Digital preservation tools and services: DC-Net overview report

Raivo Ruusalepp

Estonian Business Archives

8.3.2012

DC-Net Final Conference

Topics

- Digital preservation models and how they are changing
- Summary of DP tools and services
- Conclusions and recommendations

I. Preservation Models

Repository-centric View of DP

- We are used to implicitly assuming that digital archiving and preservation require a digital repository system
- Supported by the OAIS digital archive model (ISO 14721:2003) and numerous institutional repository system software products
- In real life, not all repositories are alike, created for the same purpose or delivering a similar range of services
- Not all of them even aim to preserve the content they are holding
- There is a feeling (or perhaps a concept) that the underlying active digital preservation involves the same or similar activities in all the different types of repositories

Weakness of in-house preservation systems

The model of long-term digital preservation where every single memory institution develops its own stand-alone system is not:

•Economically viable

•Interoperable (instead fragmented „digital silos")

•Sustainable in the longer term

Excessive fragmentation does not offer economies of scale …

We should aim for shared solutions instead

DCC Digital Curation Lifecycle Model

Source: UK Digital Curation Centre
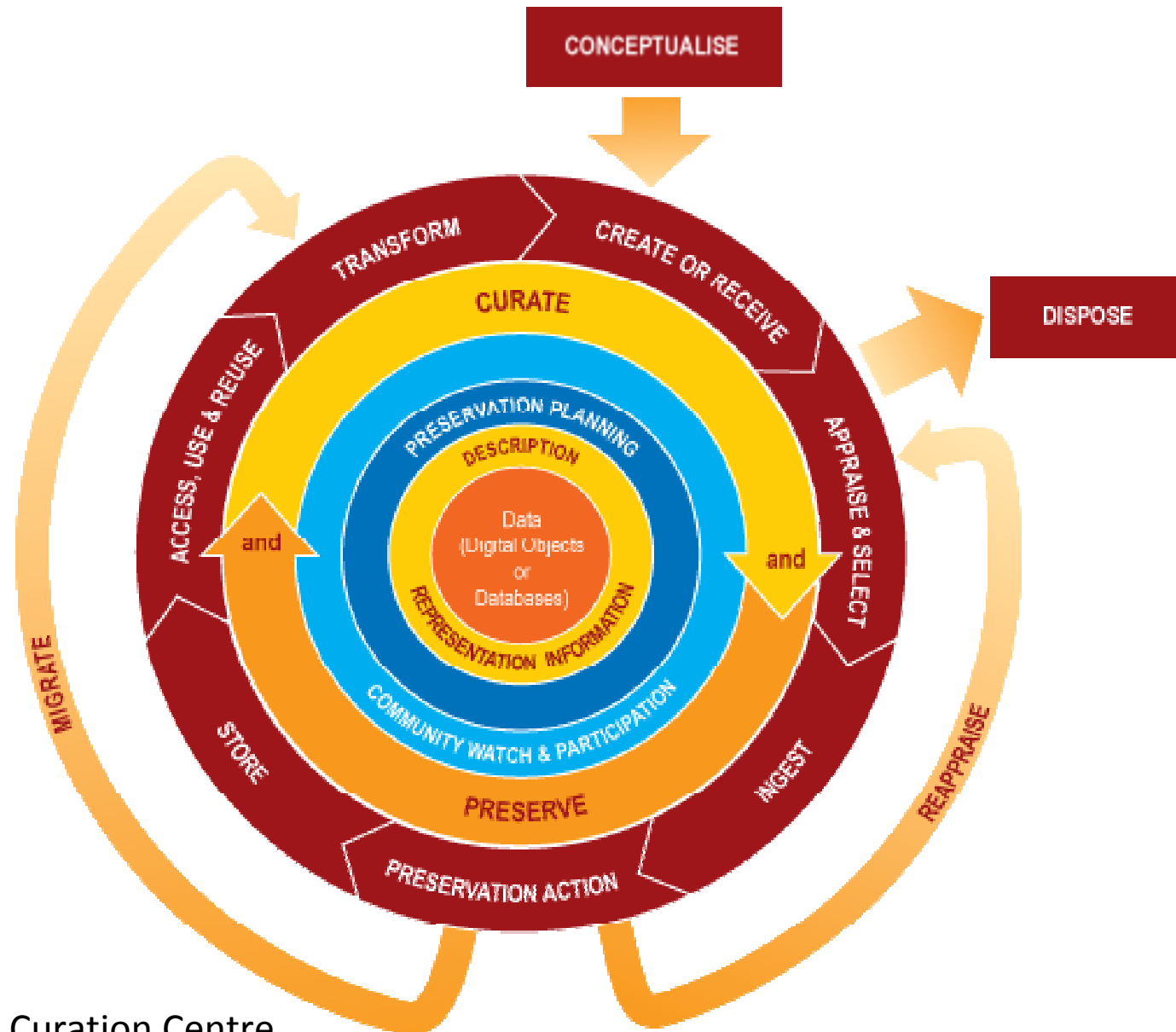
II. New trends in digital preservation solutions

