

46th Annual International Conference on Graphic Arts,  
Media Technology, Management and Education

# Metadata for characterization of digital images

**Chrysoula Gatsou , Anastasios Politis, Philippa Athimaritou**

**Technological Educational Institute of Athens,  
Faculty of Fine Arts and Design**

# Overview

- What is metadata?
- Why use metadata?
- Dublin Core
- Metadata types
- Metadata standards
- Framework of metadata for digital images
- Conclusions

# What is metadata?

- Metadata is literally “ about something that is given” ( meta- μετα is a Greek prefix meaning "after" or “about” and “data” is the plural of the Latin word datum “something given”) or more simply: **about data**
- Data associated with an information object for purposes of description, administration, legal requirements, technical functionality, use and usage, and preservation.
- Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.

- *Getty Research Institute*

# Why metadata?

- Hardware and software come and go – system-independent **metadata can be used, reused, migrated, and disseminated in any number of ways**, even in ways that we cannot predict at this moment.
- Digitization does not mean equal access. The simple act of creating digital copies of collection materials does not make those materials findable, understandable, or utilizable to our ever-expanding audience of users.
- But **digitization combined with the creation of carefully crafted metadata** can significantly enhance end-user access and interoperability.

*[introduction to Metadata, Online Edition, Version 3.0 [http://www.getty.edu/research/conducting\\_research/standards/intrometadata](http://www.getty.edu/research/conducting_research/standards/intrometadata)]*

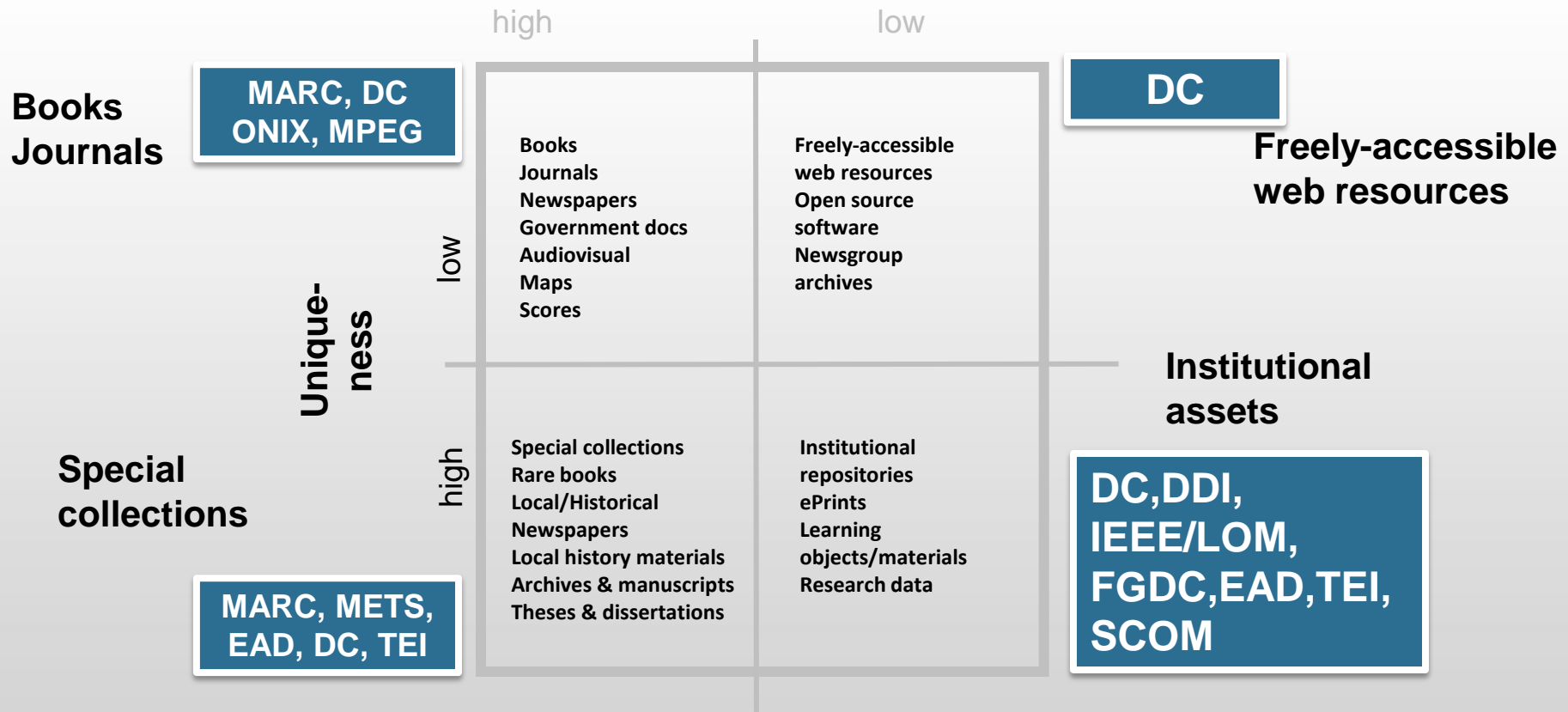
# Why metadata?

