D6.2 Report on the implementation of enhanced transfer procedures for data from EU-OS partner sites

Version 2023-10-31

for
H2020-INFRADEV-2018-1
(Development and long-term sustainability of new pan-European research infrastructures)

Research and Innovation Action (RIA)

Action Acronym:
EU-OPENSCREEN-DRIVE

Action Full Title:

“Ensuring long-term sustainability of excellence in chemical biology within Europe and beyond**”**

Grant Agreement No.: 823893

Dissemination level: public

Document Properties:

Deliverable	6.2 Report on implementation of enhanced transfer procedures for data from EU-OS partner sites
Partner responsible	EU-OS, IMG
Author(s)	Katja Herzog (EU-OS), Sarka Simova (IMG), Ctibor Skuta (IMG)

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under the grant agreement No. 823893

Table of Contents

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under the grant agreement No. 823893

1 Introduction

EU-OPENSCREEN (EU-OS, https://www.eu-openscreen.eu/) is the European Research Infrastructure Consortium (ERIC) for Chemical Biology and early Drug Discovery, which was established in 2018 and offers access to state-of-the-art high-capacity screening and medicinal chemistry services throughout Europe for the development of small molecule tools and lead compounds. EU-OS services are currently provided by >30 partner sites in 10 member countries across Europe (CZ, DK, ES, FI, LV, NO, PL, PT, SE and DE as host country). It operates an open-access database, the European Chemical Biology Database (ECBD, https://ecbd.eu/), hosted in Prague at IMG, and a central compound management facility (CCMF) in Berlin, Germany which stores quality-controls and manages the jointly-used EU-OPENSCREEN compound collection (https://www.eu-openscreen.eu/services/compound-collection.html).

Herein we report on the implementation of the data upload interface facilitating the data transfer procedures within the EU-OPENSCREEN research infrastructure. Most of the generated data within the infrastructure are uploaded into the European Chemical Biology Database (ECBD, https://ecbd.eu/). ECBD serves as the central data hub designed to FAIRify, share and disseminate the data to the wide scientific community. The data upload process represents the most crucial point on the way from individual screening sites that ensures unified data description, results curation and dissemination to other similar data resources (e.g., the ChEMBL database, https://www.ebi.ac.uk/chembl/, or PubChem, https://pubchem.ncbi.nlm.nih.gov/).

2 Report on the deliverable

2.1 Data transfer procedure

The data transfer procedure, i.e., the data upload/deposition into the ECBD, is performed through the ECBD's interactive data upload form. The data are uploaded by the local EU-OPENSCREEN screening partner sites, specifically, by selected, trained scientific users with the permission to do so. Apart from the upload of the experimental data, the form is designed to simplify data FAIRification (based on the FAIR data guiding principles have been recently developed and widely adopted to improve the Findability, Accessibility, Interoperability, and Reuse of scientific data) by utilizing ontologies and common identifiers, which streamlines the data transfer involving all collaborators in the procedure.

The data upload procedure (Fig. 1) is a three-step process, which should mitigate the risk of introduction of errors into the data and therefore, into the public domain. The individual steps are: 1) submission by the screening site; 2) acceptance by the ECBD data curator/s; and 3) confirmation by the screening site and project collaborators.

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under the grant agreement No. 823893

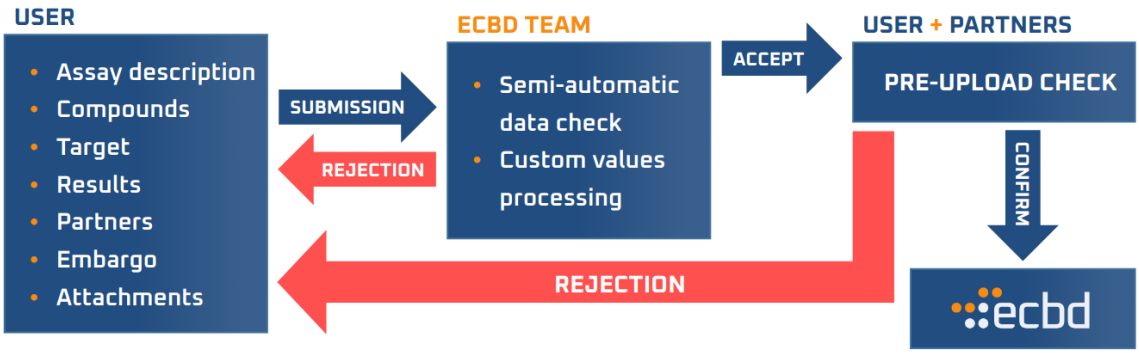

Figure 1 The data upload procedure scheme.

In the first step, the user, who is responsible for the data upload at the respective screening partner site, fills in the upload form in as much detail as possible. Besides free text, any relevant documents (such as assay protocol) can be added as an attachment. The experimental data are supplied in the form of data tables as csv/xlsx files with a predefined format. After the files are uploaded, the uploader can visually review the data in the form of a scatterplot or eventually as dose-response curves if they are part of the submission.

