

Overviewing Digital Preservation: Facets for Long-term Use of Digital Resources and Archives

Shigeo Sugimoto

Faculty of Library, Information and Media Science

University of Tsukuba

sugimoto@slis.tsukuba.ac.jp

New Title of this talk

revised based on conversation with Profs. Hara and Kameda in a car
heading toward a dinner place yesterday evening

Metadata Issues for Long-term Use of Databases across Communities

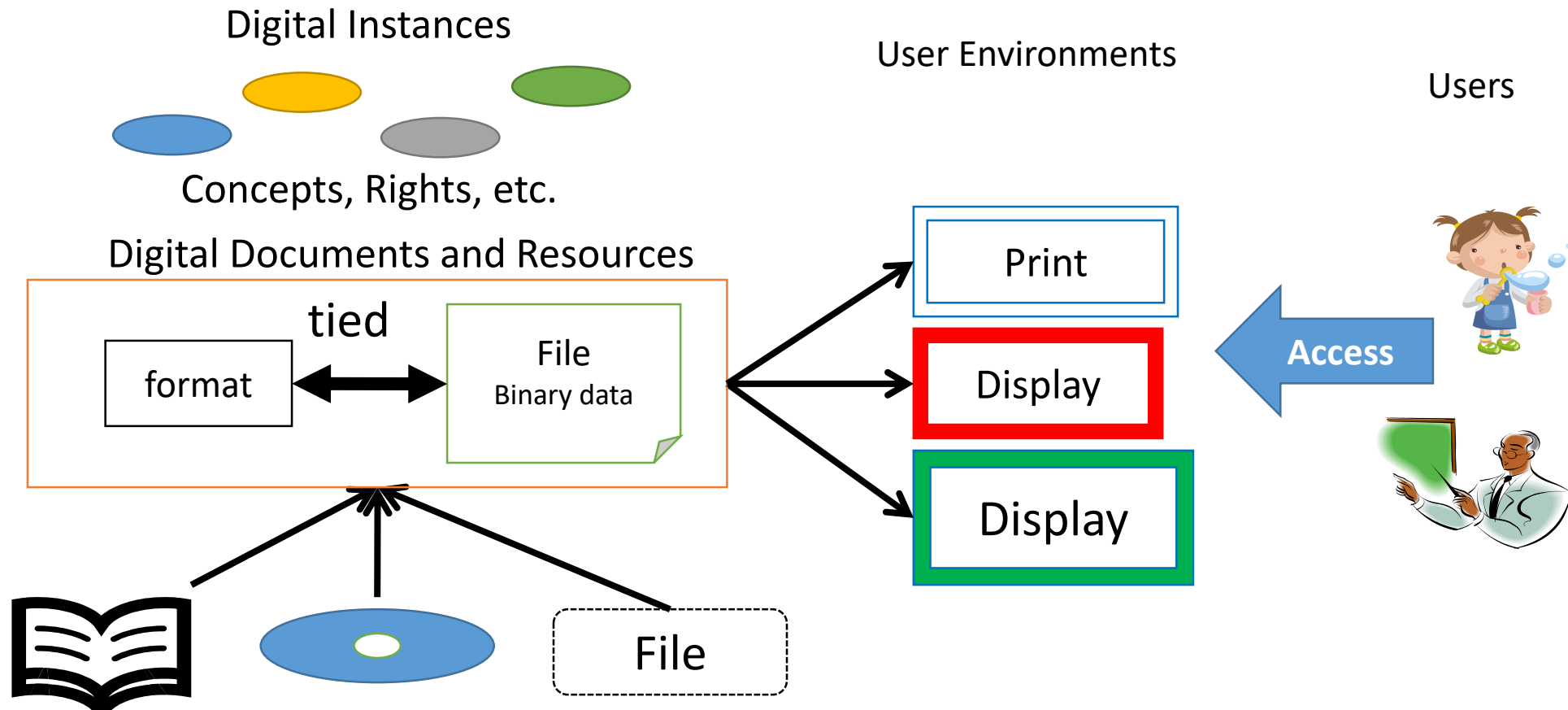
Outline

- First half of this talk is intended to share ideas and models for longevity of digital archives and digital resources, i.e. digital preservation, among the participants
- Second half is prepared to share basic concepts of metadata design for better interoperability across communities and over time and discussions on metadata design for sharing information about digital archives/databases

Part 1 Digital Preservation – an Overview

- Digital Entities on Information Environment
- Digital Preservation as Maintenance Activity
- Bird's Eye View of Digital Archives – Facets of Digital Preservation

Digital Entities on Information Environment



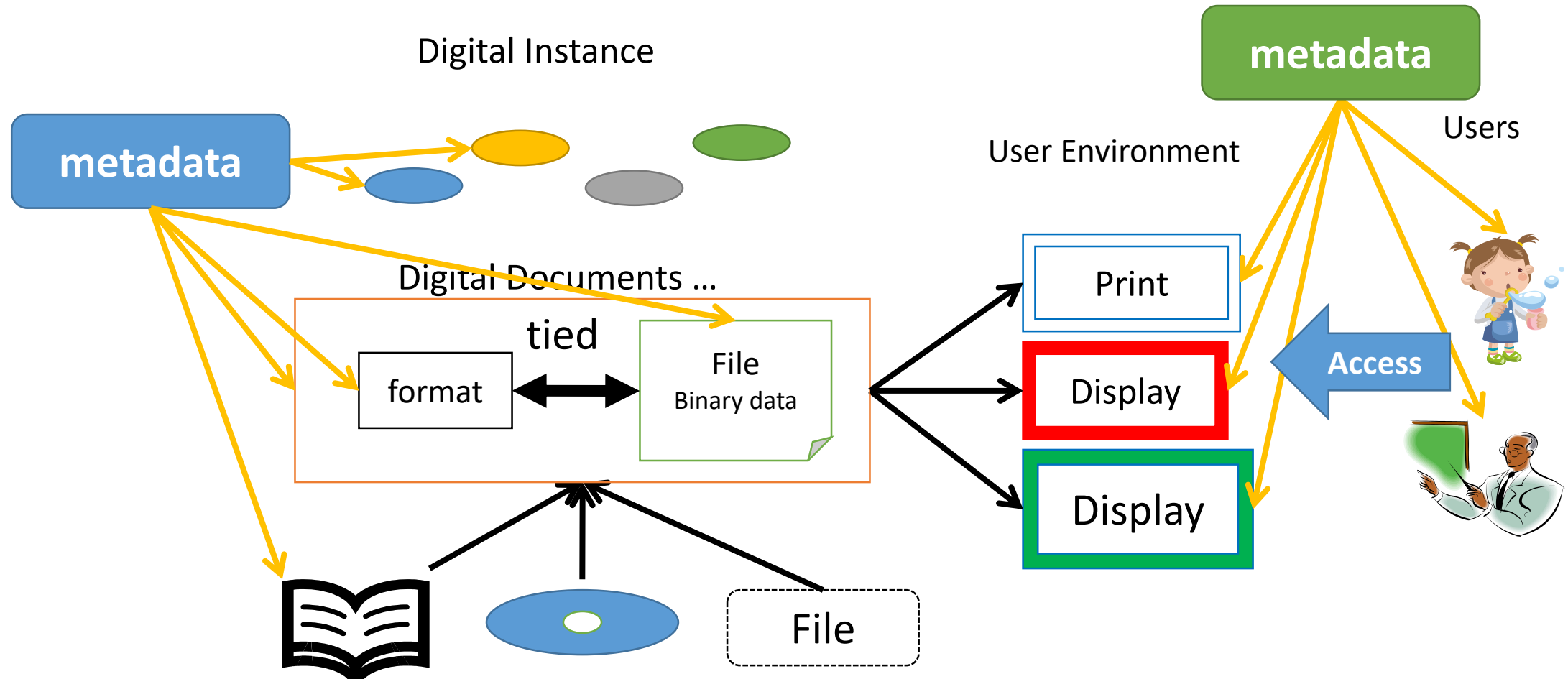
Digital Entities on Information Environment



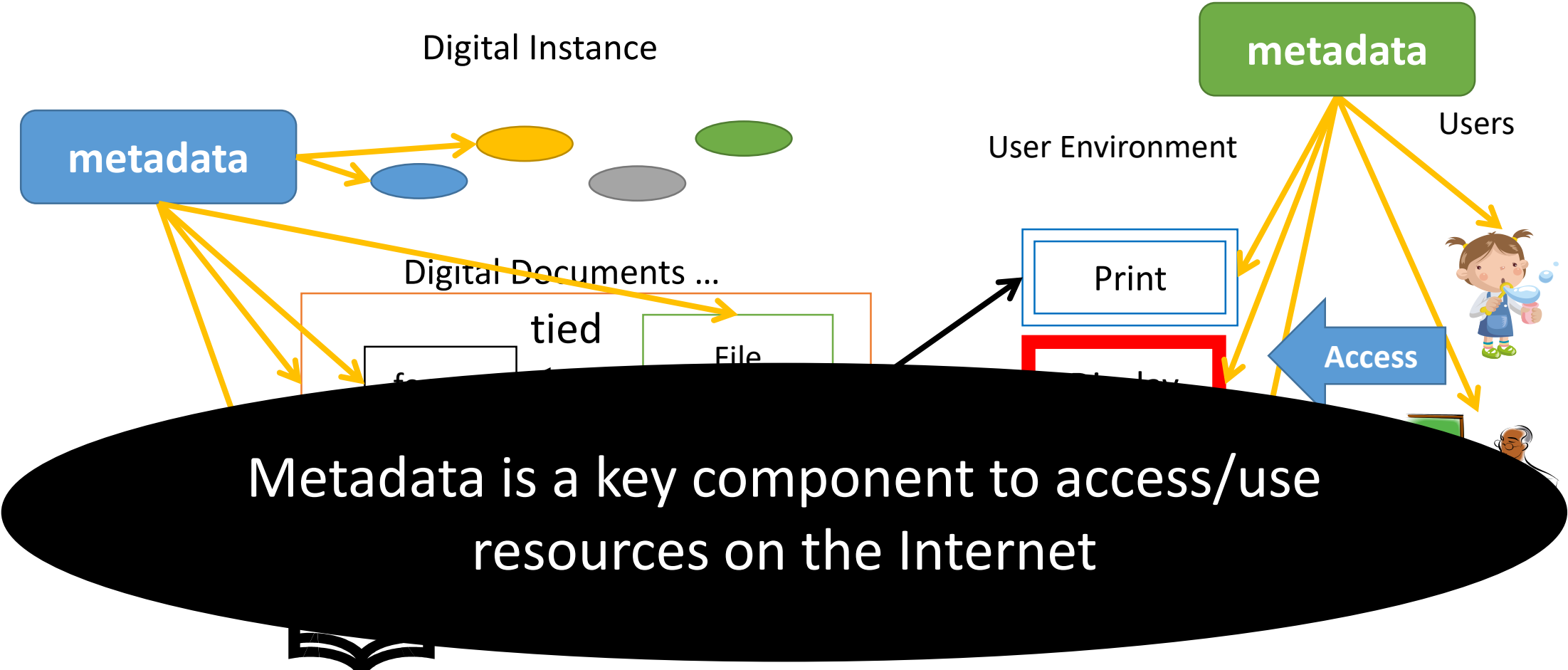
All instances are expressed as a digital entity on the Internet, e.g., books, files, digital contents, users, user environments, concepts, etc.



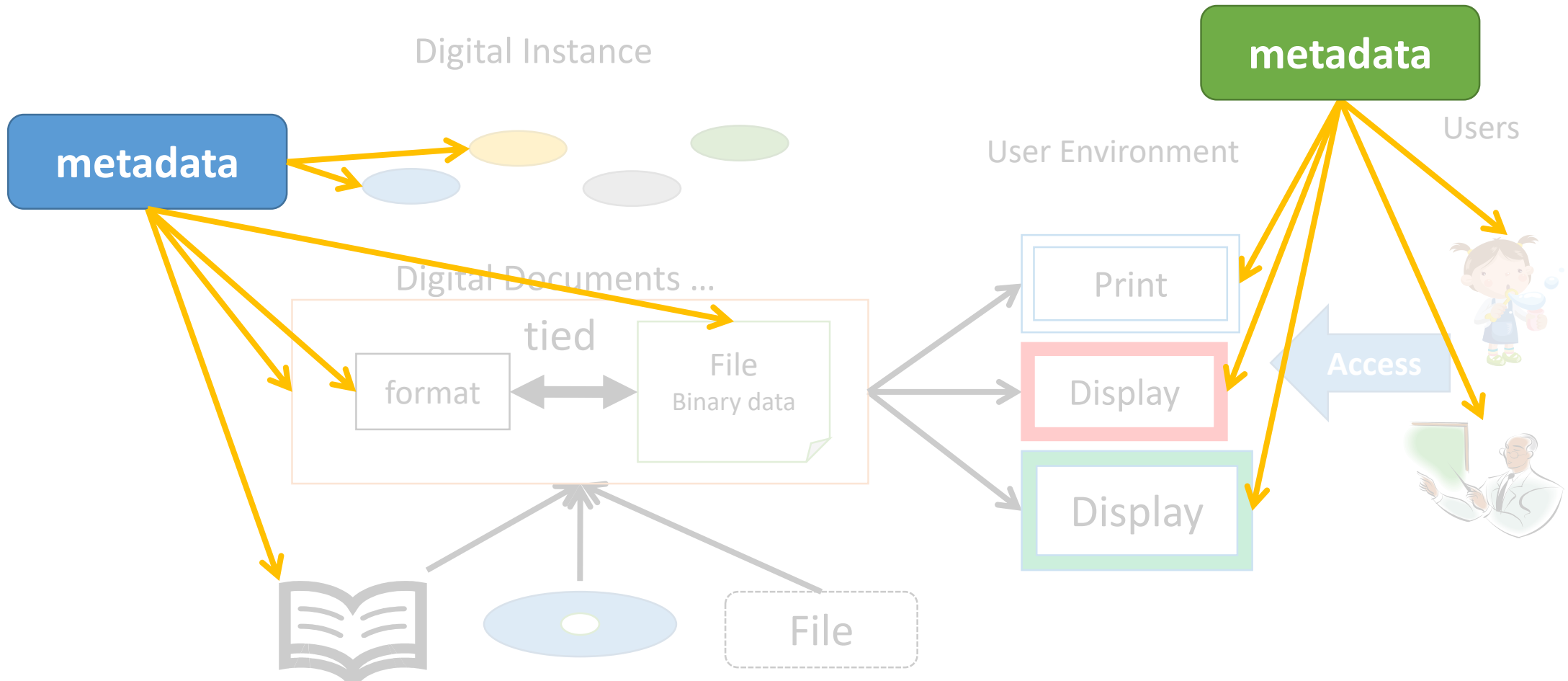
Digital Entities and Metadata on Web



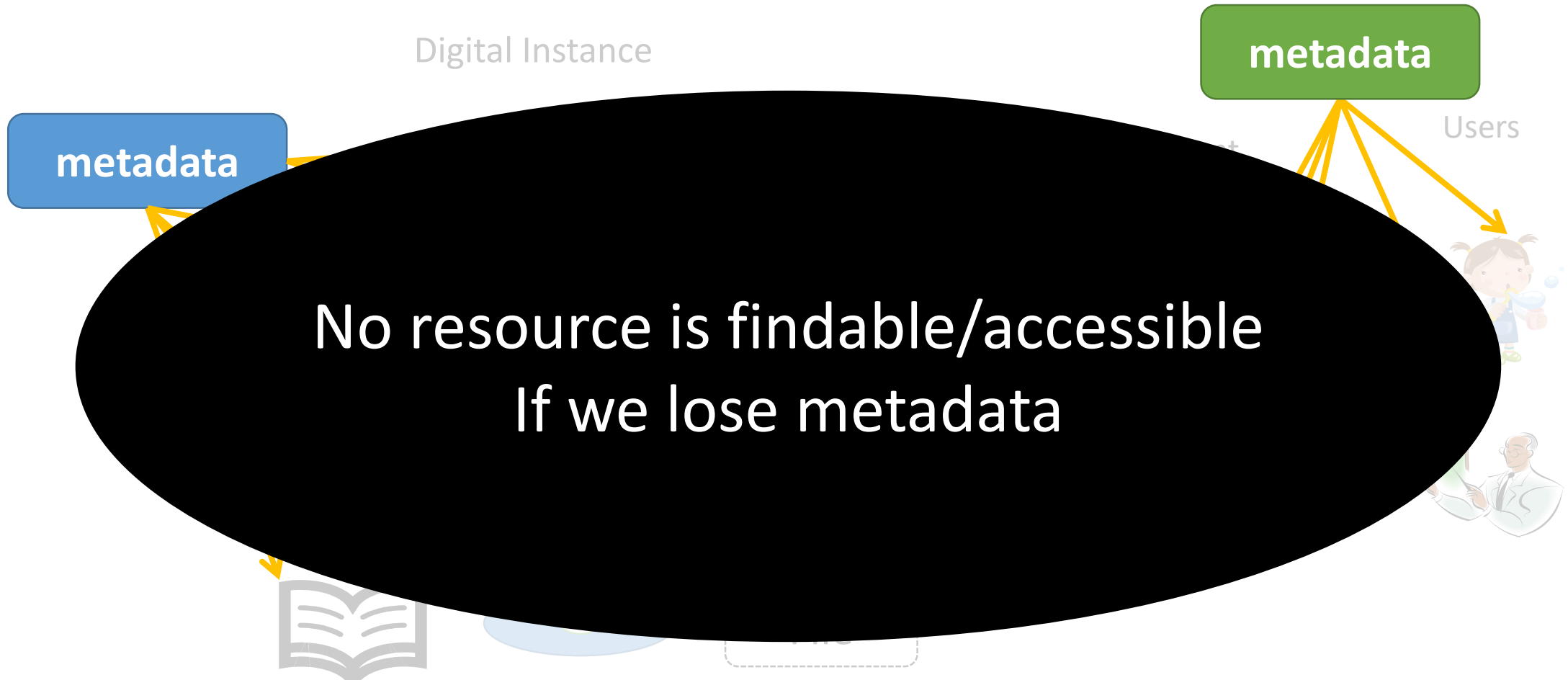
Digital Entities and Metadata on Web



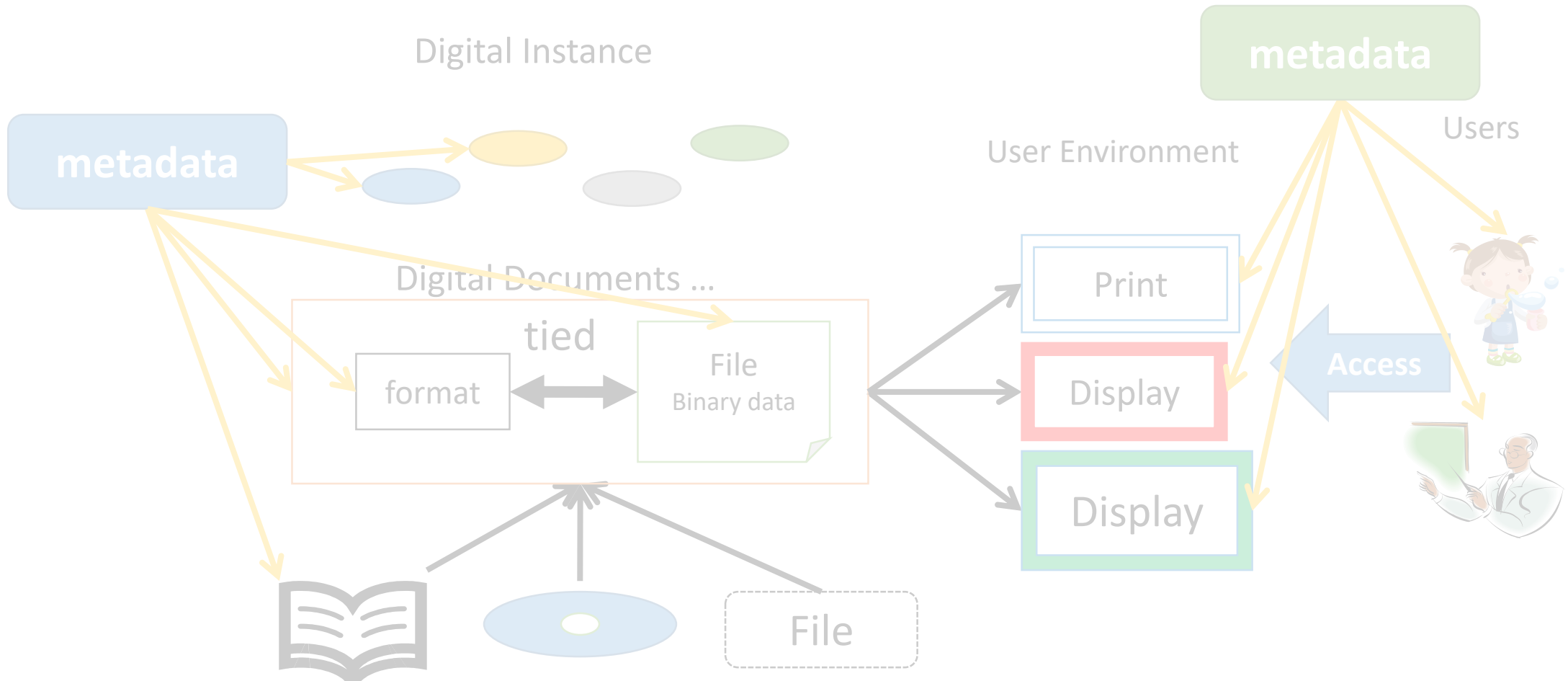
Digital Entities and Metadata on Web



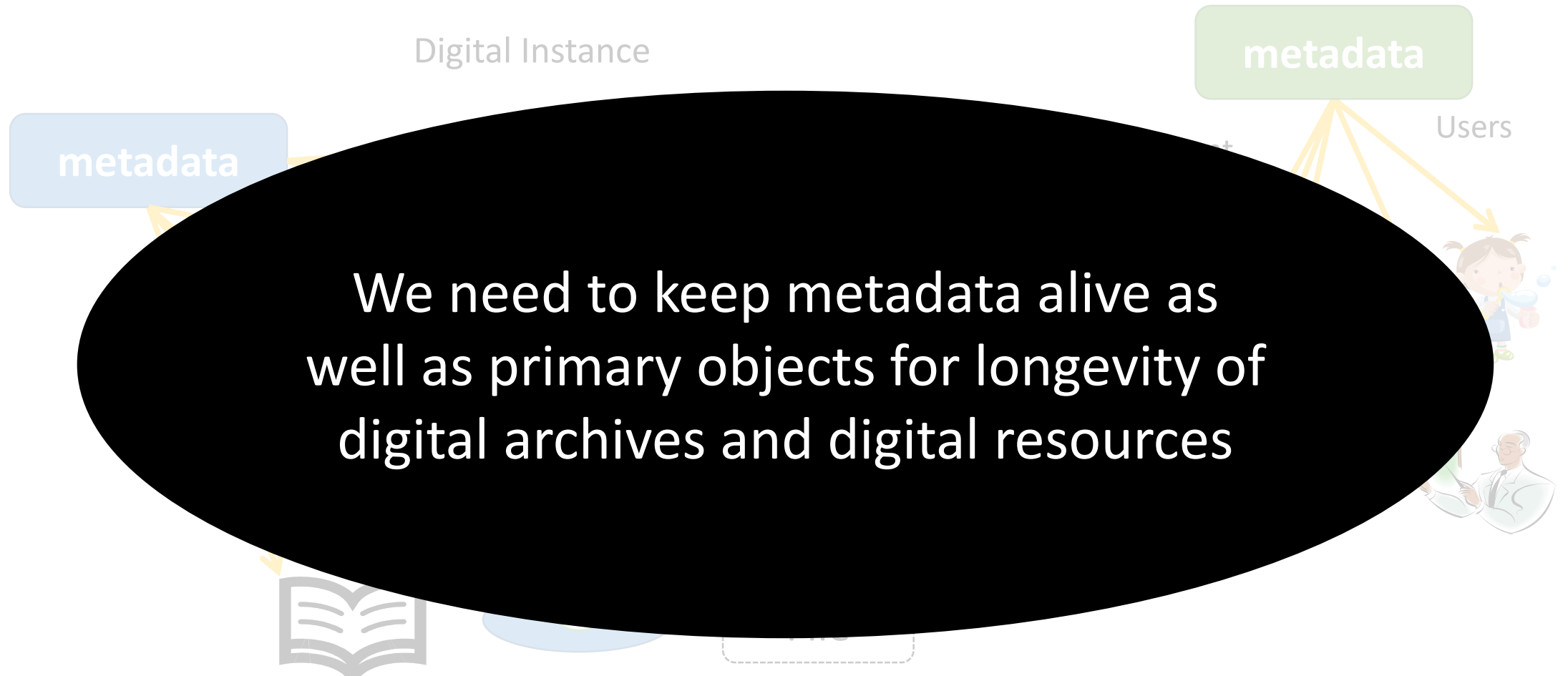
Digital Entities and Metadata on Web



Digital Entities and Metadata on Web



Metadata Longevity in Digital Preservation



Digital Preservation as Maintenance

- Digital Preservation: Keep digital resources usable over time
 - In other words, Long-term Use and Longevity of digital resources
- Some Questions
 - What is “**(digital) resource**” – medium or content?
 - What is the **difference** between “**digital** resource” and “**non-digital (or analog)** resource”?
 - What does “keep resources **usable**” mean – do we have to keep a resource usable **as it is** or **can we modify it**?
 - “as is” → bit level or functionality level?
 - “modifiable” → what features have to be preserved – visual (fonts, layout, etc.), functionality (look-and-feel such as page flipping and input mode), etc.
- Thus, lots of questions → No single answer
 - Preservation policy and strategy of an institution depend on their needs

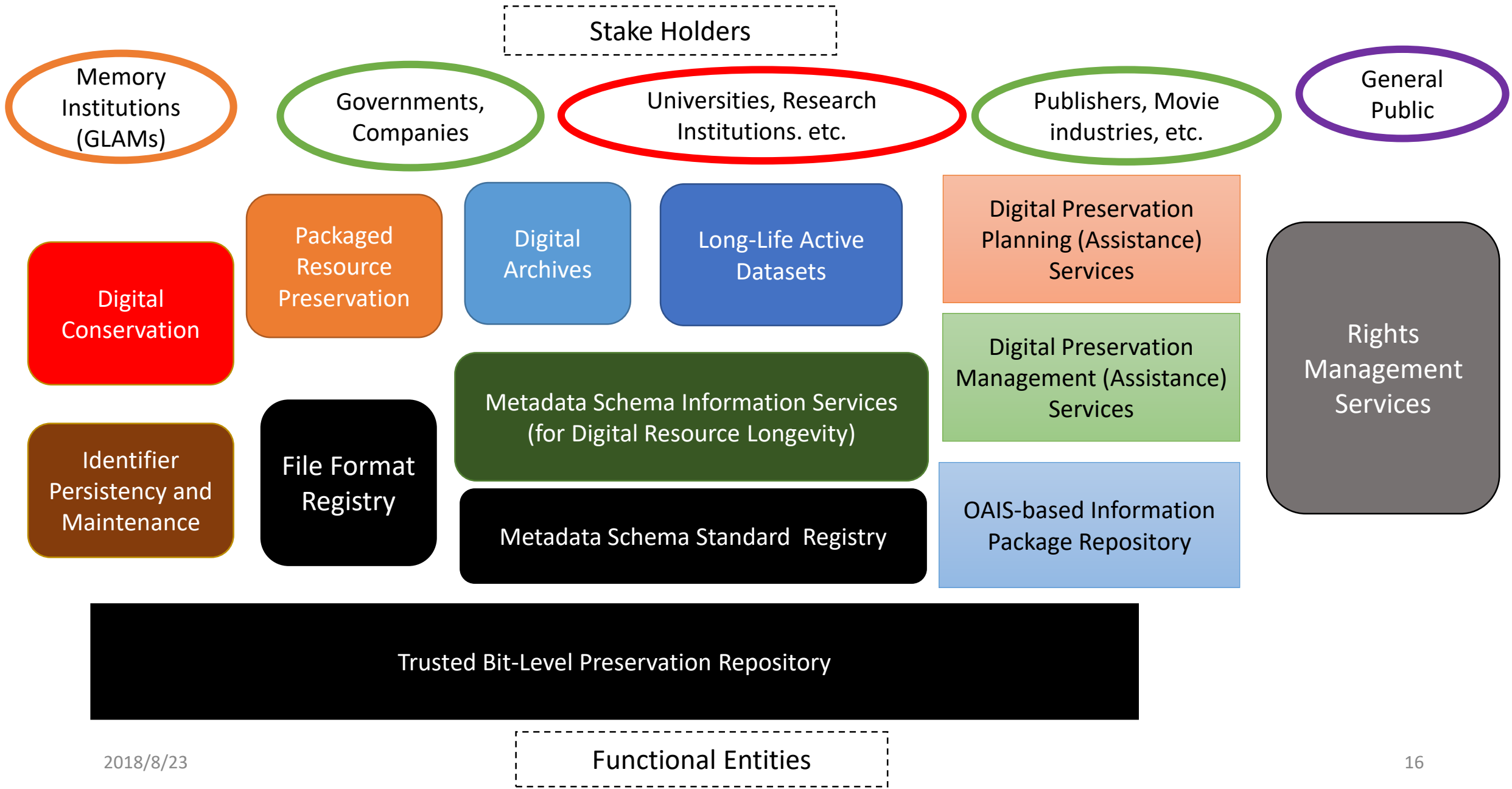
Digital Preservation as Maintenance

- “How many years do we need to keep digital resources alive?” is a frequently asked question but we don’t need to answer this question directly, because
 - all memory institutions would say “we keep the resources forever / as long as our institution exists”
 - What we can do is to hand over the resources to our next generation
 - We need to properly maintain the resources in accordance with our given social environments and technologies, which may change generation to generation
- Digital preservation is to maintain digital resources usable for long-term → long-term maintenance of digital resources

Digital Preservation – a Bird’s Eye View

- Overview of digital preservation
 - Stake holders and functional components
 - Digital preservation is “management” and need engineering point of view
- Digital Preservation is not just a longevity of digital information media but has various facets. This section will overview digital preservation from a generalized viewpoint
- This section will discuss metadata for digital preservation because metadata is a key component for digital preservation.

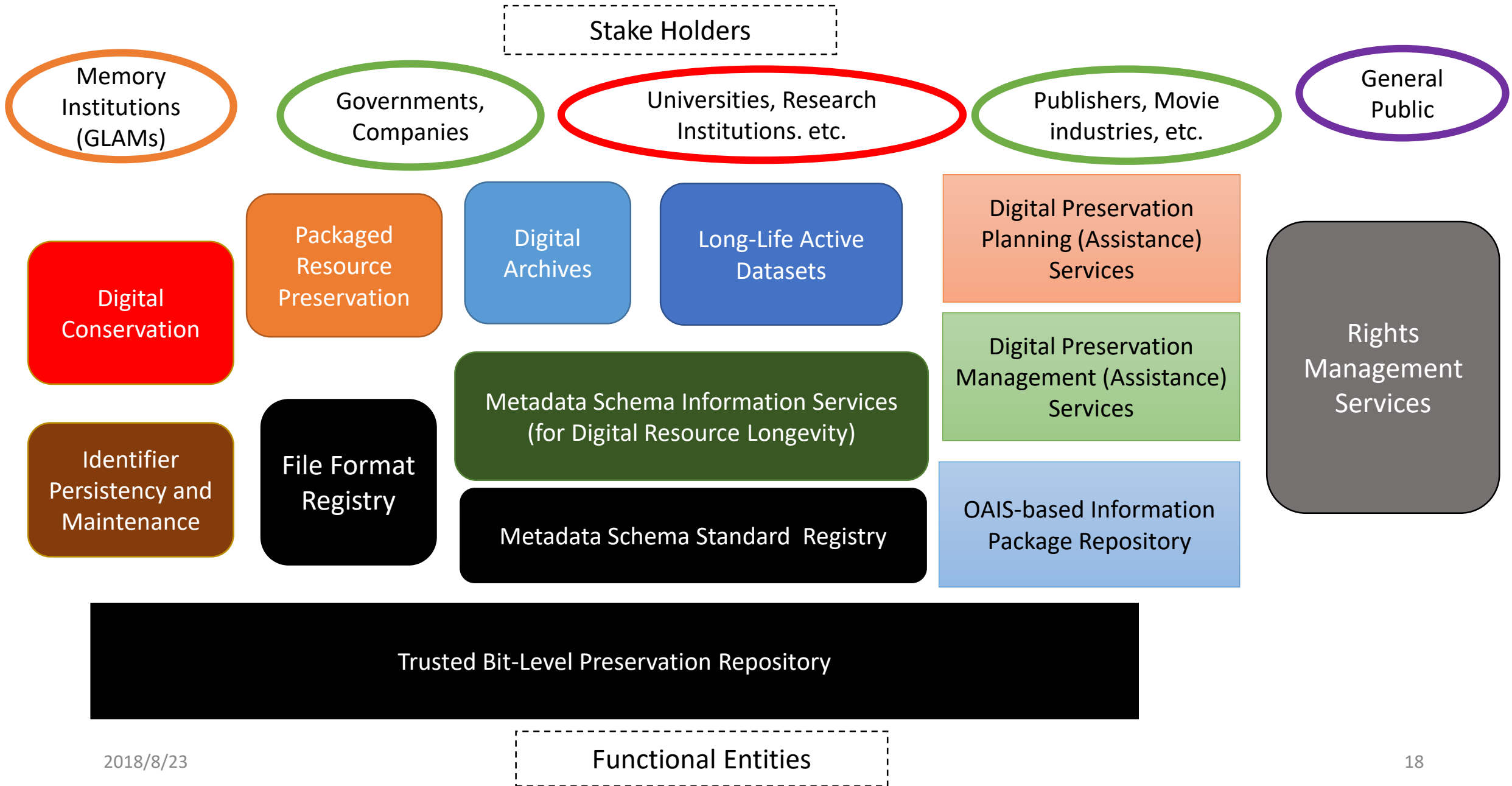
Functional Entities and Stake Holders in Digital Preservation



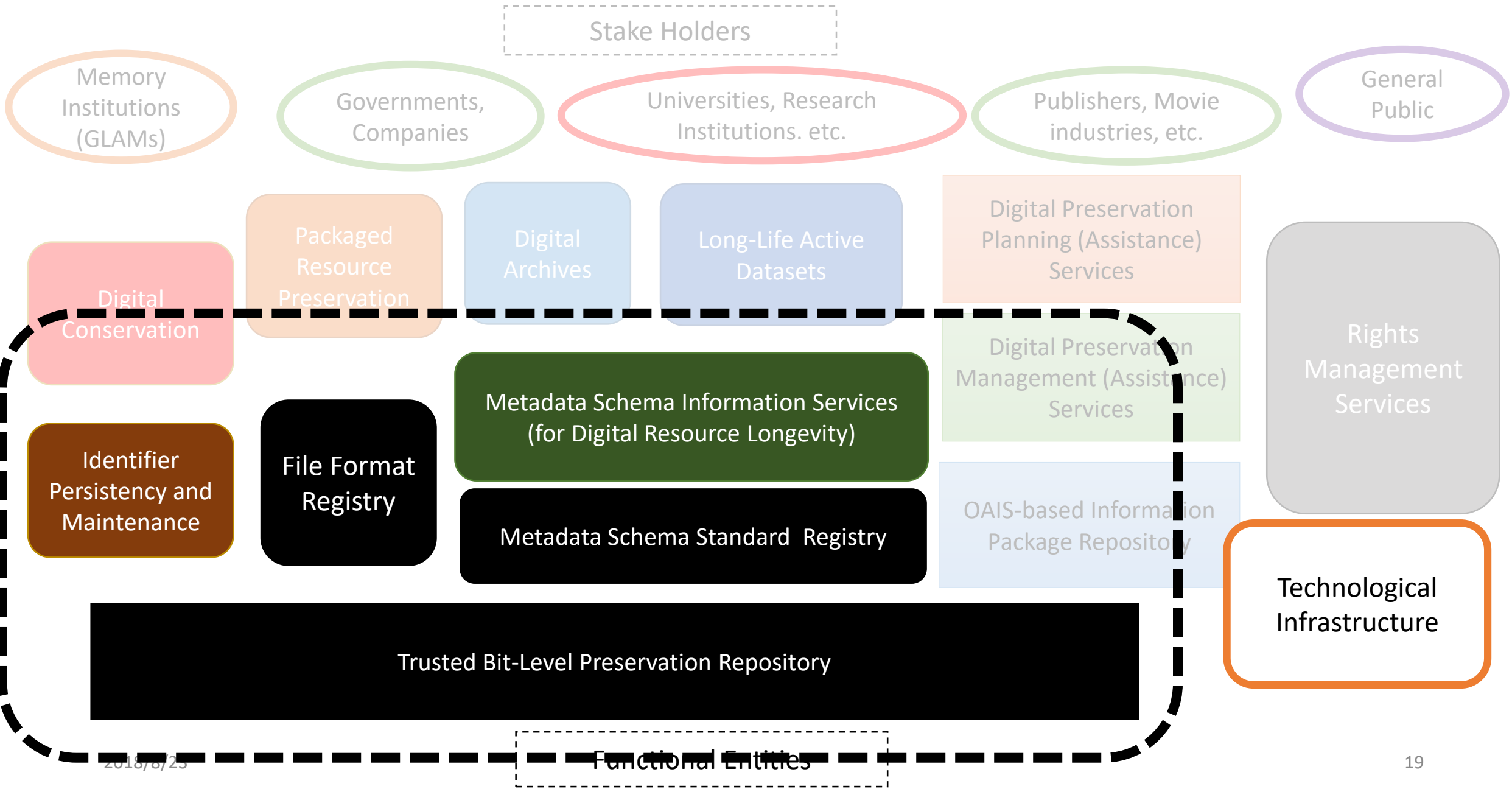
Digital Preservation – a Bird’s Eye View

- Technological Infrastructure
 - Trusted Bit-Level Preservation Repository
 - File Format Registry
 - Metadata Schema Standard Registry
 - Identifier Persistency and Maintenance
 - Metadata Schema Information Services (for digital longevity)
- Individual Content Preservation Technologies
 - Digital Conservation
 - Packaged Resource Preservation
- Long-term Data Services
 - Digital Archives
 - Long-Life Active Datasets
- Digital Preservation Services
 - Digital Preservation Planning (Assistance) Services
 - Digital Preservation Management (Assistance) Services
 - OAIS-based Information Package Repository
- Social Infrastructure
 - Rights Management Services

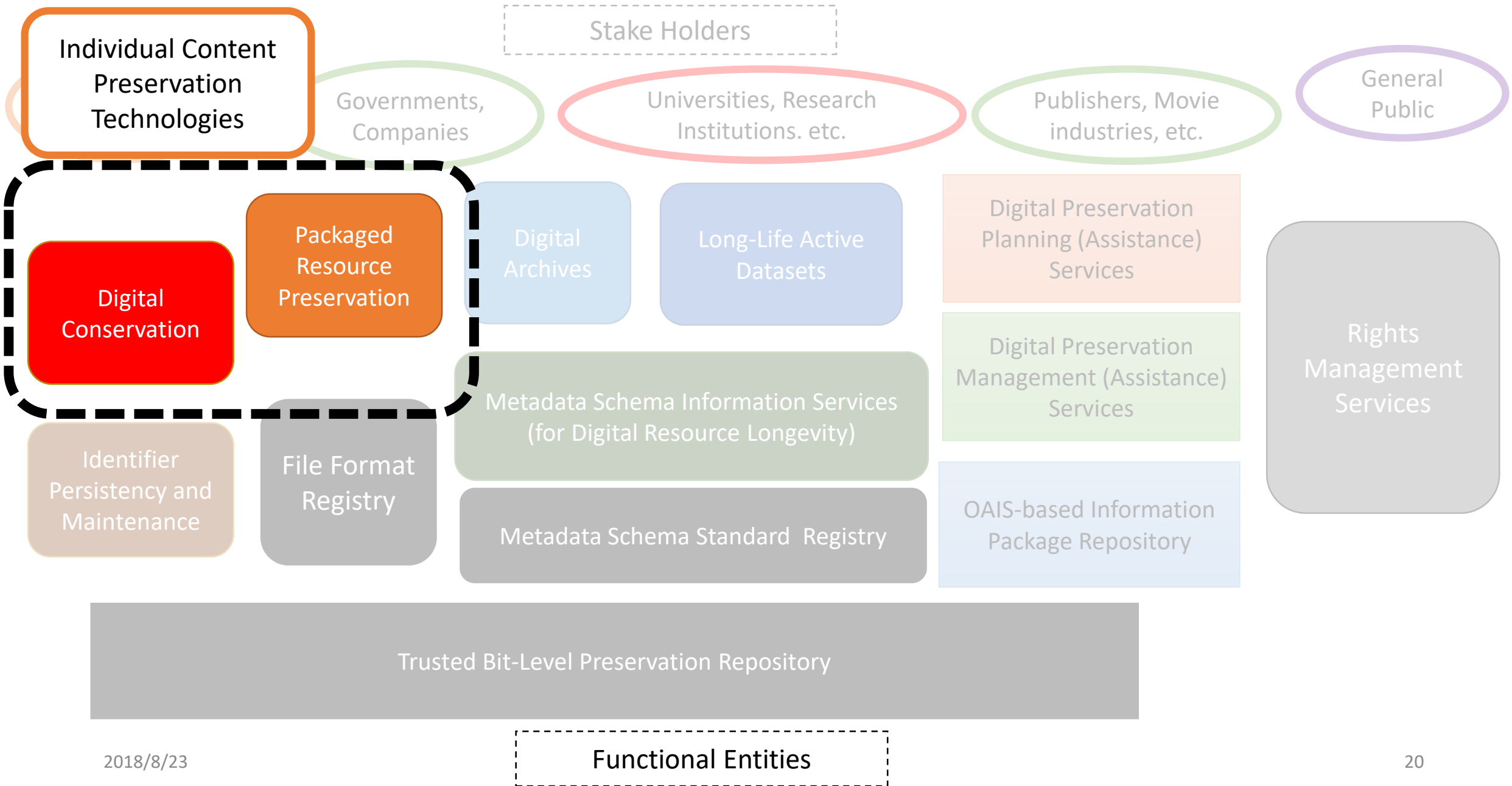
Functional Entities and Stake Holders in Digital Preservation



