## 57 SIG - Issue 490: how to model a file

MD gave a summary of the discussions that have taken place among the members of the group that have been tasked with working on the issue and proposed to introduce a property whereby to refer to the content of an information object via an external copy, via a URL or archival identifier.

The details of the property definition can be found [below](#_Pxx_has_representative_1).

**Discussion points**:

* The last paragraph needs to be redrafted as an observation, not as an instruction. Could be turned into an example instead.
* The label also needs to be reconsidered.
* It was proposed that, in order to avoid duplicating P128i, the proposed property could be turned into a guideline. The SIG did not accept this proposal, as this property is analogous to P190 (which in its turn is considered incomplete without the new property).

**Decision**:

The SIG voted in favor of adding the property, but assigned CEO, MD, and SdS to redraft its scope note to read as an observation, and also to come up with a less controversial label.

**HW**: MD, CEO, SdS

Pxx has representative copy - definition

**Pxxx has complete copy (is complete copy of)**

Domain:

E90 Symbolic Object

Range:

E25 Human-Made Feature

Subproperty of:

E90 Symbolic Object. P128i is carried by (carries): E18 Physical Thing

Quantification:

many to many (0,n:0,n)

Scope note:

This property associates an instance of E90 Symbolic Object with a complete, identifying representation of its content in the form of a sufficiently readable instance of E25 Human-Made Feature carrying it, and no other content fitting to the type of the represented object. This property only applies to instances of E90 Symbolic Object that can completely be represented by discrete symbols, in contrast to analogue information.

The kind of representing feature may be paper, but also, in particular, electronic media, regardless whether they reside internally in clusters of electronic machines, such as in so-called cloud services, or on removable media. The *observation* of the feature, be it by direct eye-sight reading or via a mechanical device transforming its content into a form comprehensible by human senses, provides the ultimate access to the content of the represented Symbolic Object.

Which kinds of symbols and their arrangement on the representing feature are to be regarded as the exact content of the represented Symbolic Object will depend on the symbolic type of the Symbolic Object represented, which should be document via the property *P2 has type*. The representing feature, being material, will typically be more specific than the symbolic form defining the identity condition for the represented.

For instance, if a text has been declared as instance of E78 Information Object with *P2 has type* "Sequence of Modern Greek characters and punctuation marks", it may be represented in a formatted file with particular fonts on a particular machine, meaning however only the sequence of Greek letters. Any additional analogue elements contained in the representing object will not regarded to be part of the represented. As another example, if the represented object has type "English words sequence", American English or British English spelling variants may be chosen to represent the English word "colour" without defining a different symbolic object.

Consequently, more than one instance of E90 Symbolic Object, but of different symbolic type, may be declared *Pxxx has complete copy* the same instance of E25 Human-Made Feature. Further, all material copies may undergo deterioration. Since the identity of the represented symbolic object depends on the material content of one or more copies, extant or not, it is advisable to have access to multiple copies, in order to sort out corrupted copies or parts of them in time. Otherwise, the last existing copy or fragments of it constitutes the last witness of its content.

In a knowledge base, typically, the represented object will appear as a URI without a corresponding file, whereas the representing one may appear by the URL of a binary encoded file existing outside the knowledge base proper, or by the archival identifier of a paper edition. A URL for identifying the copy itself in a knowledge base should only be used as long as the providers support the persistence of that copy under this URL, as it is current practice for "Linked Open Data". Associating the referred copy or copies with a checksum in the knowledge base may help safeguarding the maintainers against unexpected change of content under this URL.

Examples

* Definition of the CIDOC Conceptual Reference Model Version 7.1.1 (E73) *has complete copy* The content under <https://cidoc-crm.org/sites/default/files/cidoc_crm_v.7.1.1_0.pdf> (E25) on the sever of ICS-FORTH in Heraklion, Greece.  
  [The edition 7.1.1 of the CIDOC CRM is registered under the public URI ["https://doi.org/10.26225/FDZH-X261"](https://doi.org/10.26225/FDZH-X261), [<https://cidoc-crm.org/sites/default/files/cidoc\_crm\_v.7.1.1\_0.pdf>](https://cidoc-crm.org/sites/default/files/cidoc_crm_v.7.1.1_0.pdf)which redirects users to the representative copy under <https://cidoc-crm.org/sites/default/files/cidoc_crm_v.7.1.1_0.pdf>. [<https://cidoc-crm.org/sites/default/files/cidoc\_crm\_v.7.1.1\_0.pdf>](https://cidoc-crm.org/sites/default/files/cidoc_crm_v.7.1.1_0.pdf)ICS-FORTH as organisation is responsible for the persistence of this content under the respective URL to the DOI Foundation]

In first-order logic: