From Physical Specimen to FAIR Digital Object: Building DiSSCo's Infrastructure Vision

Sharif Islam, https://orcid.org/0000-0001-8050-0299
Sam Leeflang, https://orcid.org/0000-0002-5669-2769
Soulaine Theocharides, https://orcid.org/0000-0001-7573-4330
Wouter Addink, https://orcid.org/0000-0002-3090-1761

Since the 1990s, a global movement has emerged to digitise natural science collections and make specimen data accessible for research, policy, and education (Hedrick et al. 2020). What began as isolated local databases and museum catalogues in the 1980s has evolved into a network of data resources. The Distributed System of Scientific Collections (DiSSCo) represents a new stage in this evolution: the convergence of technical and organisational infrastructures around the concept of the FAIR Digital Object (FDO).

This presentation traces the conceptual development from digitisation and cataloguing to FAIR-by-design infrastructure thinking. While many technical components existed before DiSSCo (digitisation workflows, metadata standards, persistent identifiers, and data aggregation), the missing piece has been a conceptual and architectural framework that not only makes natural science collections data FAIR but also provides an infrastructure capable of ensuring governance and sustainability (Koureas et al. 2023). DiSSCo demonstrates how this gap can be filled by translating FAIR principles into an operational infrastructure design that ensures trust, provenance, and machine actionability.

The DiSSCo vision builds upon a series of initiatives and standards that have progressively shaped digital collection management and data sharing: from TDWG's biodiversity data standards (established in 1985), to large-scale digitisation efforts around the world (early 2010s), and the formation of the Global Biodiversity Information Facility (GBIF) and its subsequent data mobilisation activities (starting in 2001), to the publication and adoption of the FAIR Principles (2016).

At the heart of this vision lies the concept of the Digital Specimen, a digital surrogate acting as a citable and actionable proxy for a physical specimen. This builds on the Digital Extended Specimen (DES) concept (Hardisty et al. 2022; Islam et al. 2023) and operationalises it through the FAIR Digital Object framework (De Smedt et al. 2020). The FDO abstraction allows any entity to be uniquely identified via persistent identifiers (PIDs) and described in a structured, machine-actionable way. A key aspect of FDOs is that they are self-contained objects, meaning all information and relationships about the object are contained within them. This self-contained design is particularly important for Digital Specimens, as it enables linkage to related and derived data such as media, publications and molecular sequences while maintaining the integrity and autonomy of each specimen as a unit of scientific knowledge.

To achieve this, DiSSCo developed the open Digital Specimen (openDS) specification, a comprehensive data model that builds on existing community standards such as Darwin Core and its extensions, Schema.org, and W3C PROV-O. openDS serves as the practical implementation of FDO principles, embedding mechanisms for provenance, versioning, and annotation while ensuring full interoperability with existing biodiversity data standards. It

represents a common language through which Digital Specimens can be shared, cited, and linked across global infrastructures (Leeflang et al. 2023). The openDS framework also draws on Linked Data principles and JSON-LD, using defined contexts and semantics to ensure machine readability and consistency.

The development of DiSSCo's infrastructure follows a long-term and incremental trajectory, supported by multiple European projects that provided the resources and coordination to transform conceptual designs into operational components, beginning with the PID infrastructure (Addink et al. 2020). Building on this foundation, the DiSSCo Transition project defined the Minimum Viable Product for DiSSCo's core infrastructure. This work advanced the technical maturity of the system to Technology Readiness Level 6 and demonstrated integration through the DiSSCover annotations portal and the sandbox environment (Addink and Leeflang 2024; Addink et al. 2025).

Importantly, scale is now being realised. As of October 2025, DiSSCo has created over five million digital specimens and digital media objects (other object types are also defined; see schema.dissco.tech), each resolvable via DOIs or Handles, with plans to grow to 20 million combined digital specimens and media objects in the next phase (DiSSCo Tech Blog, 19 August 2025). This milestone underlines that the framework is not only conceptual but already operational at scale, and it underscores the urgency of sustained infrastructure investment when facing the estimated 1.5 billion specimens held in European collections (Johnson et al. 2023; Ahl et al. 2023; Hannco et al. 2018).

By leveraging FDOs, DiSSCo contributes to the digital transformation of natural science collections and the broader ambition of FAIR and AI-ready science. The Digital Specimen implementation of FDO serves as a showcase for how natural science objects evolve from data records into infrastructure entities bridging the physical, digital, and conceptual layers of scientific knowledge.