New trends within the DP domain

- Preservation-ready systems (Borbinha 2010)
- Integration of preservation functionality into different types of information systems
- Self-preserving objects or durable objects (Billenness 2011; Gladney 2010)
- Federated, co-operative digital archiving systems and repositories (LOCKSS, MetaArchive, DISTARNET)
- Use of cloud and grid infrastructures (Askhoj, Nagamori & Sugimoto 2011)

http://cordis.europa.eu/fp7/ict/telearn-digicult/digicult-future-digital-preservation_en.html

Influences from the IT domain

- Service oriented architectures (SoA)
  - API (Application Programming Interfaces)
  - Web services
- Service platforms
  - GRID
  - Cloud
  - SaaS (software as a service)
  - PaaS (platform as a service)
  - IaaS (infrastructure as a service)
- Microservices

Gartner Hype Cycle of Emerging Technologies (July 2011)
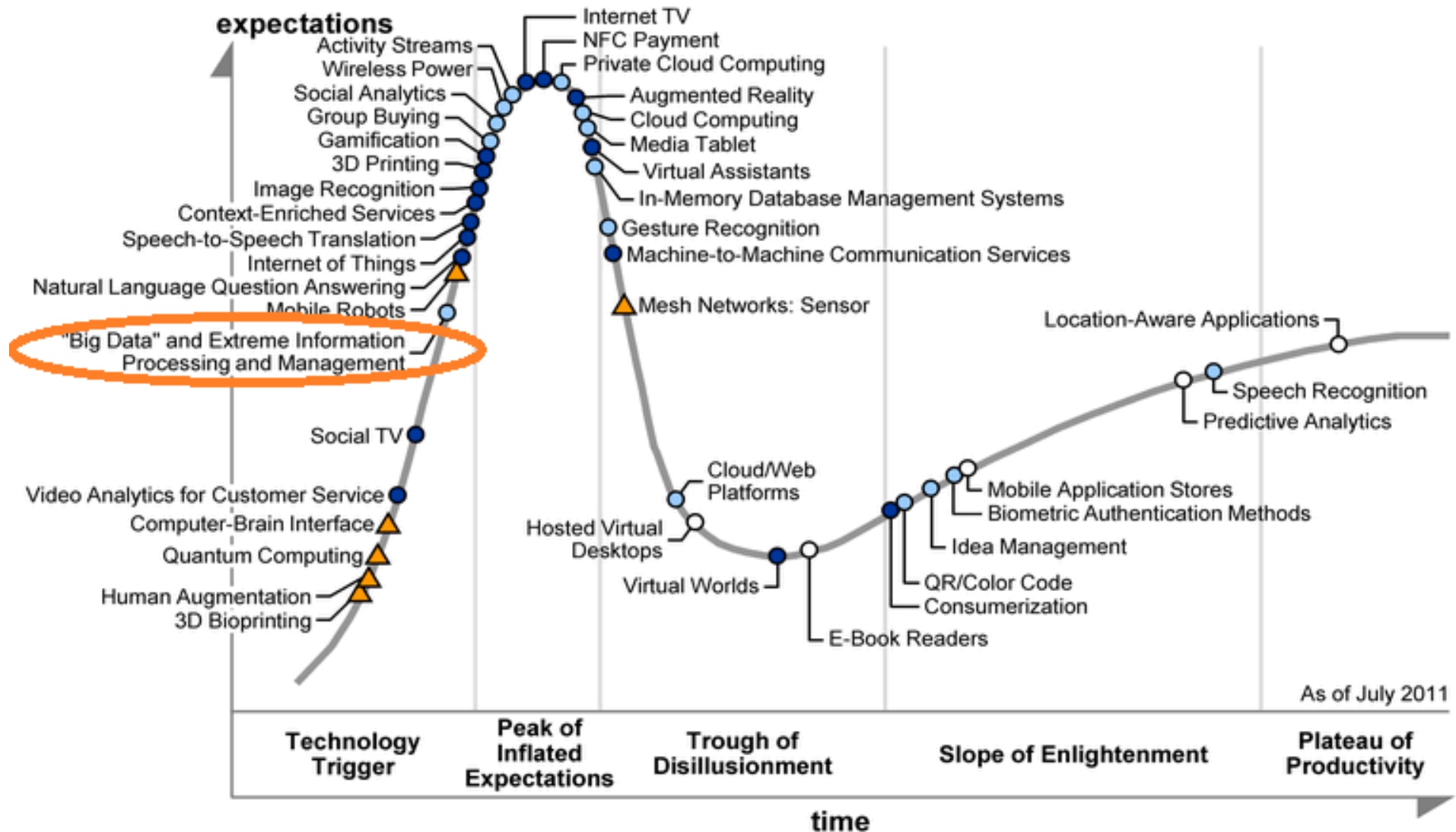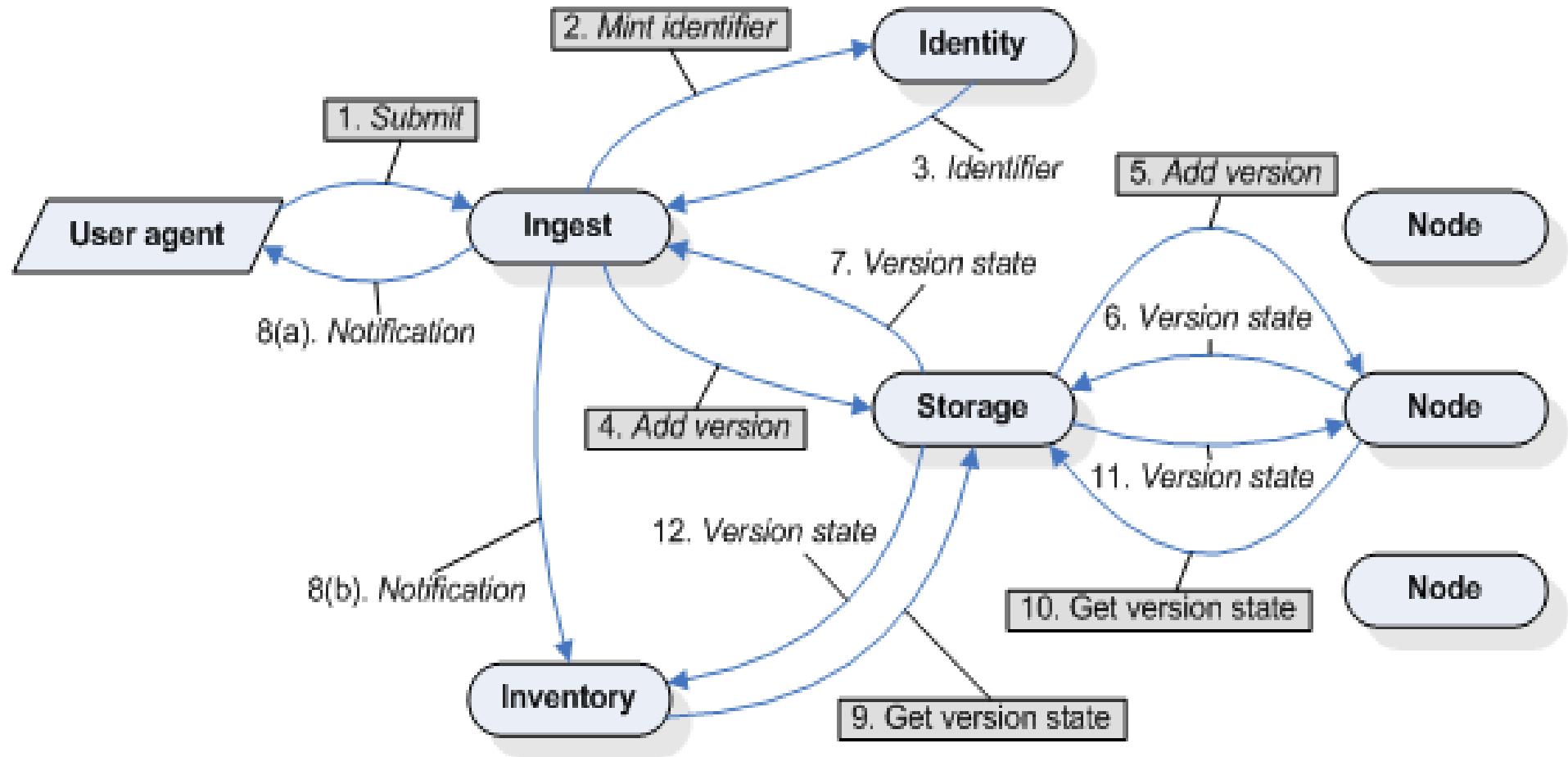
Microservices in Digital Preservation

- Micro-services are an approach to digital curation based on devolving curation function into a set of independent, but interoperable, services that embody curation values and strategies. Since each of the services is small and self-contained, they are collectively easier to develop, deploy, maintain, and enhance. Equally as important, they are more easily replaced when they have outlived their usefulness. Although the individual services are narrowly scoped, the complex function needed for effective curation emerges from the strategic combination of individual services.

[Curation Micro Services website]

Example of DP Microservices

UC Curation Center (UC3)

III. Instruments (Tools and Services) for Digital Preservation

Instruments

Tools

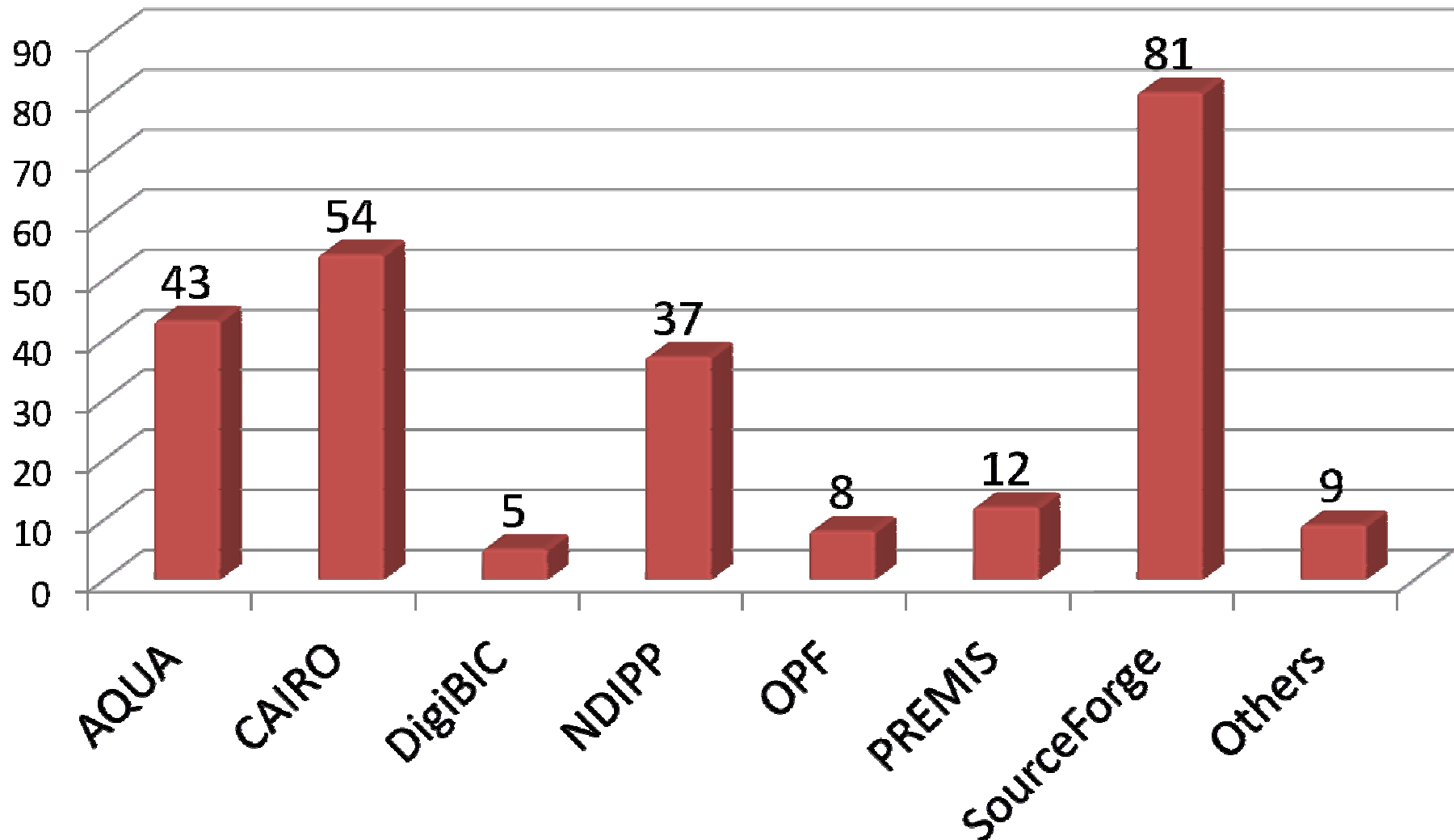•Software that has been developed in support of digital preservation tasks and made (publically) available

