DiSSCo
Distributed System of Scientific Collections

11 September 2019

Implementing FAIR Data for People and Machines: Impacts and Implications

Larry Lannom

Corporation for National Research Initiatives

Natural Science Collections: The Data Challenge

- Existing biological and geological specimens widely distributed across museums, universities, botanical gardens, etc.
- Digital information held in widely variant and largely disconnected systems
- Physical access expensive and slow
- None to few explicit links from specimens to DNA sequences, literature, ecosystem and medicinal/chemical data, ongoing census observations
- Result is significant under-utilization of an enormously important existing set of resources that are growing in importance

DiSSCo -Research Infrastructure (RI) for Natural Science Collections

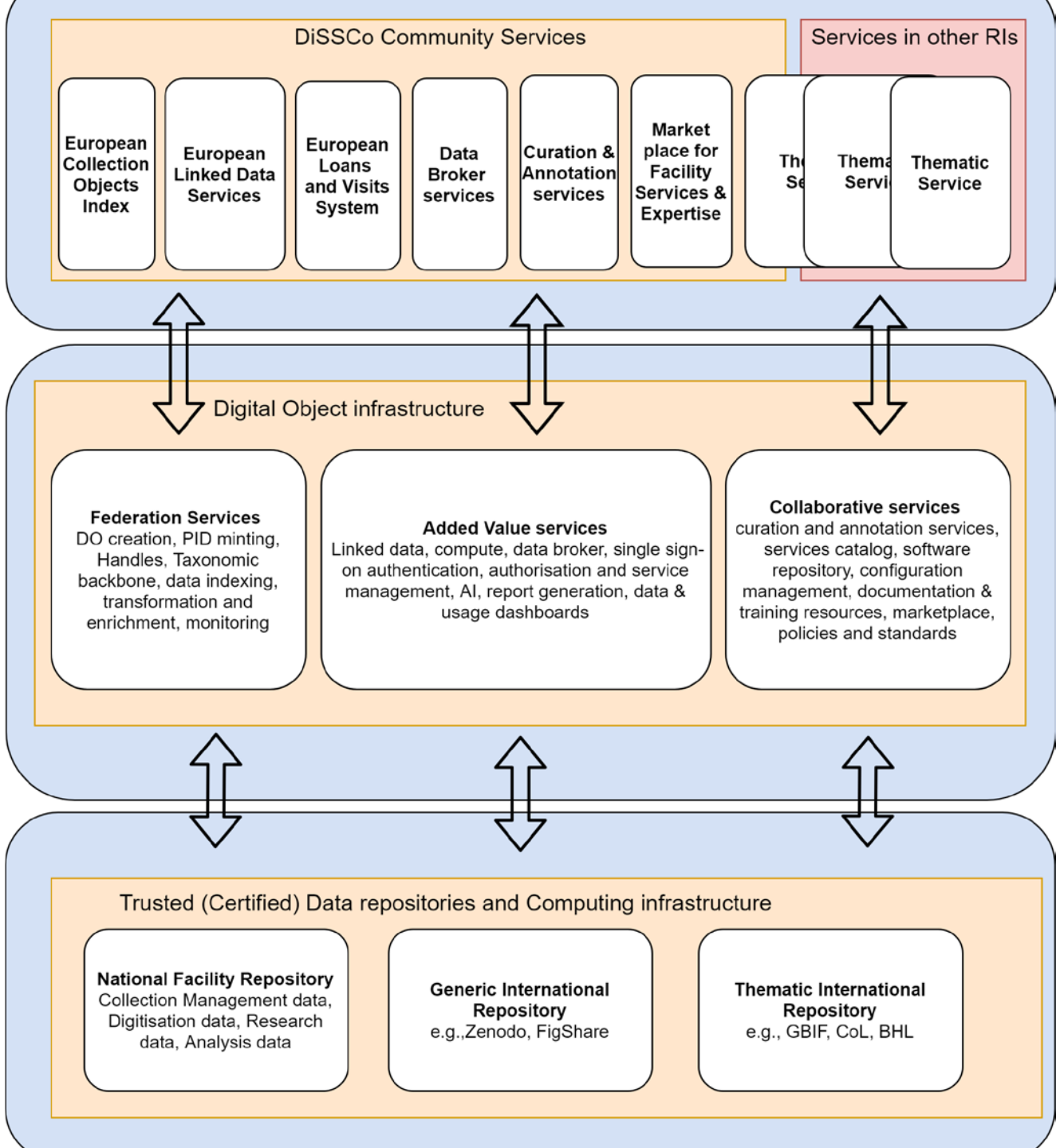
- Digital unification of all European natural science assets under common curation and access policies and practices that aim to make the data easily Findable, more Accessible, Interoperable and Reusable (FAIR)
- 1.5 Billion specimens
- 119 Collaborating Institutions
- 5000 Scientists
- 21 Countries
- ‘DiSSCo Prepare’ just funded
  - 3 years and 31 partners
  - Deliver organizational, financial and technical guiding framework for the construction of the infrastructure

The three building blocks of the
DiSSCo technical infrastructure:

