Research in Digital Libraries and DL.org
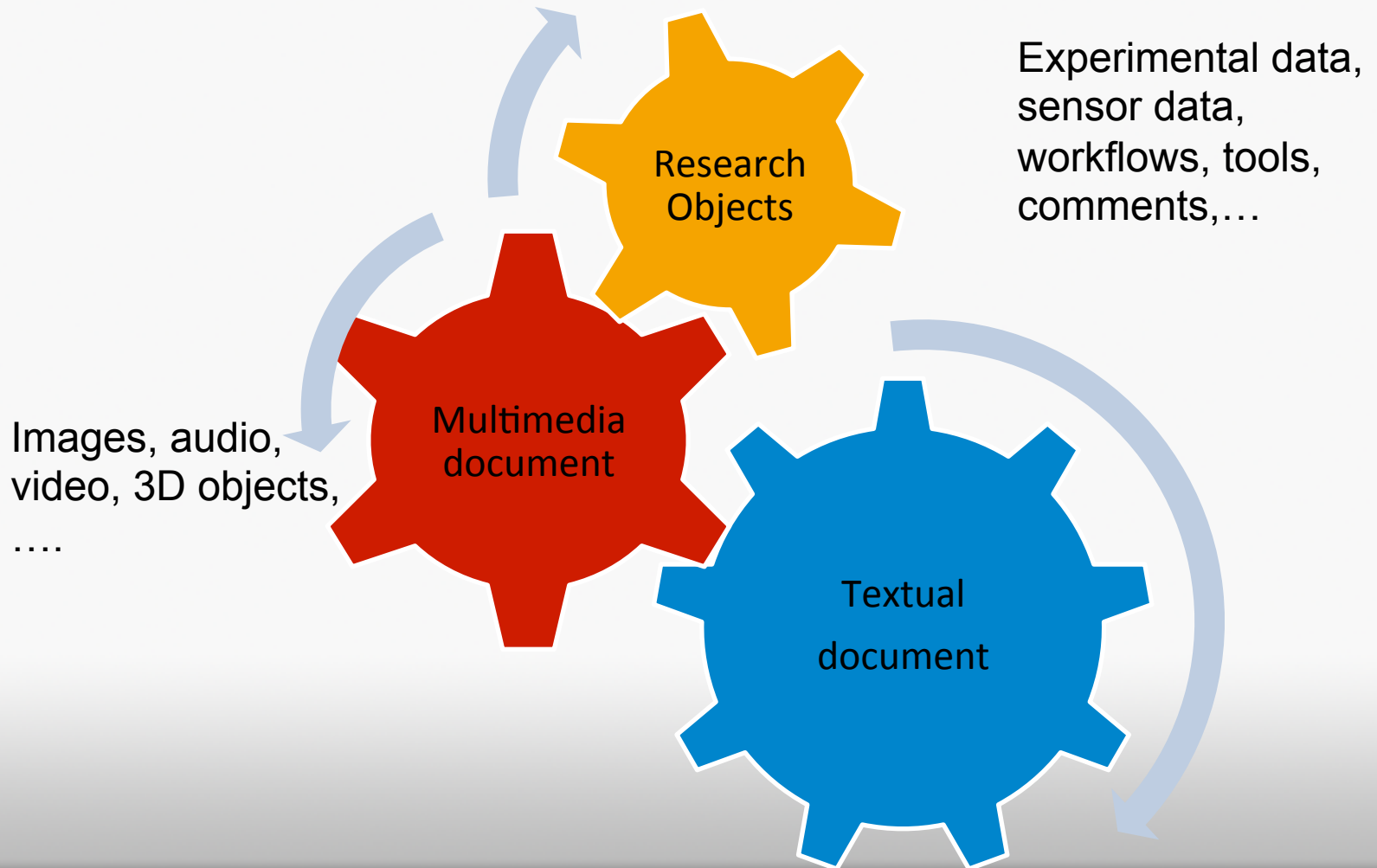
Donatella Castelli (CNR-ISTI)

Outline

- DL innovation paths
  - Scientific data & research objects
  - Content sharing
  - DL infrastructures & Virtual Digital Libraries
- DL.org project

DL Universe

- New needs, new application areas, new tecnological opportunities emerge continuosly

- Old systems and new systems cohabit together

- Several co-existing conceptions of what a Digital Library is

- Several different actor profiles are needed

Changing in the DL Universe: DL content

Research Objects

Experimental data, sensor data, workflows, tools, comments,…

Multimedia document

Images, audio, video, 3D objects, ….