Services

•Unitary software components for which there is a body which offers customer support, there are clearly defined access conditions, and user documentation is available
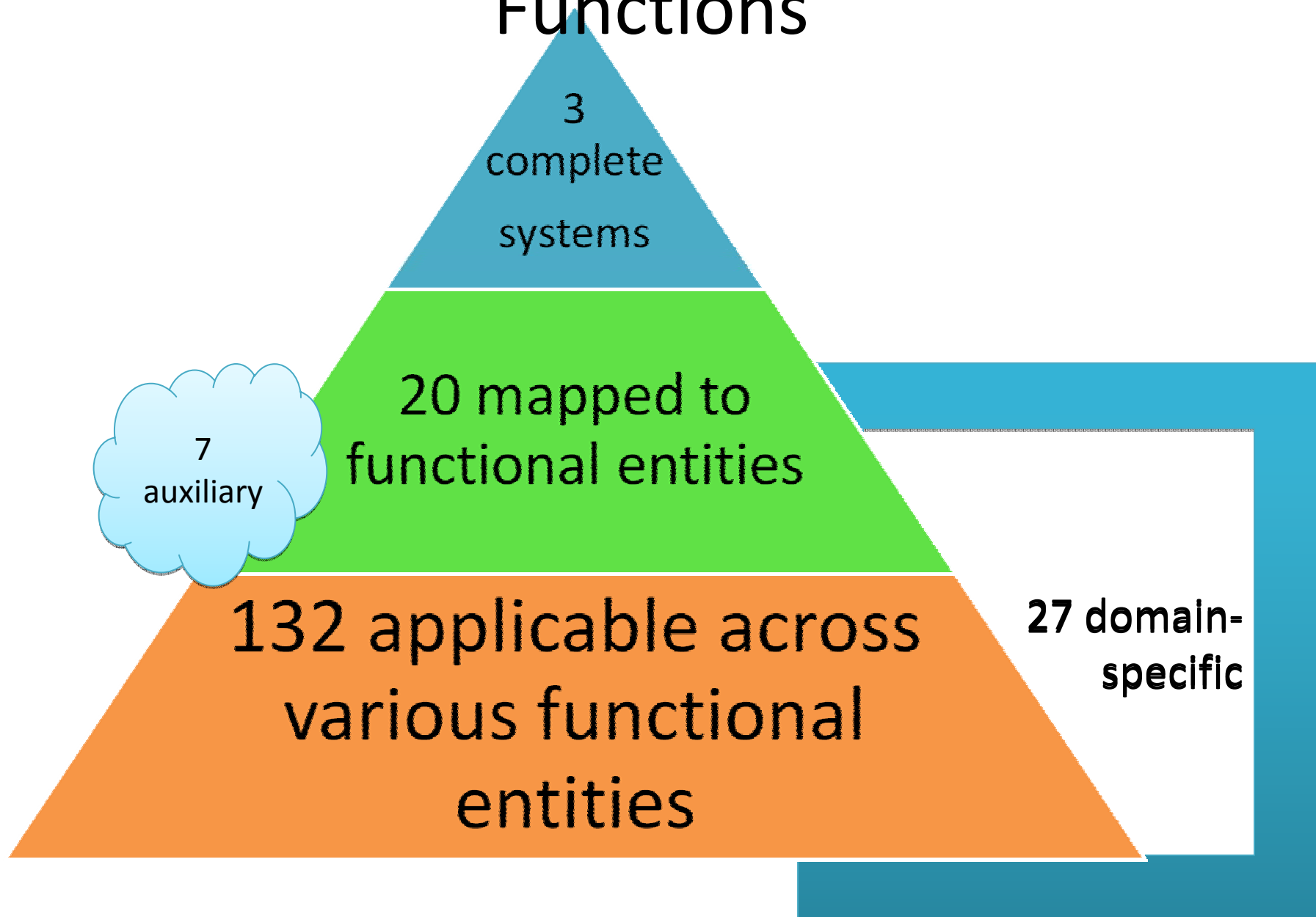
Sources

