



# Descriptive Standards for National Aggregators

5 September 2019: Library Conference Room  
University of the Witwatersrand



**Dr Roger Layton**  
[roger@rl.co.za](mailto:roger@rl.co.za)

[bit.ly/WitsNRF-20190905](https://bit.ly/WitsNRF-20190905)

PASSWORD: Digitize

Link will remain open until 8 Sep 2019

# The Facilitator

**Dr  
Roger  
Layton**

---

50 years working in IT

---

Artificial intelligence / Knowledge Representation

---

Database and Software development

---

15 years experience in heritage data and information systems

---

8 years experience in heritage training

---

Designer and developer of ETHER Base – Spectrum 5 compliant collection management system

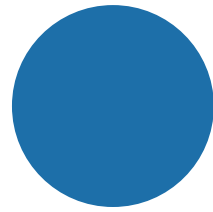
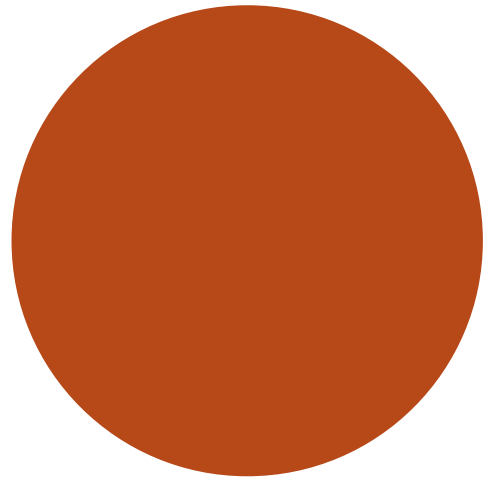
---

Project Leader and Writer of the National Policy on Digitisation 2009-11

---

# Where are we now? Where are we going?





# About The ETHER Initiative

Who we are  
What we do

# ETHER Product & Service Offerings

6

Organisational health

**ETHER GOVERNANCE**

Collection management

**ETHER DATA**

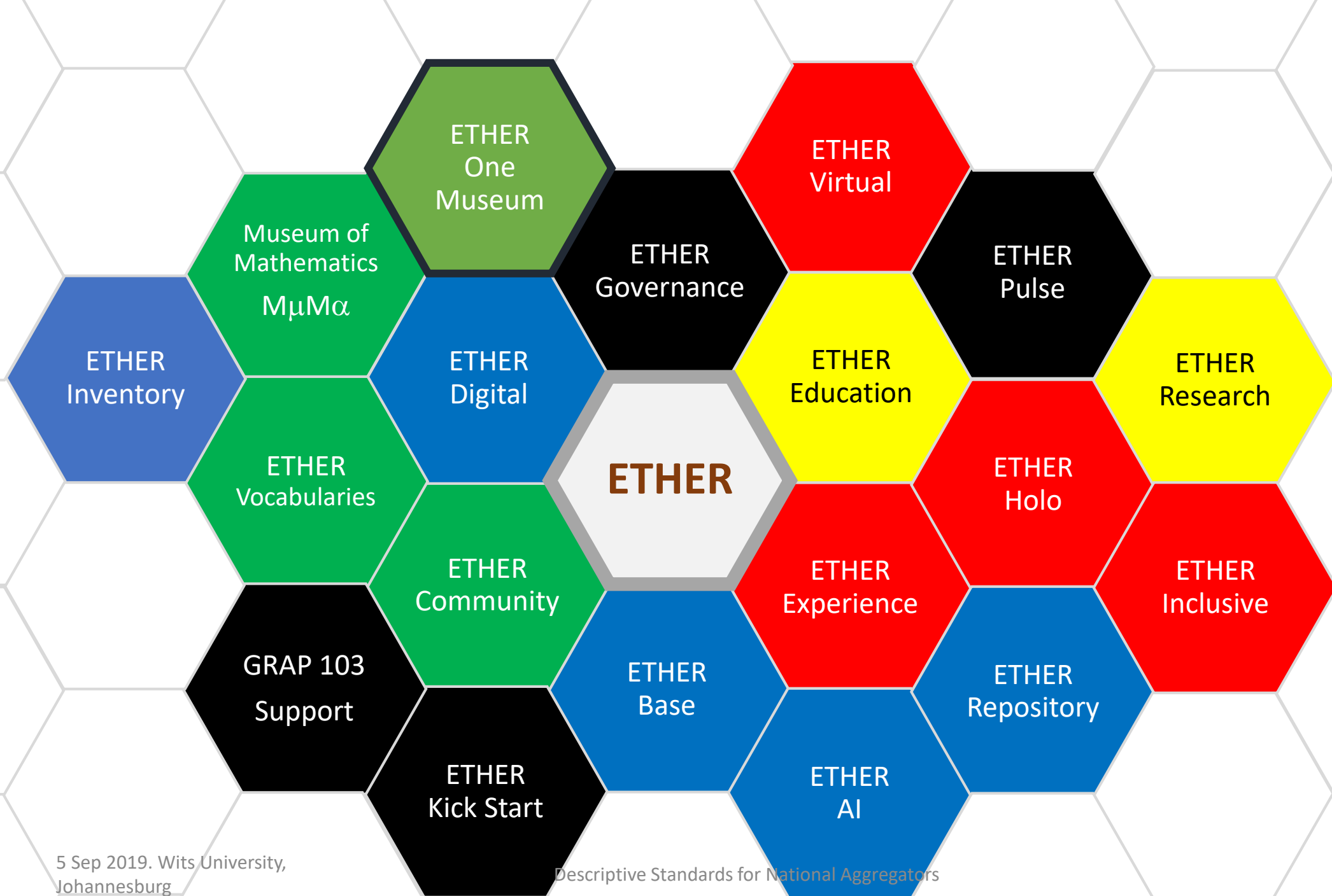
User experience

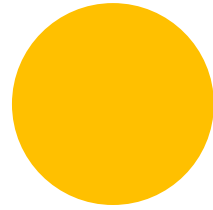
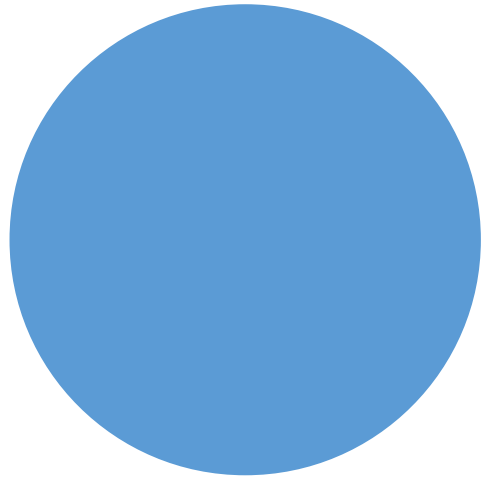
**ETHER EXPERIENCE**

Cross-cutting services

**ETHER EDUCATION**

**ETHER COMMUNITY**





# About this workshop

What I plan to do  
What I hope we will  
achieve



# What I plan to cover in this workshop

---

Background to National Aggregators and Descriptive Standards.

---

Descriptive Standards in more detail.  
Discussions on suitability and relevance.

---

Scenarios for the National Aggregator – a range of situations explored.

---

**CALL TO ACTION – what we have to do next.**

What I hope  
we will achieve

---

*There is no National Aggregator.*

---

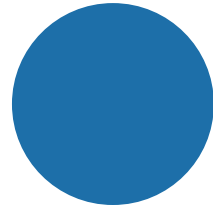
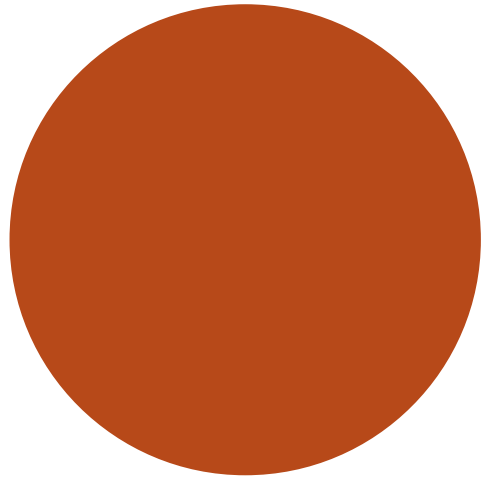
*There have been proposals, policies  
and discussions for 15 years.*

---

*South Africa is lagging far behind  
the world.*

---

**I propose that we build this  
ourselves, and that we start TODAY.**

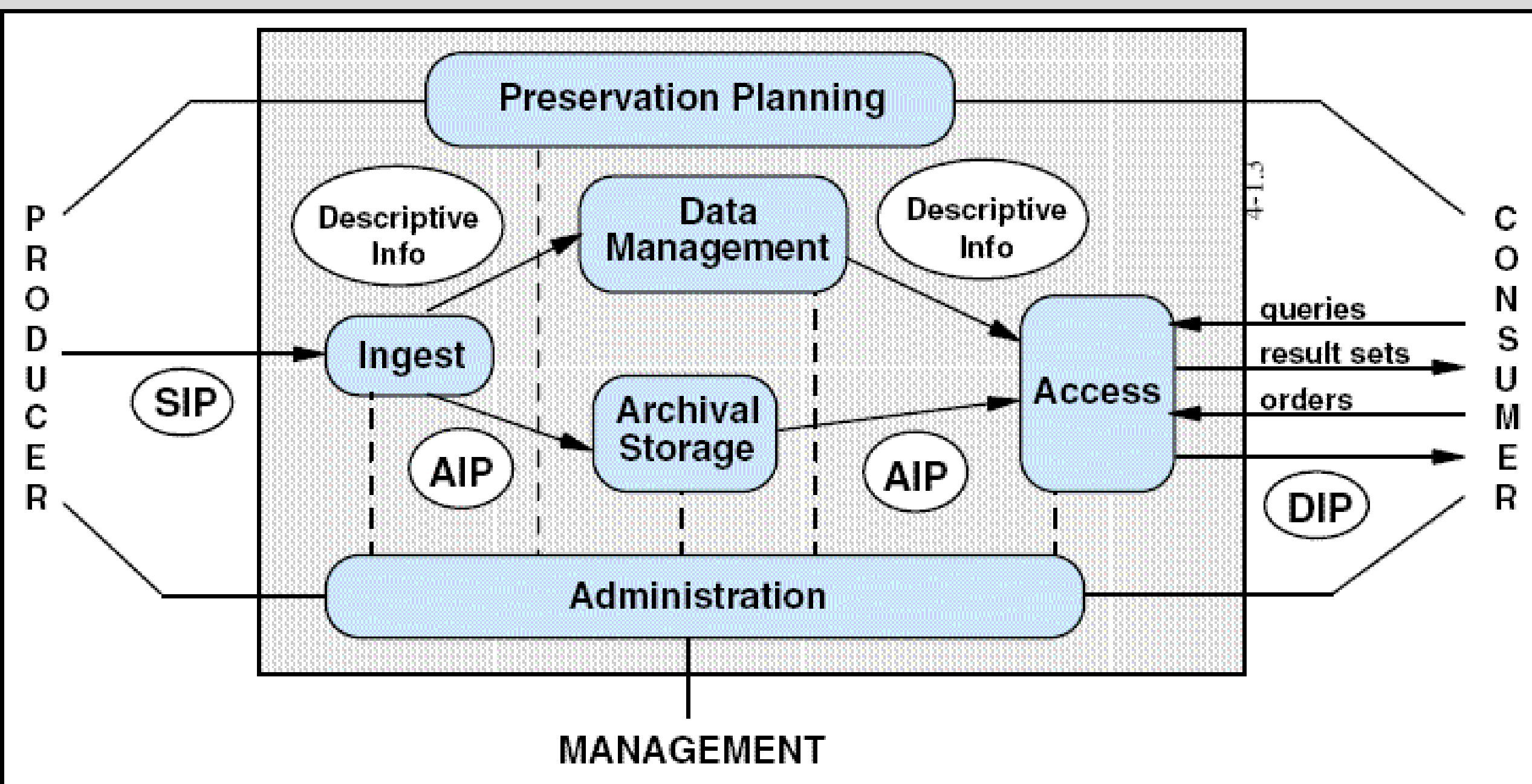


# Background

Aggregators and the problems of population, collation, and access.

What lessons can we learn.

# Context: Digitization and OAIIS



# What are we describing?

- CLASS DISCUSSION

- What do you hold, and what do you want to describe?

- Archives
- Library materials
- Museum artefacts
- Born digital
- Archaeological artefacts
- Natural science specimens
- Photographs
- Works of art
- ...

CLASS DISCUSSION

What is a descriptive standard?

Give some examples...

CLASS DISCUSSION

Why do we need  
descriptive standards?

CLASS DISCUSSION

How do we agree on  
descriptive standards?



# What is a National Aggregator?

SAHRIS: original design 2006

National Policy on Digitisation : National Digital Repositories

A single point of access to **trusted data**

- **Combine/aggregate** limited content from many sources
- Provides **linkages** back to the content at the source
- Long-term storage of the content
  - Sources provide online access to all details
  - or available in a repository / aggregator

This is a **key element** of the **Fourth Industrial Revolution**

# STRUCTURE OF THE DRAFT NATIONAL POLICY

## WHAT / WHY

### Policy

*27 policy statements*

*Governance*

*Strategy*

*Repository*

*Contracts*

*Access*

*Preservation*

*Metadata*

*Capacity*



## HOW

### Mechanisms

*Implementation recommendations  
and interventions*

## HOW TO

### Digital Heritage

### Body of Knowledge

*(DHBOOK) : Best-practice framework to create  
a Digital Institution in line with policy*



# Policy 14: Digital Masters

- Is the National Digital Repository an aggregator?
  - Providing limited basic data, to enable trusted searches back to the source.
- Our should this be a long-term repository of digital materials
  - The Legal Depository for the country.
- Digital Master
  - Provenance
  - Content and Subject Matter
  - Technical



# Policy 13: National Digital Repositories



Did not anticipate aggregators  
when this policy developed.



Rather focused on  
complete digital  
store

Preservation first  
Access second



Aggregator

Not built for preservation,  
but to provide a single point  
of access.

# Policy 13: National Digital Repositories

---

Build only a few – complex and expensive to create.

---

Fixed NDRs to be created by government.

---

Must be standards-based.

---

All digital content **MUST** be moved to an NDR.

---

Must be consistent.

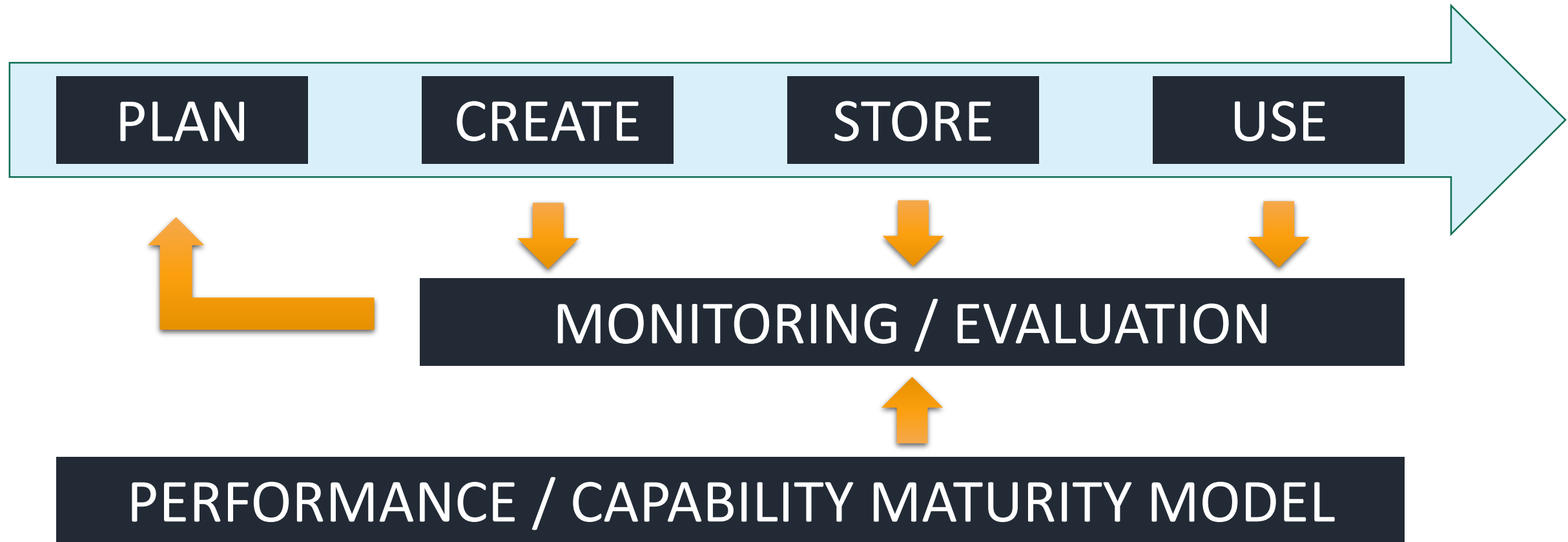
---

Must maintain a detailed register.