Textual document

Data Libraries

- ## Data Libraries

  - *Inter-University Consortium for Political and Social Research* (ICPSR), data publisher for research in Social Sciences
  - *Pangaea*: data publisher for Earth & Environmental Science

- ## Motivations for data publishing

  - Encourage data sharing
  - Data reviewed separately
  - Data normaly unpublished can be made available (e.g. negative results)
  - Better credit to scientists

Scientific Publishing

- Scientific papers linking deposited data
  - *GenBank, EMBL-Bank, DDBJ*: DNA data banks, designated archives for journals like **"*Genes, Genomes, Genetics*", "*Nucleic Acid Research*"** (Note: data deposition is mandatory before paper submission)
  - GenBank e PubMedCentral can be uniformly searched

- Data Journals
  - *Earth System Science Data (ESSD):* articles in the data section may pertain to the planning, instrumentation and execution of experiments or collection of data
    (Note: Datasets archived in *Pangaea*)

New publishing

"When the scientific world embraces data publishing, scientific journals will become less important and should focus on papers that give added value to the data instead of just describing it. For sure supplementary information under the form of figures or text will disappear. I believe that four types of papers will dominate in this version of the future: highlights, methods, reviews and opinions and discussions."

"Data Publishing and Scientific journal: the Future of the Scientific Paper in a World of Shared Data", Erik De Schutter, Neuroinform (2010)

Evolution in scientific publishing

"Data Publishing and Scientific journal: the Future of the Scientific Paper in a World of Shared Data", Erik De Schutter, Neuroinform (2010)

"Four types of papers will become dominant in data publishing"

- Highlights

  Short summaries highlighting exciting new findings & links to the data in databases where the specialists will go for deeper understanding

- Methods

  Description of the methodology and processes applied

- Reviews

  Dataset reviews with references to databases

- Opinions & discussions

  Scientific discussion in social media

Research objects

- Objects with an identity which bundle together essential information related to experiments and investigations, like, for example:

    – **Questions**: context including a description of the problem, a digest of preceding research, and optionally a hypothesis;

    – **Organisational context**: Information about ethical approval, governance policies, the investigators involved in the experiment;

    – **Study Design**: scientific workflows, web services, scripts;

    – **Data**: sources and collections of cleaned or raw data in various formats ranging from signal transducers to spreadsheets and databases;

    – **Result**s: spreadsheets, SBRML Methods;

    – **Answers**: Publications, papers, reports, slide-decks, DOIs, PUBMED ids

    + additional information to aggregate the component resources

Research objects: implications

Changes in several directions:

- Technological solutions
- Policies & procedures
- Publication workflows
- Organization roles
- Business models
- Actor roles

Content sharing

- Several providers agree to open their content and make it available under certain policies
- Content is federeated/aggregated, curated and made available to third party users
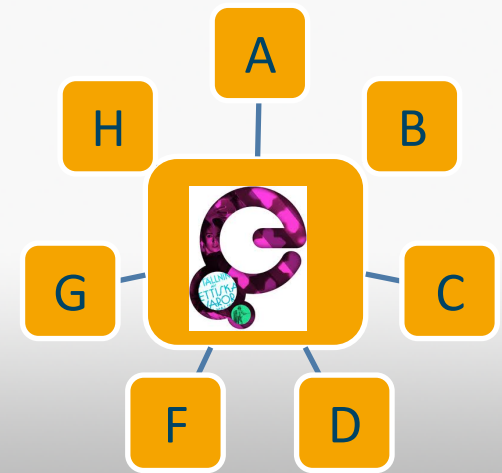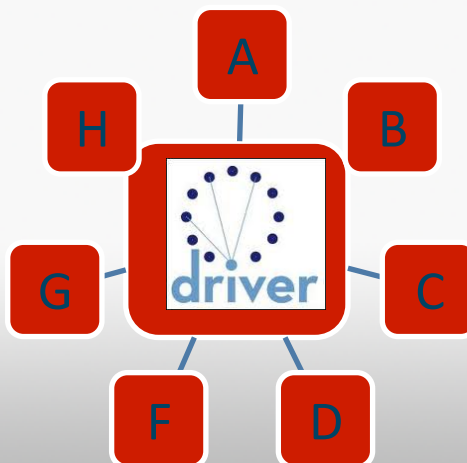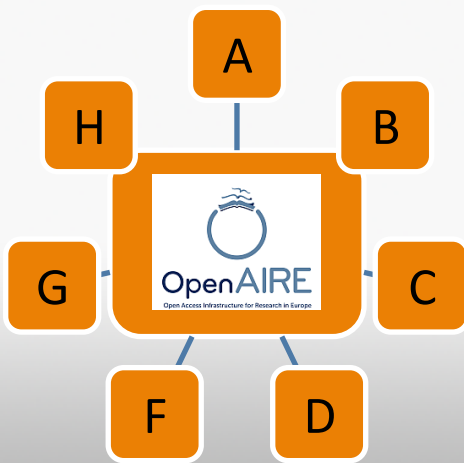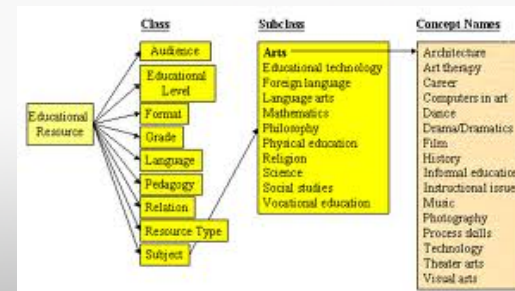
Content sharing & reuse

