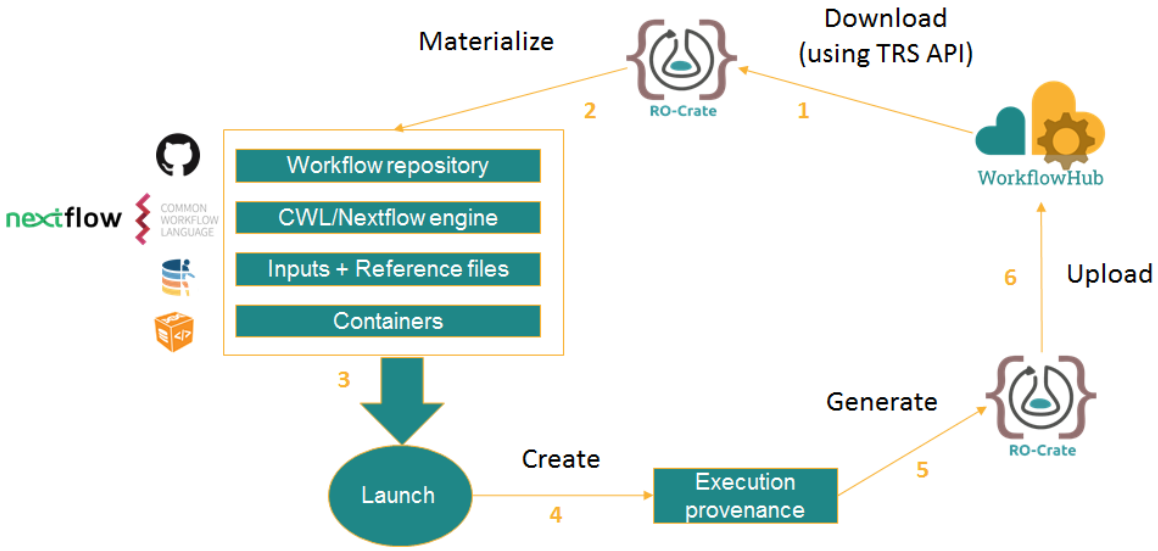

Figure 4: By adding full-circle capabilities, the workflows can re-execute from a previously generated RO-Crate. The prospective Workflow RO-Crate is retrieved from WorkflowHub (step 1), the crate contains not just the workflow definitions, but details about engines, test inputs and reference datasets (step 2), which enables launching with a compatible engine backend (step 3). The engine execution is captured as provenance (step 4), generating a retrospective Workflow Run RO-Crate (step 5), which, as test results are uploaded (by reference) to WorkflowHub (step 6) for future inspection and re-execution. There is an open question regarding how to make publicly available the resulting data sets, as those may contain sensitive data which cannot be deposited in openly available repositories like Zenodo, EUDAT B2Share and alike.

3.4.3 FAIR+Reproducible (FAIR+R) workflows: description, documentation and registration in WorkflowHub

As described in Goble et al [26] principles apply to Computational Workflows to make them FAIR+R.

[15] http://samtools.github.io/hts-specs/crypt4gh.pdf

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

The WorkflowHub[16] makes workflows Findable and Accessible by indexing workflows across Workflow Management Systems (WfMS) and repositories while providing richer standardized metadata. It enables Interoperability and Reuse through use of standard descriptions of workflows, rich metadata and the packaging of workflow components. To enable this the Hub has three primary mechanisms for the documentation of workflows with machine processable metadata: CWL, Bioschemas and RO-Crate (Figure 5).