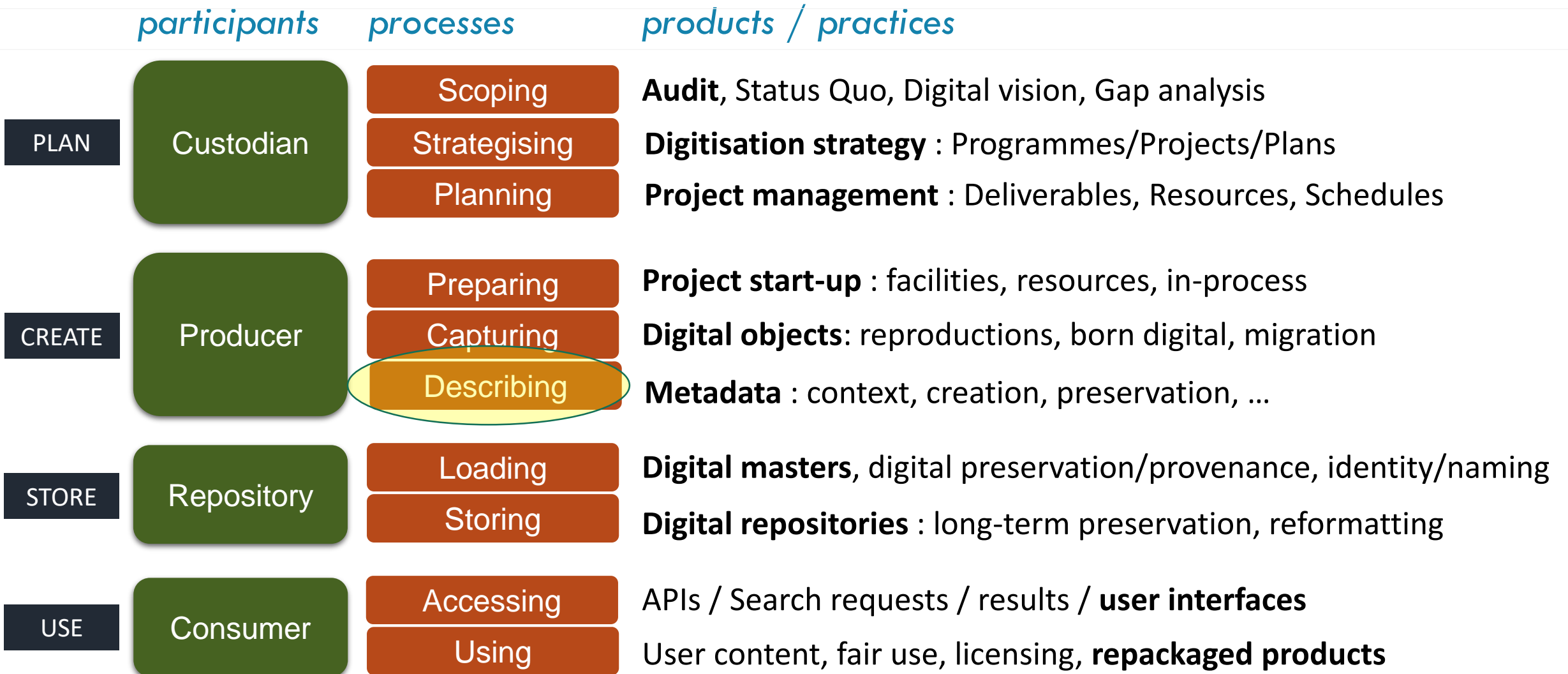
---

Must be certified to meet TDR and OAIS.

