

The Zeri & LODE project

DHDK Seminar | 6.3.2018

Marilena Daquino | marilena.daquino2@unibo.it

The Federico Zeri's Photo Archive

<http://catalogo.fondazionezeri.unibo.it>

one of the largest private photo archives focused on italian paintings (including 290.000 photographs, 46.000 volumes, 37.000 auction catalogs)

2013 – PHAROS International Consortium of Photo Archives

to create a common platform for research on images of artworks

The Zeri & LODE project

1

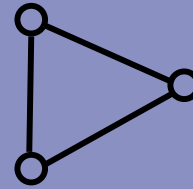
developing
ontologies for
describing
the photo archive



*reuse and creation
of ontologies*

2

data transformation
into RDF



*convert DB data into RDF
according to the ontologies*

3

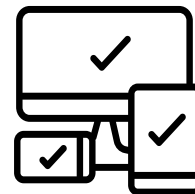
publishing
Linked Open Data



add links to the LOD cloud

4

providing
smart applications *that use such
data for real*



*a responsive website based
on a public user interface
for browsing RDF data*

What do we do today?

1

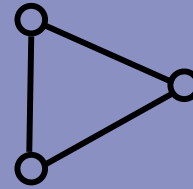
developing
ontologies for
describing
a photograph



*reuse of existing
of ontologies*

2

data transformation
into RDF



*convert DB data into RDF
according to the ontologies*

The Zeri & LODE project

Approaches to map the Zeri Photo Archive to RDF

Analysis of Content Standards



The Italian Cataloguing Standards for photographs and artworks

Scheda F (F Entry for photographs) and Scheda OA (OA Entry for depicted artworks)



Analysis of existing ontologies

The CIDOC-CRM

www.cidoc-crm.org

a standard for semantic interoperability between cultural institutions

- a domain ontology for representing Cultural Heritage Objects (museums objects)
- event-driven representation of activities related to the CHO life cycle
- general relations between cultural objects, and between cultural objects and related documentation (metadata documents, bibliography, archival sources, etc.)
- complex labelling, lacks of readability

Analysis of existing ontologies

The SPAR Ontologies

<http://www.sparontologies.net/>

a suite of modular ontologies for describing the bibliographic domain

- focus on objects that can be serialized (documents, books, journals, articles – and now also photos)
- each ontology addresses a feature of the domain (roles, citations, publication, etc.)
- bespoke relations between objects, and related documentation (metadata documents, bibliography, archival sources, etc.)
- easy to be read and reused



The HiCO Ontology

<http://purl.org/emmedi/hico>

an extension of the PROV Ontology to represent subjective attributions

- a task ontology for representing interpretations, sources, criteria
- situation-driven representation of an interpretation
- relations between cultural objects, related documentation and statements on such objects
- flexible, easy to be read and reused

Approaches to map the Zeri Photo Archive to RDF

APPROACH

ONTOLOGY REUSE

EVENT-DRIVEN

physical description,
life cycle of the object,
archival organization

use a **main domain ontology** to describe
pieces of information that different types
of cultural objects have in common

CIDOC-CRM

DOCUMENT-DRIVEN

bibliographic ref., citations,
roles in the publishing process

use **modular and simpler ontologies**
to describe information belonging
to other domains

SPAR ontologies

PROVENANCE

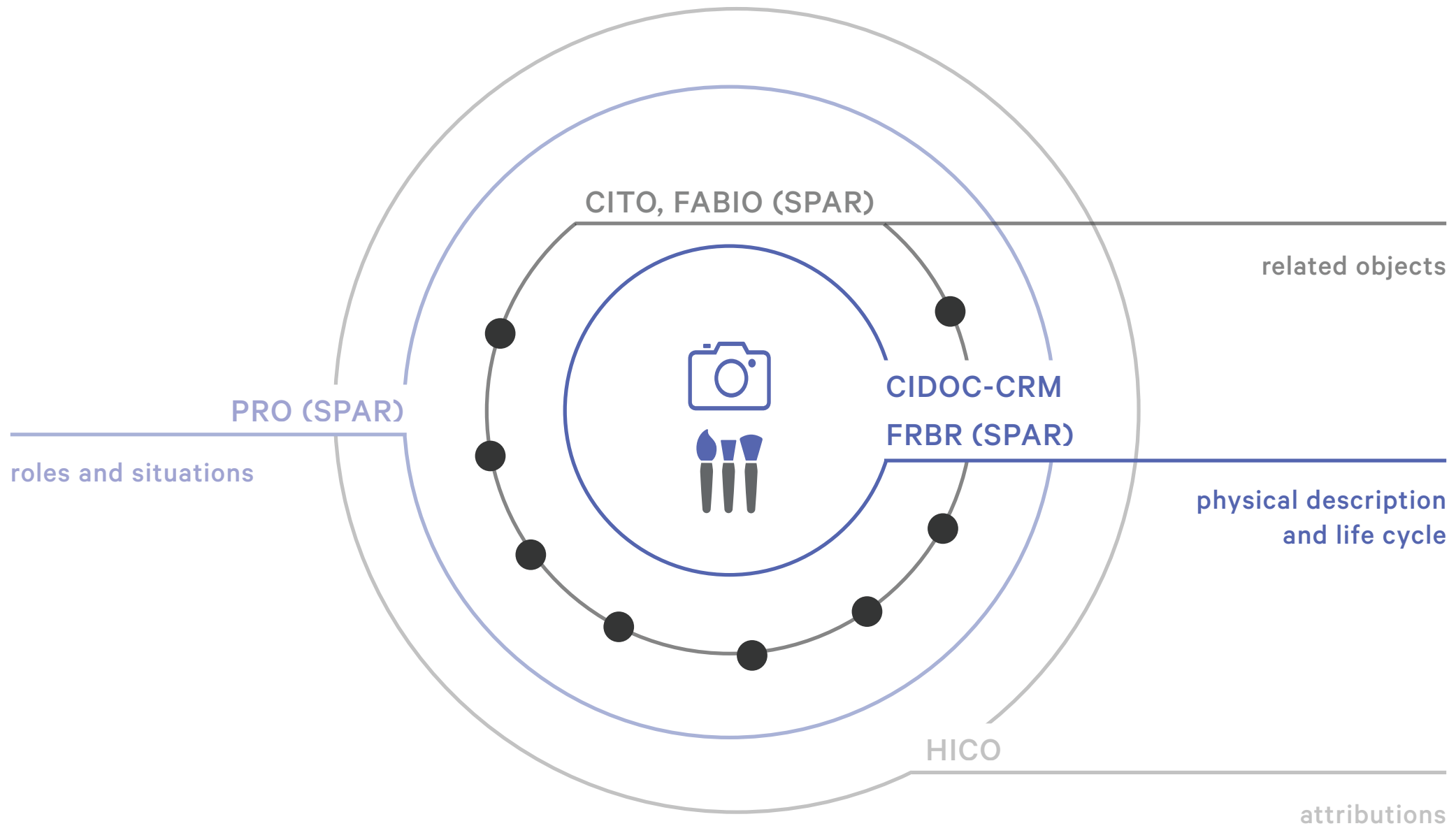
dates, title, and authorship
attributions, methodology,
sources and motivations

use **task ontologies** to preserve
complex information

HiCO ontology

a scalable approach for combining CIDOC-CRM, an OWL2DL FRBR model
and provenance assessment