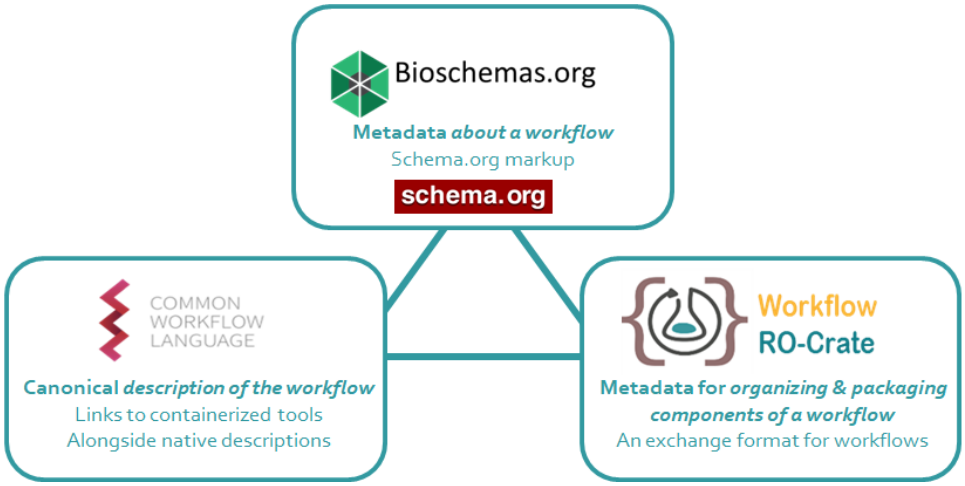

Figure 5: the three mechanisms for machine-processable metadata used by WorkflowHub for FAIR+R

The Hub is WfMS agnostic, so that workflows may remain in their native repositories in their native forms, and the workflow files can be referenced or uploaded, including arbitrary scripts. However, it is encouraged that the native workflow definitions are accompanied with a Abstract CWL (non-executable) description - giving the ability to provide the structure, composed tools and external interface in an interoperable way across workflow languages. Long term plans are to automatically generate the abstract CWL for the prominent WfMSs in EOSC-Life, as have already been demonstrated for Galaxy[17]. We see this duality as an important retention aspect of reproducibility, as the structure and metadata of the workflow can be accessed independent of its native format, even if that may no longer may be easily executable. The presence of the native format enables the reuse in the specific WfMS, benefitting from all its features.

To improve FAIR and search engine visibility, the WorkflowHub exposes Bioschemas markup of its entries, using the new Computational Workflow profile along with existing profiles like Person.

The WP also co-leads a community effort to define Workflow RO-Crate[18], as described earlier, as a workflow exchange format. RO-Crate provides the ability to package executable workflows, their components such as example and test data, abstract CWL, diagrams and their documentation making workflows more easily re-usable. The WP has also released the ro-crate-py[19] Python

[16] https://workflowhub.eu/
[17] https://github.com/workflowhub-eu/galaxy2cwl
[18] https://about.workflowhub.eu/Workflow-RO-Crate/
[19] https://github.com/ResearchObject/ro-crate-py

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

library and heavily contributes to further evolve and mature the general RO-Crate specification[20] and community.

The WorkflowHub has features for collections, versions, snapshots, metrics, and curation. It provides support for workflow teams such as credit management. Workflow management systems in EOSC-Life (Galaxy, Snakemake, Nextflow, CWL, SCIPION) are supported and their communities are working with the Hub to seamlessly and automatically support metadata collection and RO-Crate packaging, registration and import, discovery and export, and downloading and launching.

Development of the WorkflowHub was accelerated by 6 months and front loaded to address the COVID-19 crisis. Development was undertaken in an open and agile way and continues to do so. An alpha release was launched in April 2020 for COVID workflows[21], initially to support the COVID-19 Virtual Biohackathon[22], but then continuing to do so more widely. Over 25 public COVID-19 workflows have been identified, curated and registered. The Hub is registered in the EU COVID-19 Data Portal[23], listed as a service of the ELIXIR Tools Platform[24] and referenced in a Nature[25] Methods article [27]. The WorkflowHub was opened up to all workflows, including non-COVID and WP3 demonstrator contributions, in June 2020. In addition to Workflow Hub registrations by Demonstrator Project D7[26], Demonstrator D6[27] and Demonstrator D3[28], more than 30 projects[29] from EOSC and beyond have started using the hub for publicizing their workflows.