# High-Level Digitization Process Model



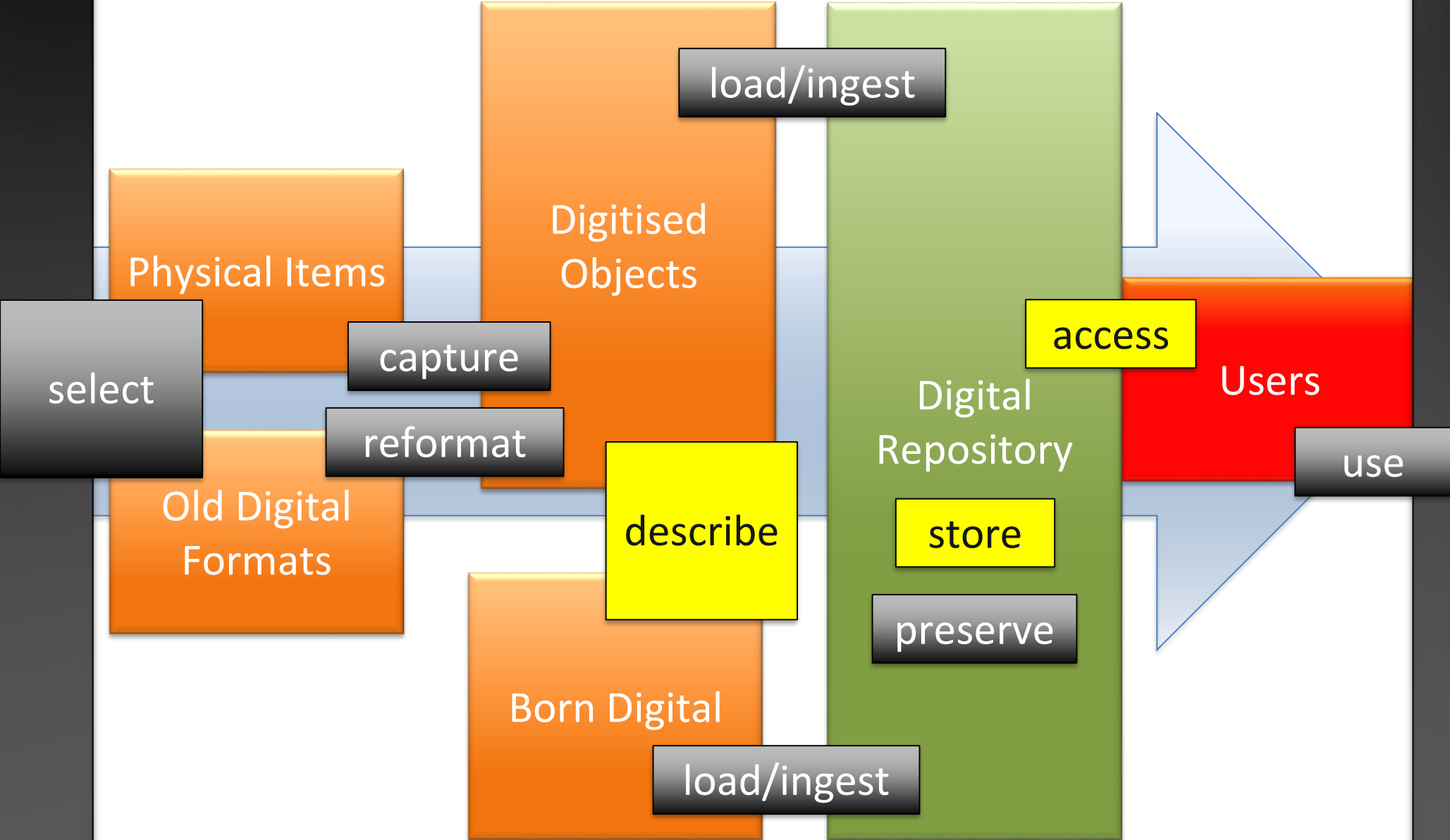
# Participants, Processes, Products



# SCOPE OF DIGITISATION

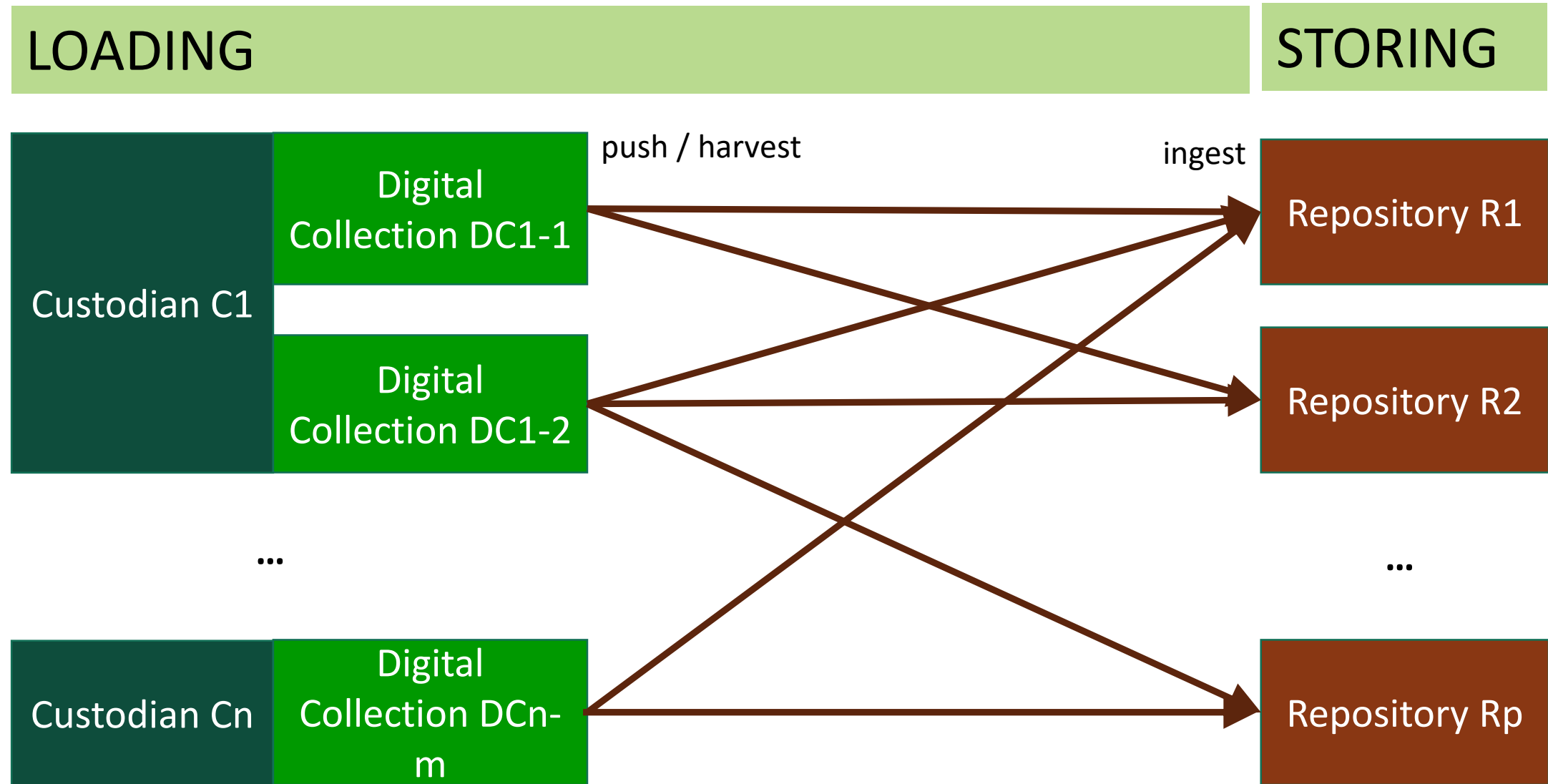
COLLECTIONS

PRODUCTS

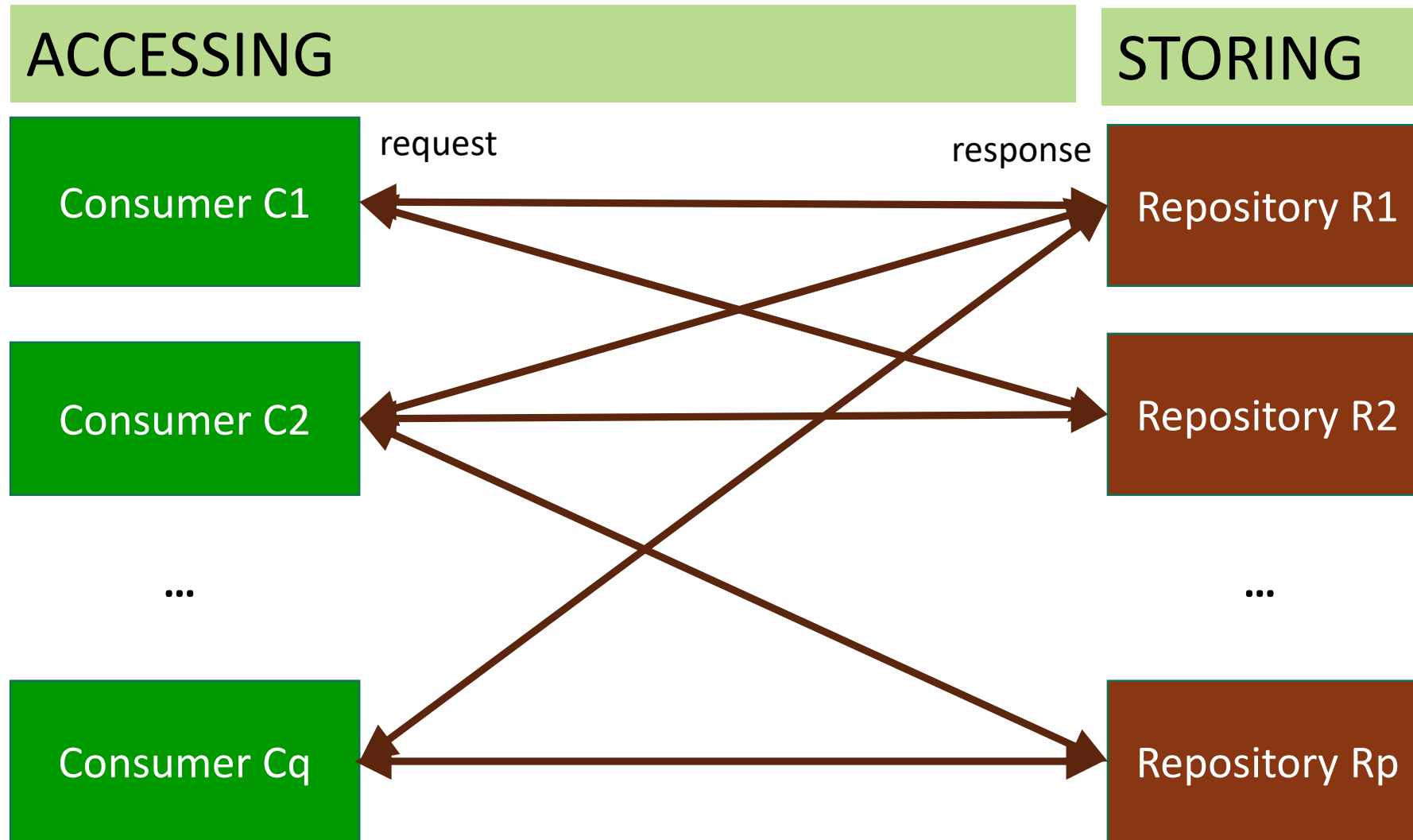




# Aggregation = Digital Providers Populating Repositories



# Aggregator Usage = Repository Contents Accessed + Used

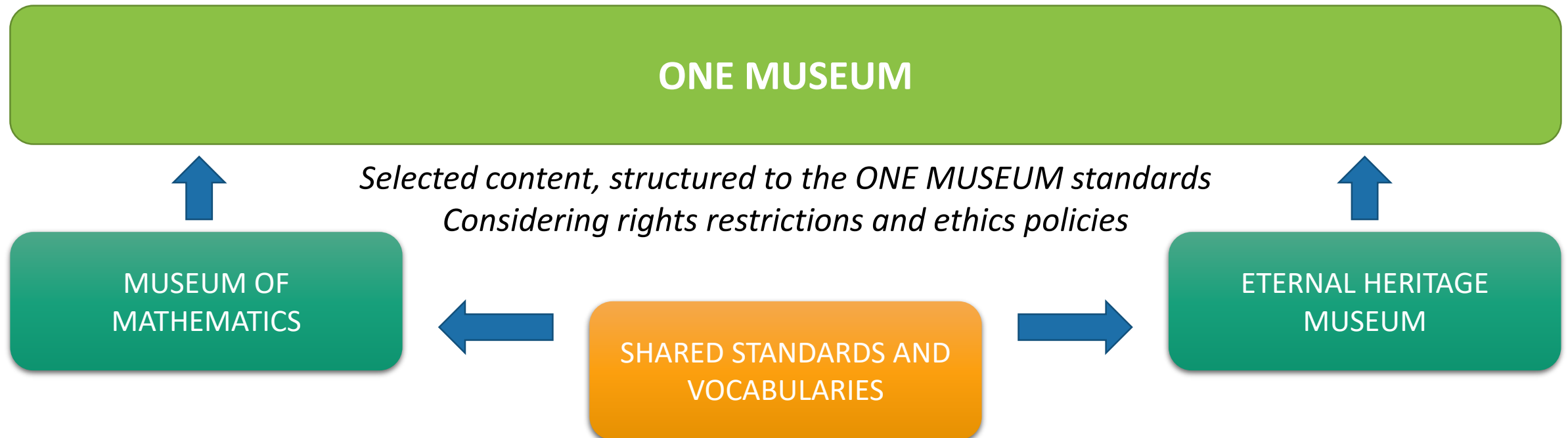


# Publishing to the ONE MUSEUM

Create **UNITY** of knowledge

Preserve **DIVERSITY** of custodians

## Combined content, from all museums



# Publishing Maturity Model

## Level 1: Custodian

**Custodian** *summary, mission, location, history, etc..*

## Level 2: Custodian + Collections

*...and collections.*

## Level 3: Custodian + Collections + Catalogue

*...and the detailed catalogue information*

## Level 4: Custodian + Collections + Catalogue + Digital

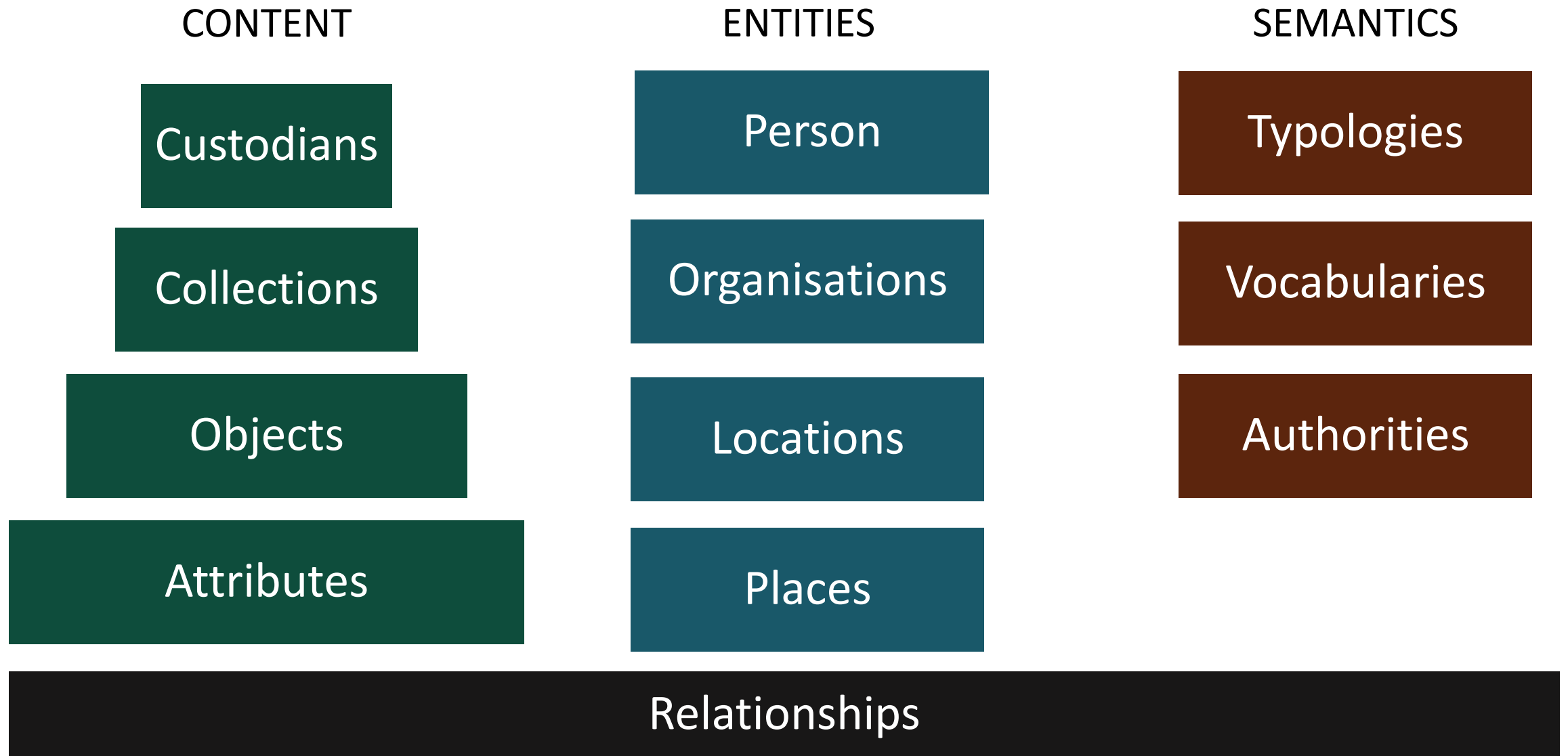
*...and digital reproductions + born digital content*

## Level 5: Custodian + Collections + Catalogue + Digital + Virtual Exhibitions

*...and curated virtual exhibitions/stories*

*Further content: user-created stories, biographies, databases, publications, research, ...*

# A Framework for Minimum Requirements



# Challenge #1: How to present information better?



# Challenge #1: How to present information better?

In this age of 3D,  
virtual reality,  
augmented reality...  
...why do we persist  
in representing all of  
our knowledge in  
words on the printed  
page?



# Challenge #2: Make content palatable to youth

- Too many repositories are stuck in the past – basic search and images, but too few connections.
- The youth are the Digital Masters of our world.
- GOOD: Turn the knowledge into collaborative games.
- BETTER: Get the youth to BUILD games based on the content.

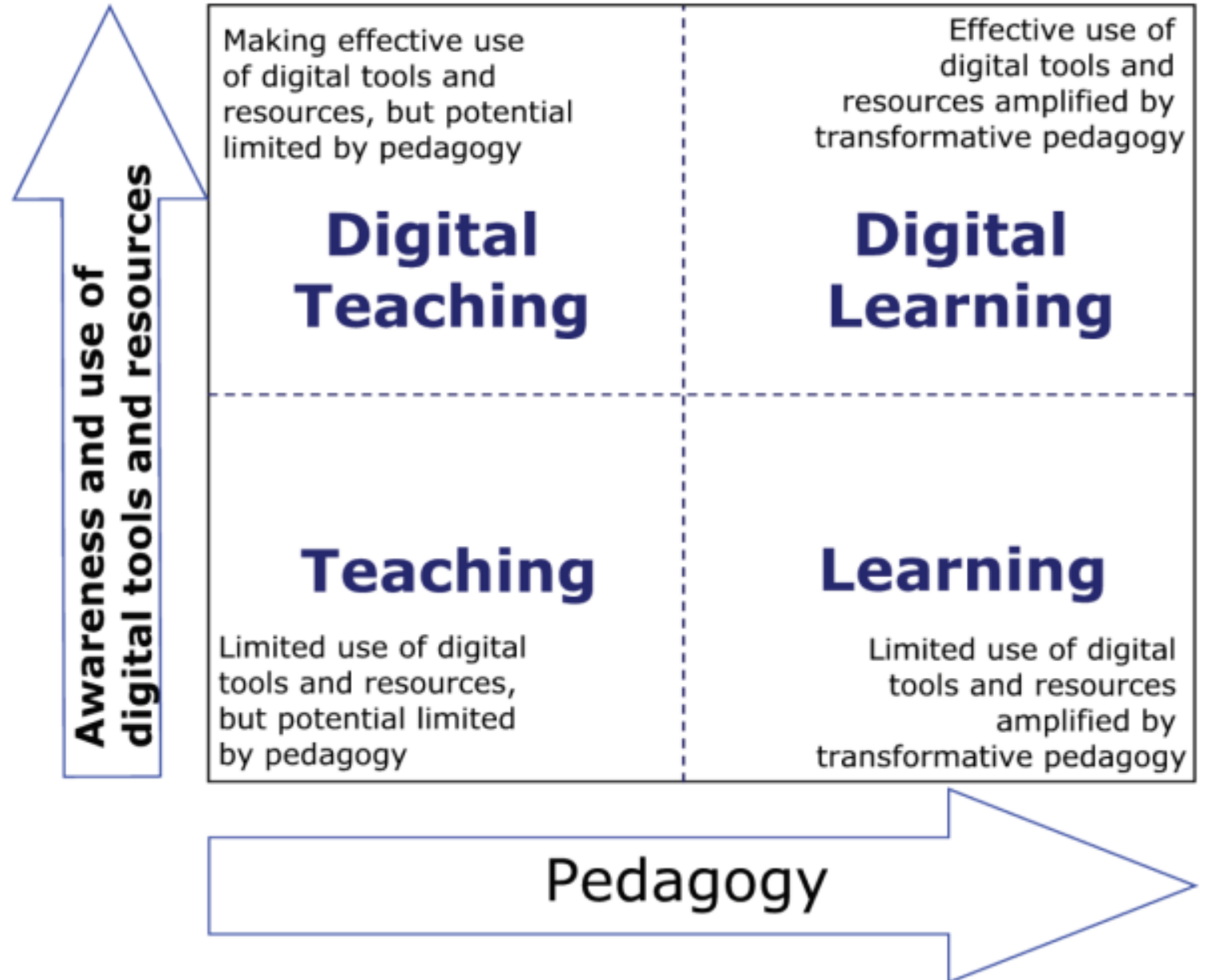






## DBE Framework on Digital Learning

Challenge #3:  
Accommodate  
