DiSSCo

Distributed System of Scientific Collections

openDS

open Digital Specimens

Alex Hardisty
26 April 2021

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

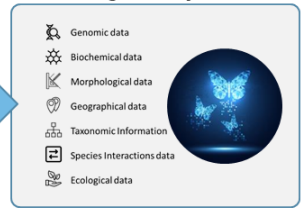
Update for CETAF ISTC & DWG, April 2021

DISSCO — Distributed System of Scientific Collections

DIGITAL SPECIMENS# (DS) PROVIDE AN ANCHORING FUNCTION FOR ALL KINDS OF DATA FROM PHYSICAL SPECIMENS

Physical Object

Digital Specimen
digital object



An actionable **information unit** on the Internet. FAIR by default

Data locked up in physical specimens is released through digitization, analytical, and computational methods.

Digital Specimens (DS) act as processable twins on the Internet for physical specimens in collections.

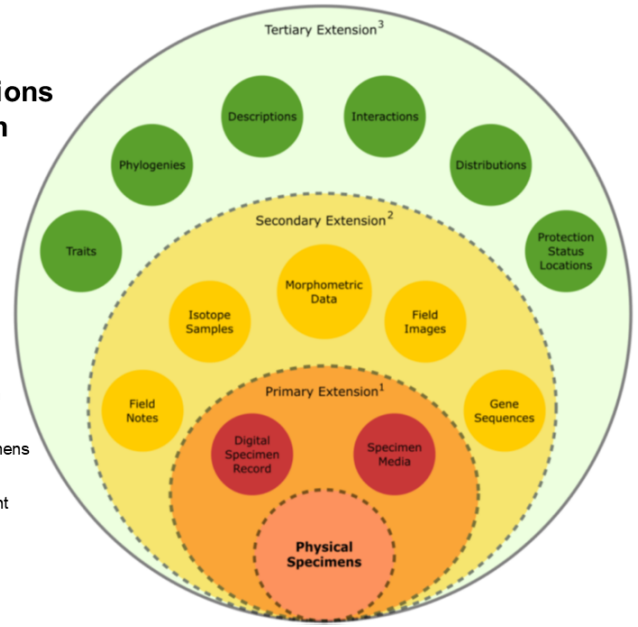
They can be manipulated remotely across a network by machines and humans.

What is a Digital Specimen? https://bit.ly/DigitalSpecimen.
Also, Lannom et al., 2020 https://doi.org/10.1162/dint_a_00034.

BIODIVERSITY COLLECTIONS NETWORK

Extending U.S. Biodiversity Collections to Promote Research and Education

"Extended specimens will form the core of a powerful Extended Specimen Network, or ESN"



Physical specimen

1 Primary Extension:
Digitized metadata/CMS record + images, etc. held by collection

2 Secondary Extension:
Data held by institution. Often disconnected from original specimens

3 Tertiary Extension:
Data in other repositories, linked by e.g., taxon or gathering event

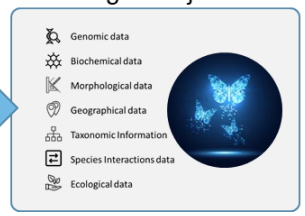
Lendemer et al. (2020) The extended specimen network: A strategy to enhance US biodiversity collections, promote research and education. doi: 10.1093/biosci/biz140.

DISSCO
Distributed System of Scientific Collections

DIGITAL SPECIMENS# (DS) PROVIDE AN ANCHORING FUNCTION FOR ALL KINDS OF DATA FROM PHYSICAL SPECIMENS

Physical Object

Digital Specimen
digital object



An actionable **information unit** on the Internet. FAIR by default

Data locked up in physical specimens is released through digitization, analytical, and computational methods.

Digital Specimens (DS) act as processable twins on the Internet for physical specimens in collections.

They can be manipulated remotely across a network by machines and humans.