- Metadata provides an essential link between the information creator and the information user
- We can ensure that this objective is met by using metadata in **accordance with international standards**

# Metadata schemas

- Dublin Core (Cross disciplinary)
- DDI (Social sciences)
- EAD (Archives)
- FGDC (Geographic)
- IMS (Education)
- LOM (Learning Object Metadata)
- MARC (Libraries)
- METS (Structural metadata)
- ONIX (Publishers & booksellers)
- TEI (Text encoding-Humanities)
- VRA (Visual resources)
- Etc. ...

# Metadata Standards in a Resource Grid



# Dublin Core

- **Simplicity**
- **Semantic Interoperability**
- **International Consensus**
- **Extensibility**
- **Metadata Modularity on the Web**

2<sup>nd</sup> W3C Conference Chicago (October 1994) Conversations at this conference led to the first meeting at OCLC in Dublin Ohio, hence its name Combination of IT and Librarians



# Dublin Core

The Dublin Core is a simple and effective way of describing a variety of network resources. It is a vocabulary of fifteen properties for use in resource description.

Content	Instantiation	Intellectual Property
Coverage	Date	Contributor
Description	Format	Creator
Type	Identifier	Publisher
Relation	Language	Rights
Source		
Subject		
Title		

Dublin Core is a de-facto standard across many other systems and standards

# Basic Types of Metadata

- Descriptive
  - Most common
    - (Title, author, summary, topic, etc.)
- Structural
  - Technical, to recreate relationships between parts of compound object
- Administrative
  - Provenance
  - Rights
- All three should be included on-line for optimal usability

# Image management system

- No single accepted standards for each type of metadata

## Descriptive metadata

- MARC, DC, MOA2, EAD, VRA, Open Archives Initiative

## Structural metadata

- LC RFP's, MOA2, DOIs

## Administrative metadata

- DIG 35, NISO draft standard, MOA2,
- in process preservation standards such as CEDARS

# Technical metadata

- Technical metadata is useful for software applications that are dealing with digital assets and should not be maintained manually.
- The available technical metadata of an asset depends largely on the file type of the asset. Examples of technical metadata are as follows:
  - the size of a file
  - the dimensions (height and width) of an image
  - the resolution (level of detail) of an image

# Graphic File Formats

... - PCX - PBM - TGA - TIFF - GIF  
- JPEG - PSD - DXF - CGM - PNG -  
SVG - RAW - WPG - FITS - BMP -  
PCD - RAS - TGA - BPS - EPS -  
PDF - PCT - WBM - FITS - XBM -  
VFF - RIB - PCX - DMP - AVS - IMG  
- ICO - JFIF - IFF - WMF - ...

# Markup languages

- Markup languages:
  - Address the structure of a document
  - Convey instructions to software that will process text to:
    - Index the text for searching
    - Render the text (e.g., for screen display or print)
    - Transform the text (e.g., for a voice synthesizer) for some output device(s)
  - The markup is generally invisible to end-users
  
- Extensible Markup Language (XML):
  - XML is a metalanguage
    - Agencies define their own XML to suit their task
      - By creating Document Type Definitions (DTDs) or XML schema

# Metadata formats within image files

- There are three metadata formats widely used in the industry:
  - EXIF
  - IPTC-IIM
  - XMP

## Metadata formats within image files

- **EXIF** means Exchangeable image file format and is the most popular metadata format used in digital Image. It provides a way of embedding a fixed vocabulary of metadata properties in a number of file formats such as JPEG, TIFF, RIFF, WAV.
- A major limitation of EXIF is that it is not supported by other popular image file formats such as BMP, GIF, or PNG.
- **IPTC** stands for International Press Telecommunication Council.
- **IPTC-IIM** , the International Press Telecommunications Council-Information Interchange Model
- **XMP** is an extensible metadata platform that is built on the standards of the Semantic Web.



# Connecting Metadata with Image Files

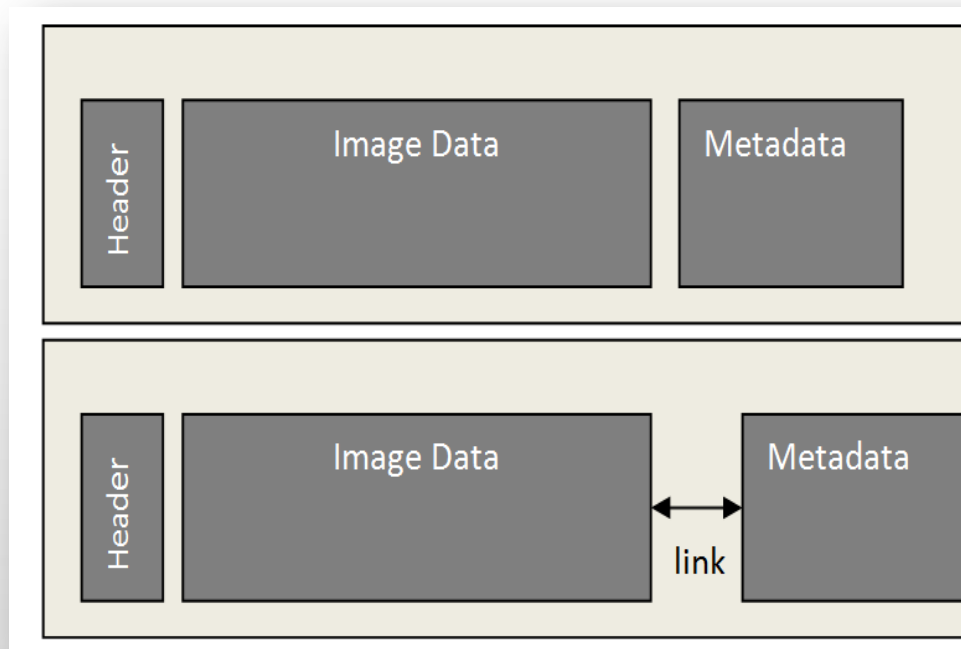
There are two ways to connect metadata with an image file

## 1. Embedding the data in the file

The linked file is similar in structure, except that metadata is externally attached to the file.

## 2. Linking the data to the file

The linked file is similar in structure, except that metadata is externally attached to the file.



*Image connection with digital files ( Adapted from DIG 35, Metadata for digital images )*

