

Digital Objects: theory and practice, critical approach to digital preservation

Gabriella Ivacs, Open Society Archives, ADA 2014

July 1, 2014.

Content

- General Concepts: digital archives/sources, recordness, preservation
- Metaphors of Information, digital object theory
- Conceptual frameworks of Interoperability: OAIS, Standards, best practices
- Case Study: Europeana – policy instrument
- Case Study: Heritage of People's Europe (HOPE)
- Break out Session

How to study digital archives

- Studies of information society, information economy, (*e.g.*, Castells, 2001, Benkler, 2006, Lessig, 2009)
- Specific technologies in particular contexts (*e.g.*, Orlikowski, 2000, Bijker, 2001)
- Internet studies (*e.g.* Zittain, 2008, Jeffrey Schnapp, 2008).
- Cultural Studies, memory studies (*e.g.* Aleida Assmann, 2008, Jan Assman, 2008)
- Information studies (*e.g.* Borgmann, 1999, Bowker, 1999)
- Instability that digital objects introduce across a variety of settings and situations (Kallinikos,
- Philosophy, normativity (*e.g.* 2012, Barham, Dreyfus, 1991)
- Normativity agency and life, Studies in History and Philosophy of Science, Part Studies in History and Science.
- Archival studies, (*e.g.* Cox, 2007, Hedstrom, 1995)
- New media studies, (*e.g.* Ernst, 2012, Manovich, 2001, Weinberger, 2007)
- RIM Business Studies, (*e.g.* Ciborra, 2007)
- Documentation studies (*e.g.* Briet)

Importance of Digital Sources

- “historians tend to assume that responsibility for preserving the past of digital data is a merely "theoretical and technical issue, tomorrow's problem or at least someone else's problem" (Rosenzweig: 2003, 759)
- The quality of sources, including their preservation as well as the ability to retrace, access and evaluate them, is decisive for historical work. (Gabriele Balbi: 2011)
- “Future presents” cannot be properly determined from a present view.
- Old and New Sources: “remains” and “testimonies”?

Hybrid Sources: “The interesting, the picayune and the profane”: 3 types

1. Old Sources: analog information, data, document
 2. Digitalized data actually comprises old sources that were transformed into digital material
 3. Born digital data refers to data that has never existed in a printed or other analogue form
- “Digital archives can be far larger, more diverse, and more inclusive than traditional archives” (Cohen 2005, 10)

Archives' Mission

- **Archives organize and manage records to protect their integrity as documentary evidence and to provide successive generations with materials for fresh interpretations of collective actions and consciousness.**

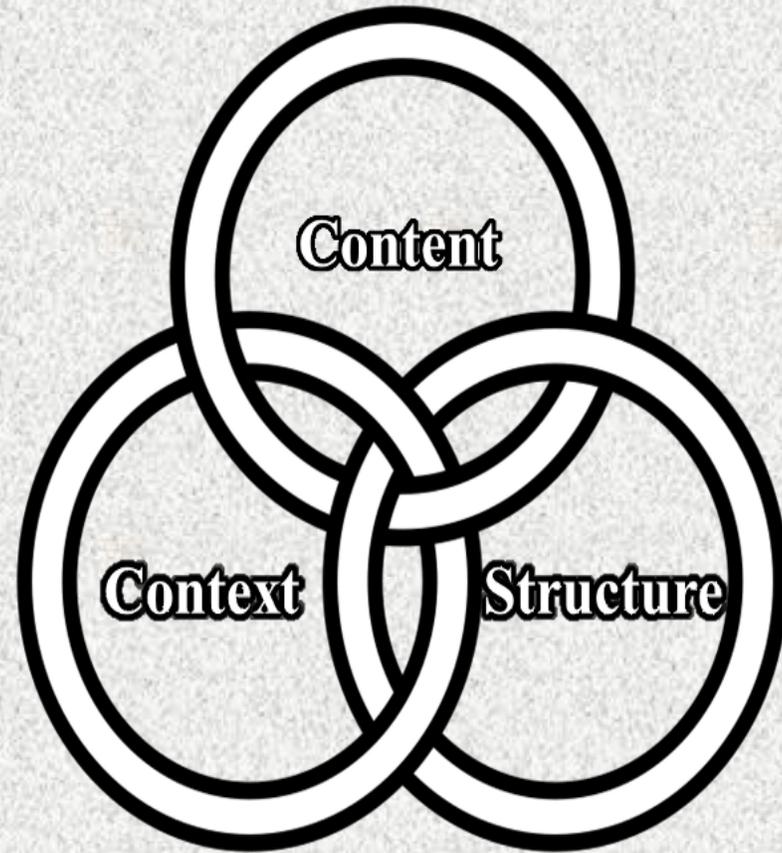
- **Institutions**
- **Individuals**



- **Respect for the principle of provenance means that records must not be separated conceptually from the broader context of their origin, creation, and use.**

Recordness

- Complete
- Fixed
- Organic
- Contextual
- Authoritative
- Unique



TRUST

Digital Preservation

Digital preservation is a formal endeavor to ensure that digital information of continuing value remains accessible and usable. It involves planning, resource allocation, and application of preservation methods and technologies, and it combines policies, strategies and actions to ensure access to reformatted and "born-digital" content, regardless of the challenges of media failure and technological change.

The goal of digital preservation is the accurate rendering of authenticated content over time.

http://en.wikipedia.org/wiki/Digital_preservation

The 3 Objectives of Digital Preservation

- Identifying objects that are in danger of digital obsolescence
- Planning and carrying out technical preservation actions
- Defining the necessary technical metadata in order to provide an environment needed for digital object delivery and access.

The Three Pillars of Digital Preservation?





Digital Objects: feasible or not to manage?

- „Authenticity and integrity could be evaluated as inference on the basis of the trustworthiness of the document/information system in which the documents/information exist”

- *(APARSEN D24.1 Report on Authenticity and Plan for Interoperable Authenticity Evaluation System)*

- Digital objects are marked by a limited set of variable yet generic attributes: "persistent representations of activities, created by participants or observers of those activities or by their authorized proxies..."

- (Geoffrey Yeo, "Concepts of Record (1): Evidence, Information, and Persistent Representations." American Archivist; Vol. 70; Fall/Winter 2007; pp 315-343)*

Digital Objects: feasible or not to manage?

- "the difference which makes a difference" to someone. In other words, information was a process of mutual alignment and was not an object to be transmitted. (Bateson,1972)
- "persistent representations of activities, created by participants or observers of those activities or by their authorized proxies..."
- „...making of granularity to a ubiquitous technological principle that relies on binary parsing is a remarkable accomplishment. It confers digital objects a distinct ontological and functional profile.”
(Kallinikos, 2010)

Digital Objects and Networks:

- Editability is intrinsic value: to rearrange elements by which the digital object is composed
 - e.g. archival item in collection
 - e.g. built-in functionality in databases and digital repositories
- Interactivity: it does not need to result in any change or modification of the digital object, it enables actions of contingent nature.
- Openness: open and reprogrammable in the sense of being accessible and modifiable by a program. Their STRUCTURE can change.
- Distributed: there is an assembly procedures by which a digital object is brought to being
 - e.g. algorithm, protocols

Metaphors about information

- These mathematical definitions were elaborated by several metaphors
- Joint Father of cybernetics: Norbert Weiner (1949) and Claude Shannon (1948)
- MODELING! to study the properties of systems independent of their physical realization: electronic circuits, brains, or organizations
- Isomorphic systems: high-level concepts such as order, complexity, information, control, and difference (Heylighen and Joslyn 2002)
- (Making Truth: Metaphor in Science. Theodore L. Brown. 2003)

The Metaphor of Refinement: Norbert Wiener, (November 26, 1894 – March 18, 1964)

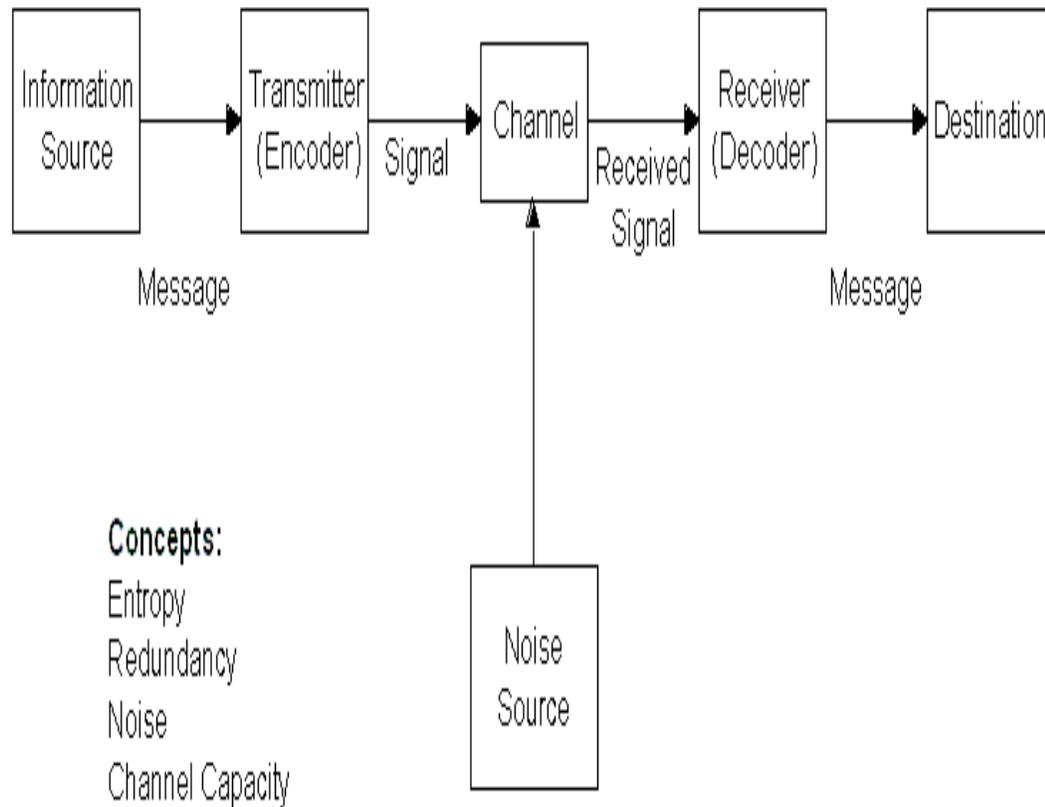


Application of the Refinement Metaphor: Information Pyramid

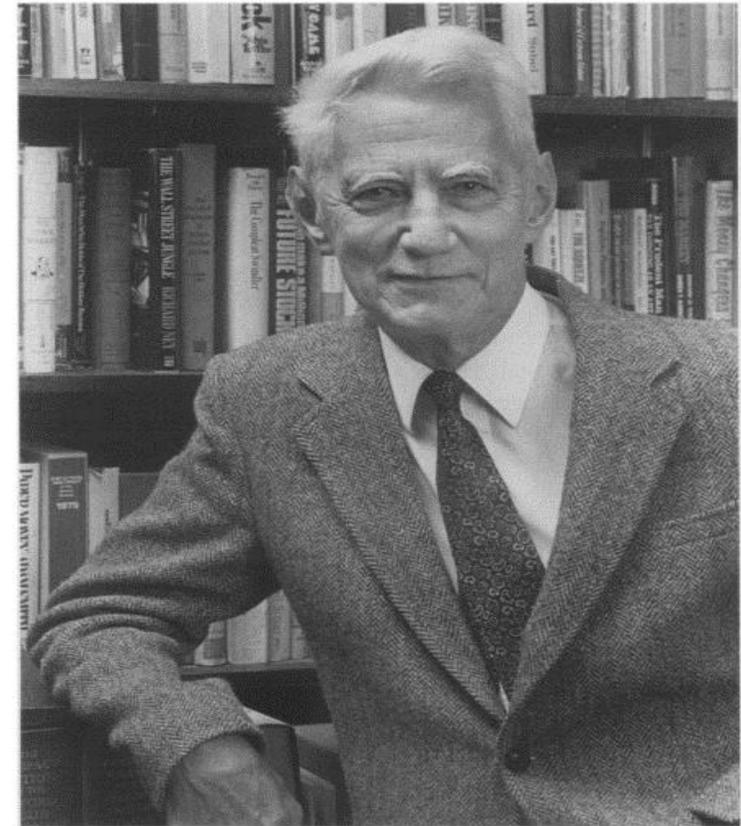
- Organization of computer systems: hardware, bitstream, middleware, structures such as application programs at the top.
- Knowledge management systems
- Scientific publications: raw data vs. edited product
- Archival collections and users: fonds/subfonds/series and curated collection, online exhibitions

The Metaphor of Invariance: Claude Shannon, (April 30, 1916 – February 24, 2001)

The Shannon-Weaver Mathematical Model, 1949



CLAUDE ELWOOD SHANNON



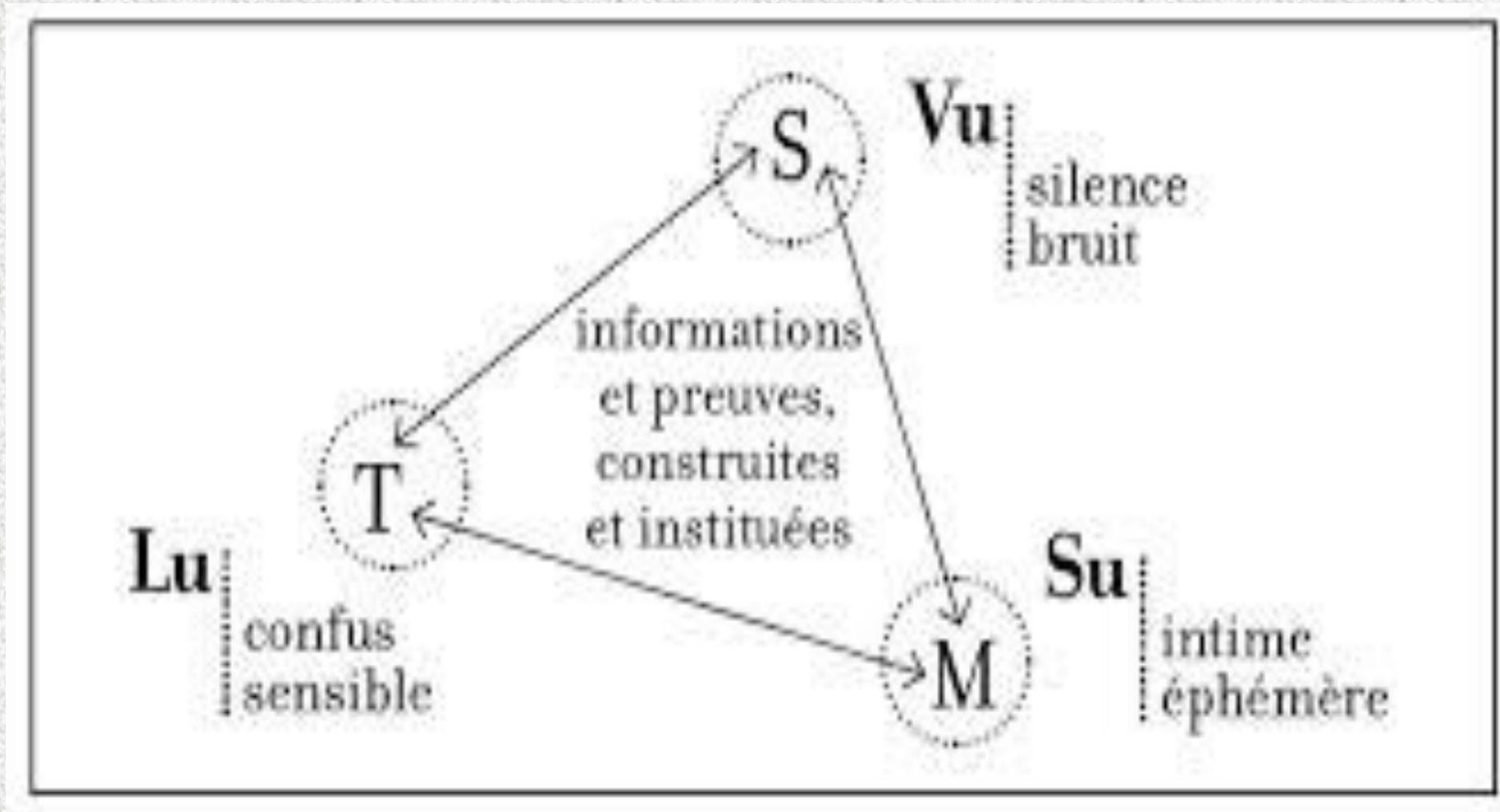
© LUCENT TECHNOLOGIES BELL LABS

30 APRIL 1916 · 24 FEBRUARY 2001

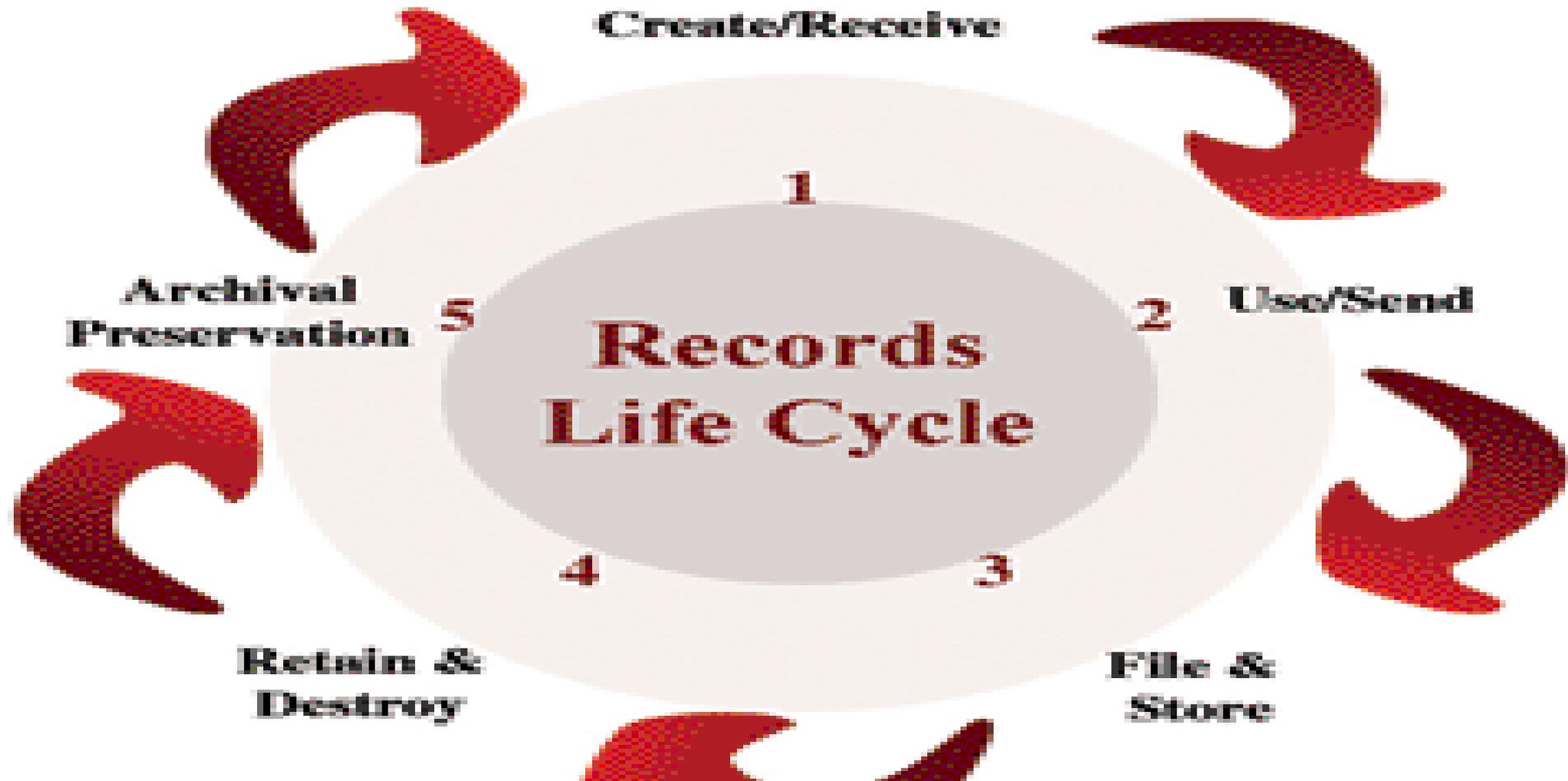
Application of Invariance Metaphor: Information as an Object

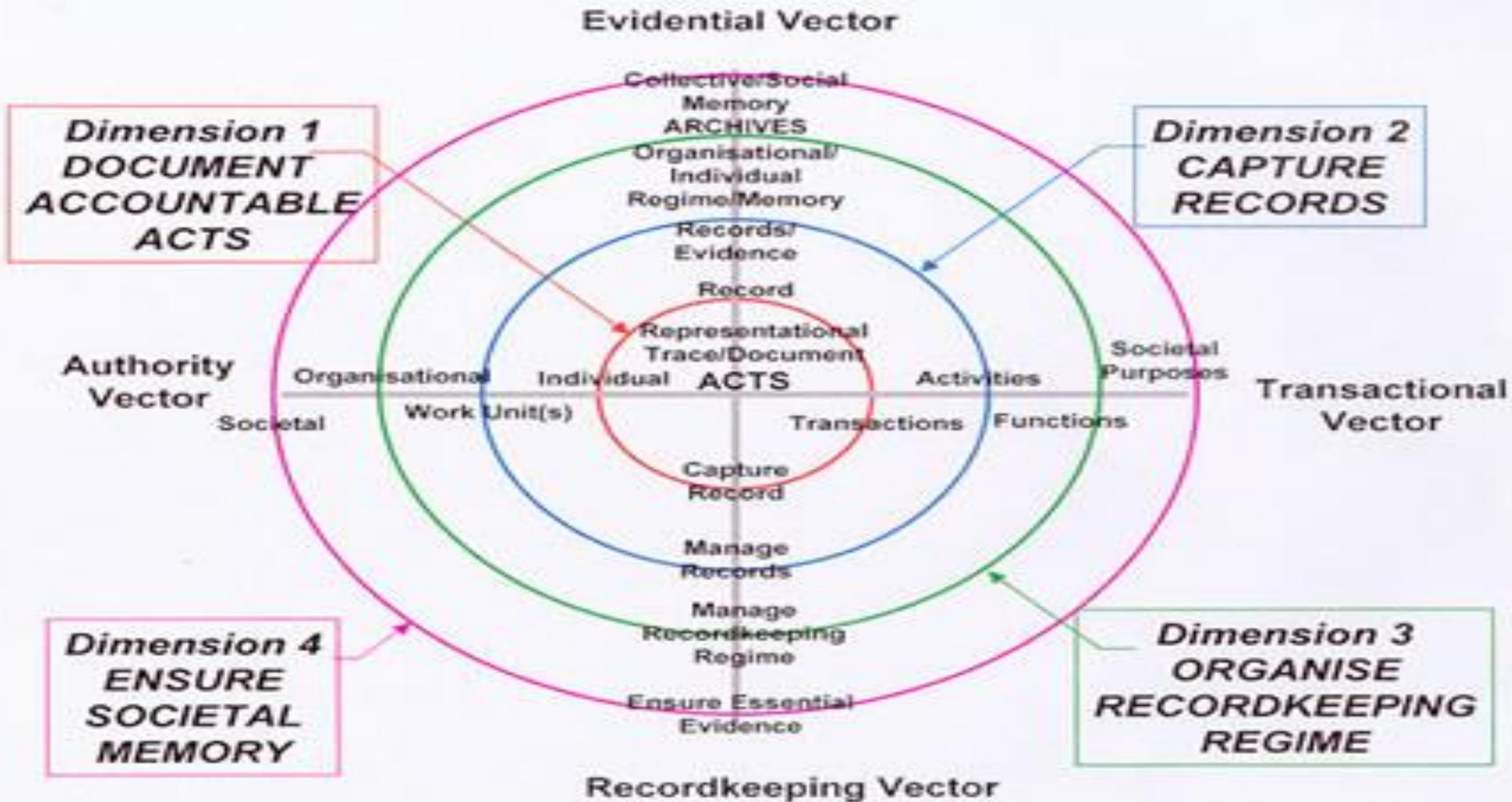
- Digitization: transition from analog to digital
- Early GIS systems
- Workflow management in business applications
- Information architecture of digital libraries and archives

The Document Metaphor: As Sign and Form (S) As Text and Content (T) and As Medium or Relationship (M) (Pédauque 2006)



The Archival Metaphor: Life-Cycle Concept





The Records Continuum model developed by Frank Upward, Senior Lecturer at Monash University.

Why metaphors are important?

- James J. Bono,(2010) “...metaphors operate as mediums of exchange between and among discourses and that, from the perspective of any one scientific discipline or field at any given moment, such metaphoric exchanges served to constitute an ecology of discourses or what might alternatively be called a synchronic discursive network”
- Jannis Kallinikos (Aaltonen, Marton, Kallinikos, 2010), provide a useful middleware to link information technology studies embedded in practices with social theories of a general nature (Castells, 2001). In his interpretation, middle range theory has to exist to allow focusing on particular practices and domain specifics without losing sight of generic processes that recur across context

Future Presents: what semantics actually do in information systems?

- Brier (2010): Semantics without context of the "conscious of living systems" lack a theory of meaning
- Cognitive science : The theory of cognitive embodiment postulates that this form of knowledge reception is not effective enough because spatial, bodily and social aspects are missing. (Dourish 2001; Anderson 2003; Cowart 2005; McCullough 2005)
- Every document always bears its three dimensions: it naturally has to be perceived as a sign and in its form, but its semantics are only understandable in the social perspective of pragmatics and the ever-changing practice of real life.

Why Interoperability matters?

- Along the lifecycle there may be several changes of custody
- Evidence about authenticity and provenance needs to be managed and interpreted by systems
- Recordkeeping and preservation systems are different
- Shared terminology, common standards and a consistent cross-domain framework of actions and procedures are needed

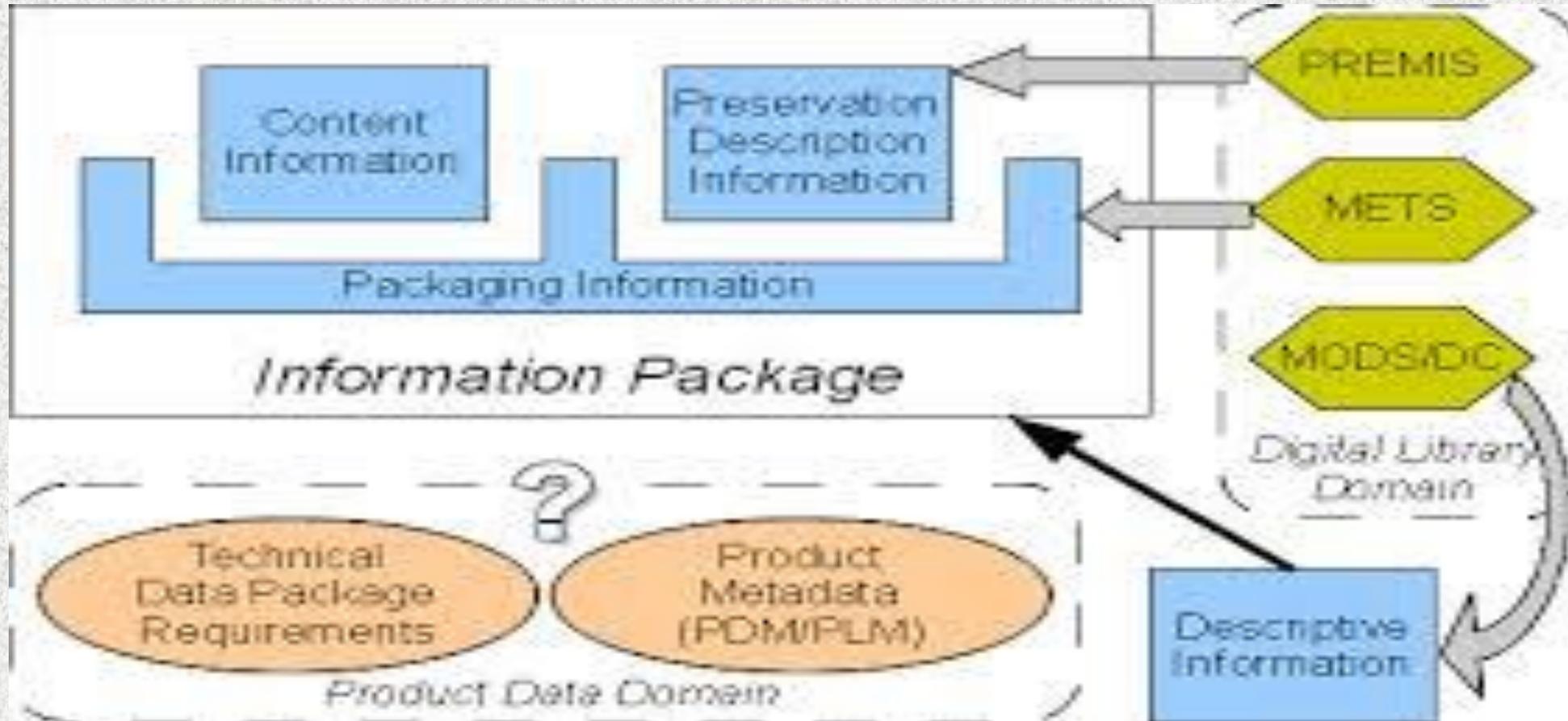
Conceptual Framework

- InterPARES 1 (1999-2001), InterPARES 2 (2002-2006): remains the main conceptual framework related to the authenticity
- CASPAR (2006-2009): provides the methodological approach for the implementation of a standardized set of tools able to integrate and document the main events and information related to the preservation function

OAIS 2012

- The Open Archival Information System Reference Model is a conceptual framework for a generic archival system which is committed to a dual role of preserving and providing access to information
- ISO standard in 2003, updated in 2012.
- Prescriptive Standard or Descriptive Practice, See ISO for standard
- CCSDS has changed the classification of Reference Models from Blue (Recommended Standard) to Magenta (Recommended Practice)

Challenges: How to preserve product metadata?



Notes:

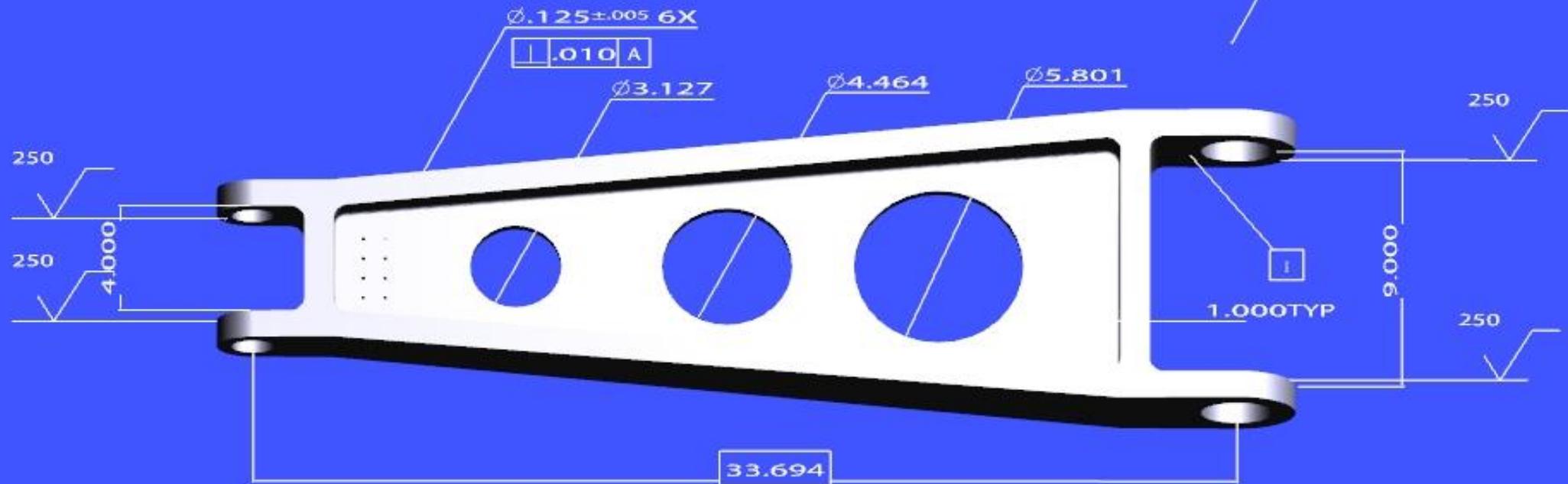
Dimensions and Tolerances per ASME Y14.5-1994 and ASME Y14.41-2003

Finish and Process codes in accordance with GAF14A

Unless annotated otherwise Tolerances shall be as follows:

- 1. A Non-biased surface profile tolerance band is $\pm .020$ mm shall be applied**
- 2. The default angular tolerance is ± 0 Degrees 30 Minutes**
- 3. Chamfer angle tolerance ± 3 Degrees**

Break sharp edges .030- .040 mm



Recordkeeping Standards

- ISO 15489-1: 2001 Information and documentation – Records management, it provides the list of actions and responsibilities relevant for the accurate and reliable records creation and keeping
- ISO 23081-1:2006 Information and documentation – Records management processes – Metadata for recordkeeping
- Recommendations and guidelines for the functional requirements for ERMS: 2001 (MoReq1), 2008 (MoReq2) and 2010 (Moreq2010)
- The recommendations for the transfer of digital objects: UN/CEFACT, Business Requirements Specification. Transfer of Digital Records, Version 1.0 (2008)

Metadata

- **Descriptive metadata:** information about the intellectual content of a digital object, which is used to aid identification and discovery of the object by the researcher. EAD, EAC, MARC21, MODS, EDM, LIDO
- **Structural metadata:** information about the relationships between digital objects, which can be very complex in a large hybrid personal archive. Structural metadata also supports the display and navigation of digital objects by users. METS
- **Administrative metadata:** information needed by the repository for the long-term management of a digital object, including information about an object's creation, technical information such as file formats, provenance information and information about intellectual property rights. PIDS, PREMIS

Administrative Metadata

- MINERVA “used for managing the digital object and providing more information about its creation and any constraints governing its use.”
- NISO: information about an object’s creation, technical information such as file formats, provenance information and information about.
- KEYWORDS: “creation”, “management”, and “access” or “use”.

METS

- **Technical Metadata <techmd>**: information regarding creation, format, and use characteristics of the files which comprise a digital object.
- **Intellectual Property Rights Metadata <rightsmd>**: information about copyright and licensing pertaining to a component of the METS object.
- **Source Metadata <sourcemd>**: information on the source format or media of a component of the METS object such as a digital content file. It is often used for discovery, data administration, or preservation of the digital object.
- **Digital Provenance Metadata <digiprovmmd>**: information on any preservation-related actions taken on the various files which comprise a digital object (e.g., those subsequent to the initial digitization of the files such as transformation or migrations, or, in the case of born digital materials, the files' creation).

Preservation Metadata

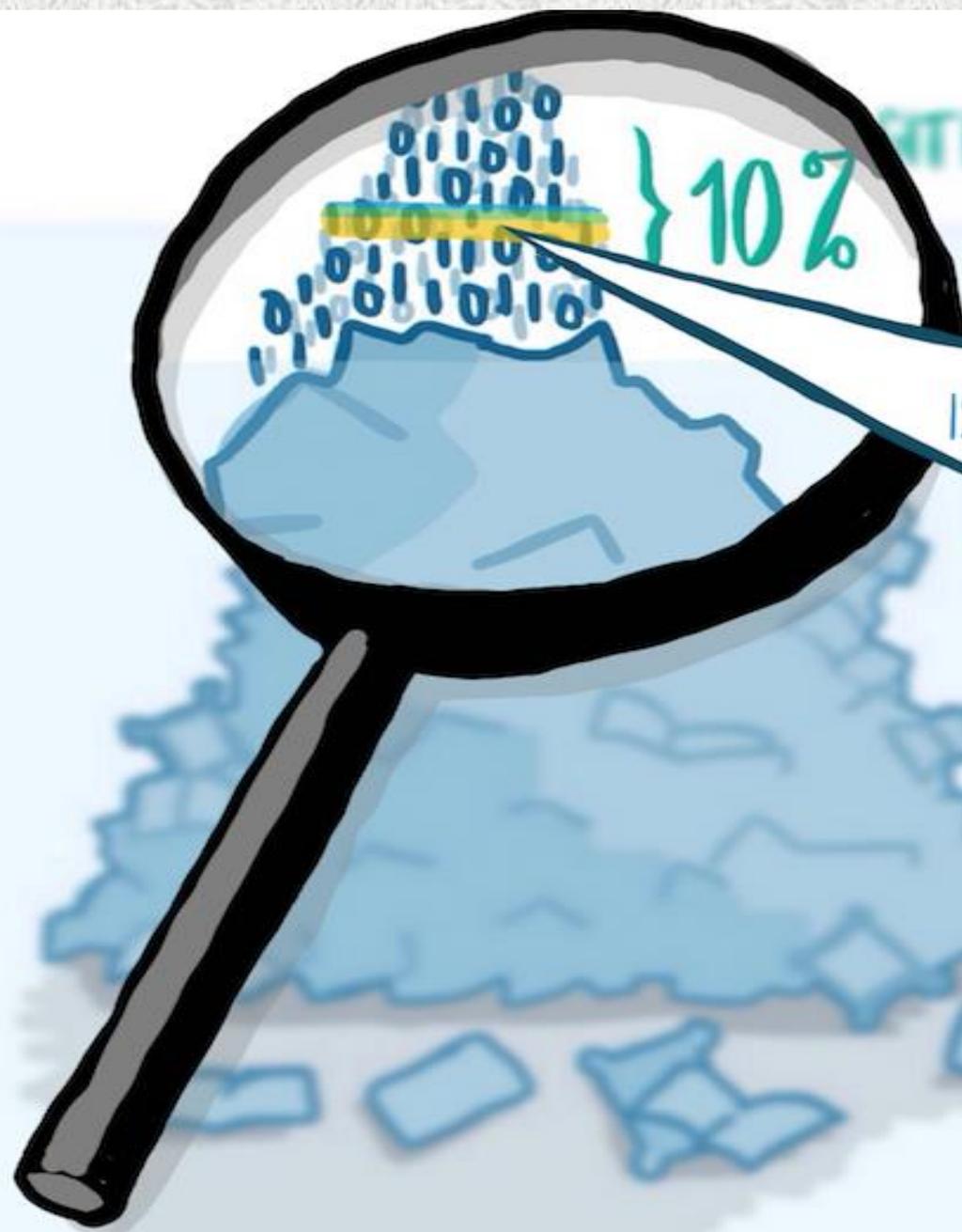
- **Provenance: Who has had custody/ownership of the digital object?**
- **Authenticity: Is the digital object what it purports to be?**
- **Preservation activity: What has been done to preserve the digital object?**
- **Technical environment: What is needed to render and use the digital object?**
- **Rights management: What intellectual property rights must be observed**