References

Addink, W., Islam, S., Dillen, M., Güntsch, A. and Theocharides, S., 2023. Deliverable D7. 1 Architecture Design for a pan-European PID system for Digital Specimens. ARPHA Preprints, 4, p.e107168.

Addink, W. and Theocharides, S., 2024. The future of referencing specimens is near: Cite the digital specimen DOI. Biodiversity Information Science and Standards, 8, e137534.

Addink, W., Koureas, D., Hardisty, A., Casino, A., Islam, S., Lannom, L., and Wittenburg, P., 2020. DiSSCo Conceptual Design Blueprint. Research Ideas and Outcomes, 6, e54280.

Addink, W. and Leeflang, S., 2024. Definition of a TRL6 Minimum Viable Product of Core Infrastructure – Milestone 14. DiSSCo Transition Project Deliverable. https://doi.org/10.5281/zenodo.15632942

Addink, W., et al. Minimum Viable Product of Core Infrastructure - Deliverable 3.1. v1.1, DiSSCo Transition Project, 27 May 2025, https://doi.org/10.5281/zenodo.15230656.

Ahl, L.I., Bellucci, L., Brewer, P., Gagnier, P.Y., Haston, E.M., Livermore, L., De Smedt, S., Hardy, H.M. and Enghoff, H., 2023. Digitisation of natural history collections: criteria for prioritisation. Research Ideas and Outcomes, 9, p.e114548.

De Smedt, K., Koureas, D., and Wittenburg, P., 2020. FAIR digital objects for science: From data pieces to actionable knowledge units. Publications, 8(2), 21.
Hardisty, A. et al., 2022. The Digital Extended Specimen concept. BioScience, 72(10), 978–988.

DiSSCo Tech Blog https://dissco.tech/2025/08/19/five-million-digital-specimens-and-images-in-dissco-millions-more-to-come/

Hardisty A, Saarenmaa H, Casino A, Dillen M, Gödderz K, Groom Q, Hardy H, Koureas D, Nieva de la Hidalga A, Paul DL, Runnel V, Vermeersch X, van Walsum M, Willemse L (2020) Conceptual design blueprint for the DiSSCo digitization infrastructure - DELIVERABLE D8.1. Research Ideas and Outcomes 6: e54280. https://doi.org/10.3897/rio.6.e54280

Hardisty, A.R., Ellwood, E.R., Nelson, G., Zimkus, B., Buschbom, J., Addink, W., Rabeler, R.K., Bates, J., Bentley, A., Fortes, J.A. and Hansen, S., 2022. Digital extended specimens: Enabling an extensible network of biodiversity data records as integrated digital objects on the internet. BioScience, 72(10), pp.978-987.

Hannco, B., Willemse, L., van Egmond, E., Casino, A., Gödderz, K. and Vermeersch, X., 2018. ICEDIG Deliverable D2. 1-Inventory of current criteria for prioritization of digitization.

Hedrick, B.P., Heberling, J.M., Meineke, E.K., Turner, K.G., Grassa, C.J., Park, D.S., Kennedy, J., Clarke, J.A., Cook, J.A., Blackburn, D.C. and Edwards, S.V., 2020. Digitization and the future of natural history collections. BioScience, 70(3), pp.243-251.

Islam, S., Beach, J., Ellwood, E.R., Fortes, J., Lannom, L., Nelson, G., and Plale, B., 2023. Assessing the FAIR digital object framework for global biodiversity research. Research Ideas and Outcomes, 9.

Johnson, K.R., Owens, I.F. and Global Collection Group, 2023. A global approach for natural history museum collections. Science, 379(6638), pp.1192-1194.

Koureas, D., Livermore, L., Alonso, E., Addink, W. and Casino, A., 2023. DiSSCo Prepare Project: Increasing the Implementation Readiness Levels of the European Research Infrastructure. Research Ideas and Outcomes, 9, p.e107220.

Leeflang, S. and Addink, W., 2023. Harmonised Data is Actionable Data: DiSSCo's solution to data mapping. Biodiversity Information Science and Standards, 7, 10.

Leeflang, S., Weiland, C., Grieb, J., Dillen, M., Islam, S., Fichtmüller, D., Addink, W. and Haston, E., 2022. DiSSCo Prepare Deliverable D6. 2-Implementation and construction plan of the DiSSCo core architecture.

Theocharides, S. et al., 2024. Cooking with DOIs: A Recipe for FAIR Digital Specimens Using FDOs. Open Conference Proceedings, 5.