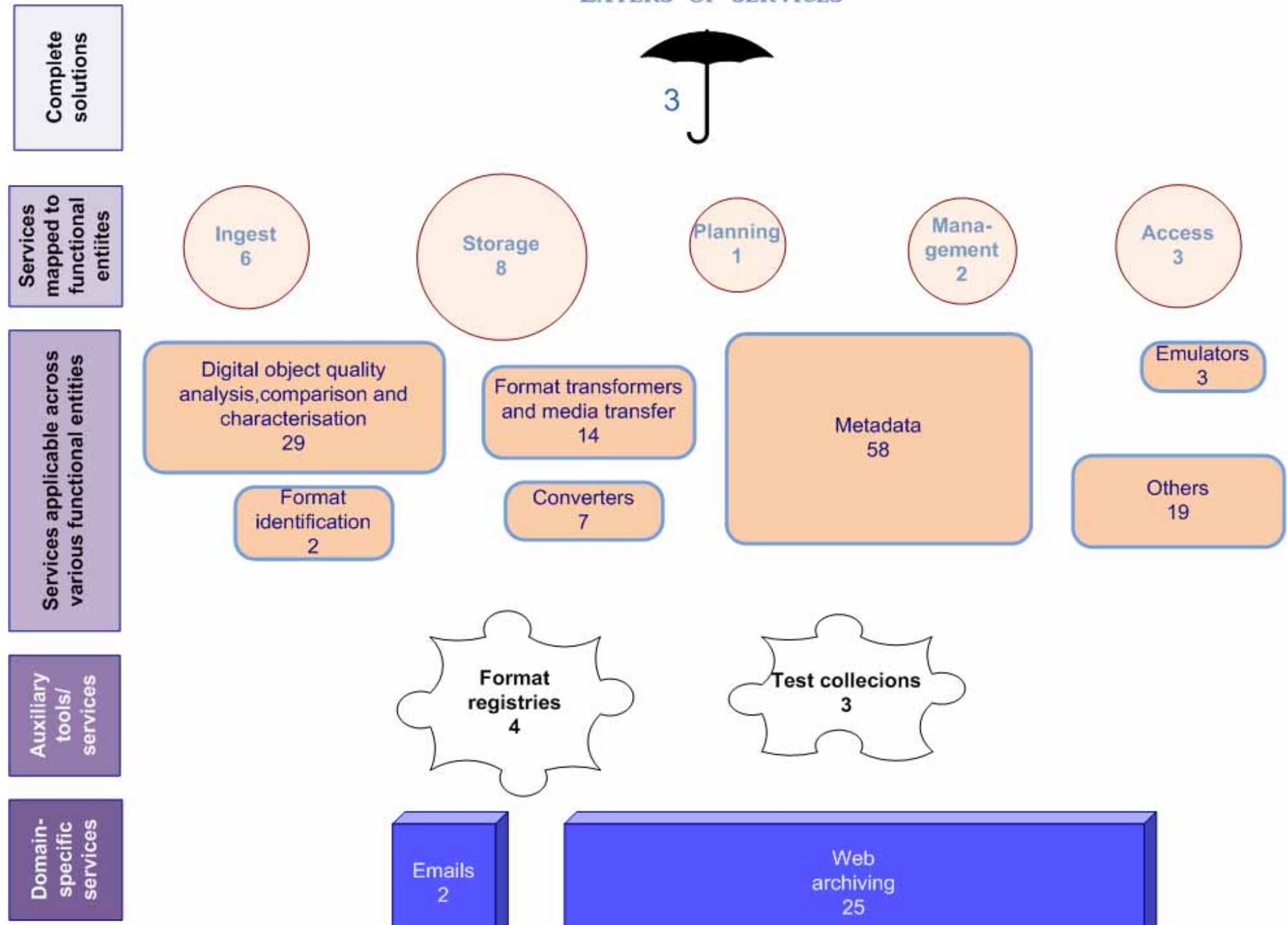- A total of 191 instruments identified