- Re-use requires that contextual information is represented explicitly

- In site solutions are based on several implicit contextual assumptions
  - Controlled vocabularies
 - Policies
  - Workflows
  - Actors
  - Attribution
  - ......

Content sharing: implications

- Changes in:
- Technological solutions

  (e.g contextual representation, provenance, interoperability)
- Activities:
  - Policy and procedures definition
  - Make explit what is implicit
  - Established policies check
  - Repository registration
  - Repository validation
- Actors roles

DL infrastructures & VDLs

- Enabling platform, operated by a responsible organisation, offering a dynamic set of resources and services in a systematic way to a wide community of actors for creating and maintaining DLs
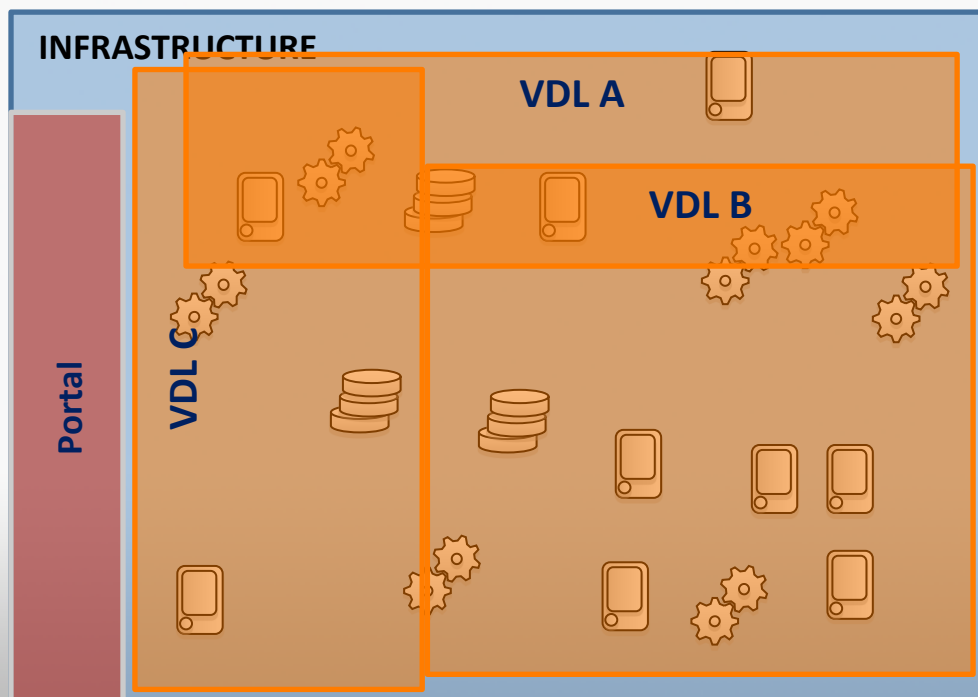
- The infrastructure acts as a broker between resource providers and consumers
- Virtual DLs (VDLs) can be be dynamically created on-demand by providing an appropriate description

DL infrastrustructure & VDLs: implications

Changes in:

- Technological solutions
- Organisational models
- Sustainability models
- Actors roles

(Infrastructure manager, infrastructure administrator, resource provider, VDL designer, VDL users, …)

Emerging needs

Large variety

- How to select the most appropriate system?
- How to compare to systems?

Complex systems

- Which are the reference methods/ principles?
- How can components be integrated and reused?
- How can components interoperate?