What is a Digital Specimen? https://bit.ly/DigitalSpecimen. Also, Lannom et al., 2020 https://doi.org/10.1162/dint_a_00034.

BIODIVERSITY COLLECTIONS NETWORK

Extending U.S. Biodiversity Collections to Promote Research and Education

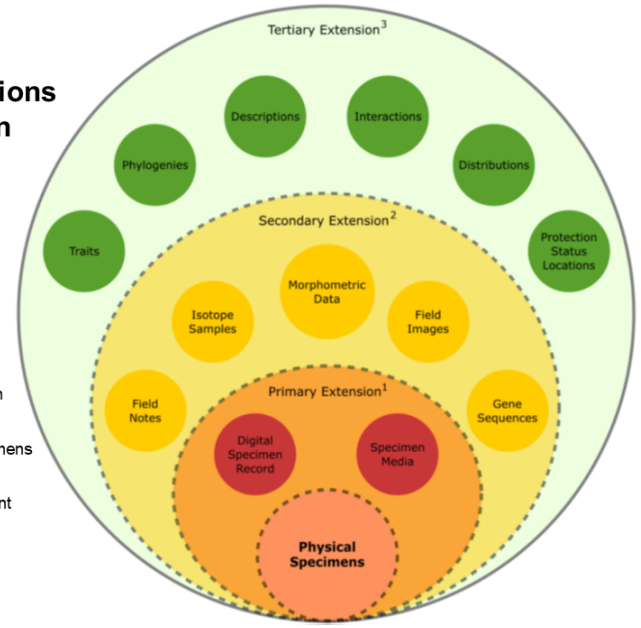
"Extended specimens will form the core of a powerful Extended Specimen Network, or ESN"

Physical specimen

1 Primary Extension:
Digitized metadata/CMS record + images, etc. held by collection

2 Secondary Extension:
Data held by institution. Often disconnected from original specimens

3 Tertiary Extension:
Data in other repositories, linked by e.g., taxon or gathering event



Lendemer et al. (2020) The extended specimen network: A strategy to enhance US biodiversity collections, promote research and education. doi: 10.1093/biosci/biz140.

Global consultation: Converging Digital Specimens and Extended Specimens - Towards a global specification for data integration