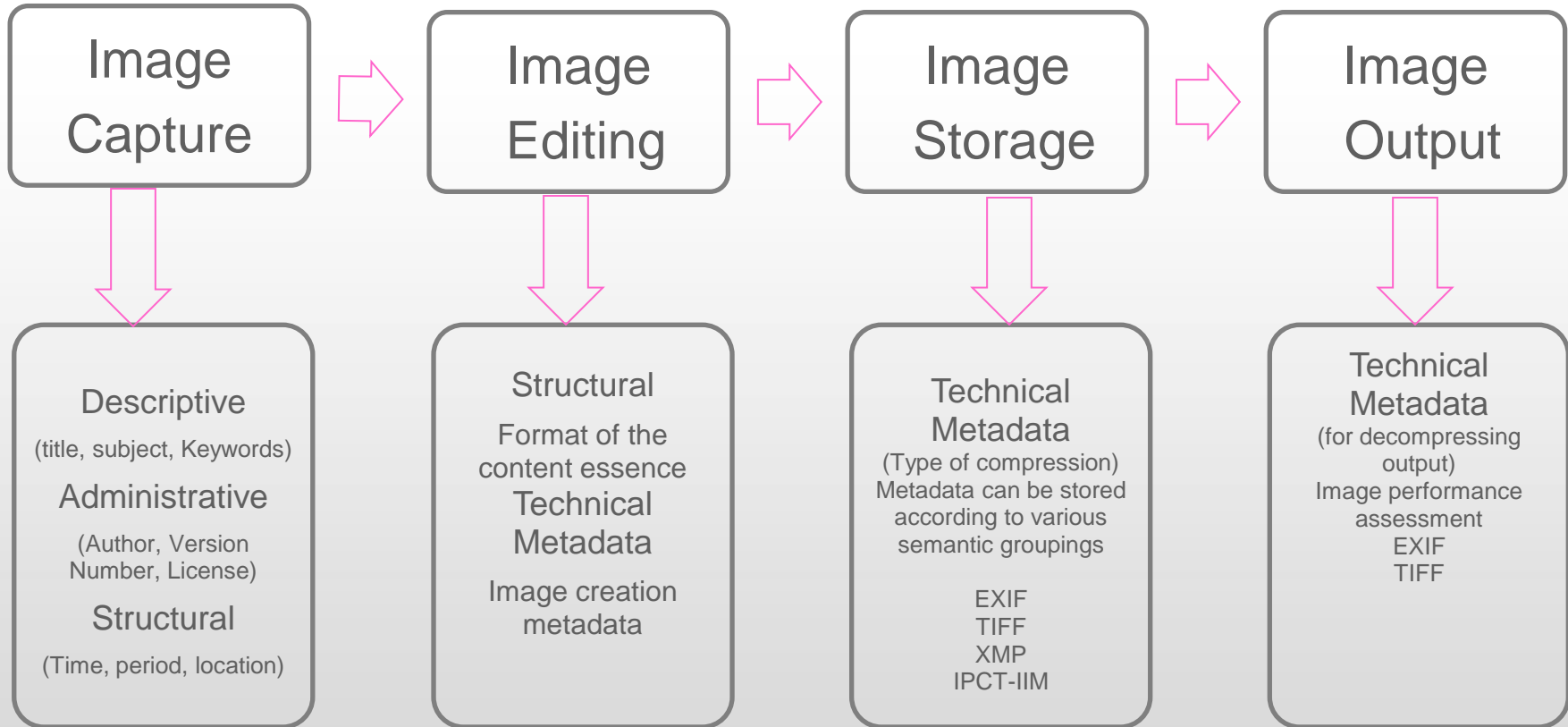
# Connecting Metadata with Image Files

- The “Header Data” is a specific set of metadata that consists of essential information to decode the image data (such as the image size, bit depth, color space and compression method, etc.)
- Metadata may include information such as the capture conditions of the camera, a description of the image content, and intellectual property information.
- The method of storing metadata within an image file format varies from format to format.
- Historically the two principle formats for storing data within the image itself were: Exchangeable Image Format (EXIF) or the International Press Telecommunications Council (IPTC)
- Embedding metadata has the advantage of protecting against loss or unavailability of a central database

# What is the “ideal solution”...?

- It depends:
  - size of database
  - expected demand for images
  - volatility of the data
  - available technical resources

# A framework of Metadata in Digital Image



# Some recommendations

- Before selecting a certain image format you need to ensure the repository software has the capacity to extract metadata, knowing that editing software doesn't corrupt the information if changes are created later on.
- The more standardized and valuable information we placed into the metadata, the greater searchable these files.

