Biospecimens in FDO world

Sara El-Gebali[‡], Rory Macneil[§], Rorie Edmunds[|], Parul Tewatia[‡], Jens Klump[¶]

‡ SciLifeLab Data Centre, Uppsala, Sweden
§ Rspace, Edinburgh, United Kingdom
| DataCite, Tokyo, Japan
¶ The Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia

Corresponding author: Sara El-Gebali (sara.elgebali@scilifelab.uu.se)

Abstract

With the advent of technological advances in research settings, scientific collections including sample material became on par with big data. Consequently there is a widespread need to highlight and recognise the inherent value of samples coupled with efforts in unlocking sample potential as resources for new scientific discovery.

Samples with informative metadata can be more easily discoverable, more readily shared and reused, allowing reanalysis of associated datasets, avoiding duplicate efforts, and providing metaanalysis yielding considerably enhanced insight.

Metadata provides the framework for a consistent, systematic and standardized collection of sample information, enabling users to identify the availability of research output from the samples, relevancy to their intended use, and a way to conveniently identify sample material as well as access provenance information related to the physical samples. Researchers need this essential information aiding their decision making process on the quality, usability and accessibility of the samples and associated datasets.

We propose to explore the practical implementation of FAIR Digital Objects (FDO) for biological life science physical samples and practically how to create an FDO framework centralized around biospecimen samples, linked datasets, sample information and PIDs (Persistent Identifiers)Klump et al. 2021. This effort is highly relevant to enhancing the portability of sample information between multiple repositories and other kinds of resources (e.g. e-infrastructures).

In this session we would like to present our current work in order **to mobilize the community** to define the FAIR Digital Object Architecture for biospecimen in life science including all infrastructure components e.g. metadata, PIDs and their integration with technical solutions.

To that end, in our community of practice we aim to:

- **What**:
    - Identify the minimum set of attributes required for describing biospecimen in biological life science (Minimal Information About a Biological Sample, MIABS) with ontological mapping for semantic unambiguity and machine actionability.

    - Identify the required attributes for registering PIDs for biospecimens and how that will operate in an FDO ecosystem. This will pave the way for a framework of coupling the descriptive metadata to the digital object in a FAIR and comprehensive manner.

- **How**:
    - Define a semantic FDO model for biospecimens.

    - Define the role of biospecimen PIDs registration information and kernel attributes and how that translates to machine actionability and programmatic decisions.

    - Define the implementation specifics for integration of biospecimen FDOs with operational infrastructure e.g. e-infrastructures, repositories and machines.

- Relevant technologies include: RO-Crate, Persistent identifiers, and metadata schemas

The recent partnership between IGSN and DataCite described below is a catalyst in this call of action to the FDO community to build a Community of Practice (CoP) specifically focused on biospecimen samples.

Community of practice:

IGSN e.V. announced a partnership with DataCite, in which DataCite's registration services and supporting technology for Digital Object Identifiers (another type of PID) are now being leveraged to register IGSN IDs, and thus ensure the ongoing sustainability of the IGSN ID infrastructure.

Importantly, the two organizations are also focusing the community's efforts on advocacy of PIDs for physical samples and expanding the global sample ecosystem. Assisted by the DataCite Samples Community Manager, the IGSN e.V. is establishing working groups (Communities of Practice) within different research domains to support development and promotion of standardized methods for identifying, citing, and locating physical samples. In particular, the partnership wishes to work with the Biosamples community to elaborate the necessary information (metadata) such that those within the community have a full

understanding of a physical sample when its descriptive webpage is accessed via its PID, see this example.

Keywords

Physical samples, PIDs, Metadata, Community of Practice

Presenting author

Sara El-Gebali, Rory Macneil

Presented at

First International Conference on FAIR Digital Objects, presentation

Conflicts of interest

References

- Klump J, Lehnert K, Ulbricht D, Devaraju A, Elger K, Fleischer D, Ramdeen S, Wyborn L (2021) Towards Globally Unique Identification of Physical Samples: Governance and Technical Implementation of the IGSN Global Sample Number. Data Science Journal 20 https://doi.org/10.5334/dsj-2021-033