Distributed System of Scientific Collections

Towards digital transformation:
Adoption of Digital Object Identifiers (DOI) as the baseline for Digital Specimens infrastructure

DiSSCo iGA Decision

Towards digital transformation:
Adoption of Digital Object Identifiers (DOI) as the baseline for Digital Specimens infrastructure

Context

European frontier research and global scientific developments anchor on the required digital transformation that drives breakthrough scientific discovery to address societal challenges. Such a transformation is already an urgent need for our economies worldwide and affects all disciplines and fields of activity, both in the public and private spheres. Governments are already aware of the necessity and promote development of mechanisms that will lead the change towards digital and open data to pursue excellence in science. Implementation of FAIR principles promote findability, accessibility, interoperability, and reuse of digital assets and these principles emphasize machine-actionability. As the very first principle it is stated that "(Meta)data are assigned a globally unique and persistent identifier". DOI as a globally adopted and maintained system of persistent and stable identifiers provides a solid basis for FAIR data, the first step towards new ways of doing science.

In Europe, within the field of natural sciences collections-related research, the implementation and use of persistent identifiers has already started (e.g. with CETAF stable identifiers for the physical objects). Identifiers for digital specimens as the digital twin of a physical specimen are a new development though, led by DiSSCo and developed through the DiSSCo linked projects. Early adoption of these identifiers is necessary for a timely start of constructing the DiSSCo infrastructure, for positioning of DiSSCo as RI of excellence in the global natural science collections community and in EOSC, and will ensure that when DiSSCo becomes operational in 2026, these identifiers are already widely adopted and implemented.

During the DiSSCo iGA2 meeting, 16 October 2020, an option analysis document for Digital Specimens Persistent Identifiers was presented. The document and supporting documentation can be found in Boardable and also here: bit.ly/DiSSCoPIDs. The Technical Team (TT) of the DiSSCo Coordination and Support Office (CSO) carried out the analysis based on recommendations included in the 'Conceptual Design Blueprint' document for the DiSSCo digitization infrastructure (produced by the ICEDIG project).

Of the 20+ options evaluated, the strongest and thus preferred option entails the use of Digital Object Identifiers (DOI) and working with the International DOI Foundation (IDF) to make these operational for DiSSCo ('Driven by DOI'). This option was chosen because of DOI/IDF achievements, implementation experience, and reputation in multiple industry sectors, as well as current DOI/IDF familiarity and uptake in the natural sciences community (already used by GBIF, in journal publications, etc.). DOI/IDF is compatible with the EOSC PID Policy and interoperability framework and the FAIR Guiding Principles. The system is standardized in International Standard ISO 26324:2012. DOI/IDF also appears to be financially viable and is likely to be globally acceptable.

Since selecting a persistent identifier scheme for digital specimens can have significant operational, technological, social, and financial implications across the infrastructure, the document was opened after iGA2 for further consultation with the DiSSCo and CETAF community, the DiSSCo Technical Advisory Board (TAB), and international stakeholders. All those actors gave their feedback. To guide this a set of open questions was provided in a survey (see Annex 2).

Feedback on the proposed option for Digital Specimens Persistent Identifiers was received from 20 stakeholders (where a group of people represented as a RI or a CETAF group counted as one stakeholder). Stakeholders included IGSN e.V., GBIF Secretariat, PLAZI, ELIXIR, DOI Foundation, Atlas of Living Australia/CSIRO, CNRI, UK Geological Curators Group, and LifeWatch ERIC. Responses from 3 CETAF groups (Earth Science Group, Legislation and Regulations Working Group, ISTC) were also received. Five responses to the survey came from individual DiSSCo participants. It is assumed that only a few members responded individually because many are already represented in the CETAF groups responses.

During iGA2 several participants expressed their interest to discuss the options analysis further, so the above mentioned consultation was followed by a seminar on 15 January 2021 (action point from iGA2), in which the consultation results were presented and further discussed with DiSSCo participants and other stakeholders. The seminar was well attended (64 participants). Notes from the session can be found here http://bit.ly/38Cmvpo. The seminar resulted in additional feedback from a wide set of interested parties, a summary of the outcomes can be found in Annex 1. The feedback indicates that DiSSCo should move forward with the proposal and provides guidance for points of attention when doing so.

Planned next steps, as already indicated in iGA2, are to:
  i.   Have discussions with IDF regarding unmet requirements (governance, operations, financing and architecture) to ensure long term sustainability;
  ii.   Become a general member of the International DOI Foundation (IDF);
  iii. Work with DiSSCo Prepare Task 6.2 partners to establish an experimental scheme of service to gain operational experience that will form the basis of a robust long-term plan for PID implementation, under the agreed governance and towards its sustainability; and,
  iv.   Have parallel discussions with the DiSSCo and natural sciences communities and (international) stakeholders with the aim of achieving global acceptance and adoption of the preferred scheme ('driven by DOI').