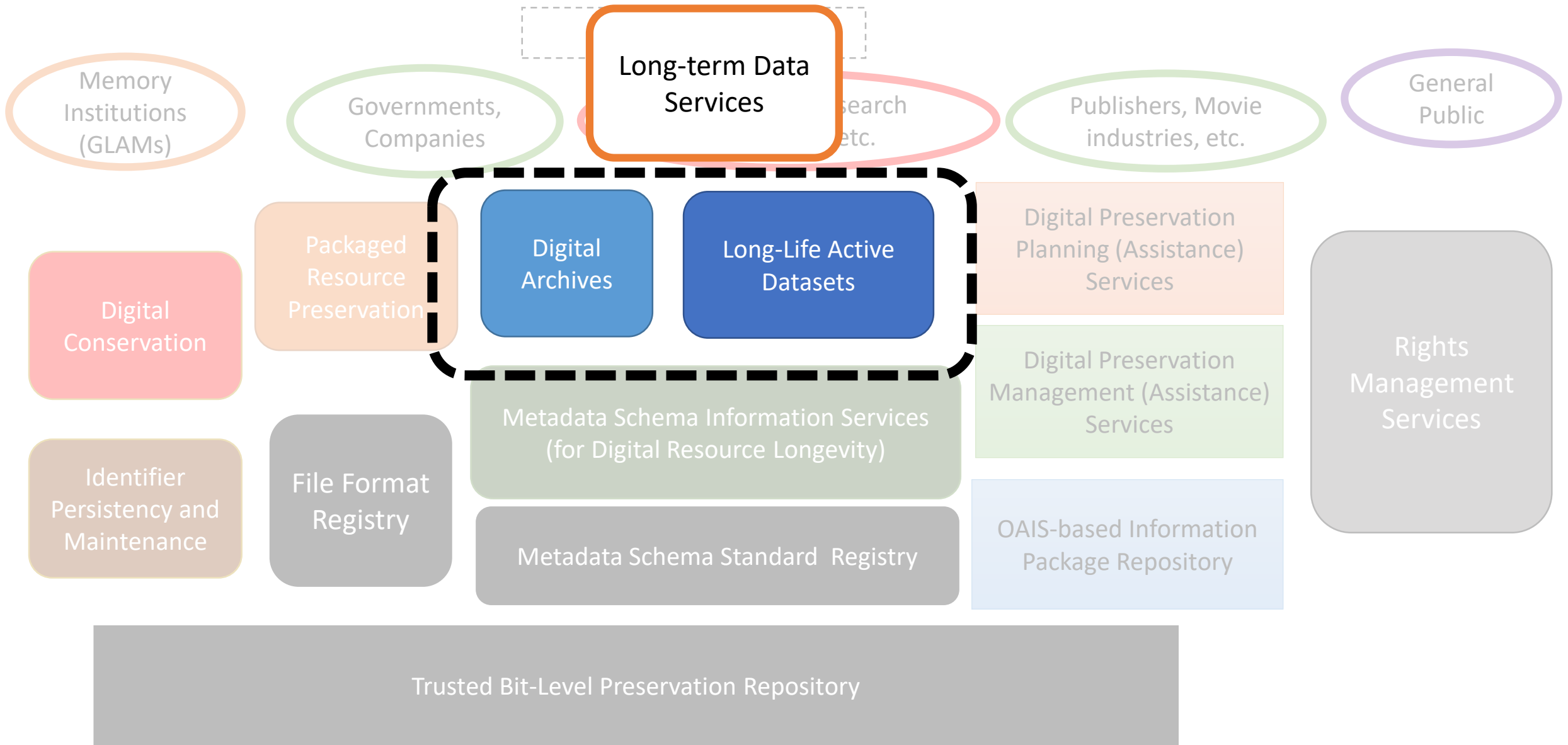
Functional Entities and Stake Holders in Digital Preservation



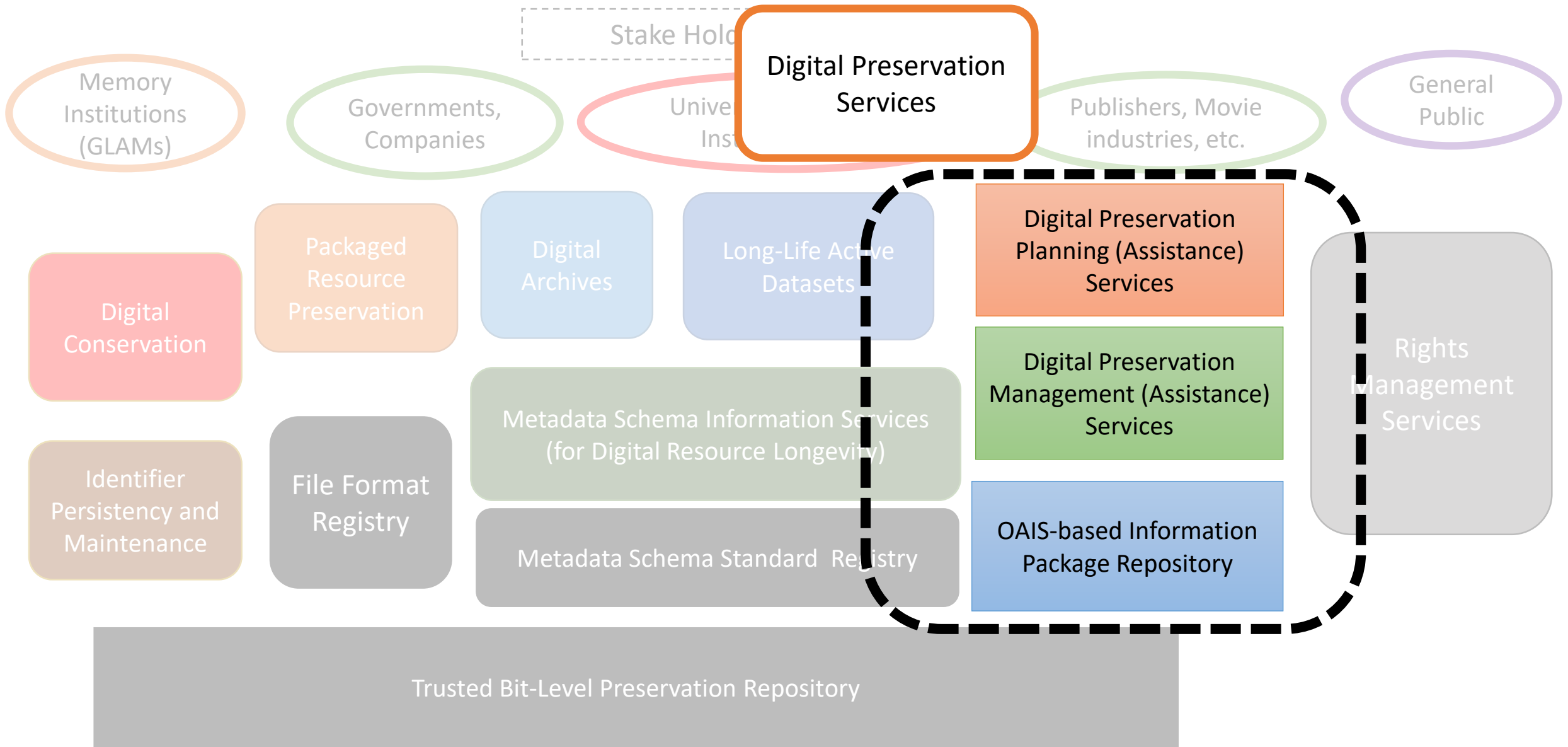
Functional Entities and Stake Holders in Digital Preservation