The WorkflowHub is currently being integrated with Life Monitor, to continuously monitor registered workflows and notify the author, and indicate to other users, if the workflow deviates from its original intended behaviour. The author has the ability to update or fix their workflow and register a new version.

Work is continuing on supporting seamless reproducibility throughout the workflow lifecycle by easing the smooth coupling of WfMSs with the Hub. Through the GA4GH TRS API, the WorkflowHub is being more closely integrated with Galaxy to provide seamless execution of Galaxy workflows through a single button click, and revisiting past results. Improved support for closer integration with the Nextflow nf-core github repositories is also planned, to make it easier to and automate the registration of nf-core workflows and recognise and register new version releases.

[20] https://w3id.org/ro/crate
[21] https://covid19.workflowhub.eu/
[22] https://elixir-europe.org/news/hacking-pandemic
[23] https://www.covid19dataportal.org
[24] https://elixir-europe.org/platforms/tools
[25] https://doi.org/10.1038/s41592-020-0886-9
[26] https://workflowhub.eu/projects/31 - workflows
[27] https://workflowhub.eu/workflows/41
[28] https://workflowhub.eu/projects/9 - workflows
[29] https://workflowhub.eu/projects

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

3.5. Development of provenance standards and guidance to FAIR principles (WP6)

Development of provenance standard documenting complete history of the data under ISO Technical Committee 276 (Biotechnology) Working Group 5 (Data Integration), registered under number ISO 23493. There are 6 parts in development: Part 1 on Requirements on provenance information management, Part 2 on Common provenance model, Part 3 on Biological material provenance, Part 4 on Data generation provenance, Part 5 on Data processing provenance, and Part 6 on Security extensions. There are ongoing discussions within the working group on provenance of database validation, too. The aim of the provenance standard is to document the history of data in the machine actionable manner, i.e., using a queryable structured information model with well-defined semantics, which can be used for automated analyses of fitness-for-purpose of the data reuse. The model is designed to work in distributed environments where provenance information can be subject to data protection and various other access limitations. The model is building on W3C PROV model, extending it substantially to support distributed generation and linking of provenance in Part 2, and developing domain-specific extensions used for describing particular aspects of provenance domains of Parts 3-5.

Personalized guidance to FAIR principles is developed as FAIRassist tool and this covers provenance informance guidance as a part of the Reusability principle of FAIR.

3.6. Support actions for sharing data, tools and workflows (WP3)

WP3 supports the collaboration between the Demonstrator teams and the technical experts in WP1 and WP2 and a number of the supported Demonstrator projects targeted improved reproducibility, as outlined in the section above.

Additionally, WP3 administers Open Calls for projects via the Digital Life Sciences Open Call, which awards funding to projects that share data, tools or workflows in the cloud. Through this, additional datasets, tools, and workflows are made available to the wider research community, increasing the potential for re-use of these materials. Projects selected through the Open Call are evaluated based on the impact of the data, tools, and workflows they make available to the wider community, as well as to their sustainability, which is a crucial factor in ensuring long-term access to the shared materials. Close collaboration with the other WPs within EOSC-Life in the maturation and evaluation phase ensures that funded projects are aligned with the efforts of the other WPs towards reproducibility, such as using the tools and registries developed by WP2.

4. Summary

The landscape of EOSC-Life activities related to reproducibility is very rich and complex. The following table summarizes it, reporting for each activity a link to the full description, the Workpackage/Demonstrator/Open Call that is responsible for its implementation, the relevant tools and the ways in which it is expected to have an impact on research reproducibility.

Activities comprise both the development of new tools and means to increase the use of existing ones. In the relative column of the summarizing table, letters in parentheses indicate whether it

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

was existing (E), developed specifically within EOSC-Life (D) or in collaboration with other projects (C).

Most activities are related to computational science, but experimental aspects are also covered by tools for the FAIRification of protocols and data as well as for the provenance of biological material, data generation and data provenance.

EOSC-Life actions are expected to warrant several aspects of reproducibility and replicability of experiments, data and computational workflows. Some of the tools have a general standing. Others are developed for a specific field or application, but may set the way for a more general use.