http://bit.ly/consultdes

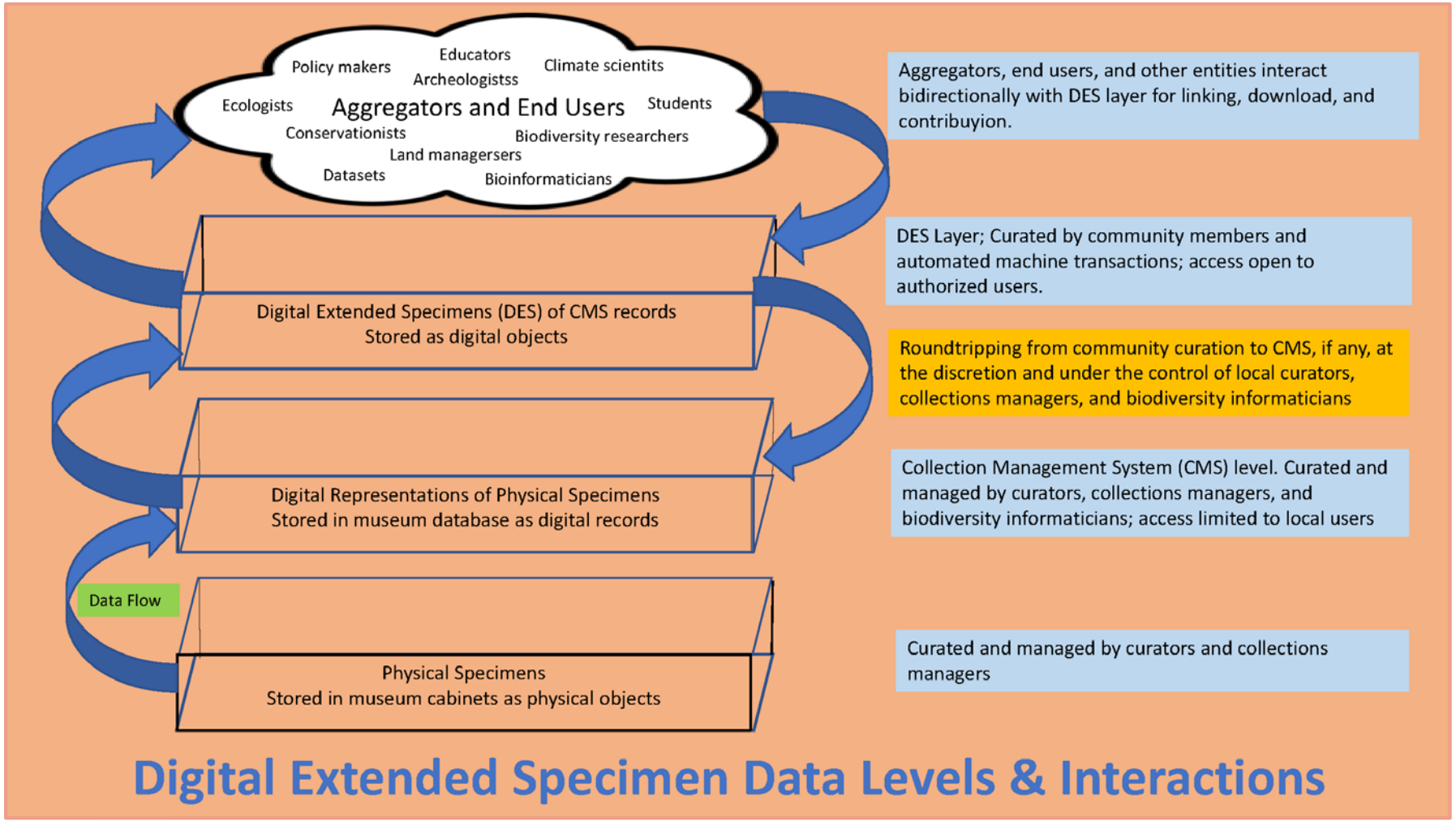
Phase 1 topics:

- Making FAIR data for specimens accessible
- Extending, enriching and integrating data
- Annotating specimens and other data
- Attributing work done
- Analysing / mining specimen data for novel applications

Phase 2 topics (May/June, * moderators needed):

- Well-founded access points and data cyberinfrastructure alignment
- Persistent identifier (PID) scheme(s)
- Meeting legal/regulatory, ethical and sensitive data obligations *
- Workforce capacity development and inclusivity *
- Transactional mechanisms and provenance
- Partnerships to collaborate more effectively

OUTCOME OF PHASE 1: STRONG CONSENSUS. POTENTIAL FOR CONVERGENCE OF DS & ES IDEAS. ACHIEVABLE.