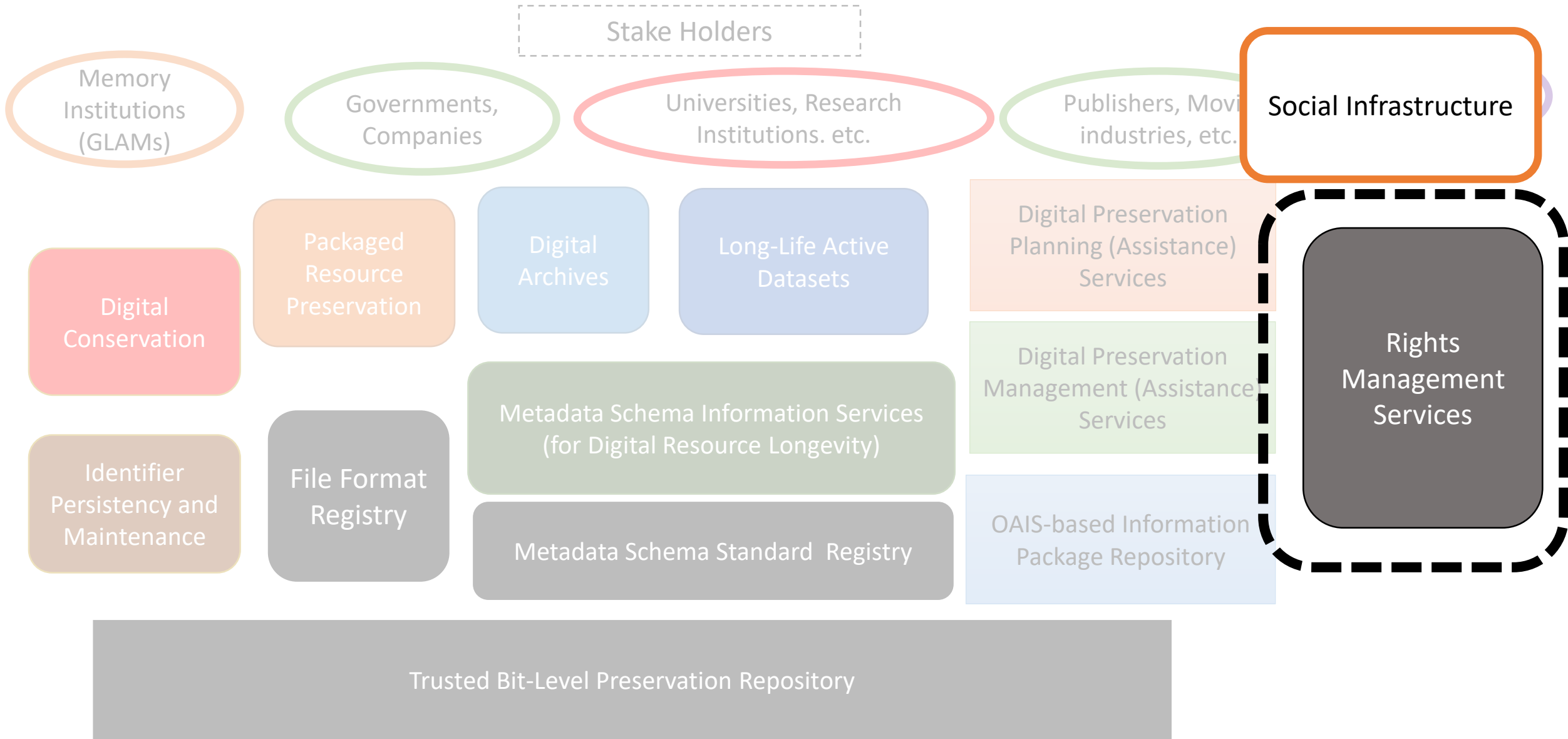
Functional Entities and Stake Holders in Digital Preservation



Functional Entities and Stake Holders in Digital Preservation



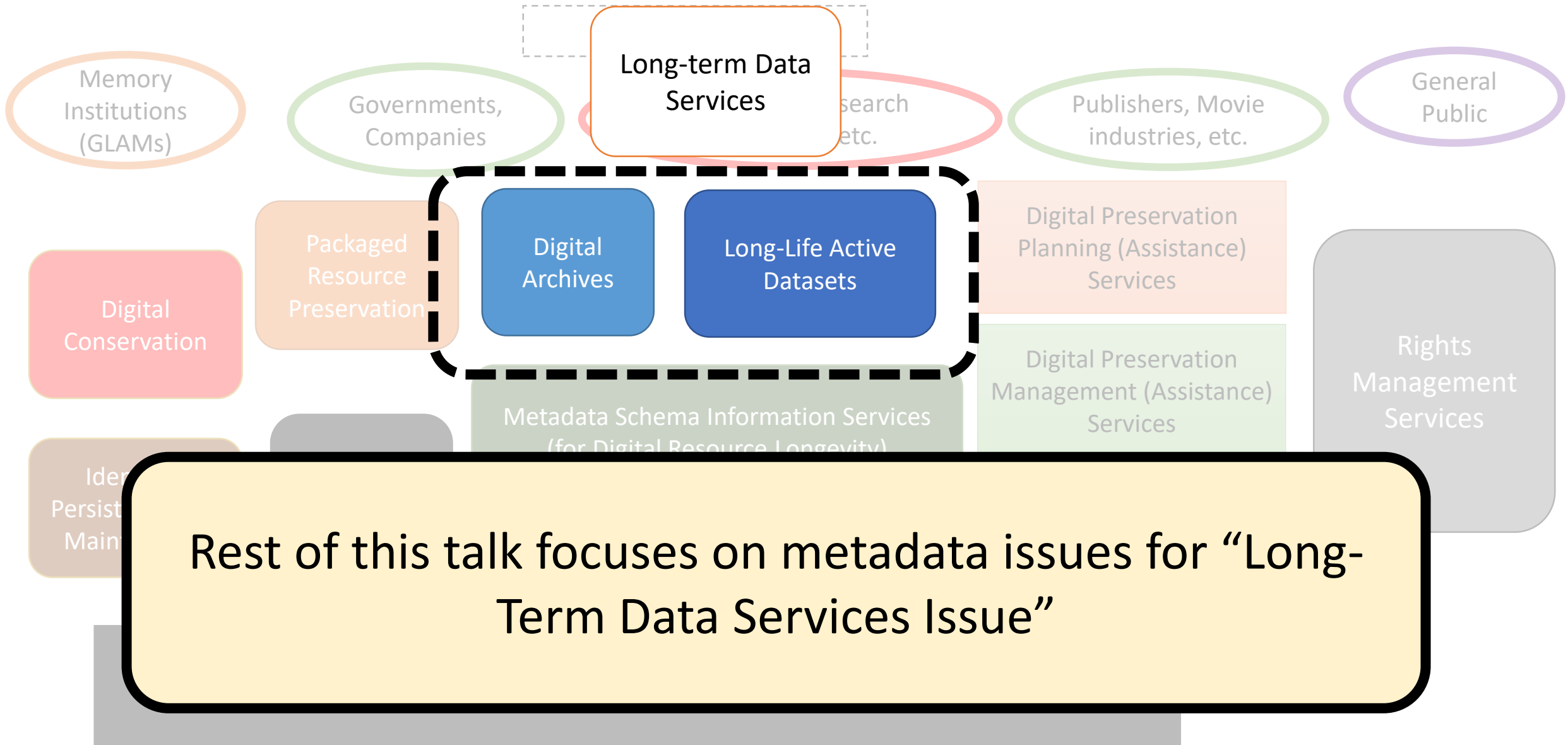
Functional Entities and Stake Holders in Digital Preservation



Part 2 Metadata Issues

- Metadata in Digital Archives
- Metadata Interoperability

Functional Entities and Stake Holders in Digital Preservation



Metadata in Digital Archives



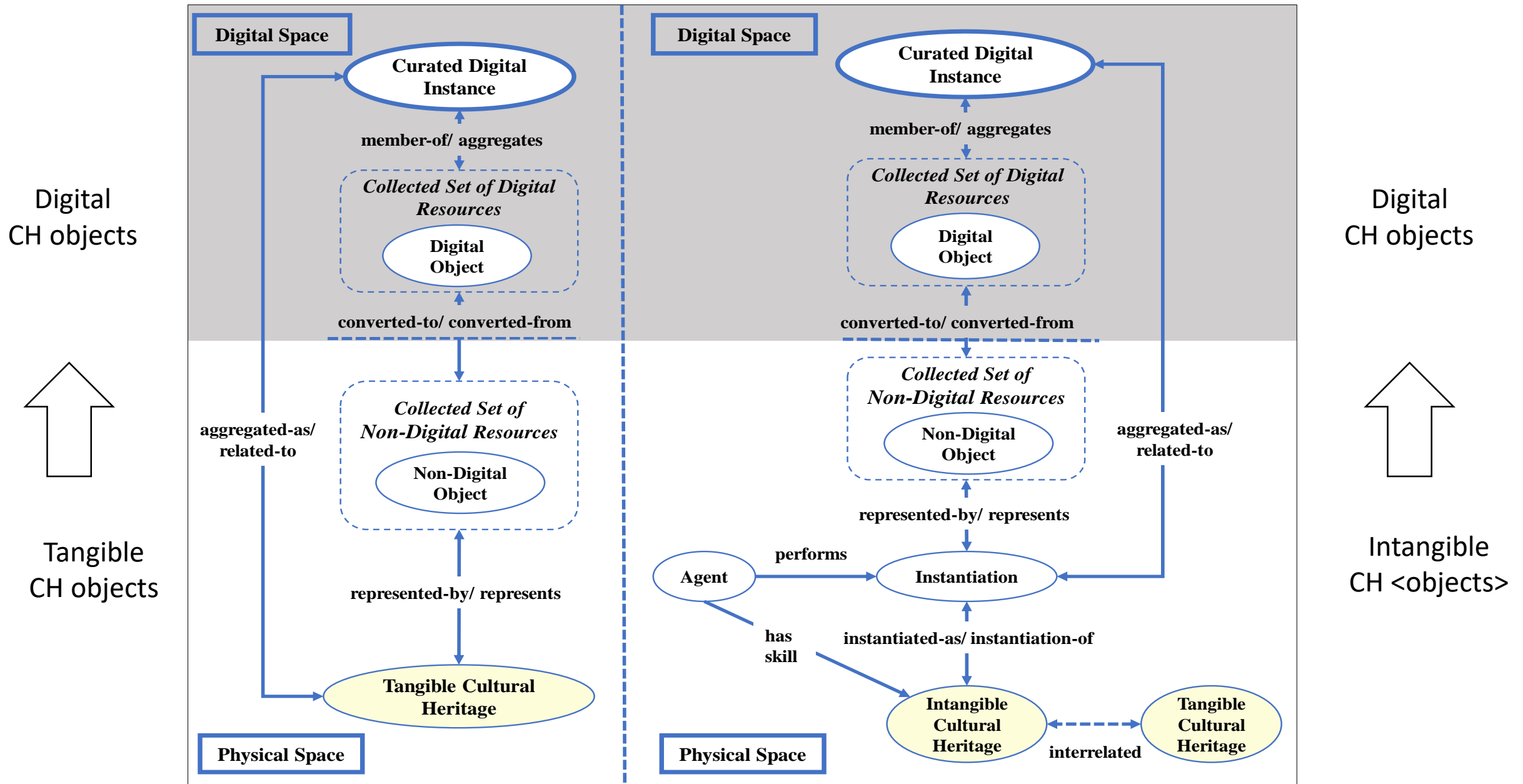
A Generalized Model of Digital Archives from a paper by Chiranthi Wijesundara, Winda Monika, & Shigeo Sugimoto published at ICADL 2017