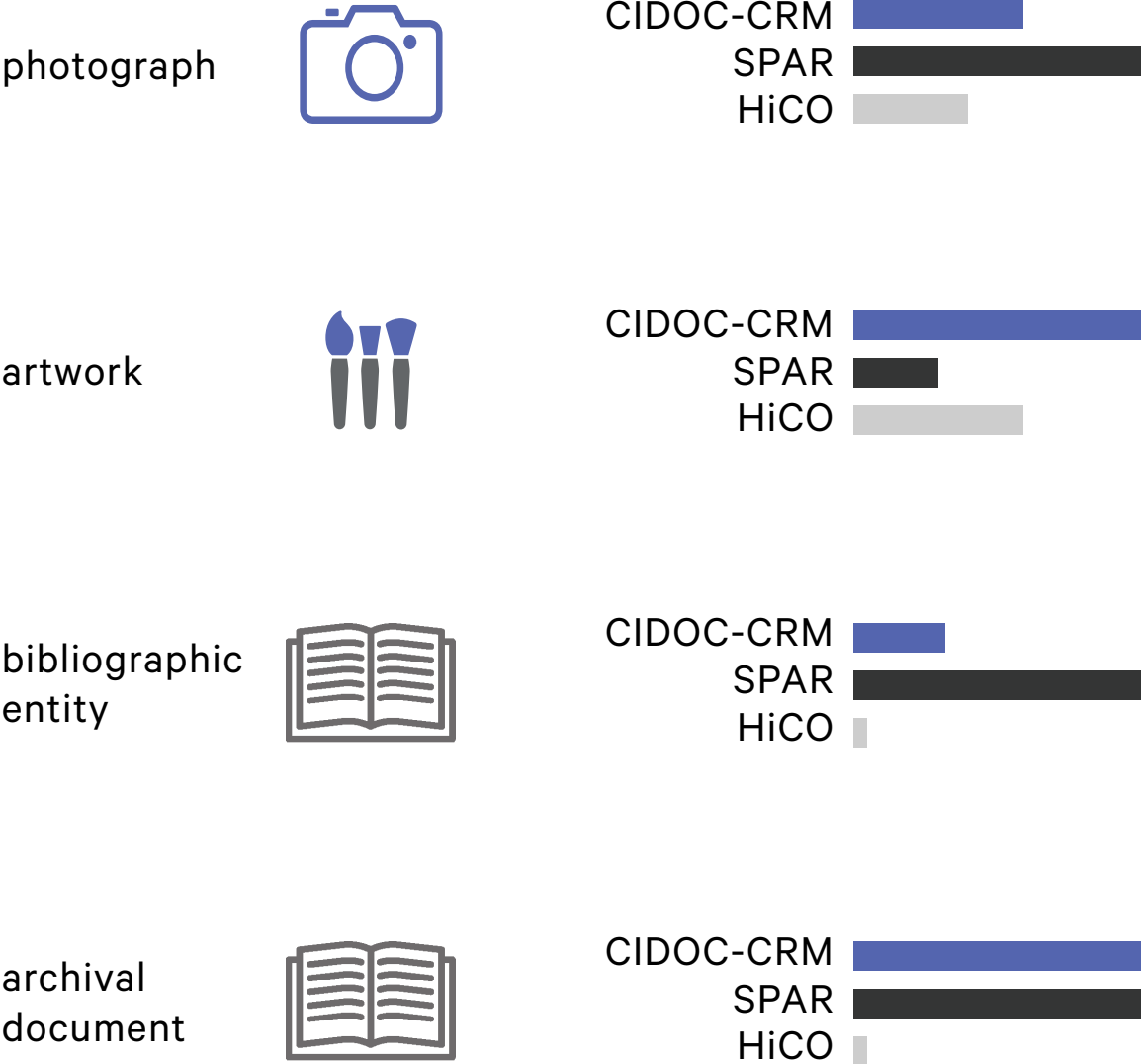
Reuse of existing ontologies



Ontology usage

OBJECT

ONTOLOGY USAGE

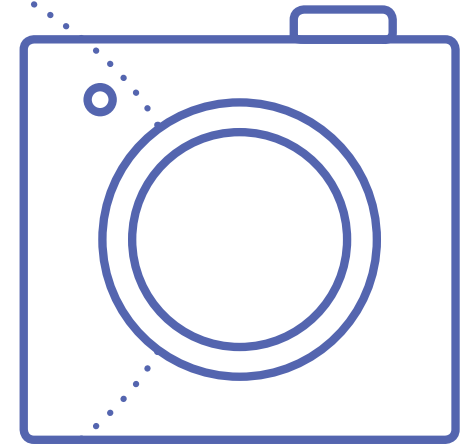


F Entry Ontology

<http://www.essepuntato.it/2014/03/fentry>

CIDOC-CRM

<http://www.cidoc-crm.org/>



MAPPING FtoRDF

<https://dx.doi.org/10.6084/m9.figshare.3175273.v1>

RDF/ttl example - F entry

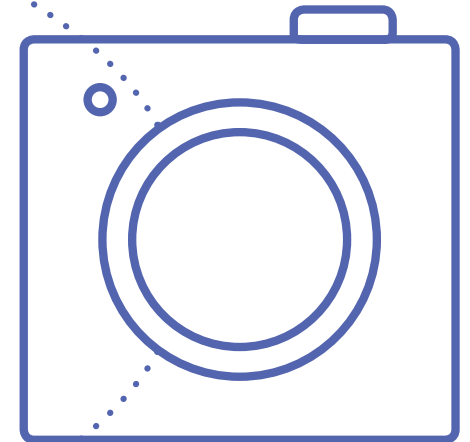
<http://dx.doi.org/10.6084/m9.figshare.3175252.v1>

OA Entry Ontology

<http://purl.org/emmedi/oaentry>

CIDOC-CRM

<http://www.cidoc-crm.org/>



MAPPING OAtoRDF

<https://dx.doi.org/10.6084/m9.figshare.3175057.v1>

RDF/ttl example - OA entry

<https://doi.org/10.6084/m9.figshare.3175048.v1>

The Zeri & LODE project

How we developed the ontologies for the Zeri Photo Archive

SAMOD Methodology

<http://dx.doi.org/10.6084/m9.figshare.3189769>

an agile methodology for ontology development:
small steps in an iterative workflow

- *competency questions for each small piece of information*
- *development of **modelets***
- *a data-centric approach: immediate test on existing data to guarantee consistency*
- ***refactoring** of terms by using established models; create only lacking terms*
- *produce the ontology documentation*

The Zeri & LODE project

Let's do it (again)!

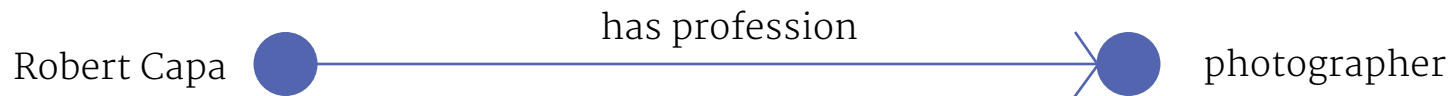
This is a simplified example

We'll skip some steps for the sake of simplicity
(i.e. check data consistency with SPARQL queries, create new terms, produce the ontology documentation)

ps/ Developing an ontology is not a “creative activity”

pps/ several representations are possible

The basics

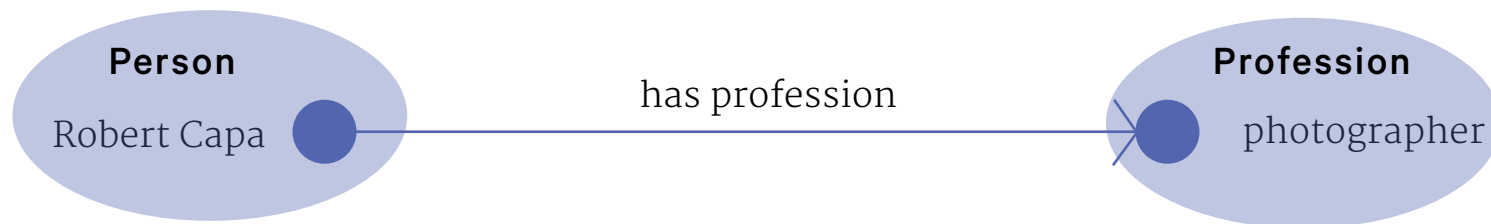


A triple is defined by a subject, a property, and an object

The **subject** is always identified by a URI (e.g. <http://example.org/robert-capa>)

The **property** is always identified by a URI (e.g. <http://example.org/hasProfession>)

The **object** can be either a URI or another datatype (e.g. a string, a number, a date, etc.)



Subject and objects belong to classes

A **class** is the definition of a stable quality shared among entities (being a person, being a place, etc.) and is identified by a URI. (e.g. <http://example.org/Person>)

A **property** is an attribute of an entity (has name, is part of, etc.) and is identified by a URI.

An **objectproperty** links two URIs. A **dataproperty** links a URI to another datatype.



Subjects and objects that are identified by URIs are called **individuals** of a class

A cataloguing entry describing a photograph

https://github.com/marilenadaquino/computational_thinking/blob/master/Seminar/fentry.csv

- have a look at the .csv
- group headings by the following 5 main areas

DESCRIPTION OF THE PHOTOGRAPH

ARCHIVAL CLASSIFICATION OF THE PHOTOGRAPH

CREATION OF THE PHOTOGRAPH

DESCRIPTION OF THE ARTWORK

AUTHOR OF THE ARTWORK

>> open a text editor, paste titles of areas, list .csv headings under the proper title of area

Thematic areas

DESCRIPTION OF THE PHOTOGRAPH photograph ID, Object, Photo title, Medium or materials, Height, Length, Unit of measurement, File, Negative number

ARCHIVAL CLASSIFIC. OF THE PHOTOGRAPH Archive series, Container number, Container heading, Folder number, Folder heading, Shelfmark, Region / Federal State, District, Town / Municipality, Repository, Address, Collection

CREATION OF THE PHOTOGRAPH Photographer, Reason for attribution, Source of attribution, From, To, Reason for dating, Source of dating, Shot date

DESCRIPTION OF THE ARTWORK artwork ID, Subject, Object depicted

AUTHOR OF THE ARTWORK Artist's name, Cultural context

Pick the following headings

photograph ID, Object, Photo title

Repository, Archive series, Container heading, Folder heading

- for each heading define which question it answers
 - what is the ID of the photograph?
 - what kind of object is preserved?
 - what is the title of the photograph?
 - where is the photograph preserved? etc.
- identify entities, labels and relations:
 - real entities (e.g. objects, people, places)
 - labels, text associated to each entity (e.g. titles, names, IDs) – NB always objects
 - relations between entities, and between entities and labels

Produce a spreadsheet like the following

heading	competency question	subject	object
photograph ID	what is the ID of the photograph?	photo(entity)	ID (label)
Folder Heading	where is the photograph preserved?	photo(entity)	folder(entity)
Folder Heading	what is the folder heading?	folder(entity)	heading(label)

>> open excel or similar, paste the headings of the table, list .csv headings under the proper heading

Define competency questions

Expected result

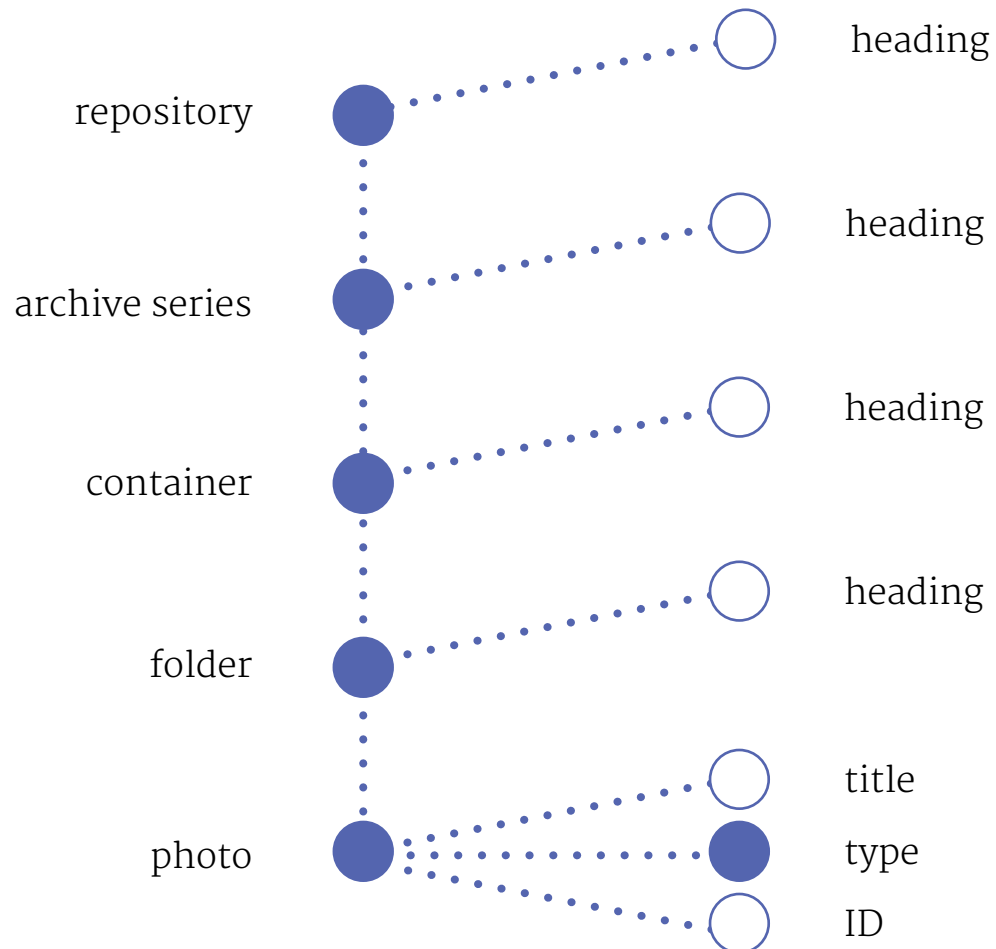
heading	competency question	subject	object
photograph ID	what is the ID of the photograph?	photo(entity)	ID(label)
Object	what kind of object is preserved?	photo(entity)	type(entity)
Object	what is the name of the type?	type(entity)	type(label)
Photo title	what is the title of the photograph?	photo(entity)	title(label)
Folder Heading	where is the photograph preserved?	photo(entity)	folder(entity)
Folder Heading	what is the folder heading?	folder(entity)	heading(label)
Container Heading	where is the folder?	folder(entity)	container(entity)
Container Heading	what is the container heading?	container(entity)	heading(label)
Archive series	where is the container?	container(entity)	series(entity)
Archive series	what is the series name?	series(entity)	heading(label)
Repository	who is the keeper of the series?	series(entity)	repository(entity)
Repository	what is the keeper name?	repository(entity)	heading(label)

Why is “type” an entity?

- when a quality is shared among several entities, a good practise is to create a controlled vocabulary of entities rather than strings
- positive, negative, polaroid, etc. are all types of objects you can find in a photo archive
- we define them as individuals of the same class
- uniquely identify relevant qualities allows you to have more links between data (a string can be only object of a triple, never a subject)
- other datasets may reuse the same term (type) to describe their objects, thus we can easily integrate data

Define competency questions

Expected result



Define classes and properties

- for each entity define the class that better fit the definition of the object
- several entities may share the same class
- one entity may belong to different classes
- strings do not need a class
- for each relation define a property
- the class of a subject is also called Domain of a property
- the class of the object is also called Range of a property
- similar properties may be reduced to the same one

Revise the previous table

The diagram illustrates the revision of a table. A horizontal table with three columns is shown. Above the first column, the text '... subject' is present. Above the second column, the text 'predicate' is present. Above the third column, the text 'object' is present. A curved arrow labeled 'modify' points from the first column to the second. A curved arrow labeled 'add a column' points from the second column to the third. The first column contains the text 'photo(className)', where 'className' is circled in blue. The second column contains the text 'hasId', which is also circled in blue. The third column contains the text 'ID (label)'. A vertical line extends from the blue circle around 'className' down to the text 'use CamelCase syntax to name a Class. First letter in lowercase'. Another vertical line extends from the blue circle around 'hasId' down to the text 'use CamelCase syntax to name a Class. First letter in uppercase'.

... subject	predicate	object
photo(className)	hasId	ID (label)

use CamelCase syntax to name a Class.
First letter in lowercase

use CamelCase syntax to name a Class.
First letter in uppercase

A python script to transform csv to RDF according to a modelet

https://github.com/marilenadaquino/computational_thinking/blob/master/Seminar/fentryToRDF.py

- save .csv and .py files in the same folder
- open the .py script in PyCharm (or another editor with a python interpreter)
- look at the function **getDataAccordingToMyModelet**
- add parameters of the function, uncomment, launch the interpreter (NB. include only names of classes and properties, not the complete URI)

An RDFtoSVG visualizer

<http://rhizomik.net/html/redefer/rdf2svg-form/>

- open the generated myModeletFentry.rdf file
- copy and paste your .rdf file into the text form
- select n3 as input format
- click on Submit

You'll see a graphical representation of your triples

Analysis of existing ontologies

OAD Ontology

<https://labs.regesta.com/progettoReload/wp-content/uploads/2013/08/oadNew.html>

DCTerms

<http://dublincore.org/documents/2012/06/14/dcmi-terms>

- map your classes to OAD and/or DCTerms properties

NB. pay attention to properties range!

And reflect on the nature of repository...

- substitute terms of your model with the new ones in the spreadsheet

A python script to transform csv to RDF according to existing ontologies

https://github.com/marilenadaquino/computational_thinking/blob/master/Seminar/fentryToRDF.py

- look at the function **getDataAccordingToOtherOntologies**
- add parameters of the function, uncomment, launch the interpreter (NB. this time include the full URI of classes and properties)

An RDFtoSVG visualizer

<http://rhizomik.net/html/redefer/rdf2svg-form/>

- open the generated finalFentry.rdf file
- copy and paste your .rdf file into the text form
- select n3 as input format
- click on Submit

You'll see a graphical representation of your triples

The Zeri & LODE project

DHDK Seminar | 6.3.2018

Marilena Daquino | marilena.daquino2@unibo.it