The Europeana Story

- *Initiated* by [Jacques Chirac](#), President of France, together with the premiers of Germany, Spain, Italy, Poland and Hungary to the President of the European Commission, [José Manuel Durão Barroso](#), in April 2005. The letter recommended the creation of a virtual European library, to make Europe's cultural heritage accessible for all.
- [European Commission's Information Society and Media Directorate](#), strategy, i2010: communication on digital libraries, which was published on 30 September 2005
- Europeana was called the European Digital Library Network (EDLnet) and was aimed at building a prototype of a cross-border, cross-domain, user-centred service. It was funded by the European Commission under its eContentplus programme,
- The prototype was launched on November 20, 2008
- In February 2009, the successor of EDLnet – Europeana version 1.0

Some Facts about Europeana: cultural heritage portal

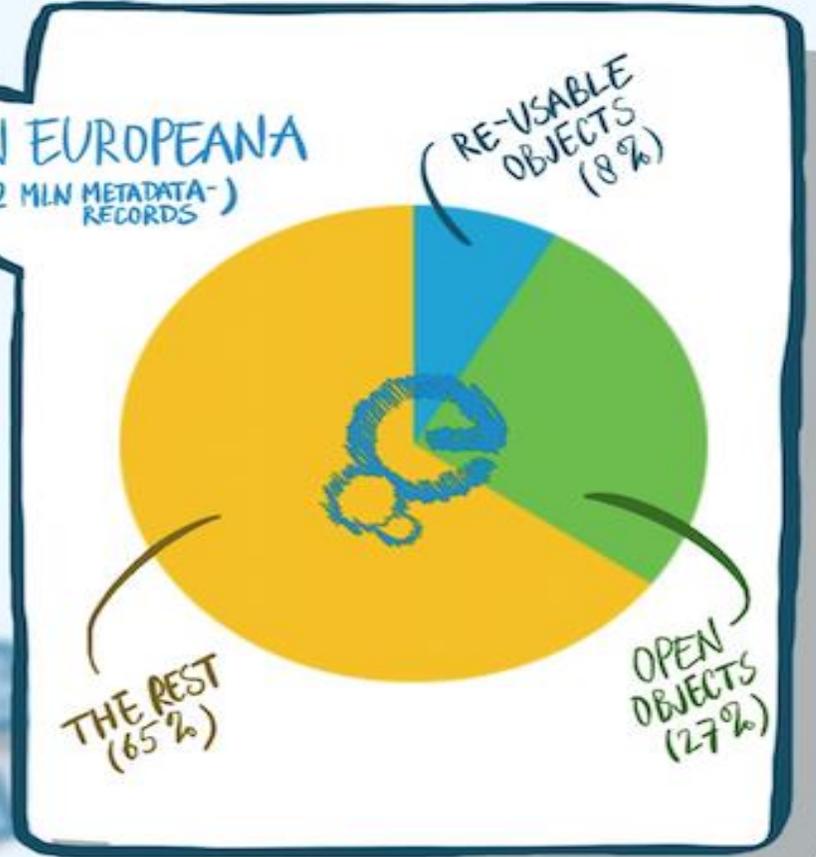
- Europeana is the result of the work of 2,500 institutions: to increase to 25,000
- 30 million digital object shown by or shown at institutional web sites
- 120 aggregators
- The 10% of the objects are photographs and moving image.
- However, only one third of that (34%) is actually available online, and only 2-3% is available for real creative re-use (for example in social media, via APIs, for mash-ups, etc.)
- Funded through Econtent and CIP PSP program



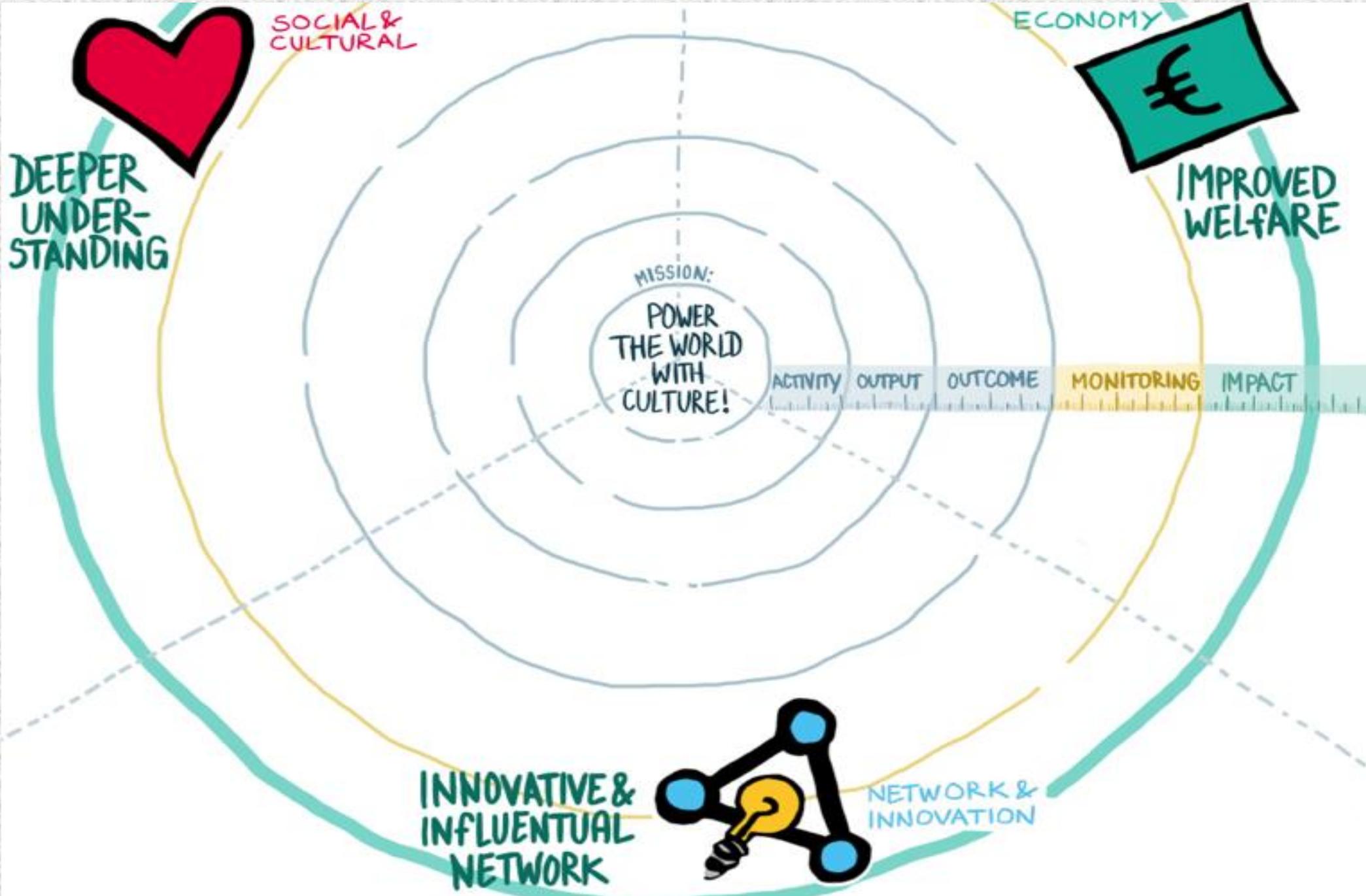
ITIZED (~500 MIL OBJECTS)

} 10%

12% IN EUROPEANA
(32 MILN METADATA-
RECORDS)