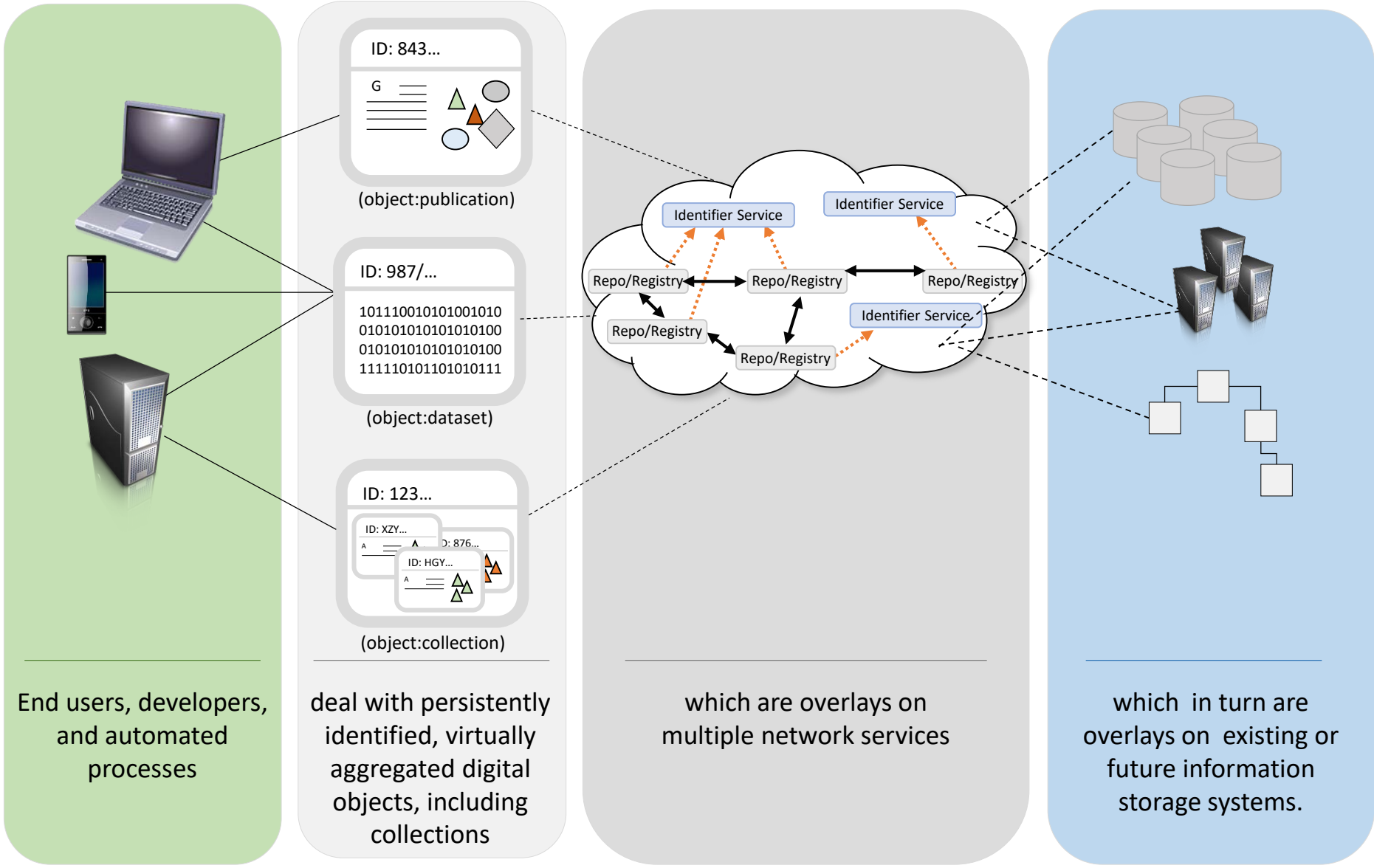
3. Community Services

2. Digital Object Infrastructure

1. Repositories with data



Global Digital Object Cloud (GDOC)



End users, developers, and automated processes

deal with persistently identified, virtually aggregated digital objects, including collections

which are overlays on multiple network services

which in turn are overlays on existing or future information storage systems.

Where Are We Now and What is Needed?

- Achieving FAIR not optional
  - Too many societal problems need good science now
- Good news: FAIR widely accepted
- PIDs and other levels of indirection becoming an accepted part of the landscape
- Need to solve and get past the technical issues and hand-off the research infrastructures to the researchers
  - In other IT regimes, developers can achieve thorough understanding of the issues; this is the not case with science
  - Research Infrastructures must be usable at a high level by researchers – not every small variation should need a developer
  - Tooling must allow researchers to script work flows, define data models, and in general drive and manage the infrastructure

Science is a ‘light’s better’ endeavour in that research effort is **not directed at areas where the work is technically infeasible**. Research is directed where real, interpretable results may be obtained.

We do, in fact, conduct research where the light’s better.

But, when the light changes, so does science.

With better illumination, we look in new areas.

We find new things…