the shift to  
digital learning



## Challenge #4: What grain size for the digital content

### COURSE GRAIN:

- Basic information only: Such as OBJECT-ID standard.
- Just sufficient for a simple search.
- Full data stored at provider.

### FINE-GRAIN:

- As much details as can be made available.
- LIDO / METS / EAD.
- May no longer need provider database – this is their cloud storage.
- Can act as a full backup – enough to reconstruct the original database.

# Challenge #5: Who will the aggregator belong to?

Government: Dept  
Communications /  
DAC / DST

Group of  
Government  
Departments

A new government  
SOE

A single university

A collective of  
universities

The National  
Library / National  
Archives

SADI

A new  
independent NPO

DISCUSS

## Challenge #6: Unique Identity / Identifier

Essential that every item has a unique identifier

Which will live with this item forever

- GUID / UUID – link to a data resources
- DOI – link to a web resource
- URI – generic unique label – may include GUID

How many unique identifiers exist?

- Source provider – unique within institution / collection – accession#
- Original source
- Newly created identifier
- Pre-established identifier – ISBN / ISSN

# Challenge #7: Hierarchical objects

- Most objects are structured into parts:
  - Archives
  - Books / chapters
  - Journals / volume / number / articles
  - Collection / items
- It is important to preserve these relationships between the source and the repository.
- And also to preserve other relationships:
  - Sequence: next/previous
  - References / citations: to other items





## Challenge #8: What's in it for the providers?

### CARROT

- Improved access to resources
- Do not need to build their own search engines
- Participation in a national and global initiative
- Unlock the collections

### STICK

- If do not participate may be left behind
- How much do we have to pay?
- How much work to prepare our data?

