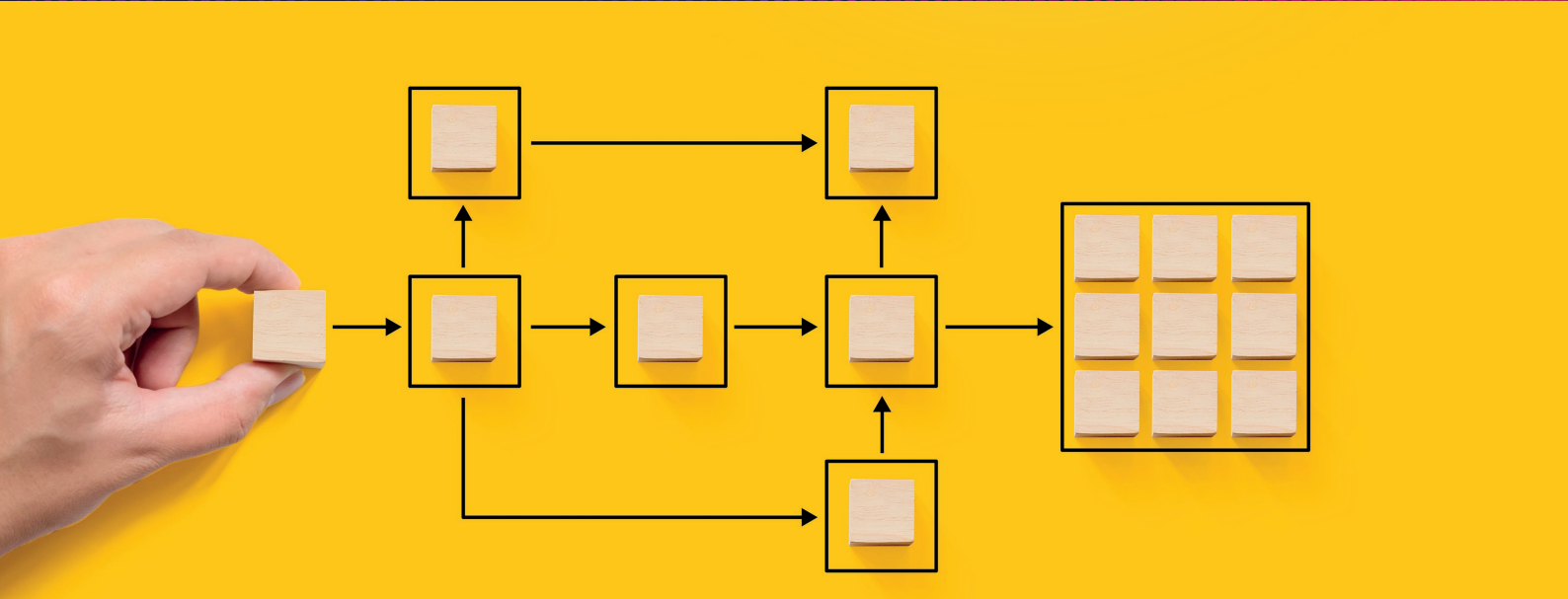
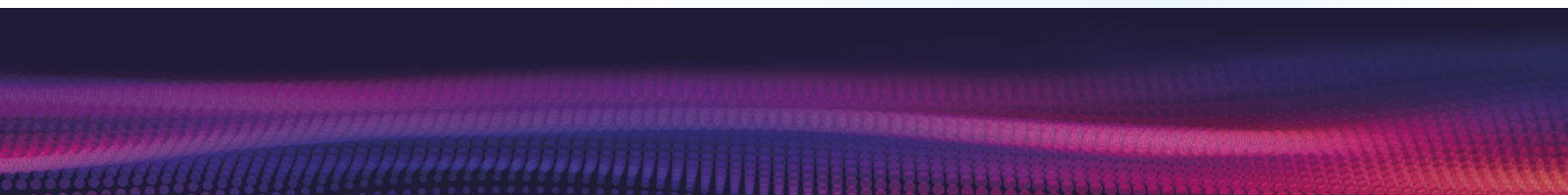


Prototyping the JACQ Herbarium Specimen Digital Asset metadata model



Authors names, affiliations and ORCIDs:

- Petr Novotný, Charles University, Prague, Faculty of Science, Herbarium
PRC [0000-0002-9655-6278](https://orcid.org/0000-0002-9655-6278)
- Patrik Mráz, Charles University, Prague, Faculty of Science, Herbarium
PRC [0000-0002-1415-070X](https://orcid.org/0000-0002-1415-070X)
- Zdeněk Vaněček, Charles University, Prague, Faculty of Science,
Herbarium PRC [0000-0002-8704-9768](https://orcid.org/0000-0002-8704-9768)

Summary:

The support action introduced successful applicants to three tools developed by the FAIRCORE4EOSC (FC4E) project¹. These include the Data Type Registry (DTR)², vocabulary service, and Metadata Schema and Crosswalk Registry (MSCR)³.