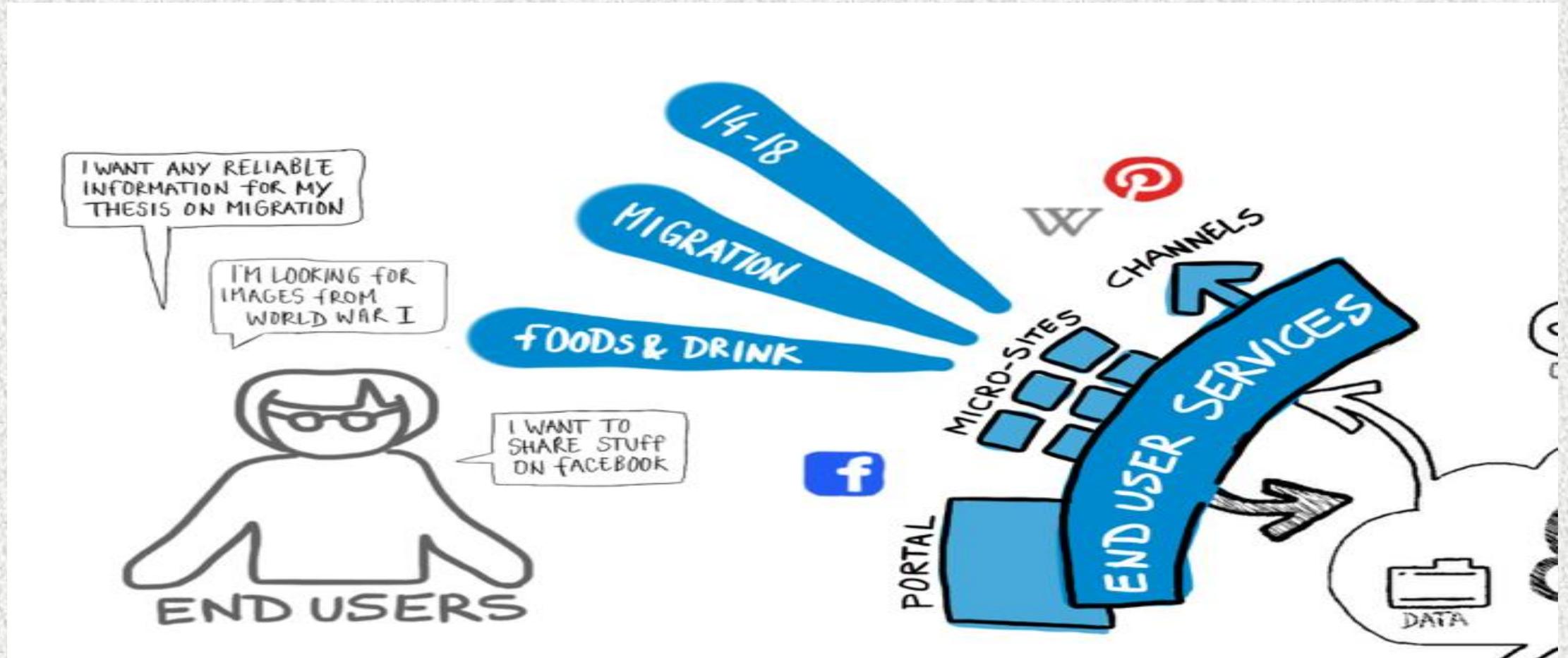
DATA = METADATA + OBJECTS



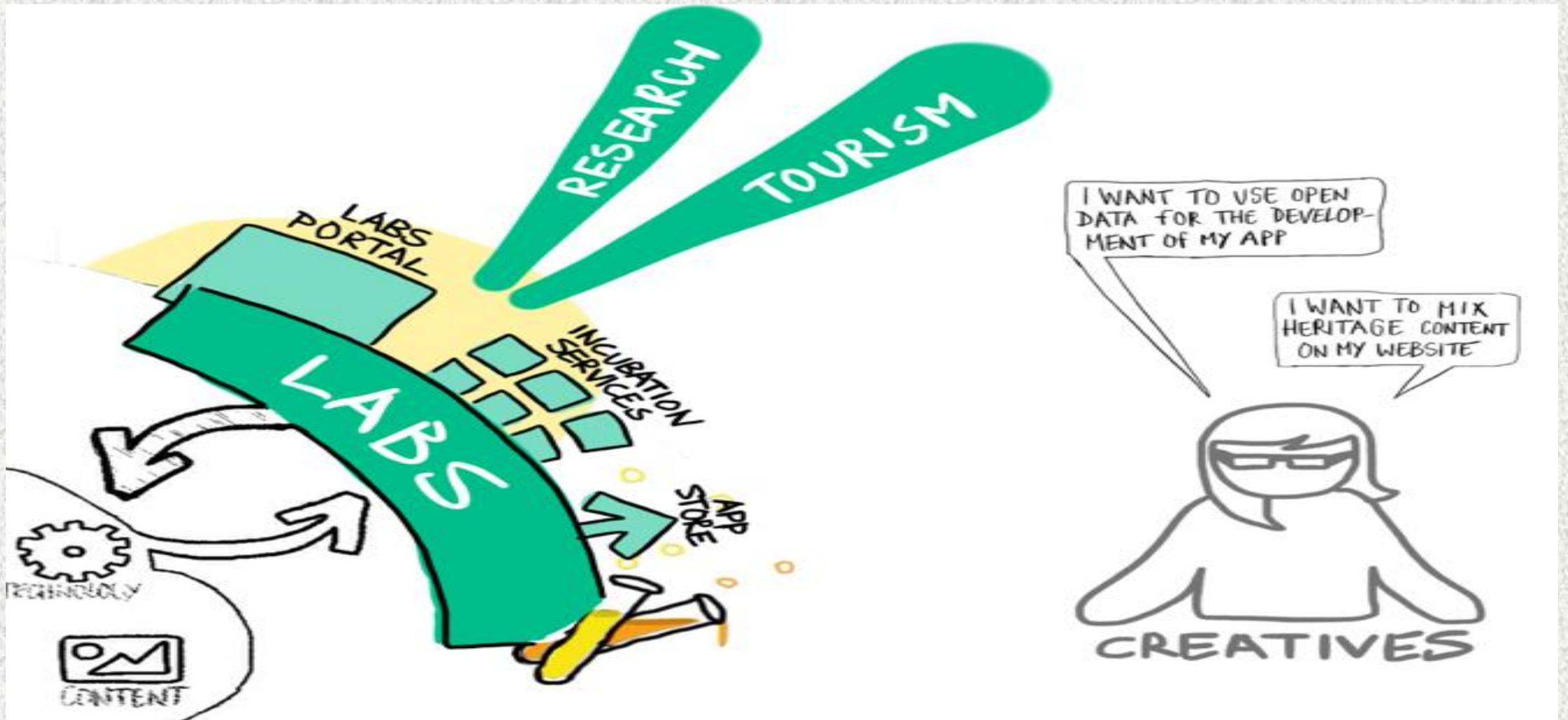
Europeana Vision: Professionals



Europeana vision: End user



Europeana Vision: Creatives, B2B

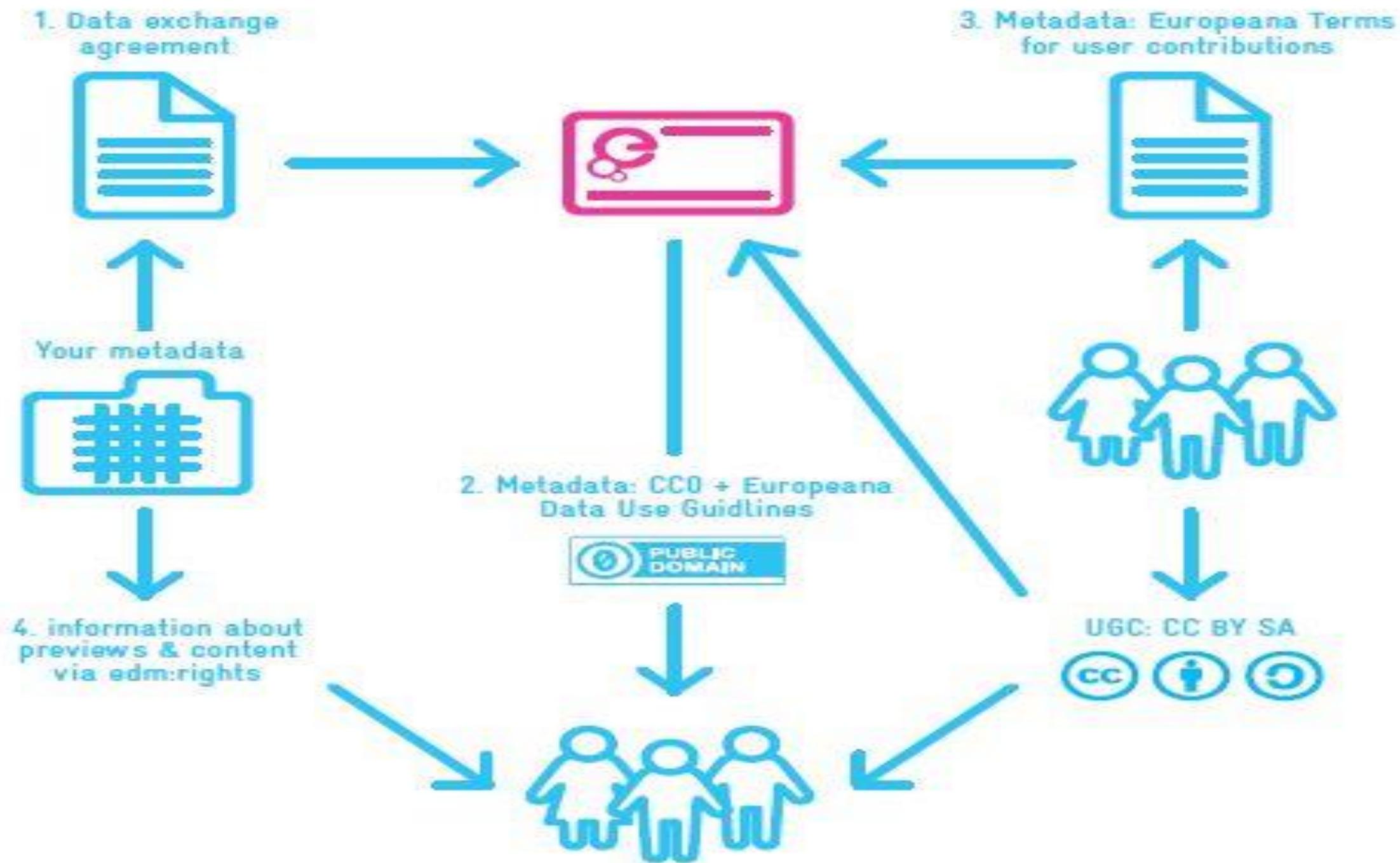


Propoposed Europeana Business Model

<p><i>Partners</i></p> <p>European Commission</p>	<p><i>Activities</i></p> <p>End user Creative Partner services</p>	<p><i>Value proposition</i></p> <p>Free & Open</p> <p>Innovation</p> <p>Easy & reliable</p> <p>R&D</p>	<p><i>Relation</i></p> <p>engaged</p> <p><i>Channel</i></p> <p>Platform</p>	<p><i>Clients</i></p> <p>End Users</p> <p>Cultural Inst.</p> <p>Member States</p> <p>Creatives</p> <p>EC</p>