After the initial submission, ECBD data curators are notified and can review the data. In general, the clarity of free-text description is assessed, and the used ontological terms are reviewed accordingly. In cases where custom terms are employed instead of the ontological ones, i.e., the user did not find a fitting term in the used ontology, the ontology is double-checked and the term is either replaced by the ontological one or aligned with other similar custom terms, if they are available. The experimental data validation involves a manual check of the data format and value ranges (measured/normalized values, concentrations, etc.) and the check of the correspondence between the description of the activity determination method and the stated compound activity results (e.g., active, inactive, inconclusive, etc.).

If the data curators do not identify any issues within the submission, it is accepted, and the submitter as well as all collaborators stated within the submission are notified in order to perform the final confirmation. Once they all confirm the submission, the data can be uploaded into ECBD. In case the curators find one or more issues within the submission, it is reopened for the submitter to make any necessary changes.

2.2 Training and documentation

To guide users through the data submission process, we have included descriptive texts within each section (Fig. 2) of the upload form, providing step-by-step instructions. Additionally, an interactive guide (Fig. 3) is available to offer further assistance and ensure error-free submissions. For users' convenience, we also provide data format templates, examples and documentation (Fig. 4), illustrating the specified formats and structures for different types of data (experimental values file and results file). Moreover, in November 2020, we organized a webinar on data upload procedures available upon request to all EU-OPENSCREEN partner sites, and we also offer introductory calls for beginner users (most often for first-time uploaders). Generally, our aim is to make the upload process as straightforward as possible, with all the necessary information, including format templates, examples, and support calls, readily accessible.

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under the grant agreement No. 823893

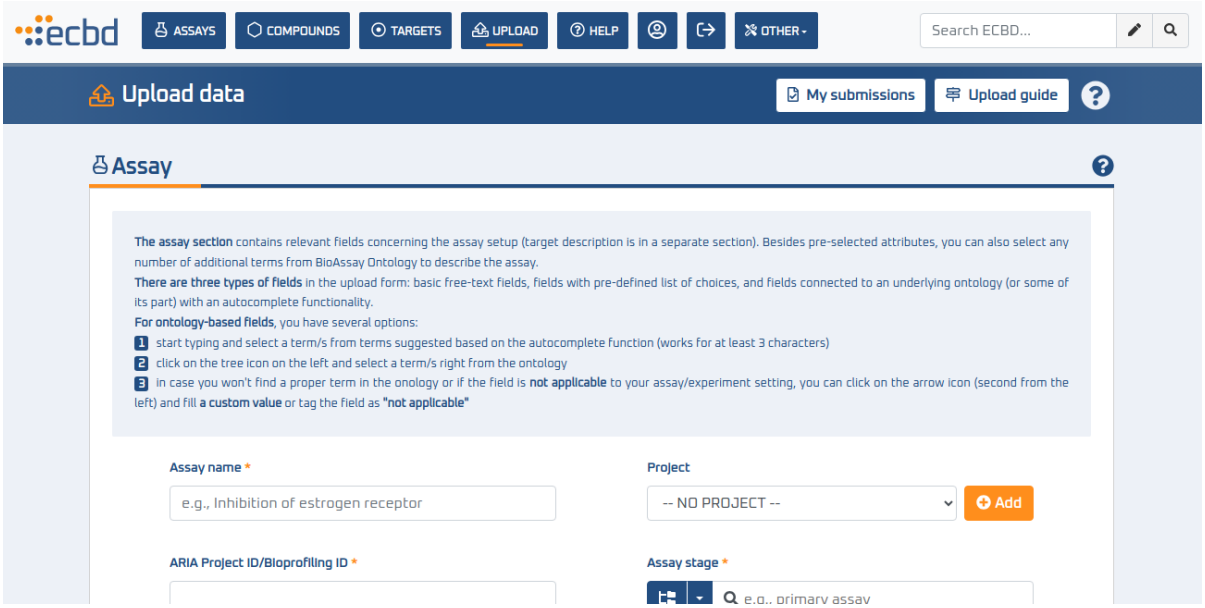

Figure 2 Helping descriptive text for the Assay section.