In this support action the participants will use one or more of the FC4E tools to enhance the machine-actionability of their existing schema and crosswalk specifications. This could mean for example the registration of schema from public APIs to the MSCR and creation of mappings from those service specific data models to common formats such as Datacite and DCAT or creating a metadata schema using enriched schema elements registered in the DTR.

This FAIR Implementation Story outlines the specific aims, actions and experiences of Charles University, Herbarium PRC in relation to their participation in this support action⁴.

-
- 1 <https://faircore4eosc.eu/>
 - 2 <https://faircore4eosc.eu/eosc-core-components/eosc-data-type-registry-dtr>
 - 3 <https://faircore4eosc.eu/eosc-core-components/metadata-schema-and-crosswalk-registry-mscr>
 - 4 <https://fair-impact.eu/managing-data-schemas-vocabularies>



■ ■ Introduction:

The PRC herbarium at the Faculty of Science, Charles University in Prague, is one of the largest university herbaria in the world, holding over two millions preserved specimens, moreover, the Czech Republic ranks fourth globally in the number of specimens per capita. A wide range of institutions are involved in the ongoing digitisation of these valuable scientific records, producing high-resolution scans that are made publicly accessible. Within the EOSC-CZ platform, we are preparing a repository that will ensure the long-term preservation of these data and support their FAIR qualities.

During the support action, our focus was on designing a metadata schema tailored to the needs of a national repository for these objects. They are currently not stored by international platforms such as JACQ or GBIF, which primarily manage structured specimen metadata but lack infrastructure for storing associated digital assets. Our goal was to conceptualize a model that links these digital representations with existing specimen metadata using persistent identifiers and internationally recognized standards.

■ ■ Approach taken:

During the support action, we held regular meetings within our team, composed of three members representing both the herbarium and national infrastructure development. These sessions were dedicated to present tools (DTR, VS and MSCR) and helped us in clarifying the core objectives of our planned metadata model and understanding the technical and conceptual challenges of implementing FAIR principles for digital representations of herbarium specimens in broader context.

As we familiarise ourselves with the tools offered during the support action, we began creating early drafts of our intended model using them. This hands-on process provided valuable insights in two directions: it helped us formulate constructive feedback for the developers (mostly concerning UX clarity and the need to better understand the purpose of specific features), and it also offered us the opportunity to reflect on our own model's structure — ensuring that it would be both formally valid and practically usable.

Throughout the programme, we refined a conceptual structure of the schema and drafted an initial machine-readable version using YAML and JSON Schema syntax, setting the stage for further development and implementation. We have presented the prototype to several Czech herbaria that are working with us on the repository and we will continue to work on its finalization and use in next months.

■ ■ Challenges encountered and addressed:

During the development of the prototype metadata model, we encountered various conceptual challenges that needed to be addressed. To address these challenges, we facilitated joint meetings with designated representatives of partner herbaria. Through open dialogue and feedback, we tried to reach at a minimum a preliminary consensus to be confirmed in the final decision-making phase.

One of the exemplary and important issues was the need to decide whether we would design the model as strictly JACQ-related or go in the direction of greater generalization to allow use also for herbaria that cannot or do not want to share data via JACQ and publish their data independently (and refer different type of persistent IDs).

Another major challenge was deciding how much granularity to allow in the model. From a theoretical point of view, it is possible, for example, for microscopic sections made from specimens to ontologically describe the type of



staining, microscopic imaging methods, etc... Such attributes are well described in existing schemes (e.g. Biological Imaging Methods Ontology). However, in practical terms, for the vast majority of the available material within our working group, this information is not available or only to a very limited extent.

In both cases presented, we have opted for a “data-first” approach, i.e., setting both granularity and generality according to the data we have available instead of trying to cover the whole set of theoretically possible cases and data fields.

■ ■ Impact:

Working on this project has given us a much better insight into the space for FAIRification of our practices. We realized that it is not only possible to publish a metadata model once, but also to collaboratively manage either partial data types or the whole schema. With the use of the Crosswalk functionality, we find that even managing new versions of the metadata schema can be done in a clear and understandable way for our users.

Thanks to our work during this support action, we have a prototype metadata model which is now being commented on by the collaborating institutions and will lead to the approval of the final model at the end of 2025. This will be used for the national repository developed within the EOSC CZ node and will thus serve at national level. Its potential extension to the entire JACQ community would affect the currently rather fragmented architecture of digital object sharing in the individual herbaria of this consortium.





@fairimpact_eu



company/fair-impact-eu-project/



fair-impact.eu