## Challenge #9: Who will pay for the aggregator?

Will you invest in this?

How much control do you want for your investment?

Are you already building your own?

**CLASS DISCUSSION**



# Challenge #10: Are you inventorized? GRAP 103?

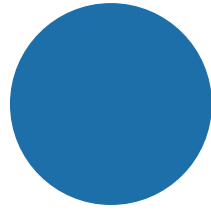
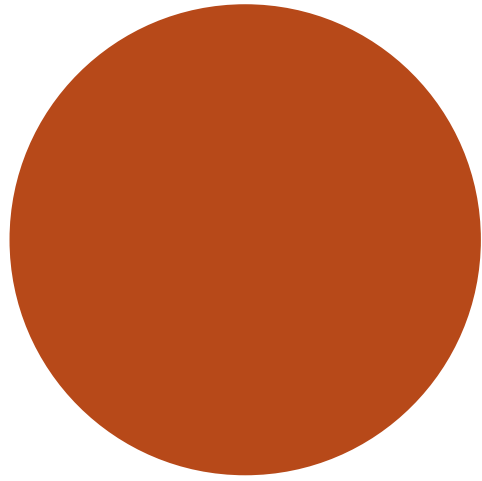
For many organisations, GRAP 103 has become a significant effort of compliance? However, this is also a call-to-action in getting the house in order.

## **INVENTORY: NO SEMANTICS**

- What you have as heritage assets
- Where it is
- What condition it is in
- What is its value – from professional valuer

## **CATALOGUE: FULL SEMANTICS**

- These are the descriptive standards.



# Descriptive Standards

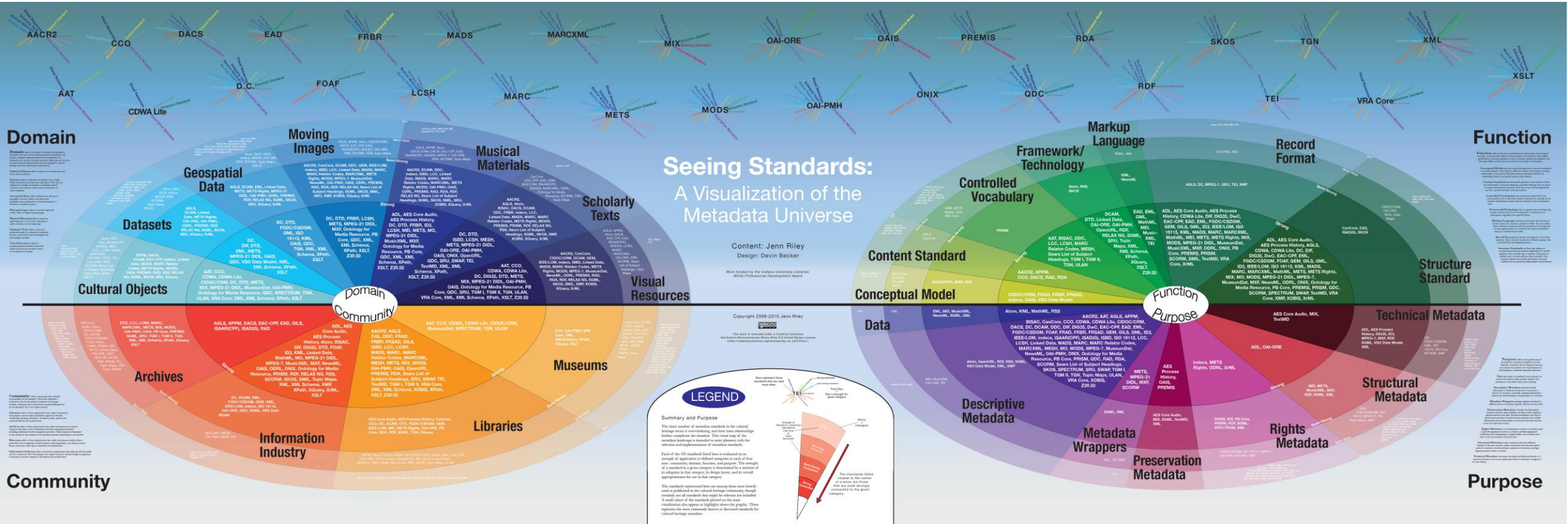
What are we describing?  
In what level of detail?  
And why?

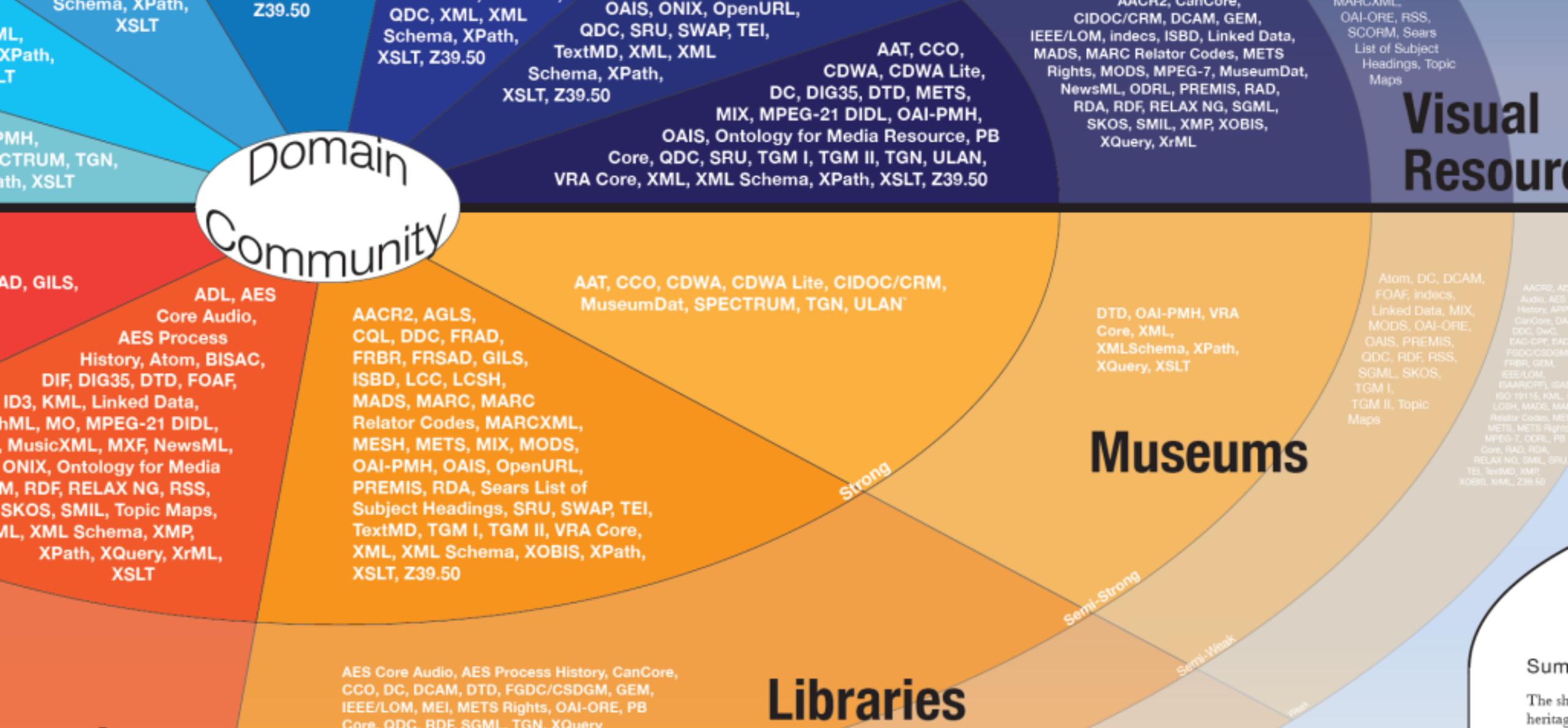
---

“The nice thing about standards is that you have so many to choose from.”

---

*Andrew Tenenbaum. Computer Networks, 2nd ed., p. 254.*





# Classification of Descriptive Standards



Schemas



Vocabularies



Authority Files

# Schema Standards

- What fields are needed to represent an item
  - ...or an element of an item
- Common fields
  - Title
  - Name
  - Description
  - Date Created
  - Author
  - ...



# Breaking down a schema

- Individual fields have semantic commonalities
- Simple data field
  - Author: link to Persons authority file
  - create new fields for other authors and for other roles
- Generic data field: Roles
  - **Role = Author, link to persons**
  - Can add as many as possible
  - And can add new roles





Descriptive standards are SEMANTIC

### CCO: Catalogue of Cultural Objects

- Generic lists of fields and their definitions
- For all kinds of cultural objects

### CDWA:

- Works of art – more specific
- Less generic

All indicate WHAT should be stored

But do not indicate HOW

# CCO: Cataloguing Cultural Objects

- The AACR2 of cultural objects
- paintings, sculpture, objects of material culture, architecture/built works, installation art, performance art, manuscripts, photographs, prints, etc.
- Suitable for museums, archives, libraries
- Follows from other metadata standards:
  - AAT, TGN, etc.
- Maps to Dublin Core and MARC 21 and crosswalk to other standards

- Object names
- Creator – including other roles
- Physical characteristics – size, condition
- Stylistic / chronological – cultural origins / dates
- Location / geography – where now? Where discovered? Made?
- Subject – what is depicted
- Class - thesaurus
- Description – a range of descriptive notes
- View – what view of the work is displayed (e.g. building)

# Syntax standards

- Excel
- XML
- CSV

These only depict the manner in which the information is presented for later usage. For example, to transfer to a repository for ingest into the content.