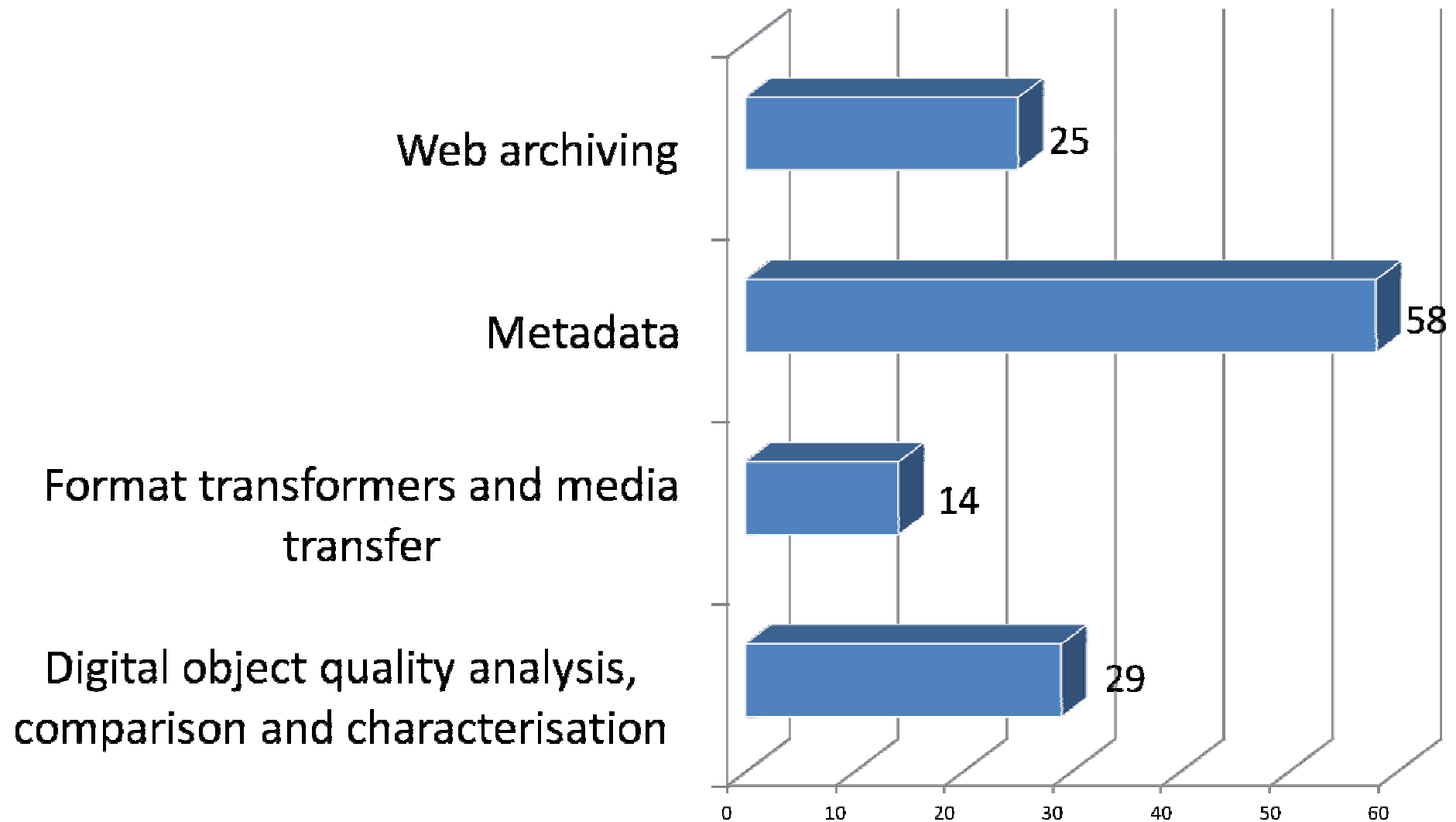
Mapping the DP Tools to OAIS Functions