We also highlighted the four elements of FAIR principles, with particular emphasis on reuse, which is the goal of the Open Science Cloud and leads to the possibility of extending and generalizing the conclusions of a study across different situations (experimental settings, models, organisms, populations, …).

Gap/s	Action	Description in text (page)	Author	Tools (**E**xisting, **D**eveloped in EOSC-Life, developed in **C**ollaboration)	Expected or observed Impact	Performance indicator(s)
Non-overlapping hits from different screenings (i.e. in COVID-19 assays) Poor description of assay design and data processing Difficulty in comparing experiments and integrating results	Assay metadata - cell-based assays	Replicability in COVID-19 drug repurposing; FAIR chemosensitivity assays (p 6-7)	D1, WP1	chEMBL (C), ECBD (C), MICHA (D)	Facilitate experimental design and reporting Testing how differences in COVID-19 assays affect replicability (under way)	Number of submissions to chEMBL, ECBD and MICHA
Insufficient description of the structure models used in studies Structure models are continuously updated and newly entered in the databases	FAIRification and versioning - protein structure data and analyses	FAIR protein structures (p 7-8)	WP1 OC2020	PDB-REDO (D)	Accurate data provenance tracking Entries and versions described in a machine-readable format Long-term archiving, sustainability	Number of submissions to PDB-REDO
Inability to install and execute software and/or workflows on different	Computational environment of a workflow	Reproducible workflows, Installation and execution	WP2	BioConda (C), containerisation	computational reproducibility, independent of context	Number of submissions to BioConda

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

computational environments		of software and workflows (p 9-10)			of deployment	
Fragmentation of research outcomes across many resources	Workflow definition	Reproducible workflows, Packaging workflows (10-11)	WP2	Research Output packaging: RO-Crate (C), Workflow RO-Crate (C).	Computational reproducibility, data reuse	Number of submissions to RO-Crate
Non-standardised language(s) in computational workflows	Workflows description, agnostic to workflow management systems	Reproducible workflows, Workflow specification (p 11)	WP2	Common Workflow Language (E), Computational Workflow Bioschemas	Computational reproducibility, findability to improve reuse	Number of publications referencing Common Workflow Language
Difficulties in complying to GDPR regulations for analysis of sensitive data	Analysis of un-disclosable data - high level workflow execution service for variant calling in genomic data	A particular workflow specification for analysis of undisclosable data (p 11-13)	WP2, D7	Singularity (E), WorkflowHub (D), RO-Crate (C), crypt4gh (E)	Computational reproducibility, sensitive data integration and reuse to improve replicability	Number of publications referencing crypt4gh
Difficulty in assessing bottlenecks and presenting standardised overview of computational workflows	FAIR+Reproducible workflows	FAIR+Reproducible (FAIR+R) workflows: description, documentation and registration in WorkflowHub (p 13-15)	WP2	WorkflowHub (D), Life Monitor(D), COVID-19 WorflowHub (D)	Computational reproducibility, findabilty	Number of submissions to WorkflowHub
Lack of provenance of biological materials (including samples) and data	Provenance of biological materials, data generation, data processing	Development of provenance standards and guidance to FAIR principles (p 15-16)	WP6	Provenance standard under ISO 276 (ISO 23493) (D)	Quality, reusability	Number of publications referencing ISO 276 (ISO 23493)
Data is not findable, accessible,	FAIR data principles	Development of provenance	WP6	FAIRassist tool (D), FAIR cookbook (C)	Findability, accessibility, interoperability,	Number of publications

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

interoperable nor reusable	guidance	standards and guidance to FAIR principles (p 15-16)			reusability	referencing FAIRassist tool and FAIR cookbook
Poor research data management	Guidance on sharing data, tools and workflows	Support actions for sharing data, tools and workflows	WP6	Research Data Management kit (RMD Kit) (C)	All aspects described above	Number of publications referencing RDM Kit