# Object-ID: A minimal standard for objects

## Nine categories of information

- **Type** of object
- **Materials** and techniques
- **Measurement**
- **Inscriptions** and markings
- **Title**
- **Subject**
- **Date** or period
- **Maker**

## Four processes

- Taking **photographs** of the object
- **Informing** the above mentioned categories
- Writing a **short description** including additional information
- Keeping the constituted documentation in a **secure place**

<https://pro.europeana.eu/post/what-is-digital-storytelling-and-what-has-it-got-to-do-with-cultural-heritage>

“Think [online exhibitions](#) that use narrative text to weave together curated content - images, texts, video “

[https://www.europeana.eu/portal/en/exhibitions/heritage-at-risk#ve-anchor-intro\\_15449-js](https://www.europeana.eu/portal/en/exhibitions/heritage-at-risk#ve-anchor-intro_15449-js)

Add a search term



Browse ▾

Explore 57,362,447 artworks, artefacts, books, films and music from European museums, galleries, libraries and archives

Europaana

Rynek Starego Miasta w Warszawie. Strona Dekerta (Północna), przed 1939, Henryk Poddębski  
Biblioteka Cyfrowa Politechniki Warszawskiej

Public Domain

EXHIBITION

5 Sep 2019 | MITA University,  
Johannesburg

Descriptive Standards for National Aggregators

55

FEEDBACK

Computer games × Add a search term

LANGUAGE: English ×

REFINE YOUR SEARCH

COLLECTIONS

- All Items
- 1914-1918
- Archaeology
- Art
- Fashion

More ▾

MEDIA

1 - 12 of 97 results

Per page: 12

GRID

LI

More than a game: The computer game as fictional form (2003) | Atkins, Barry

Whether you love them or loathe them, look back with wistful nostalgia to the days of Pong and Space Invaders, or regard the whole phenomenon with blank incomprehension, there is no doubt that computer and video games now occupy a significant place in contemporary popular...

Game

View at OAPEN Foundation

Text







Computer games shop in Gillygate, Pontefract. Photograph courtesy of the Pontefract and Castleford Express.


Computer games shop in Gillygate, Pontefract. Photograph courtesy of the Pontefract and Castleford Express.

SHARE

DOWNLOAD

CAN I USE IT?

No 

[View more at Wakefield Council](#) 

## People


**Creator:** [Pontefract and Castleford Express](#)

## Classifications

**Type:** [photograph](#), [Photograph](#)

**Subject:** [Shopping](#), [Shopping - Household goods](#)

## Extended Information

Close all 

### Time

Creation Date: 1980/1989

---

### Provenance

Identifier: WAKGMP : P2002.1178

Institution: Wakefield Council

Provider: AthenaPlus

Providing Country: United Kingdom

First Published In Europeana: 2015-11-18

Last Updated In Europeana: 2016-07-12

---

### References And Relations

Dataset: 2048087\_Ag\_EU\_AthenaPlus\_CollectionsTrust

---

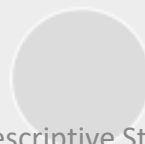
### Explore By Colour



#2F4F4F



#A9A9A9



#DCDCDC



#696969



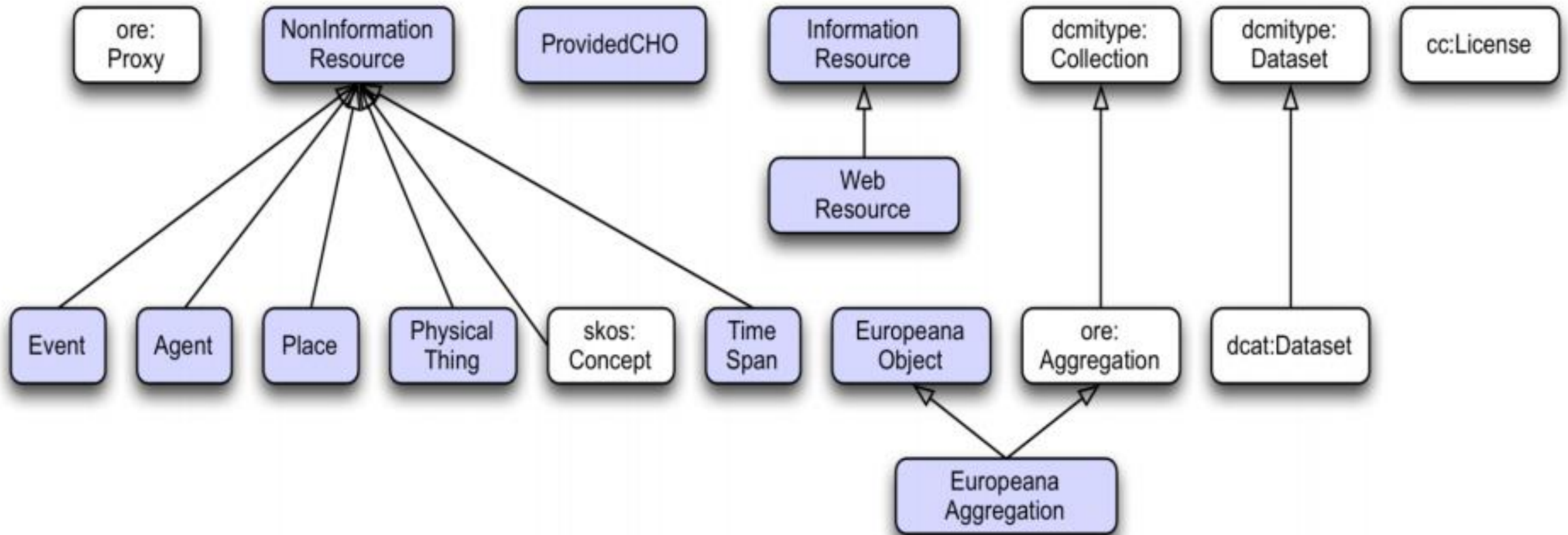
#C0C0C0



#000000

Descriptive Standards for National Aggregators

# Europeana Data Model



# Spectrum 5

Best-practice procedures for collection management.

Comprehensive set of information units on which databases should be developed for all collections and items, of all types (cultural, natural, ...)

<https://collectionstrust.org.uk/spectrum/>

# Spectrum 5 information groups

- Object identification
- Object description
- Object location
- Object requirements
- Object production
- Object history / association
- Object field collection
- Rights / rights in / rights out
- Object condition / assessment

- Object conservation
  - Object usage
  - Object valuation
  - Object collection review
  - Object audit
  - Object owner / user contributions
- Standards for Persons, Organisations,  
Places, Locations, Address,  
References / Citations

This is specific for natural history specimens.

## Object status

---

### Definition

A statement of the standing of a natural science specimen or other object in relation to others in existence.

### How to record

Use a single term, without punctuation or capitalisation. Maintain a list of standard terms.

### Examples

type; paratype; holotype; paralectotype; copy; forgery

### Use

Record once only for an object.

### Information group

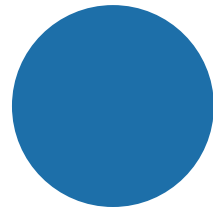
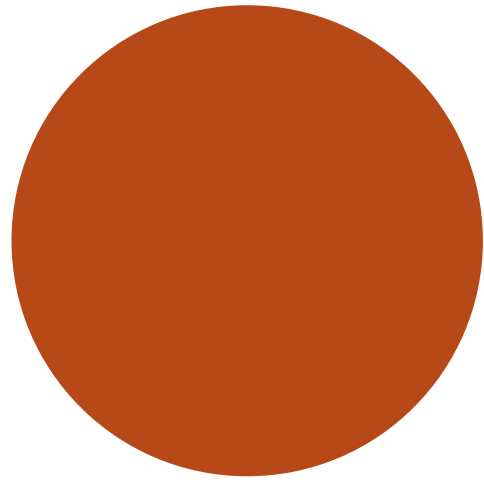
Object description information

# The SAHRIS Data Coding Standard

Will open as PDF files:

- Data Coding Standard
- Vocabulary Management





# Scenarios

Try it using gedanken experiments.

# Gedanken experiments

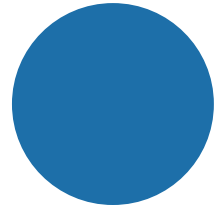
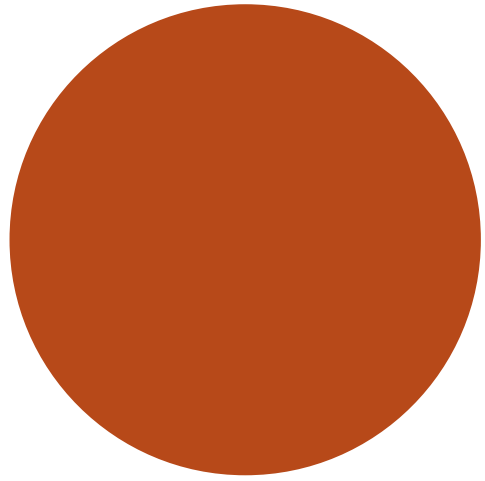
- Imagine that the **national aggregate already exists** and is operating.
- You will now **analyse this aggregate in groups** to describe what you see in your mental imaginings.
- Divide into groups, and **each will focus will be one specific element** of the aggregator design concerning the data standards.
- All groups are concerned with how data from various sources can be combined to appear to be all unified into a single model.
- It is assumed that each of you have different data management systems for your collections, and yet all need to move this into a single aggregate structure. How will you reach a compromise on how to combine, and how to ensure that there is no duplication.

You have 30 minutes for this - followed by 30 minutes of report back.

We will then use this to inform the development of the National Aggregator in the call to action.

# SUMMARY OF SCENARIOS

- Dates and Time Periods
- Places
- (Persons, Organisations) as Agents, Naming
- Measurements
- Vocabulary Management – as shared resources
- Customizing the Search and Query Interface
- Using AI: in the Ingest Engine / in the Search Engine
- Access by schools – reusage and citation
- Digital storytelling
- Composite objects (which contain a variety of types of objects from different collections)
- Provenance



# Call to Action

Let us build it.

# Let us built it – Stage 1 Starts TODAY!

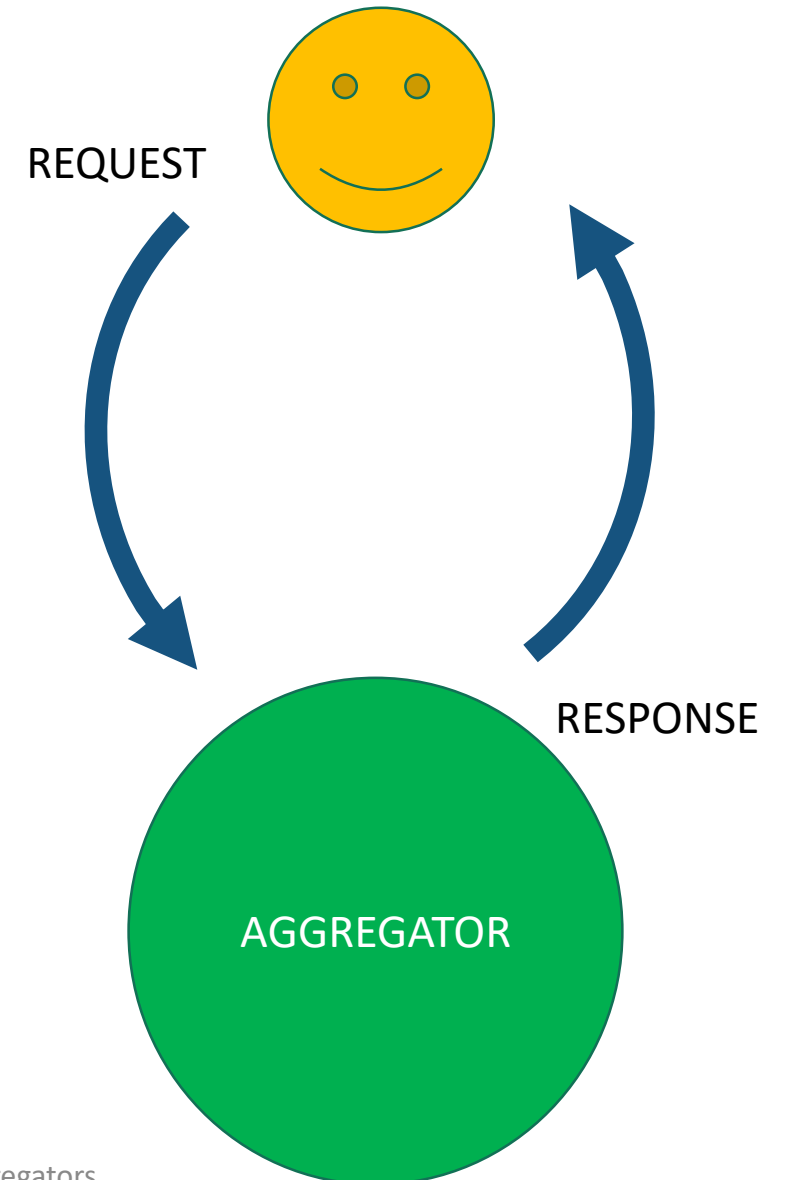
- DISCUSSION Whose responsibility is it to create the national aggregator?
  - Government
  - Private enterprise
  - Universities
  - Some collaboration which can be formalised (there is money and resources needed for this to work).
  - Europeana : initiated and funded by the EU as their digital library

# Getting Started: Identify Stakeholders

- Inclusive approach to stakeholders - any institution with an interest in aggregated publishing – with resources to share
  - Libraries / Archives / Museums / Natural History / Heritage Sites
  - Private / Education / Research / Government
- All user groups who need trusted content
  - University students
  - Scholars / learners
  - Researchers
  - Interested persons / public
  - Government users / policy makers

# Scope: What functions will it provide?

- Scoping – specify in terms of
  - BLACK-BOX APPROACH TO SCOPING
    - What messages / commands it will receive - content
    - What response it will give to each message - content
  - GLASS-BOX APPROACH
    - How it will respond to the messages
    - What internal structures it has to respond
    - Data models
    - Data processing and operations
  - PROTOCOLS
    - How messages are delivered – formats / channels
    - How responses are returned – formats / channels





# Let us use a gedanken experiment again

## DISCUSSION

- Enumerate all of the messages which the aggregator can receive
- Examples – population of the aggregator
  - Register a new institution / custodian
  - Update details of institution (name, contact persons, etc.)
  - Add a new collection
  - Add collection details
  - Add new item – descriptive standards / images / videos
  - Bulk loading of collection / items
- Examples – usage
  - Conduct a search and get results
  - Open up a single result
  - Store a result
  - Link back to source institution for further details

# Stage 1: Project Planning

- Establish a Team
- Identify initial responsibilities
- Establish initial objectives – low-hanging fruit
- Identify resources
- Identify sources of funding and how money will be spent
- Establish milestones and a measurement system
- SWOT analysis
- Business Case – benefits and beneficiaries
- Establish readiness of institutions to participate
  - Political will, internal readiness, resources and skills
- Define a maturity model – for the aggregator and for participating institutions

# Stage 2: Rapid Pilot Creation

- Cloud Service Provider
- Database model (AIP)
- Loading process / ingest (SIP)
- Web-based search and results (DIP)
- Pilot institutions – reduced content to be provided
- Implement maturity model
- Develop a test set and test project
- 3 months

# Stage 3: Institutional adoption

- Each institution which will be providing their data into the National Aggregator
- Must prepare their own data
  - Inventorise their data using sufficient information and finding aids
  - Develop a tool to extract data and publish into a suitable format
- Send data packages and receive confirmations or errors
- Continue this process
- Automate this process

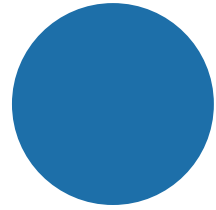
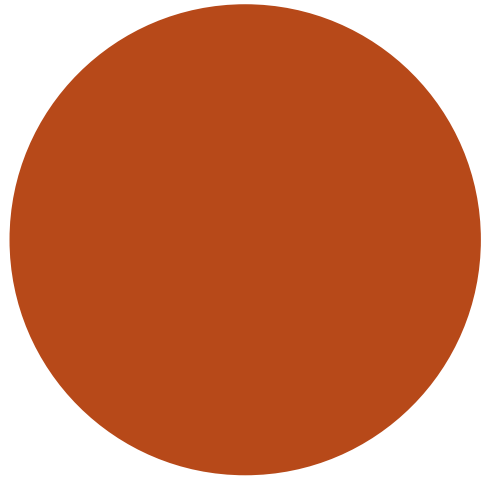
# Stage 4: Open for Business

- Open for the users to commence the usage of this
- Needs a minimum size of content to make this useful for the users
- The maturity model will assist here
  1. Finding Institutional sources
  2. Finding collections
  3. Finding items
  4. Finding digital assets
  5. Finding digital stories
  6. Building and publishing digital stories

In this way, content is available FROM THE START in aggregated form (collections and not items).

# Stage 5: Expand and Sustain

- Expand the capabilities in the aggregator.
- Capacity development of providers to move through the maturity model.
- Build a committed community so that this is not dependent on just a few individuals and organisations.
- Identify all possible users – including schools – of this being a TRUSTED repository of knowledge.
- Meet the TDR requirements in full over the long term.



End

Questions and Comments