<p><i>Costs</i></p> <p>10 million Euro/year EF + Network</p>		<p><i>Benefits</i></p> <p>1 attention</p> <p>2 service fee</p> <p>3 project funding</p> <p>4 revenue</p>		

Europeana Licensing Framework

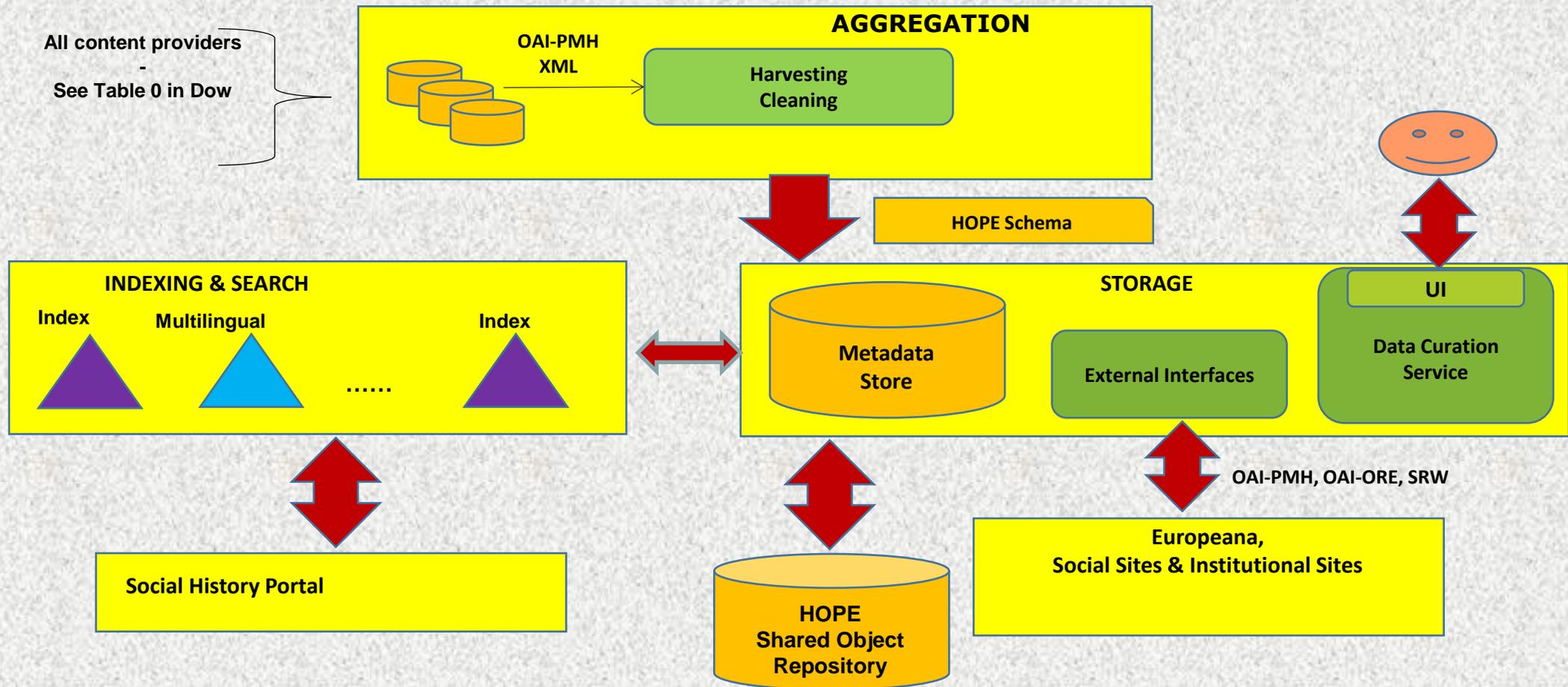
- The Europeana Licensing Framework
<http://pro.europeana.eu/documents/858566/7f14c82a-f76c-4f4f-b8a7-600d2168a73d>
- Guidelines for the Rights in objects submitted to Europeana
http://pro.europeana.eu/documents/900548/1037382/Europeana_rights_201202.pdf
- Europeana Data Exchange Agreement
<http://pro.europeana.eu/documents/900548/8a403108-7050-407e-bd00-141c20082afd>



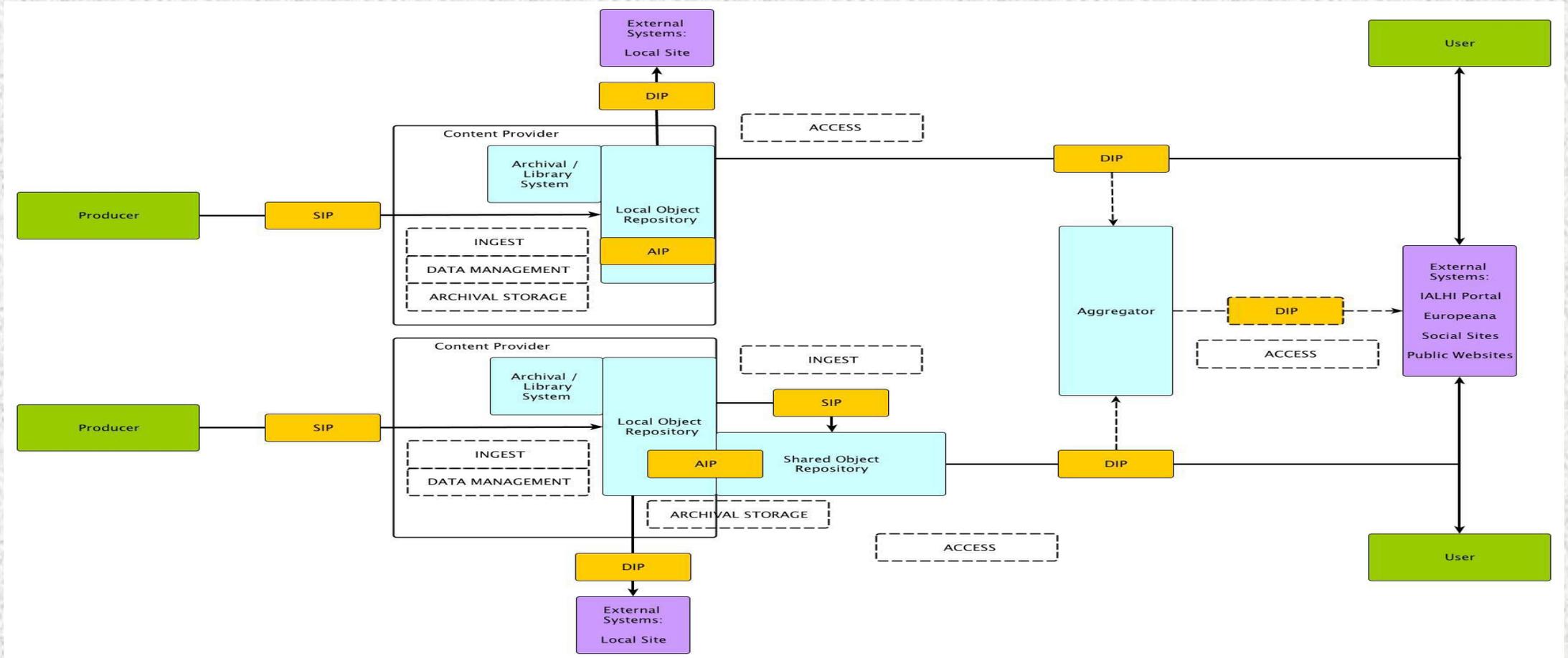
HOPE Project (2010-2013): Short-Term or Long –Term Access?