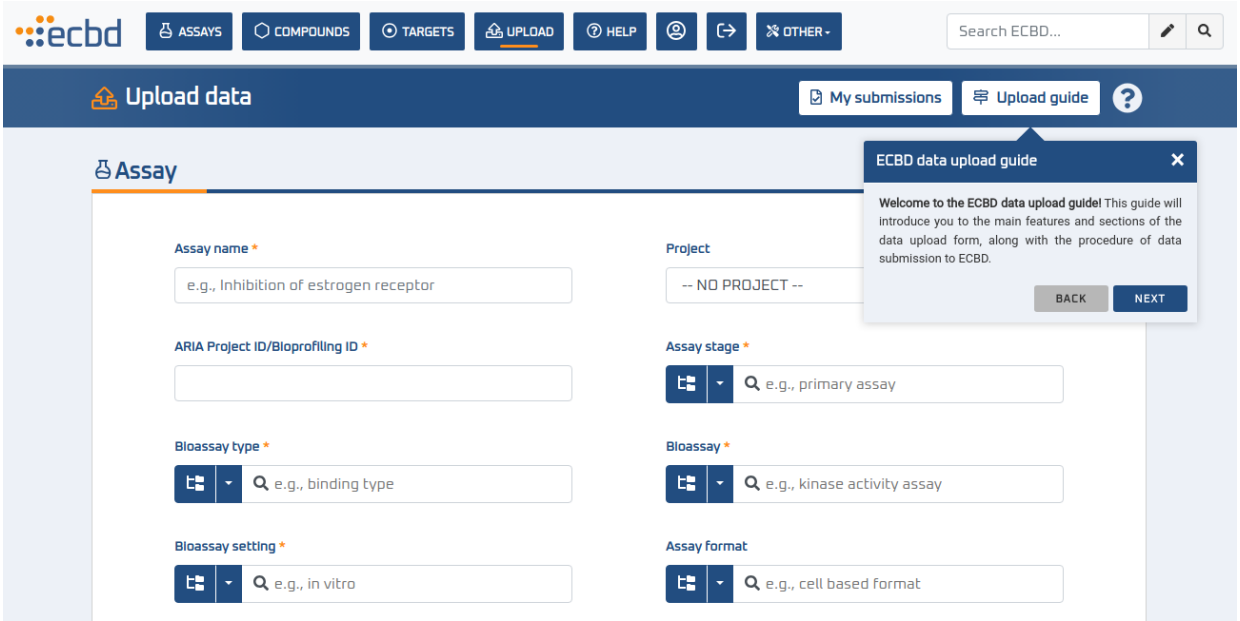

Figure 3 Interactive data upload guide.

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under the grant agreement No. 823893

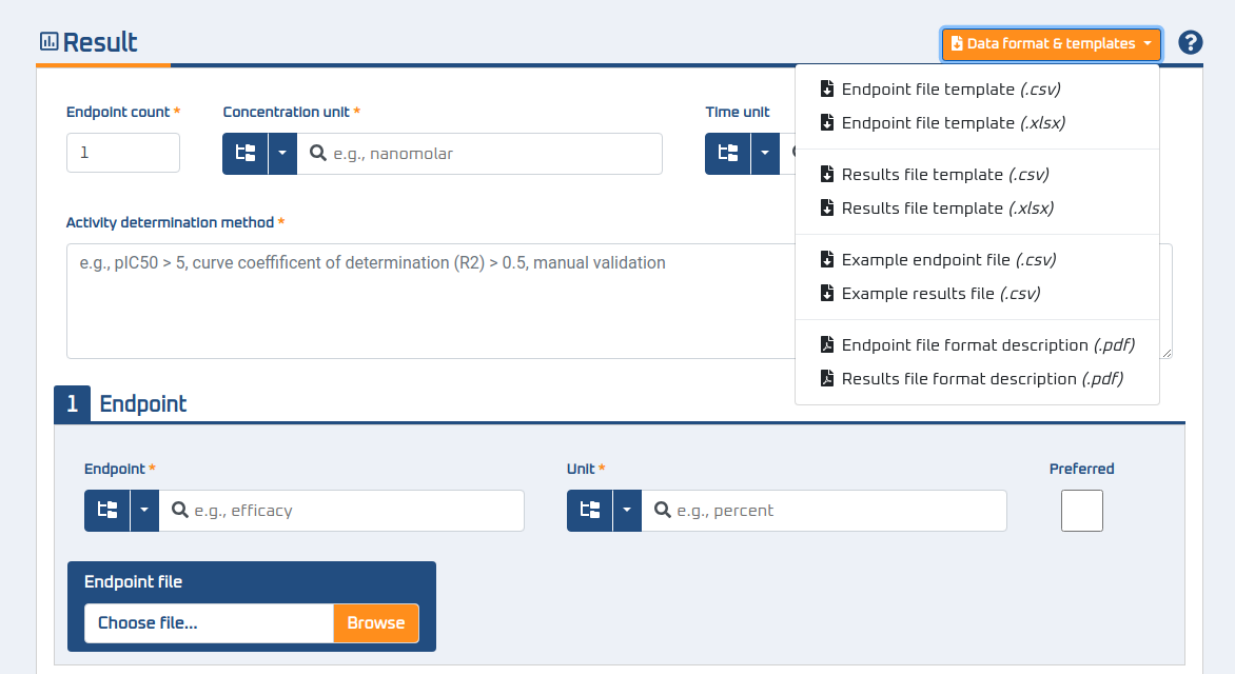

Figure 4 Results section with the file templates and examples.

3 Conclusion

A three-step data upload procedure, which mitigates the risk of introduction of errors into the data, has been established during the DRIVE project. Descriptive texts to guide users through the data submission process have been added to the ECBD. Online training for data uploaders have been organised. Webinars on data upload procedures is available upon request to partner sites as well as introductory calls for first-time users.

This project has received funding from the European Union´s Horizon 2020 research and innovation programme under the grant agreement No. 823893