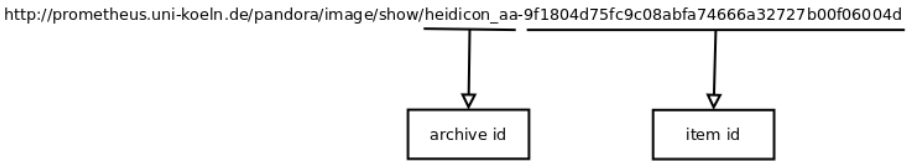
Prometheus2CRM

Creating LD-based description for a selected web resource

Step I – Selected Web Resource

Prometheus – the distributed digital image archive for research and teaching

The site features an aggregator for several **image archives** related to **cultural heritage**. The individual images are each embedded in a page containing describing data about it. They can be accessed using an URL-based scheme containing an **archive id** and an **item id**.



The received web site is of **XHTML** format and carries the semantics for the data in a **class attribute** of a containing <td> tag.

F.e. <td class="location-field"> means that the contained data is the location of the item.

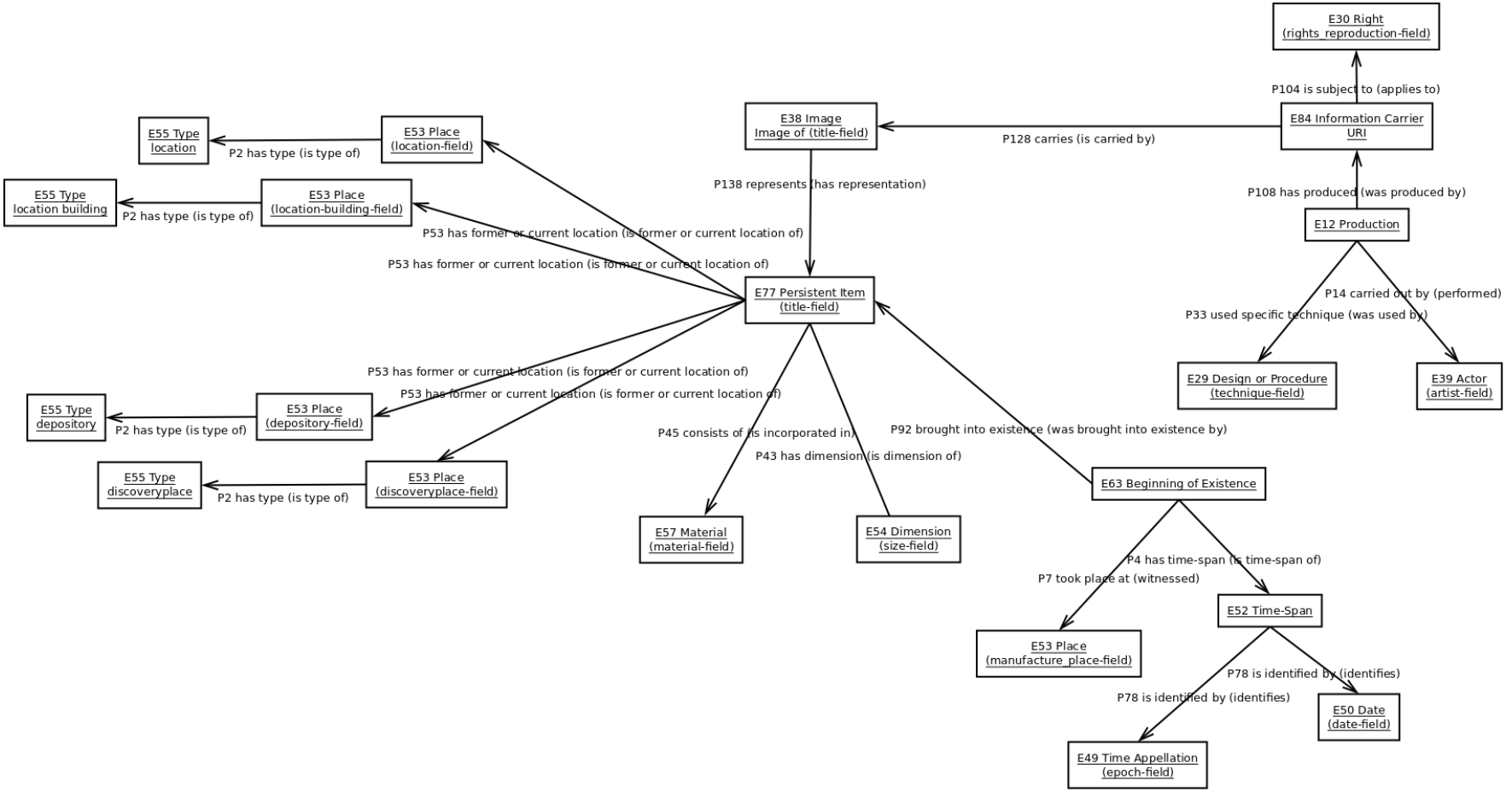
Using this scheme we can extract the data including the original semantics.

Step II – Page parser – RDF version

After extracting the data we create the corresponding RDF version. The modeling scheme we use for this uses an existing **ontology**, the **CIDOC CRM**, that was designed to describe data in the context of cultural heritage. We used an existing implementation in OWL-DL 1.0, the **Erlangen CRM**, that can be found at: http://erlangen-crm.org/current/.

The RDF triples created are instances (individuals) of CRM classes, and are connected using CRM properties. We tried to map the values of the class tags of the source to CRM classes. Additional information is added by connecting the instances via properties. This was only implicitly present in the original data and is made explicit by our modelling. However, not all data contained in the original web sites can be mapped, so there is some information modelled only as text, f.e. Descriptions or proprietary item ids of an archive. These are connected to instances using the P3_has_note property.

The detailed mapping can be seen in the following diagram. The P3_has_note textuals are omitted for better viewability.


E30 Right
(rights_reproduction-field)
P104 is subject to (applies to)
E84 Information Carrier
URI
P128 carries (is carried by)
E38 Image
Image of (title-field)
E55 Type
location
P2 has type (is type of)
E53 Place
(location-field)
E55 Type
location building
P2 has type (is type of)
E53 Place
(location-building-field)
P138 represents (has representation)
P108 has produced (was produced by)
E12 Production
P53 has former or current location (is former or current location of)
P53 has former or current location (is former or current location of)
E77 Persistent Item
(title-field)
P14 carried out by (performed)
P33 used specific technique (was used by)
E29 Design or Procedure
(technique-field)
E39 Actor
(artist-field)
P53 has former or current location (is former or current location of)
P53 has former or current location (is former or current location of)
E55 Type
depository
P2 has type (is type of)
E53 Place
(depository-field)
P45 consists of (is incorporated in)
P43 has dimension (is dimension of)
P92 brought into existence (was brought into existence by)
E55 Type
discoveryplace
P2 has type (is type of)
E53 Place
(discoveryplace-field)
E63 Beginning of Existence
E57 Material
(material-field)
E54 Dimension
(size-field)
P4 has time-span (is time-span of)
P7 took place at (witnessed)
E52 Time-Span
E53 Place
(manufacture_place-field)
P78 is identified by (identifies)
P78 is identified by (identifies)
E50 Date
(date-field)
E49 Time Appellation
(epoch-field)