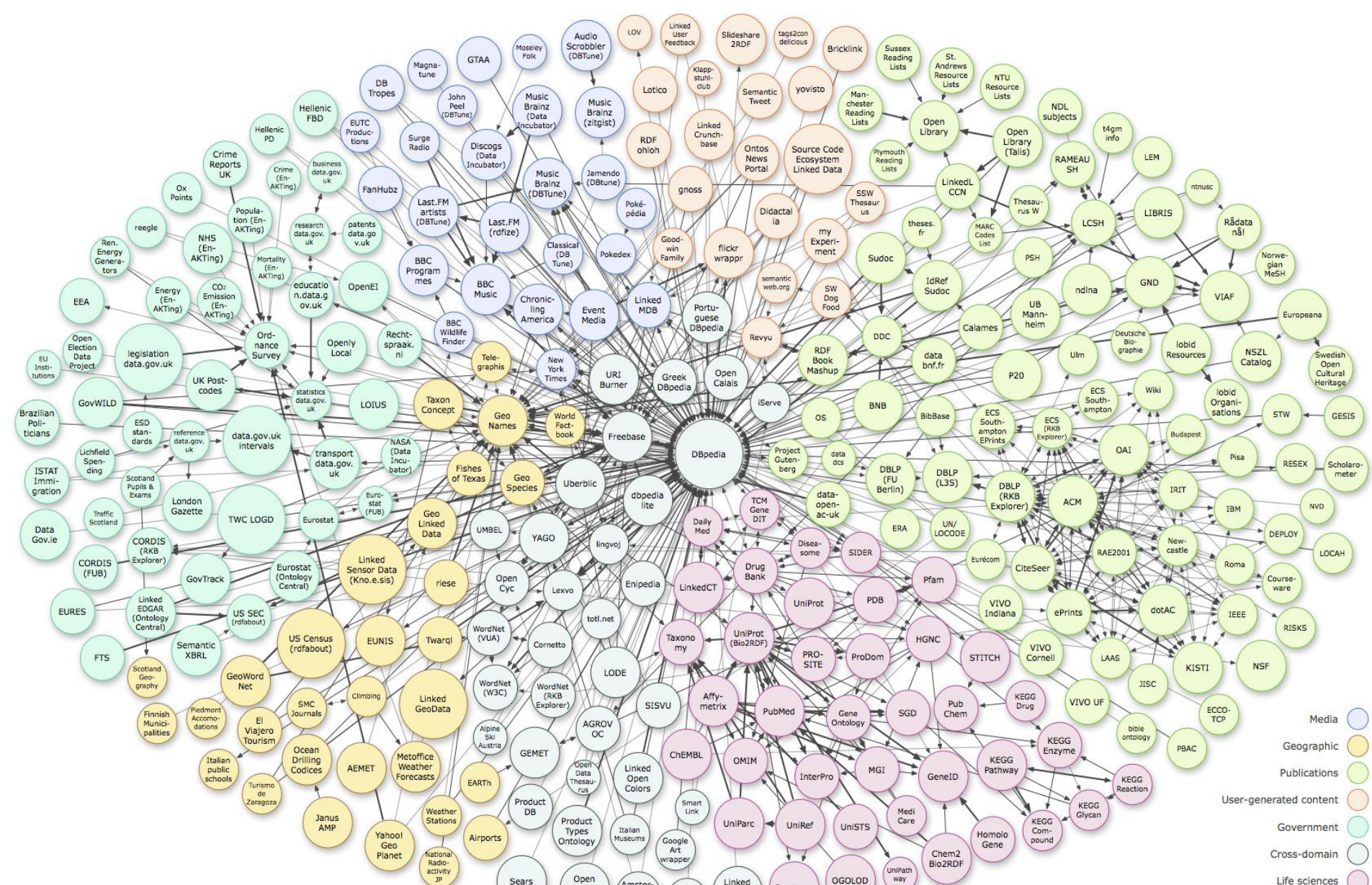
But now...




Infrastructure is moving to cloud ...

Network effects in end - user services ...

Consolidation leads to data driven services  
and experiences ... (analytics,  
recommendation, ...)





As of September 2011   

Each node in this cloud diagram represents a distinct data set published as Linked Data. The arcs indicate that RDF links exist between items in the two connected data sets. Heavier arcs roughly correspond to a greater number of links between two data sets, while bidirectional arcs indicate the outward links to the other exist in each data set. (Richard Cyganiak and Anja Jentzsch, 2011)

# To conclude...

- Metadata for digital images **serves several purposes**. It can be administrative, descriptive, identifying and categorizing a digital recourse.
- With digital metadata, the image becomes important not just for **today but also in the future**.
- This paper has provided a framework of contextual information that is based on a synthesis of a diverse and extensive body of literature about metadata for digital images.
- I believe this represents an important step in the ongoing evolution of thinking and practice in the management of digital asset metadata.



# THANK YOU!

Dr. Chrysoula Gatsou  
cgatsou@teiath.gr

Technological Institute of Athens

