

**Experts group on archival description**  
**Girona meeting**  
**October 15-17<sup>th</sup> 2014**  
**Notes**

**October 15<sup>th</sup>**  
**Morning discussion**

**Discussion about the ontology.**

The ontology is named “Records in context”, which illustrates our desire to describe archives and their context.

**Discussion about higher-level entities.**

The ontology is placed in a broad framework to ensure semantic operability with other cultural ontologies. A close look at CIDOC-CRM and upper ontologies revealed that they divided the world into two categories: “persistent entities” and “temporal entities” (this term from CIDOC-CRM seems inadequate and confusing, and is replaced by “occurrent entities”).

The DOLCE ontology (Descriptive Ontology for Linguistic and Cognitive Engineering) was a good source of inspiration: it is an upper ontology developed by an Italian research center, the Laboratory for Applied Ontology.

Following this distinction, records belong to continuant entities, as well as some of our conceptualized notions.

The occurrent entities category encompasses things that occur in time: processes, occurrences, happenings.

The distinction between continuant and occurrent entities forms the basis of our ontology.

Another distinction, inspired by upper ontologies, which extends to all knowledge domains, lies between “cultural” and “physical” objects. Among many references, it can be found in Jean-Jacques Rousseau's theories.

The ontology is a multiple inheritance class system. e.g. “Person” depends both from “Agent” and “PhysicalObject”/“NaturalPhysicalObject”/“Biological”.

**Discussion about “Agent” and its sub-entities: “Group”, “Person”, “Persona” and “ProxyAgent”.**

- “**CorporateBody**” should apply to a formally organized group, a legally registered body, having an identifier. It does not really encompass a socially cohesive “group”, like a village or a clan.

Discussion from Bogdan and Florence about the terminology. Should we prefer “legal body” to “corporate body”?

We should be careful about introducing new terminology. A definition of a corporate body can be found in ISAAR-CPF and should be considered: “An organization or group of persons that is identified by a particular name and that acts, or may act, as an entity. Also includes an individual acting in a corporate capacity”.

Florence. Sometimes we follow a pragmatic approach when applying a “corporate body” character to something. For instance, there is not a firm legal establishment for all medieval institutions and we consider them as corporate bodies if they have a name, had a stable existence and did produce some records.

Daniel thinks important not to refine so much into details if it does not really matter.

- **“ProxyAgent”** would cover a range of things which perform (semi)autonomous activities. It is a growing category of things: bots, software agents, space crafts... that generate information and archives, and “do act on behalf of ...”.

There is a need to be precise about “ProxyAgent”'s scope. For example, in France, a ship is also a corporate body, and the log is considered as the ship's produce, not the captain's.

- **“Persona”** would be used to describe a public identity. Corporate bodies can have a public identity and so “Persona” does not depend from “Person”.

Example of use case: two people write together a book under the same identity.

Gavan. Does “Persona” come for a role? i.e. to make a distinction between a professorship of physics in a university and several people who held the post throughout time. The answer is no, as this professorship should represent a different type of entity, nearer to an activity.

### **Discussion about “InformationRealization”.**

Depends from “ContinuantEntity”/“CulturalObject”/“CulturalPhysicalObject” and has “Record” as a sub-entity.

Daniel introduces this notion as being an abstraction. The record realization is an abstraction. It refers to its intellectual content, to its message. It reflects a concern to distinguish “what is intellectually there” and “what is physically there”, i.e. the documentary form. This concern could be put in relation with FRBR's “Expression” and “Manifestation”. We have to see if it fits correctly with diplomatics needs. Indeed, Stefano points out that a record can have different intellectual meanings depending from the physical form taken.

Even if a notable mass of archives does not seem to “be diplomatic” or to require such a degree of specification (Vitor), it appears all the same that every record follows some rules of representation and, as such, has a diplomatic key (Bogdan). It would not be the same need for details, but there should be enough possibility in the ontology to deal with “diplomatically demanding” archives as well as with standard records.

Some questions raised by this discussion: How are we to address a situation where books are part of a set of archives, if we do have to address this problem with our ontology? Is a copy of a letter a carrier or a medium?

Daniel ends this discussion for the moment, by concluding that we will have to structure the relationships between the intellectual and physical aspects, to leave the possibility to express complex situations as well as simple ones.