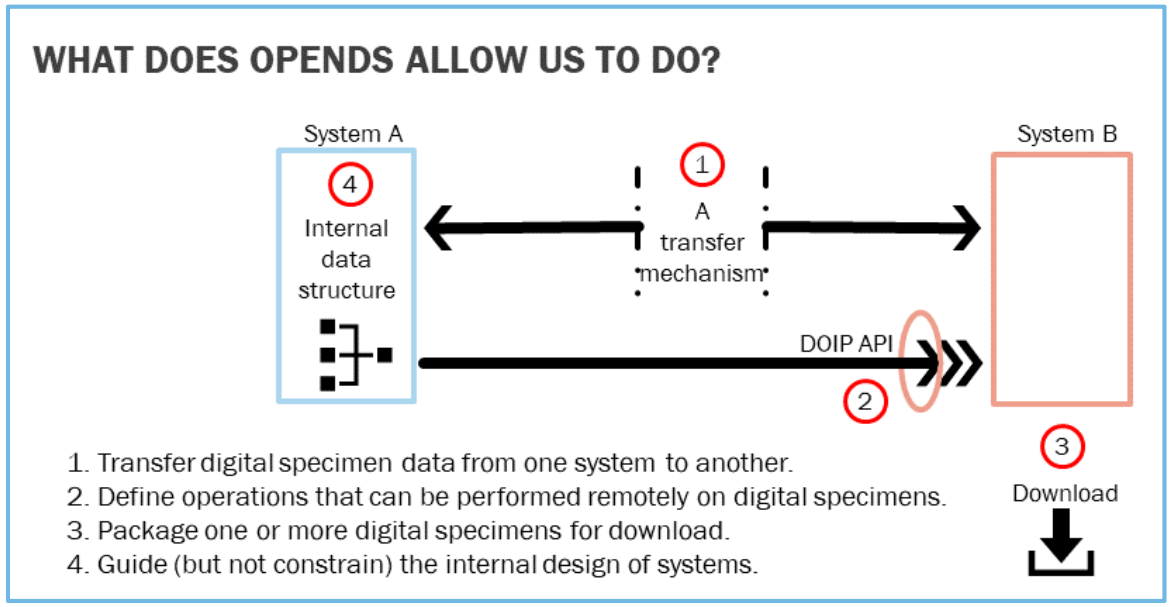
Work underway on two explanatory articles

1. Towards a common concept for Digital extended Specimens on the Internet.

2. Beyond Collection Management Systems: A Global Digital Object Architecture Design to Address Extended Specimen Based Research

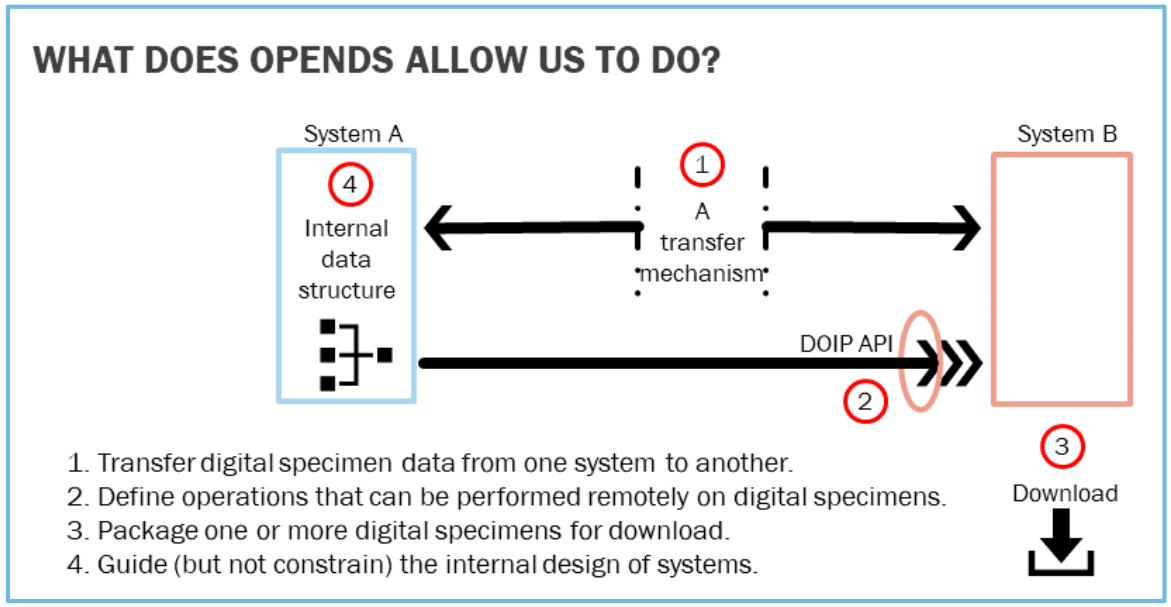
Source: Gil Nelson (University of Florida, iDigBio). Final version will be improved and corrected.