Clarify entities included in organization processes of digital archives of tangible and intangible cultural heritage



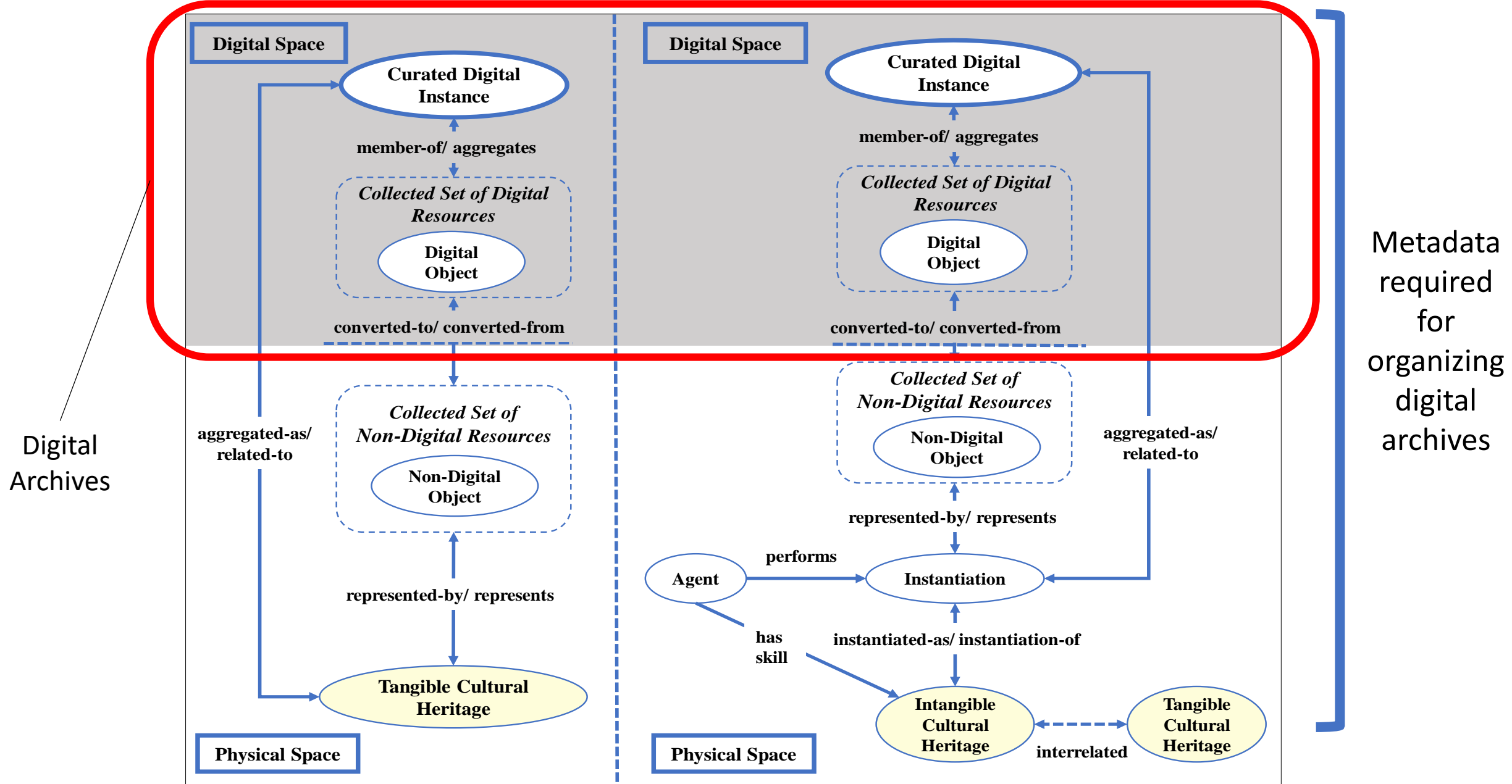
A Generalized Model of Digital Archives

CHDE model presented at ICADL 2017, Chulalongkorn, Nov.2017



A Generalized Model of Digital Archives

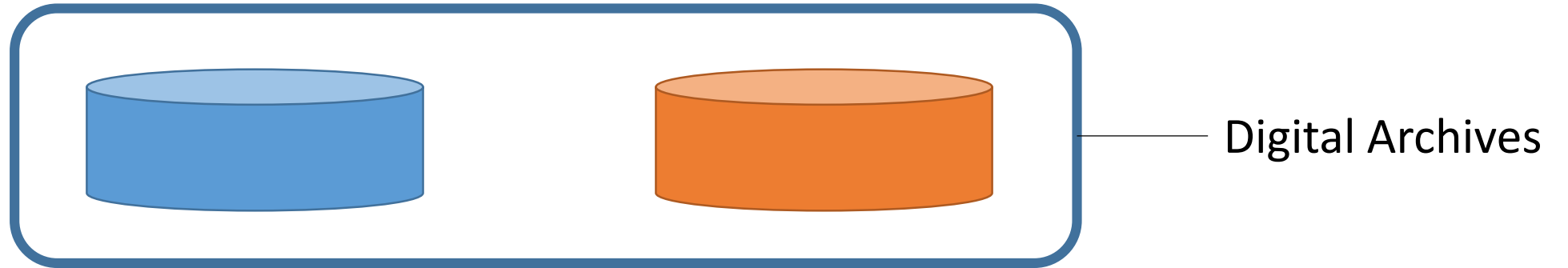
CHDE model presented at ICADL 2017, Chulalongkorn, Nov.2017



A Generalized Model of Digital Archives

CHDE model presented at ICADL 2017, Chulalongkorn, Nov.2017

Very Abstracted View of Digital Archives

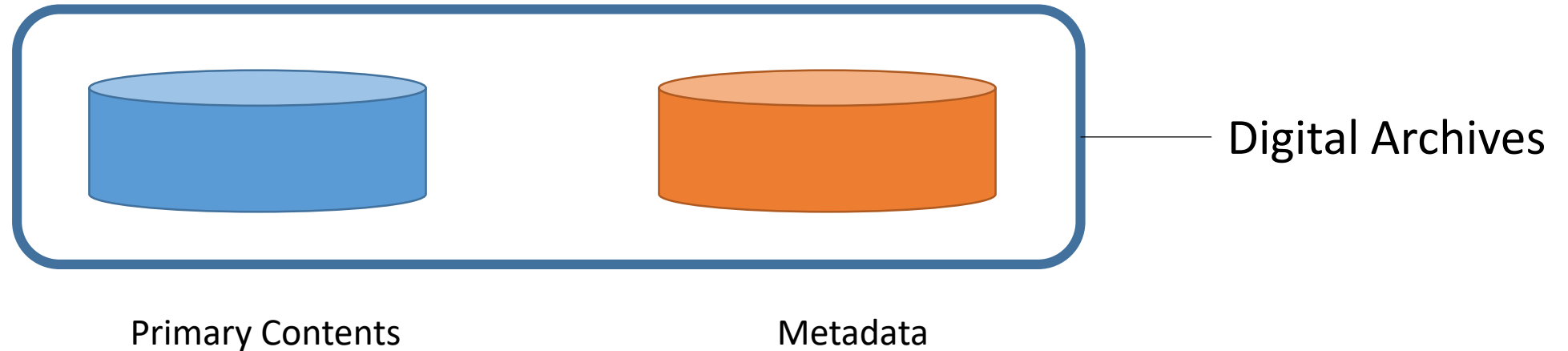


Primary Contents
Digital Images, Sounds,
Documents

Metadata
Description about

- Original CH
- Digital surrogates
(Primary Digital Contents)
- Administrative information
- etc.

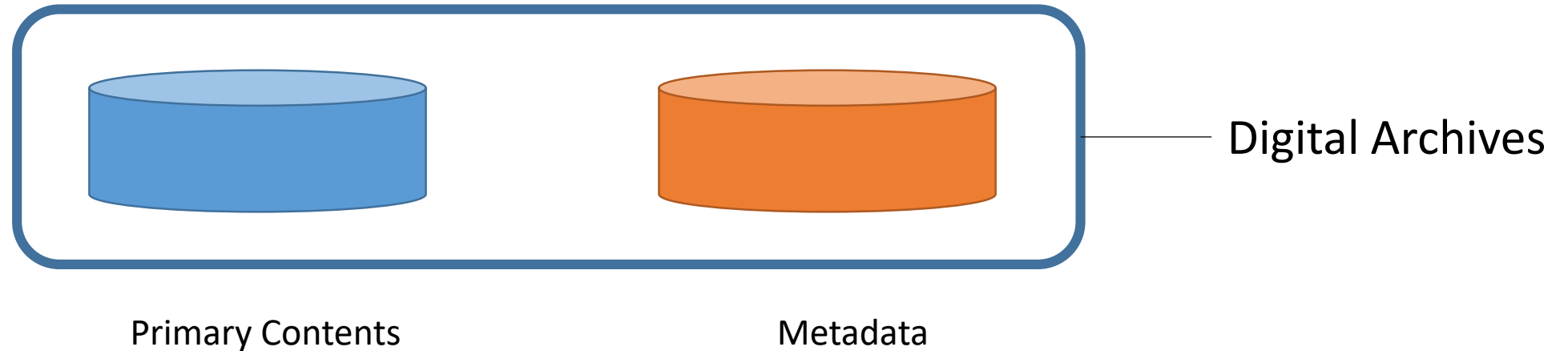
Very Abstracted View of Digital Archives



Longevity of both Primary Contents and Metadata is essential for digital archives

- (Primary Digital Contents)
- Administrative information
- etc.