3 complete systems

7 auxiliary

20 mapped to functional entities

132 applicable across various functional entities

27 domain-specific

LAYERS OF SERVICES

Complete solutions

3

Services mapped to functional entitites

Ingest
6

Storage
8

Planning
1

Mana-
gement
2

Access
3

Services applicable across various functional entities

Digital object quality analysis,comparison and characterisation
29

Format identification
2

Format transformers and media transfer
14

Converters
7

Metadata
58

Emulators
3

Others
19

Auxiliary tools/ services

Format registries
4

Test collecions
3

Domain-specific services

Emails
2

Web archiving
25

IV. Conclusions of the Report

Conclusions

- The number of tools currently offered for various preservation-related tasks is significant, majority of them offered as open source

- In most cases tools originate from short-term research projects and are openly published tools rather than supported services

- There is an overlap of what the tools developed by different institutions do – we duplicate our effort

- Choosing and implementing these tools will require considerable technical expertise that is often not available in (smaller) cultural heritage institutions

From Instruments to Infrastructure

- Mechanisms and metrics for **evaluating** the existing tools do not exist
  - **Benchmarks** for service **maturity** and **quality**
  - A **registry** of preservation tools and services with clearly applied metrics is needed
- A **roadmap** for connecting the existing tools into an e-Infrastructure should be outlined
- **Preservation** as an **infrastructure** would make it cost effective and easier to implement for smaller institutions
  - Successful examples of relying on Grid and cloud infrastructures already exist both in cultural heritage and neighbouring domains – this should be built on and developed further
  - Will this provide the level of trust DCH requires?

Thank you for your attention!

Questions?

raivo@eba.ee