THE ZERI PHOTO ARCHIVE TO LINKED OPEN DATA

MEMORIAV
LE PATRIMOINE PHOTOGRAPHIQUE
A L’ÈRE NUMERIQUE
15-16 novembre 2018
Lausanne / IDHEAP - Unil

MARILENA DAQUINO
University of Bologna
marilena.daquino2@unibo.it
@emmedaquino
OUTLINE

PHOTO ARCHIVES
[some] problems with data

ZERI PHOTO ARCHIVE

LINKED OPEN DATA
approaches & benefits

ZERI & LODE
LOD for photo archives
COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.
Sharing requires a lot of effort.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs

COMMUNICATION

STANDARDS

every institution adopts national / community / technological standards that fit-for-purpose.

PROPRIETARY SOFTWARE

sometimes not even cataloguers can access their own data and reuse them in other contexts.

CATALOGUING PROCESS

NOT UPDATED

the cataloguing process is an expensive and time-consuming activity. Data may be affected by information quality issues over time.

MISSING/PARTIAL

missing/partial information can affect the reliability of an institution and draw people to look somewhere else for more complete information.

COMMON PROBLEMS IN GLAMs
users easily discover your website

users can find what they are looking for (even when they do not know what it is)

users retrieve more information than they expected (serendipity)

users look for other information clearly connected to the original source (follow your nose)

users can answer complex research questions

users can formulate new hypotheses
only experts know how to find you

users do not understand whether you have what they are looking for or not

users cannot use properly aid tools you give them

users get frustrated

users do not know how you organised your knowledge and what they can discover

users search in Google
THE ZERI PHOTO ARCHIVE

- an art historical photo archive
- the bequest of Federico Zeri to the University of Bologna
- a strong focus on subjects (artworks)
- the first Italian photo archive creating detailed LOD

http://data.fondazionezeri.unibo.it
WHAT ARE LOD
AND WHAT THEY CAN DO FOR PHOTO ARCHIVES
the online cataloguing record of a photograph @Zeri

the online cataloguing record of the artwork depicted in a photograph @Zeri

an image attached to the record

web documents linked together by means of hyperlinks
ANONYMOUS, PERUZZI BALDASSARRE
- SEC. XVI - TRE GRAZIE

A photograph depicting an artwork

Machine-readable data linked by meaningful relations
the photograph is taken by an anonymous photographer

machine-readable data linked by meaningful relations
the photograph is a gelatine silver print

machine-readable data linked by meaningful relations
THE RDF MODEL

from a flat/tree data structure to a graph representation

photo N. 8675
described in

artwork id. FZ127
depicts

photographer
created by

has material

shot

taken by

EVENT

STILL IMAGE

artwork id. FZ127
WORK

photographer
PERSON

gelatin silver
MATERIAL
IDENTIFY persistent URIs for identifying concepts rather than documents

e.g. people, objects, materials, titles, subjects

CONSISTENCY ACROSS INSTITUTIONS
REUSE TERMS FROM SHARED VOCABULARIES
**THESAURI**

ICONCLASS, AAT
identify subjects, materials, techniques

**MULTILINGUISM**

e.g. gelatin silver print
http://vocab.getty.edu/aat/300128695

**AUTHORITIES**

ULAN, VIAF, TGN, geoNames
people, places, subjects again

**DISAMBIGUATION**

e.g. Oxford (1 out of 17)
http://www.geonames.org/2640729
ALIGN RECORDS BETWEEN DATA SOURCES
DATASETS
- open data now!

WIKIDATA, DBPEDIA, others
link same objects in different places

RECONCILIATION

e.g. “Peruzzi Baldassarre, Three Graces”
in Zeri photo archive, Berenson Library, Frick
Art Reference Library
INTEGRATE semantic and technical interoperability allows to aggregate data sources

INFORMATION QUALITY
ISSUES  VISIBILITY
ENRICHMENT

Gather data from other sources and save time

@Zeri

@DBpedia

→

Artist’s biography

Baldassarre Tommaso Peruzzi

http://dbpedia.org/resource/Baldassarre_Peruzzi

Baldassarre Tommaso Peruzzi (7 March 1481 – 6 January 1536) was an Italian architect and painter, born in a small town near Siena (in Ancaiano, frazione of Sovicille) and died in Rome. He worked for many years with Bramante, Raphael, and later Sangallo during the erection of the new St. Peter’s. He returned to his native Siena after the Sack of Rome (1527) where he was employed as architect to the Republic. For the Senese he built new fortifications for the city and designed (though did not build) a remarkable dam on the Bruna River near Giuscarico. He seems to have moved back to Rome permanently by 1535.

@en
CURATION

compare data sources to update data and speed-up the cataloguing process

harvester of authorship attributions from photo archives
AGGREGATION

create several access points to your data and increase discoverability of collections

Europeana includes Zeri’s LOD

PHARROS includes Zeri’s images and will include Zeri’s LOD
CREATE

develop expressive services and discover latent knowledge

ANSWER SOPHISTICATED QUESTIONS
mix data from different sources to create new applications

Italian Ministry of Cultural Heritage reuses Zeri’s data

psss! we can reuse their data as well!
MASHUP

mix data from different sources
to create new applications

provide tools for supporting the decision-making process
of users
reuse internal (or external)* information for making new hypotheses

identify photographers’ names in time in the Zeri’s authority file

https://w3id.org/zericatalog/name/anderson-roma/1853-1960
https://w3id.org/zericatalog/name/james-anderson/1853-1877
https://w3id.org/zericatalog/name/stab-d-anderson/1877-1938
https://w3id.org/zericatalog/name/d-anderson/1877-1938
https://w3id.org/zericatalog/name/guglielmo-anderson/1938-1960
https://w3id.org/zericatalog/name/alessandro-anderson/1938-1960
https://w3id.org/zericatalog/name/giorgio-anderson/1938-1960

define rules for automatically attributing the date of a new* photo recording one of those forms of name
LINKED OPEN DATA

1. Leverage external sources for your purpose
2. Facilitate others to reuse your data > gain visibility
3. Discover new knowledge

EXCHANGE AND CREATE KNOWLEDGE

CONCLUSION
THANKS

MEMORIAV
LE PATRIMOINE
PHOTOGRAPHIQUE
A L’ERE NUMERIQUE
15-16 novembre 2018
Lausanne / IDHEAP - Unil

MARILENA
DAQUINO
University of Bologna
marilena.daquino2@unibo.it
@emmedaquino