References

1. Fidler, F.; Wilcox, J. Reproducibility of Scientific Results. In The Stanford Encyclopedia of Philosophy; Zalta, E. N., Ed.; Metaphysics Research Lab, Stanford University, 2018.
2. Committee on Reproducibility and Replicability in Science; Board on Behavioral, Cognitive, and Sensory Sciences; Committee on National Statistics; Division of Behavioral and Social Sciences and Education; Nuclear and Radiation Studies Board; Division on Earth and Life Studies; Board on Mathematical Sciences and Analytics; Committee on Applied and Theoretical Statistics; Division on Engineering and Physical Sciences; Board on Research Data and Information; Committee on Science, Engineering, Medicine, and Public Policy; Policy and Global Affairs; National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science; National Academies Press: Washington, D.C., 2019; p 25303. https://doi.org/10.17226/25303.
3. Union, P. O. of the E. Reproducibility of scientific results in the EU : scoping report. http://op.europa.eu/en/publication-detail/-/publication/6bc538ad-344f-11eb-b27b-01aa75ed71a1 (accessed Feb 23, 2021).
4. Plesser, H. E. Reproducibility vs. Replicability: A Brief History of a Confused Terminology. Front Neuroinform 2018, 11. https://doi.org/10.3389/fninf.2017.00076.
5. Drummond, C. Replicability Is Not Reproducibility: Nor Is It Good Science. Proceedings of Evaluation Methods for Machine Learning Workshop at the 26th Annual International Conference on Machine Learning (ICML), Montreal, Canada, 2009. 4.
6. Munafò, M. R.; Nosek, B. A.; Bishop, D. V. M.; Button, K. S.; Chambers, C. D.; Percie du Sert, N.; Simonsohn, U.; Wagenmakers, E.-J.; Ware, J. J.; Ioannidis, J. P. A. A Manifesto for Reproducible Science. Nature Human Behaviour 2017, 1 (1), 1–9. https://doi.org/10.1038/s41562-016-0021.
7. The EQUATOR Network | Enhancing the QUAlity and Transparency Of Health Research https://www.equator-network.org/ (accessed Apr 20, 2021).
8. FAIRsharing https://fairsharing.org/ (accessed Apr 10, 2021).