Sharing Digital Archives across Communities and over Time

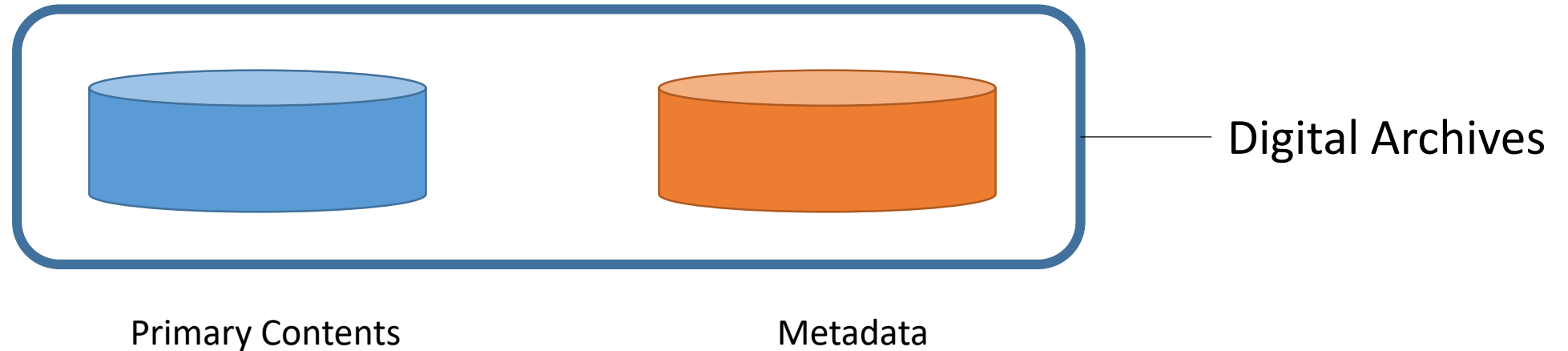


Longevity of both Primary Contents and Metadata is essential for digital archives

Metadata Longevity

- Keep metadata database alive over time
- Maintain metadata schemas over time
- Maintain metadata terms over time

Sharing Digital Archives across Communities and over Time



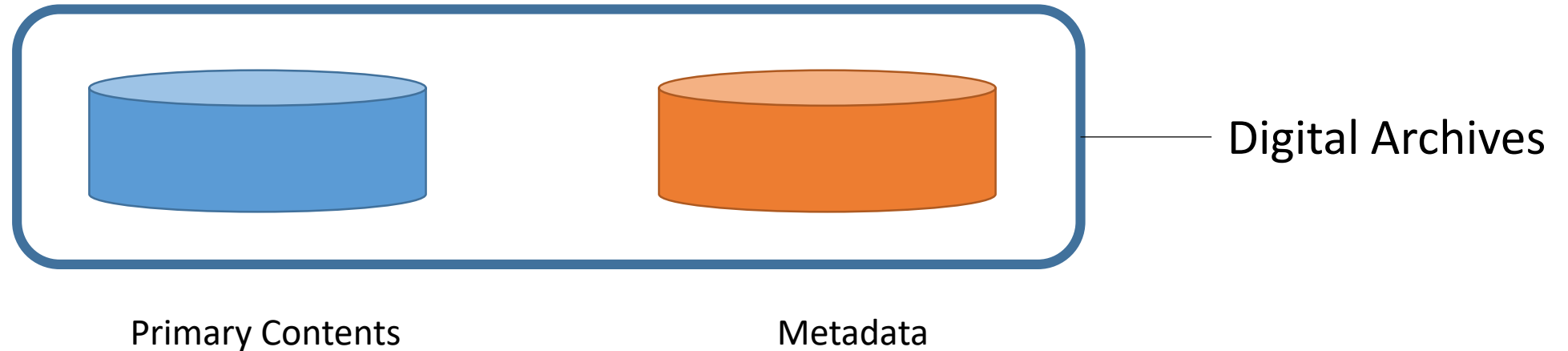
Longevity of both Primary Contents and Metadata is essential for digital archives

Metadata Longevity

- Keep metadata database alive over time
- Maintain metadata schemas over time
- Maintain metadata terms over time

This talk focuses on metadata interoperability and sharing

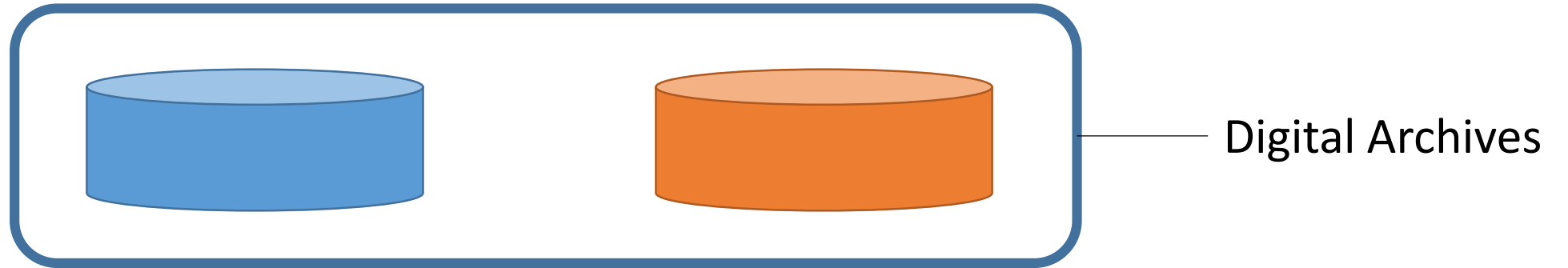
Sharing Digital Archives across Communities and over Time



Longevity of both Primary Contents and Metadata is essential for digital archives

Providers of digital archives are responsible to keep both primary contents and metadata alive over time. However, digital archives are sometimes lost because of limitation of financial and human resources.

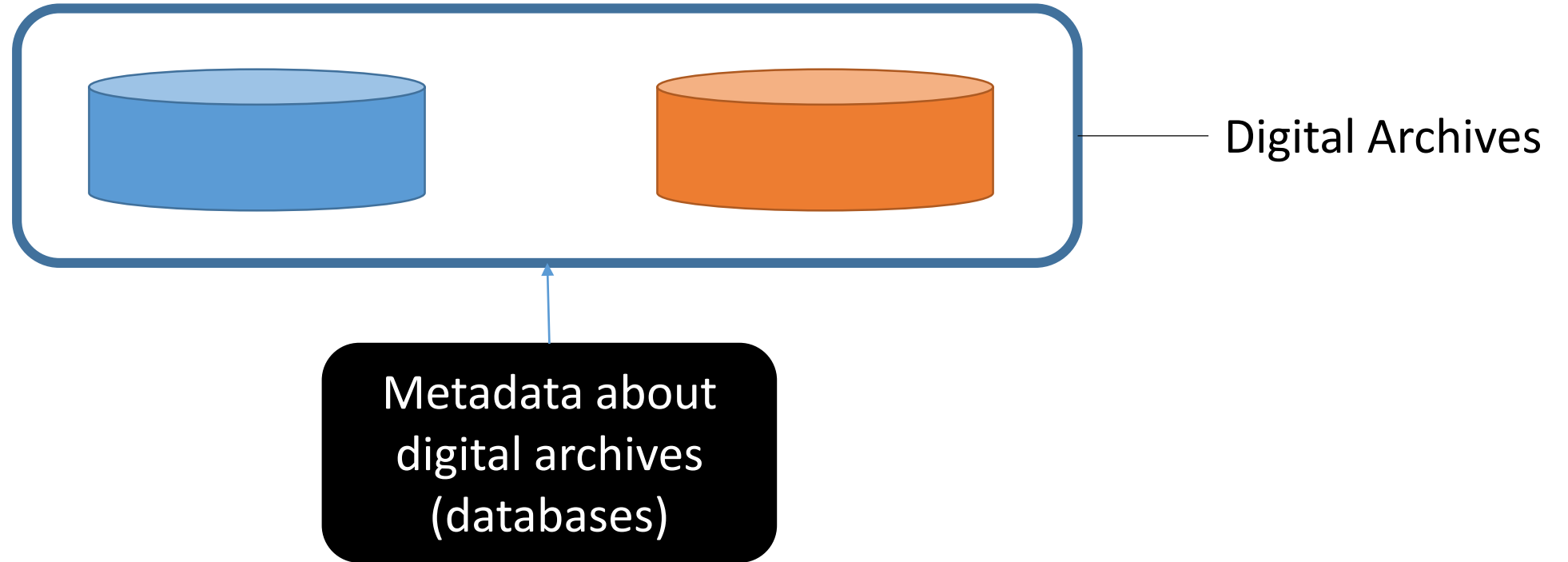
Sharing Digital Archives across Communities and over Time



A simple step forward – sharing information about digital archives in our community

Approach - sharing the digital archive information, i.e., metadata about digital archives

Sharing Digital Archives across Communities and over Time



Metadata schema may be defined by each digital archive provider but metadata schemas should be interoperable across providers – a well-known metadata interoperability problem

Metadata Interoperability

- Conventional methods to make metadata interoperable
 - A) All participating providers create metadata based on a standard schema
 - B) Create a mapping table across schemas used by participating providers
- Comparison
 - A) allows easy integration of metadata from all providers but limited flexibility to the providers and the providers need to get consensus on the metadata schema as their standard
 - B) allows providers to create their own schemas based on their needs but creating mapping tables is a heavy task
- In both methods, building a **data model** before property-level (or element level) definition/mapping is required

Metadata Interoperability

- What is data model of metadata?
 - Data model for an application problem defines entities to be described in the application and relationships among the entities
 - Data model should define classes/types of the entities
- A lesson learned in a research on the Media Art Database project
 - Media Art Database composed of four component databases – Manga, Anime, Game and Media Arts (<https://mediaarts-db.bunka.go.jp/>)
 - Each component database schemas were initially defined separately
 - Rich sets of elements were defined in these schemas without explicitly defined data models
 - Found difficulties in mapping their elements even though it looks that Manga, Anime and Game are quite close

Metadata Interoperability

- Metadata standards which are rather new include definition of their data models
- Data models play important roles for achieving metadata interoperability across different schemas (= communities)
- Dublin Core Application Profiles
 - Generalized model for designing interoperable metadata
 - Separation of structural and semantic features
 - Structural features in three levels – Description Set, Description and Statement (+ structural constraints)
 - Semantic features – Semantic definition of terms should be given separately from structural features

Metadata Model for Digital Archives/Databases – Discussion for Tomorrow

- A catalogue record for digital archives and databases (DA/DB)
- What Entities?
 - Digital Archive and Database
 - Agent involved in creation and/or operation/maintenance
 - Other entities?
- Facets for describing a DA/DB?
 - Content description – title, abstract, subjects, etc.
 - Stakeholders
 - Technological features – DBMS, user interface, API, etc.
 - Rights
 - Other facets?

Wrap Up

- Digital preservation is a maintenance activity of digital resources and digital archives including their metadata
- Digital preservation (or longevity of digital resources and archives) requires metadata interoperability
- Conventional element-level mapping is required for interoperating metadata of different schemas but model-level mapping before element-level mapping is recommended