Multiple actor roles

- Which roles are need? Librarians, administrators, managers, ICT experts, stakeholders?
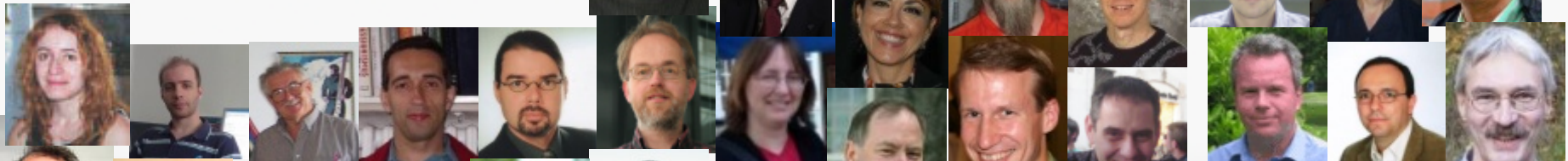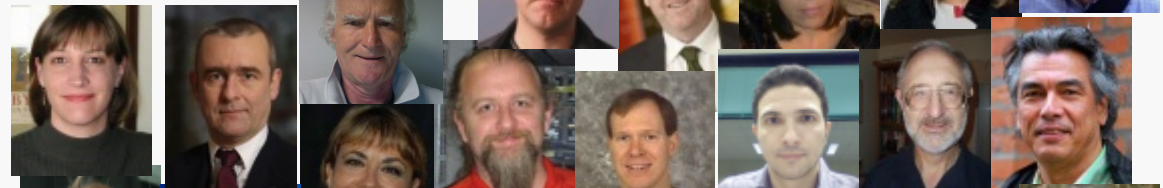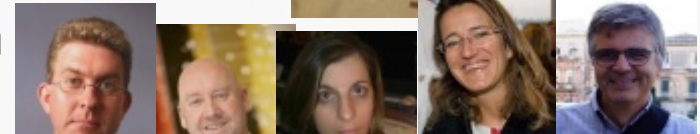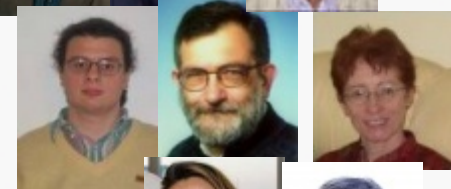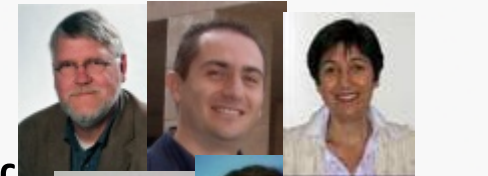
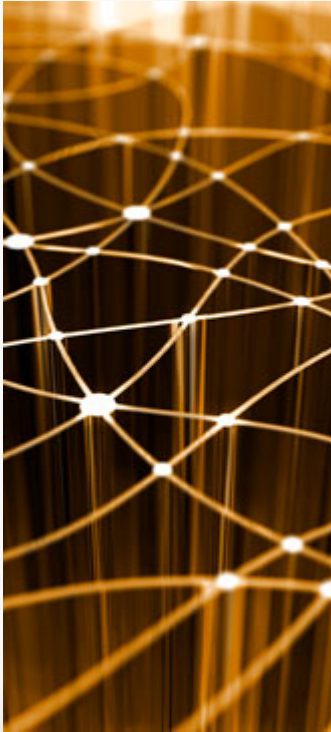**How much of this need is already covered in DL Education?**

DL.org

Coordination Action on Digital Library Interoperability, Best Practices and Modelling Foundations

Collaborative effort

- 45 participants
  - Leading technical **experts** representatives of **key on-going projects** and **initiatives**

  - **Researchers** in the specific domain
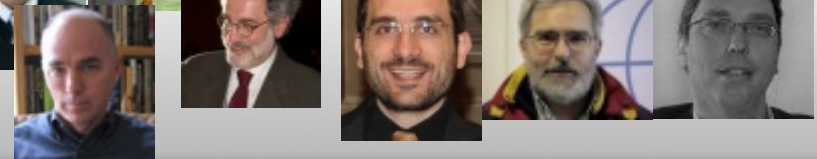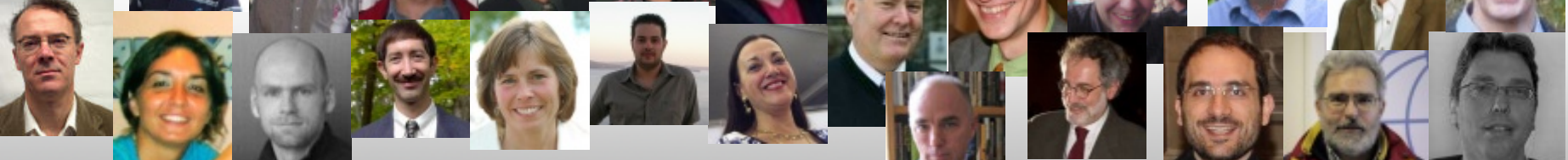
DL.org products

- *"Digital Library Manifesto"*

- *"Reference Model - In a Nutshell"*

- *"Digital Library Conformance Checklist"*

- *"Digital Library Technology & Methodology Cookbook"*
- *"DL.org Digital Library Manifesto:
  A Primer for Library and Information Professionals"*

DL.org products

More information at:

http://www.dlorg.eu/