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

9. Sansone, S.-A.; McQuilton, P.; Rocca-Serra, P.; Gonzalez-Beltran, A.; Izzo, M.; Lister, A. L.; Thurston, M. FAIRsharing as a Community Approach to Standards, Repositories and Policies. Nature Biotechnology 2019, 37 (4), 358–367. https://doi.org/10.1038/s41587-019-0080-8.
10. Kenall, A.; Edmunds, S.; Goodman, L.; Bal, L.; Flintoft, L.; Shanahan, D. R.; Shipley, T. Better Reporting for Better Research: A Checklist for Reproducibility. Genome Biology 2015, 16 (1), 141. https://doi.org/10.1186/s13059-015-0710-5.
11. New code completeness checklist and reproducibility updates https://ai.facebook.com/blog/new-code-completeness-checklist-and-reproducibility-updates/ (accessed Dec 24, 2020).
12. Reproducibility Checklist | AAAI 2021 Conference.
13. Pineau, J. ReproducibilityChecklist-v1.2. 1.
14. Freire, J.; Fuhr, N.; Rauber, A. Reproducibility of Data-Oriented Experiments in e-Science (Dagstuhl Seminar 16041). Dagstuhl Reports 2016, 6 (1), 108–159. https://doi.org/10.4230/DagRep.6.1.108.
15. Data Security Toolkit https://www.ga4gh.org/genomic-data-toolkit/data-security-toolkit/ (accessed Apr 3, 2021).
16. The Minimum Information About a Bioactive Entity (MIABE) | HUPO Proteomics Standards Initiative https://www.psidev.info/miabe (accessed Apr 19, 2021).
17. MICHA https://micha-protocol.org/ (accessed Apr 19, 2021).
18. Tanoli, Z.; Aldahdooh, J.; Alam, F.; Wang, Y.; Seemab, U.; Fratelli, M.; Pavlis, P.; Hajduch, M.; Bietrix, F.; Gribbon, P.; Zaliani, A.; Hall, M. D.; Shen, M.; Brimacombe, K.; Kulesskiy, E.; Saarela, J.; Wennerberg, K.; Vähä-Koskela, M.; Tang, J. Minimal Information for Chemosensitivity Assays (MICHA): A next-Generation Pipeline to Enable the FAIRification of Drug Screening Experiments. bioRxiv 2020, 2020.12.03.409409. https://doi.org/10.1101/2020.12.03.409409.
19. Bank, R. P. D. RCSB PDB: Homepage https://www.rcsb.org/ (accessed Apr 19, 2021).
20. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. Nucleic Acids Research 2000, 28 (1), 235–242. https://doi.org/10.1093/nar/28.1.235.
21. PDB-REDO https://pdb-redo.eu/ (accessed Apr 19, 2021).
22. Joosten, R. P.; Salzemann, J.; Bloch, V.; Stockinger, H.; Berglund, A.-C.; Blanchet, C.; Bongcam-Rudloff, E.; Combet, C.; Da Costa, A. L.; Deleage, G.; Diarena, M.; Fabbretti, R.; Fettahi, G.; Flegel, V.; Gisel, A.; Kasam, V.; Kervinen, T.; Korpelainen, E.; Mattila, K.; Pagni, M.; Reichstadt, M.; Breton, V.; Tickle, I. J.; Vriend, G. PDB_REDO: Automated Re-Refinement of X-Ray Structure Models in the PDB. J Appl Crystallogr 2009, 42 (3), 376–384. https://doi.org/10.1107/S0021889809008784.
23. Crystallography: Protein Data Bank. Nature New Biology 1971, 233 (42), 223–223. https://doi.org/10.1038/newbio233223b0.
24. Eoghan Ó Carragáin; Carole Goble; Peter Sefton; Stian Soiland-Reyes. A Lightweight Approach to Research Object Data Packaging, 2019. https://doi.org/10.5281/zenodo.3250687.
25. Common Workflow Language, v1.0, 2016. https://doi.org/10.6084/m9.figshare.3115156.v2.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

26. Goble, C.; Cohen-Boulakia, S.; Soiland-Reyes, S.; Garijo, D.; Gil, Y.; Crusoe, M. R.; Peters, K.; Schober, D. FAIR Computational Workflows. Data Intelligence 2020, 2 (1–2), 108–121. https://doi.org/10.1162/dint_a_00033.
27. Marx, V. When Computational Pipelines Go 'Clank.' Nature Methods 2020, 17 (7), 659–662. https://doi.org/10.1038/s41592-020-0886-9.

Delivery and schedule

The delivery is delayed: No

Adjustments made

In answer to the recommendations from the review report of the 2nd periodic report several adjustments have been made to clarify how the deliverable provides a framework for the evaluation of the impact of EOSC-Life activities on reproducibility. The first step in such a process is the understanding of the tools and services produced and made available, the gaps they are filling up and what is their expected impact. In the table, we have added two columns where we specify in a more accurate way, for each of the described activities, the gaps that it intends to fill and the way in which we will assess its impact at the end of the project. To clarify the structure of MICHA, we have added a graphical overview of the project. Also, the summary has been modified, to describe these additions.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Appendices

Appendix 1.

An experiment or observation gives rise to one or more outcomes. For instance, if a coin is tossed it falls on head or tails. If the lifespan of a cell or organism is measured, a real number is obtained. If the experiment or observation is deterministic, it will give rise to a single outcome. If the experiment or observation is random (i.e. random error is present), the set of all possible outcomes is called the **sample space** (ref. Feller W. An Introduction to Probability Theory and Its Application, 1950).

An experiment or observation makes sense only when **eligibility criteria** (i.e. initial conditions) and **operative procedures** are rigorously defined (or, at least, properly identified). For instance, the toss of a coin is performed under accurate spatial and temporal coordinates and the life span of a cell or organism depends on the environmental conditions. We will indicate eligibility criteria (i.e. initial conditions) and operative procedures with the symbol D. Operationally, this can mean different things in different sciences. In clinical trials, this means, at minimum, a detailed study protocol and a description of measurement procedures. In laboratory science, how key reagents and biological materials were created or obtained can be critical (ref Goodman SN et al. What does research reproducibility mean?[30], June 2016 Vol 8 Issue 341 341ps12).