Considering the mentioned context, the DiSSCo iGA hereby instructs the DiSSCo Coordination and Support Office (CSO) to:

1. <u>Move forward with the proposed 'driven by DOI' solution</u>.
   Initiate DOI's registration and resolution services and procedures, with a custom metadata schema and cost model. Ideally, this can be in collaboration with international stakeholders: data aggregators such as GBIF, iDigBio, ALA, but also research infrastructures such as LifeWatch, ELIXIR, and infrastructure providers such as IDF and CNRI. Work with DiSSCo partners to facilitate this implementation by providing information (note that support in collection management systems would be advised but optional; operation of PID registration procedures as part of digitization processes by partners themselves would also be optional; and no changes would be required in curation of physical objects or their labels or barcodes).

2. <u>Become a general member of the International DOI Foundation (IDF)</u> using Naturalis Biodiversity Center as interim legal proxy and <u>for up to 3 years from 1 April 2021</u>, aiming at transitioning to a new DOI registration agency (RA) within those 3 years. The CSO shall work with key partners from our community (e.g. GBIF, iDigBio) to initiate a global membership RA model. Costs of IDF general membership should occur at the already negotiated fixed fee for each year (USD 10k for year one, USD 20k and USD 35k for year 2 and 3 respectively). An opt-out clause shall be included in the membership contract.

3. <u>Provide workshops or seminars</u>, for example organised through CETAF, addressing the community of practice to further explain the plans and to specifically tackle possible concerns for curatorial and management operations, as detected in the seminar. This will ensure uptake by a broader community.

Annex 1: Summary of outcomes from the consultation & seminar

- Most stakeholders fully support the proposal to use DOIs and several advantages were mentioned.
- Concerns with the proposal were about costs, attainability by small users and risk for wrong expectations: people may think this is a solution to problems it doesn't address such as a guarantee that DOIs will always resolve to data.
    - Regarding costs it was explained that every PID solution has a cost and since the proposal can leverage from existing DOI infrastructure provided by IDF and CNRI, the costs can be lower than with other solutions. Also, it will be a negligible fraction of total digitization costs and insignificant compared to DiSSCo infrastructure costs related to indexing, bandwidth and storage. Partnering with the International DOI Foundation (IDF) will be necessary though to establish an efficient cost model for large numbers of PIDs (a model different from that used by DataCite).
    - Attainability by small users is seen as an implementation issue. Individual researchers will not be charged for registering or using DOIs for digital specimens.
    - Availability of data is not solved by DOIs and is the shared responsibility of institutions and the DiSSCo RI.
- It was mentioned that different specialisms should be allowed to use the PIDs that reflect their needs and buy ins, but this is already part of the digital specimen design that allows for using any existing PID system to identify the physical specimen, including IGSNs or CETAF identifiers, etc.
- PIDs like DOIs allow for linking data and it was mentioned that this can be a risk for e.g. privacy and ABS. Every technological solution can be misused however and needs measures to safeguard from that. This is not a problem specific to the proposed solution.
- More work is needed to explain to users (curators, researchers) what will be the impact of the Digital Specimen concept in their daily work. This requires further work in DiSSCo Prepare project to work out how to implement community curation, storage, versioning etc.
- A successful implementation and global adoption will be a challenge, it requires careful planning and special attention should be paid to issues dealing with its everyday implementation, such as versioning, scalability, and community engagement.

Annex 2: Survey questions for the consultation on Digital Specimens Persistent Identifiers (PIDs) for the operation of DiSSCo

* 1. What type of stakeholder are you representing?
- A DiSSCo member
- A CETAF group
- A GBIF Node
- Other (please specify)

* 2. Name of your organisation or group

* 3. Your name

* 4. Your email address

5. How do you think a 'Powered by DOI' solution for digital specimens can be achieved for DiSSCo? What are the main risks you see?

6. A 'Powered by DOI' solution for digital specimens could be used around the world (i.e., promoted for global adoption). What are your suggestions to make this achievable?

7. How can people (scientists, collectors, taxonomists, curators, data managers, analysts, etc.) be encouraged to routinely use DOIs for digital specimens e.g., when writing about or referring to specimens? What incentives would help?

8. What do you think will be the impact on research of a 'Powered by DOI' approach for digital specimen?

9. What do you think will be the impact on data sharing and data curation of a 'Powered by DOI' approach for digital specimen?

10. Are there other comments or thoughts you would like to share with us about the proposed 'Powered by DOI' approach?

* 11. Data processing consent
- I agree with my information to be processed by the DiSSCo Coordination and Support Office and Technical Team

12. Data sharing consent
- I agree with public sharing of my (anonymized) feedback.