OPENDS,



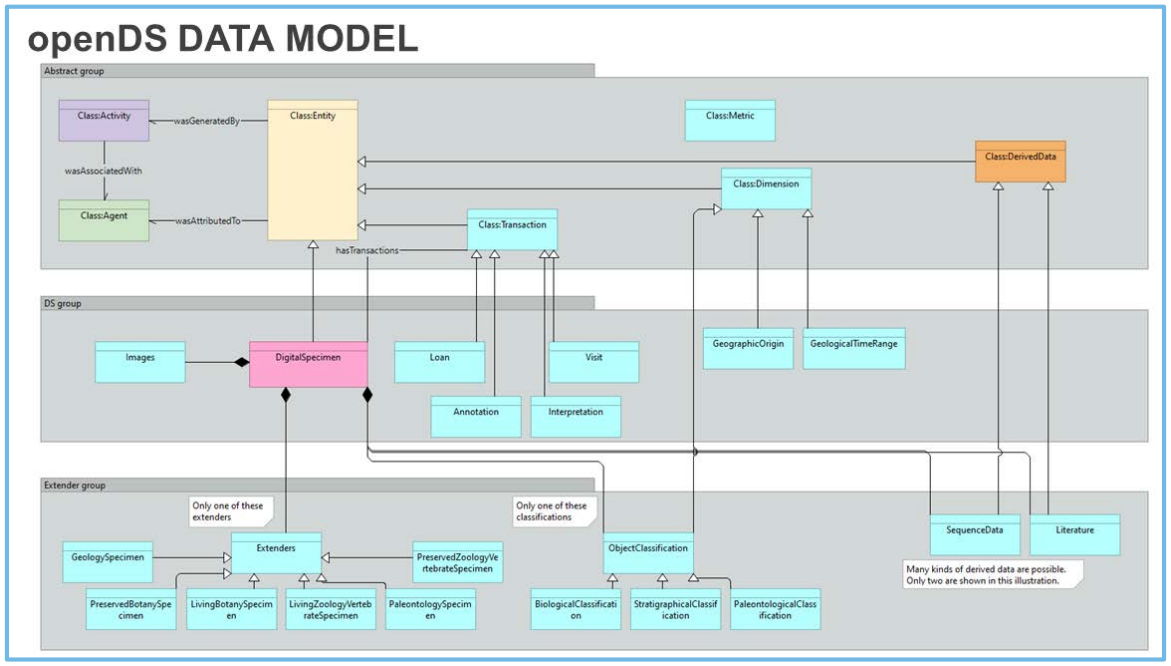
- Standardized for exchange between computer systems and for interoperability between software modules. Standardized to allow operations to act on DS remotely and to allow machines to process DS, as well as humans.
- FAQ: http://bit.ly/opendsfaq.
- Github: https://github.com/DiSSCo/openDS

OPENDS, PAST 12 MONTHS



- Standardized for exchange between computer systems and for interoperability between software modules. Standardized to allow operations to act on DS remotely and to allow machines to process DS, as well as humans.
- FAQ: http://bit.ly/opendsfaq.
- Github: https://github.com/DiSSCo/openDS

(SEEKING CONTRIBUTIONS)



- Positioning openDS relative to existing: Darwin Core and ACBD/EFG, RDA Rec. Attribution Metadata & W3C PROV, OBO Foundry ontologies, heritage sector stds (CIDOC CRM)
- First JSON schemas appearing

Now need two running implementations to interoperate with one another

THANK YOU

- To contribute to the work: https://github.com/DiSSCo/openDS
- To become involved, email: hardistyar@cardiff.ac.uk

- **Questions? Discussion.**