Four pillars of reproducibility have been recognized:

1. **Methods reproducibility** refers to the provision of enough detail about eligibility criteria (i.e. initial conditions) and operative procedures D so **the same experiment or observation** could, in theory or in actuality, be exactly repeated. Firstly, the definition of methods reproducibility does not concern the outcome. In other terms, methods reproducibility could be satisfied even if the outcome is different in a perfect replication of the original experiment. Secondly, a perfect replica of the original experiment is obtained, if and only if, the same experimental or observational units are used and the same initial conditions (e.g. patient weight, coin position, clock time) are satisfied.
2. **Results reproducibility** refers to obtaining the same outcome from the conduct of **an independent experiment or observation** whose initial conditions and operative procedures are as closely matched to the original experiment as possible (i.e. at least in theory the same D are used). The definition of results reproducibility concerns the outcome. The meaning of "same" outcome depends on the deterministic or random nature of the experiment or observation. If the experiment is deterministic, the outcome is distinctively determined by the initial conditions. If the experiment is random, the distinctiveness of the outcome should be judged apart from random error. Operationally the distinction between deterministic and random experiments could make unclear the criteria for considering outcome to be "the same".
3. **Inferential reproducibility** refers to the drawing of qualitatively similar conclusions from either an independent replication of a study or a reanalysis of the original study. Inferential reproducibility is not identical to results reproducibility or to methods reproducibility,

[30] https://doi.org/10.1126/scitranslmed.aaf5027

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

because scientists might draw the same conclusions from different sets of studies and data or could draw different conclusions from the same original data (Goodman et al., 2016[31]). Inferential reproducibility is related to the soundness of research claims. Scientific questions are not settled on a particular date, by a single experiment, nor are they settled irrevocably. We speak of the weight of evidence (ref: Rosenbaum PR Observational Studies, Second Edition, 2010[32]). Viewed through this lens, the aim of repeated experimentation is to increase the amount of evidence, measured on a continuous scale, either for or against the original claim.

4. **External reproducibility (i.e. external validity)** refers to obtaining an **equivalent outcome**. Outcome y' from the conduct of **an independent experiment or observation** whose eligibility criteria (i.e. initial conditions) and operative procedures D' are more or less different to that D of the original experiment. Notice that outcome y' may be different from the original outcome y. For instance, the original experiment aims to study tumor shrinkage in a murine model. The outcome is assessed using a caliper (i.e. tumor volume is expressed as mm3). An independent experiment is performed on humans and tumor shrinkage is assessed using RECIST criteria version 1.1. Outcome y' is mathematically and/or statistically related to the original outcome y (e.g. 30% tumor shrinkage in murine models is statistically related to the objective response rate in humans).

The mathematical and statistical relationship between (D, y) and (D', y') could be more or less strong. The weakest relationship is the qualitative (i.e. only the sign is preserved): if the outcome measure y increases then the outcome measure y' increase. As a stronger relationship, the measurement order magnitude could also be preserved (i.e correlation): if $y1 < y2$ then $y1' < y2'$. The strongest relationship is that of surrogacy: the treatment effect in (D', y') is at least partially explained by the treatment effect in (D, y). In this case, we say that y is a **surrogate measure** of y'.

External validity is deemed very important, as the true impact of an experiment relies not only on its internal validity (the possibility to identify and fairly estimate causal pathways between investigated variables), but also and particularly on its generalizability. In fact, there is often a trade-off between internal and external validity, for instance when excess efforts are made to control biological and experimental variability. This improves results reproducibility at the expense of external validity.

[31] https://doi.org/10.1126/scitranslmed.aaf5027
[32] http://doi.org/10.1007/978-1-4757-3692-2

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.