Hierarchical objects Task force - Case study of Europeana 1914-1918

1. **Name:** Robina Clayphan, Europeana

2. **Background information:**

Europeana is the aggregator and archive for cultural heritage digital resources in Europe. It harvests metadata for cultural heritage objects that have a digital representation and makes these available via a common portal and distributes the data via an API. It harvests data from other aggregators and a few individual institutions.

In addition to harvesting data, Europeana creates its own "collections". It generates themed exhibitions from provider data and it has had a programme of collecting data and digital objects directly from the public. The latter was the Europeana 1914-1918 initiative http://www.europeana1914-1918.eu/en .

3. **Definition of hierarchies for the domain**

Europeana does not yet have its own definition of what a hierarchy is. As an aggregator of data from different domains, it must find a way to make sense of the variations that are defined by other institutions. We apply our own metadata format to harvested data and this is used in the portal and API. At the moment this is the ESE format which has many limitations.

Our particular challenge therefore is to define the best representation of data in EDM to support the display of hierarchical data in the portal. We must offer guidance to providers as to how to map their hierarchical data to EDM in order to get a good representation of their objects. In particular, we need to find a way to display hierarchically structured information in the portal so that users can easily see what relationships exist between objects found in the portal.

As a contribution to this Task Force we want to look at the data collected during the Europeana 1914-1918 initiative as many of the objects collected can be seen to be hierarchical in nature.

4. **Use Case Scenario**

Europeana 1914-1918
The collection is made up of contributions from members of the public who bring in memorabilia of various kinds to be digitised. The story behind the items is recorded, the items digitised and metadata created.

One contribution consists of:
1 metadata record
1 story (which is actually a text string in the metadata)

1 or more items
  - each item is represented by one or more digital file.

The relationship between the digital files that make up one item will vary depending on the nature of the items.
For example:

- A photograph may have only one file
- a post card will have two files – one each for the front and the back. In this case the sequence may not be of great significance but the relationship of the two files is critical.
- A set of several postcards – each individual postcard will have two files. The sequence of the post cards may or may not be of significance, but the relationship between the files for the front and the back of each card is.
- A letter – may have several digital files one or two for each page and the sequence is important.
- A diary – may have many files and the sequence is important.
- A physical object – may have several images of the different aspects and the sequence may not matter.

One contribution may consist of all or some of these items. There is therefore a sibling relationship between different items in one contribution in addition to the relationship between the files making up the individual items.

Examples:

1. http://www.europeana1914-1918.eu/en/contributions/4028
This contribution consists of one story with one notebook from which all 48 pages have been scanned in sequential order. This represents a short hierarchy with one parent (the story) and 48 children. The item is an object that needs to be represented in Europeana as a story with a group of 48 related files with a defined sequence.

2. http://www.europeana1914-1918.eu/en/contributions/3958
This contribution consists of three booklets with different types of content. Not all of the content of each booklet has been digitised. This contribution represents a deeper hierarchy with one story at the top level, three notebooks at the next level, each of which has children of its own at the item level.

At the moment these contributions are not represented in a hierarchical fashion due to the limitations of the metadata format used.

5. Problems and limitations

The diagram below shows how the data is modelled in the current implementation. It is sub-optimal due to the limitations of the metadata format available.
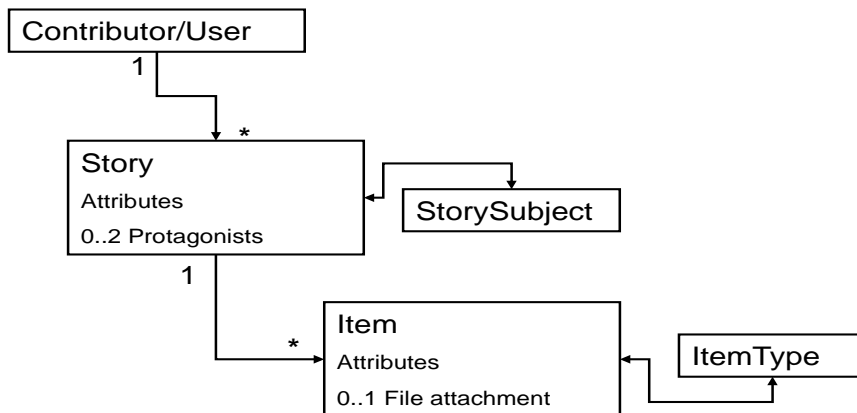
What we had to do…

No distinction between the Item class and its digital representation

Persons the stories are about as attrubutes and not as a class

We've improved the controlled vocabs. Structure though

The new model in practice

6. Proposed solutions for Europeana

This diagram represents a preferred model for the data and should be used as the basis for a representation in EDM.
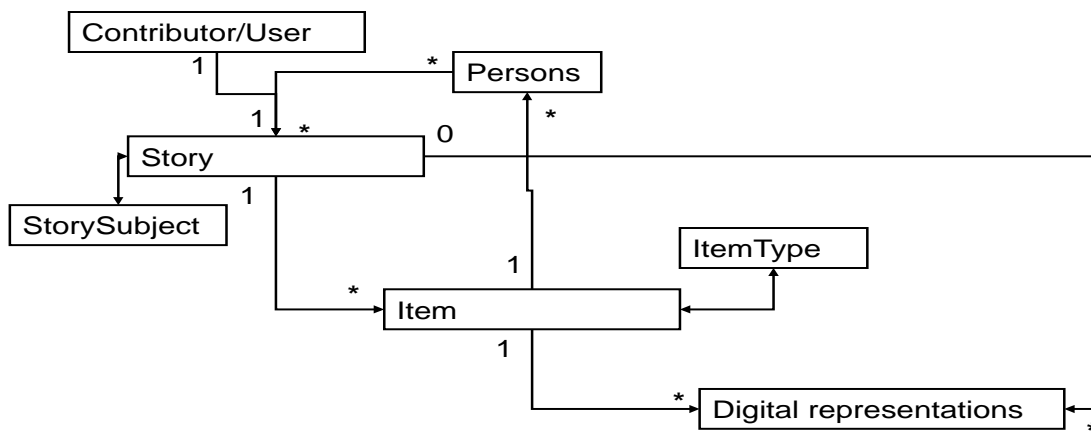
What we wanted to do…

Have story, items, digital representations, users and the persons the stories and items are about as different but related classes.

Improved storytelling data model