- 12 social history institutions: small and medium size
- 1 M digital objects and 4 M metadata records
- HOPE Best Practices
- Shared Object Repository (SOR)
- Sustainable network and infrastructure

HOPE Architecture



HOPE Implemented Architecture based on OAIS



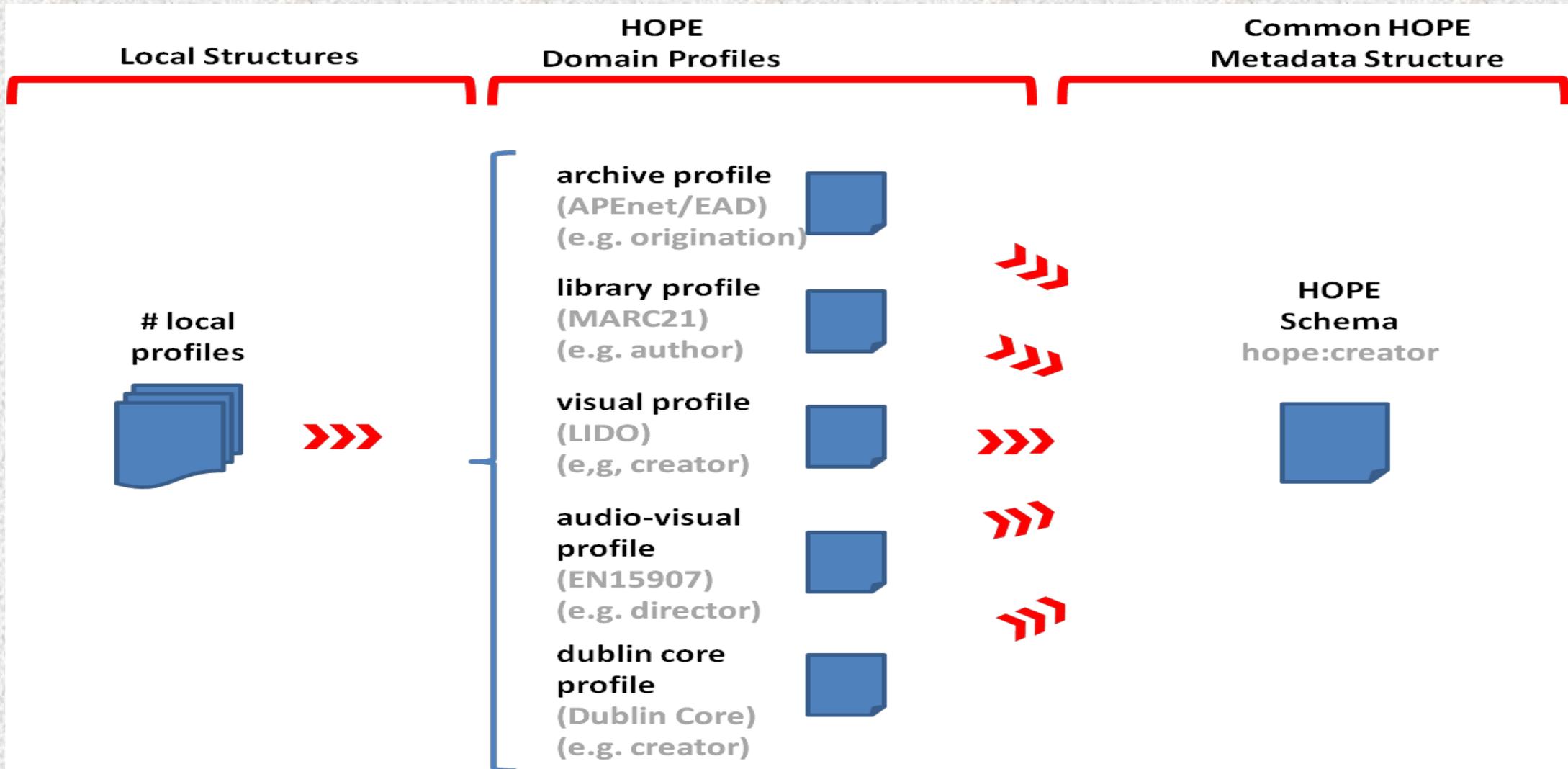
HOPE Best Practices

- 2010 June-September: Survey, Assessment
- 2010 September-November: interviews, milestone report
- 2011 January: [architecture, high level design](#)
- 2012 június: TDR [best practice report](#) and case studies, [data model](#)
- 2012 november: [public HOPE wiki](#) to content providers and [Budapest workshop](#)
- 2013 April: project closing

D2.6 Best Practice for Trusted Digital Object Repositories

EXECUTIVE SUMMARY	5
INTRODUCTION	7
1. Framework(s) of the Social History Institution	11
1.1. Organizational Framework: Governance and Viability	11
<i>Case Study: Strategic Planning at the Internationaal Instituut voor Sociale Geschiedenis (IISG)</i>	18
1.2. Legal Framework: Due Diligence	18
<i>Case Study: Compliance at the Schweizerische Sozialarchiv (SSA)</i>	21
1.3. Technical Framework: Systems and Practices	22
<i>Case Study: Adapting Practice at the Archiv der sozialen Demokratie (AdsD) of the Friedrich-Ebert-Stiftung (FES)</i>	27
<i>Case Study: Digital Object Management Reborn at the Fundação Mário Soares (FMS)</i>	28
1.4. Framework(s) of the Social History Institution: References	29
2. The HOPE Federated Repositories	31
2.1. Content Profile	32
2.2. Designated Community	34
2.3. The Federated Model	35
2.4. The Federated Model: HOPE Compliant Local Object Repositories	38
2.5. The Federated Model: Persistent Identification and the HOPE PID Service	39
2.6. The Federated Model: Transforming and Disseminating Data through the HOPE Aggregator	41
2.7. The Federated Model: Secure Storage in the HOPE Shared Object Repository	42
2.8. Sustainability of the HOPE Federated Repositories	46
2.9. The HOPE Federated Repositories: References	46
3. Managing Objects Through Administrative Metadata	48
3.1. Administrative Metadata	48
3.2. Administrative Metadata in HOPE: Current Status	51
3.3. Administrative Metadata in HOPE: Recommendations	54
3.4. Administrative Metadata: References	56
3.5. PREMIS	56
3.5.1. PREMIS: Conformance	57
3.5.2. PREMIS: References	57
3.6. Persistent Identifiers (PIDs)	58
3.6.1. PIDs: Benefits of a PID System	58
3.6.2. PIDs: Characteristics of a PID System	59
3.6.3. PIDs: Selecting a System	60
3.6.4. PIDs: Local PID Policies	64
3.6.5. PIDs: PID Workflows and Maintenance	68
3.6.6. PIDs in HOPE: Recommendations	69
<i>Case Study: Amsab-Instituut voor Sociale Geschiedenis (Amsab-ISG) Implements PIDs</i>	70
3.6.7. PIDs: References	71
3.7. File Naming	72
3.7.1. File Naming: Characteristics	73
3.7.2. File Naming: Elements	73
3.7.3. File Naming: Directory Conventions	76
3.7.4. File Naming: Workflow	76
3.7.5. File Naming in HOPE: Recommendations	76
3.7.6. File Naming: References	77
3.8. Technical Metadata	78
3.8.1. Technical Metadata: Selecting Standards	78
3.8.2. Technical Metadata in PREMIS	78
3.8.3. Technical Metadata: Collection and Storage Workflows	80
3.8.4. Technical Metadata in HOPE: Recommendations	81
3.8.5. Technical Metadata: References	82
3.9. Fixity	83
3.9.1. Fixity: Checksums	83
3.9.2. Fixity: Message Digest Algorithms	84
3.9.3. Fixity: Digital Signatures	84
3.9.4. Fixity: Workflows	85
3.9.5. Fixity in HOPE: Recommendations	85
3.9.6. Fixity: References	86
<i>Case Study: Open Society Archives (OSA) Manages Administrative Metadata</i>	87
CONCLUSION	89

Interoperability: D 2.2 Data Model



Hope requirements: policies and infrastructure

- Each CP defines its own digital object and digital collection, mandatory to produce access copy of the! PDF, MPEG, JPEG,
- Digital collections should be online either *shown at* or *shown by*
- Harvesting protocol should be implemented, OAI-PMH
- Persistent Identifier: both metadata and object should have, possibly digital collection by using HOPE PID System.
- Master copy of the digital object could be deposited with SOR
- Only a limited set of file formats could be used, mostly based on open standards
- Access is regulated by the local repository (LOR), but no use of technical barriers: view only, watermark
- LOR collects technical and administrative metadata
- SOR integrated functions: shared authentication, ingest, API, checksum, Dashboard, Object Depot, Conversion Platform

Sustainability?

- Technical Service. Who takes care of the aggregation? New collections, new CPs, changes in descriptions?
- Maintenance of the infrastructure, cost model?
- Why has the capacity to train new comers about best practices?
- Governance structure: IALHI network or externals as well?
- Research and Development

Understanding archives as a process rather than object

- Why and how is it that the ephemeral endures?
- And what does the constant repetition and regeneration of information effect?
- But what about the implications that new digital media have for our very understanding of the "past"?
- What does it mean to historicize content that is "constantly revised, deleted and 'saved over'" (Ankerson 2011)?

Break out Sessions

- Group 1

You represent the European Commission and would like to support seamless, short term access to archival collections which EU information policies should be reviewed, reinforced perhaps to be changed in order to reach this goal?

- Group 2

Please list the main audit toolkits/certificates for trusted digital repositories, which one would you recommend to a small cultural heritage organization? Why?

- Group 3

Discuss the link between Open Government and trustworthy public records, paying particular attention to digital records. 2. Identify key concepts and phrases that could be used to raise awareness of and advocate digital records management and preservation.