

Conference Abstract

Making Heterogeneous Specimen Data 'FAIR': Implementing a digital specimen repository

Abraham Nieva de la Hidalga[‡], Alex Hardisty[§]

[‡] Cardiff University School of Computer Science and Informatics, Cardiff, United Kingdom

[§] Cardiff University, Cardiff, United Kingdom

Corresponding author: Abraham Nieva de la Hidalga (nievadelahidalga@cardiff.ac.uk)

Received: 12 Jun 2019 | Published: 18 Jun 2019

Citation: Nieva de la Hidalga A, Hardisty A (2019) Making Heterogeneous Specimen Data 'FAIR': Implementing a digital specimen repository. Biodiversity Information Science and Standards 3: e37163.

<https://doi.org/10.3897/biss.3.37163>

Abstract

The definition of a digital specimen is proposed to encompass the digital representation(s) of physical specimens from natural science collections. The digital specimen concept is intended to define a representation (digital object) that brings together an array of heterogeneous data types, which are themselves alternative physical specimen representations. In this case, the digital specimen (DS) holds references to specimen data from a collection management system, images, 3D models, research articles, DNA sequences, collector information, among many other data types. The proposal is to create persistent relationships between the DS and other categories of digital objects (e.g. resource types mentioned above, collections, storage platforms, organisations, databases, and provenance data). Complying with FAIR data principles (findability, accessibility, interoperability, and reuse), i.e., achieving data 'FAIRness', eases data integration, which is needed for cross-disciplinary linking and combination of data from different domains, making the DS as a comprehensive package of information about a specimen.

Implementation and access to a digital specimen repository (DSR) as a Digital Object Architecture (Sharp 2016) component demonstrates the alignment of the DS concept and FAIR data principles (Wilkinson et al. 2016, Kahn and Wilensky 2006). The DSR fulfills four roles: data producer, resource manager, data publisher, and collaboration space. As data producer, the DSR allows acquisition and curation (indexing, storage) of DSs linking

primary data, models, analyses, and other digital object types. As resource manager, the DSR manages access to distributed platforms, ranging from acquisition networks (digitisation stations, museums, herbariums) to processing services, advanced computational resources, data asset storage systems, and specialised servers. As data publisher, the DSR provides access to data assets from national and transnational data archives. As collaboration space, the DSR supports users' accessing, sharing and (re)using data assets, and derived data products and services. Adopting the collaboration space and data publisher roles, the DSR implements interfaces that expose the DSs to the research community, fulfilling the FAIR findability, accessibility, and reuse principles. Adopting the data producer and resource manager roles, the DSR creates meaningful and persistent relationships required to link DSs and other types of digital objects, fulfilling the FAIR interoperability principle.

A prototype DSR based on the Cordra digital object repository has been deployed (Corporation for National Research Initiatives (CNRI) 2018, Reilly and Tupelo-Schneck 2010). The advantages of Cordra are: rapid deployment, customisable object model, creation of relations between digital objects, and application program interfaces for programmatic access. Rapid deployment of the DSR provides a tangible target for discussing the implementation of the DS concept. The customisable object model enables the refinement and enhancing of the definition of DS in response to feedback from colleagues who have accessed the DSR and used its contents. Creating relations between digital objects enables flexible linking to digital objects stored in different repositories. Accessing the DSR programmatically through APIs enables extending the use of the repository in different platforms (e.g. mobile devices) as well as integration with other repositories and services. As well as supporting a HTTP-oriented API, Cordra implements Digital Object Interface Protocol (DONA Foundation 2018), allowing the definition of operations to act directly on selected DSs in the repository.

The DSR prototype has been demonstrated by providing access to the repository administrative interface and with a custom interface designed to facilitate access by different user groups, such as collection curators, researchers, teachers, and students. The client interface has been designed to demonstrate a subset of the functionalities derived from user stories, which describe software features from the end-user perspective. Demonstrating the DSR capabilities as proposed, will inform the refinement of the design of the DS model and provide early feedback about the needed software features.

Keywords

digital specimen repository, digital specimen, natural history collection, digitisation, FAIR

Presenting author

Abraham Nieva de la Hidalga

Presented at

Biodiversity_Next 2019

Funding program

Horizon 2020 Framework Programme of the European Union

Grant title

ICEDIG – “Innovation and consolidation for large scale digitisation of natural heritage”
H2020-INFRADEV-2016-2017 – Grant Agreement No. 777483

Hosting institution

Cardiff University

References

- Corporation for National Research Initiatives (CNRI) (2018) Cordra digital object repository and registry software. <https://www.cordra.org>. Accessed on: 2019-2-04.
- DONA Foundation (2018) Digital object interface protocol specification. https://www.dona.net/sites/default/files/2018-11/DOIIPv2Spec_1.pdf. Accessed on: 2018-11-12.
- Kahn R, Wilensky R (2006) A framework for distributed digital object services. International Journal on Digital Libraries 6 (2): 115-123. <https://doi.org/10.1007/s00799-005-0128-x>
- Reilly S, Tupelo-Schneck R (2010) Digital object repository server. D-Lib Magazine 16 <https://doi.org/10.1045/january2010-reilly>
- Sharp C (2016) Overview of the Digital Object Architecture (DOA). <https://www.internetsociety.org/resources/doc/2016/overview-of-the-digital-object-architecture-doa/>. Accessed on: 2019-4-02.
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 URL: <https://www.nature.com/articles/sdata201618>