### **Discussion about “Set” and its sub-entity “ArchivalRecordSet”.**

Depends from “ContinuantEntity”/“CulturalObject”/“ConceptualObject”.

Daniel explains that the first notion retained, “record aggregation”, seemed too common a word, very used (e.g. “data aggregators”...), that also suggested something too physical. He substituted this expression with the term “Set”, with the postulate that the members of a set share at least one property (e.g. They derive from the same creator).

It is a mathematical notion. Salvatore stresses the point that there could be some problems applying a mathematical notion to our archival sets: there could be properties specific of the set that can not apply to the records that the set contains.

See Salvatore's email about NESTOR project.

Vitor raises the problem of the collection, which does not fit as an “archival record set”. Bogdan distinguishes two cases for the collection: it can be a group of elements assembled as such by a person (so it is not a problem to characterize it as an archival group) or it can be a group of elements that were assembled by the archivist, which is a more complex situation to deal with.

Bill suggests to add another sub-entity for collections.

Importance to adopt a less purist approach not to leave domains of reality aside.

Use case of “miscellaneous” sets, stressed out by Vitor. Sometimes, the only property of the records that are placed in a single set is simply that they do not fit into any other set... Some finding aids have a purely thematical organization. This raises also the problem of how to deal with the series (Bill).

Daniel points out that a record may be a member of more than one set. The physical management of the arrangement of a set might be coexistent with the intellectual set, but it is not frequent. Dispersed records could be put in a “conceptual set”.

## **Afternoon discussion**

### **Remark about the ontology's name.**

Jaana. What about the use of the plural: “records in contexts”?

Daniel. It raises a parallel question: should we put “records” in the singular? It would be ambiguous in respect to diplomatics. As for “context”, it is an elastic notion, which does not negate an eventual plurality.

### **Work organization.**

Timeline: there is still 2 years from now. Absolute deadline for a call for comments on next March 1<sup>st</sup>.

Methodology between work packages.

The ontology group and the conceptual model (CM) group really constitute one group.

The principles group will check terminology and definitions' clarity.

See CIDOC-CRM's introduction for naming entities, etc.

We could put multilingual labels and definitions from the beginning. It is possible in an ontology to put as many languages as we want (see the SKOS labels for example, with preferred and non preferred labels, all relating to a single concept).

Principles group: creation of an Excel online database for class names and property names. Two columns for names in English and French. This is the minimum for the ICA. Then other languages. Definitions in English and French.

This will be useful for the efficiency of the call for comments.

Creation of a terminology database for the glossary.

Usefulness of converting it in SKOS formalism, RDF model for vocabularies?

## October 16<sup>th</sup> Morning discussion

### Discussion about the top level entities of the conceptual model.

Daniel introduces five entities which should be the top level entities of our conceptual model:

- Agent;
- Function;
- Mandate;
- Record;
- Record set.

All of them are persistent entities. The comparison between the national models by WP4 showed more or less a consensus on these top level entities.

Vitor. Do we need five entities? What about a unique entity which would cover “record” and “record set”?

Daniel. They are not referring to the same reality.

Stefano. The set's properties are different from the properties of each single record. The set's properties are not the sum of the properties of the members of the set.

Record is not a subclass of record set.

Vitor. The set is not always an archival record set (cf in the ontology, “ArchivalRecordSet” is “Set”'s sub-entity). It can be any set.

Daniel. A set can also be a construction by a user, an historian, made from various single records from different sets.

Stefano. From a recordkeeping perspective, sets can be accumulated with single records. It is a different approach from when describing “sets” of historical archives: “spontaneous” sets.

→ Should “set” be divided between a subclass for “archival record set” and “record manager set”?

Bogdan. “Set” in the ontology is too generic. It does not represent what we are dealing with. We could qualify the naming, for example to say it is a “record set”.

Daniel. We can introduce another class under “set” as “record set”. It is useful in its abstractness to keep “set” as a higher level.

Daniel. I do not anticipate if we can have the need to make sets of agents, of functions, etc.

Conclusion from Daniel: “record” and “record set” are two different sorts of things, intimately related, but different.

### First discussion about how to express relationships.

To read the introduction of CIDOC-CRM proved very useful. There is a lot of interesting postulates and thoughts in paragraphs introducing its methodology and choice of formalism.

In CIDOC-CRM, there are two ways to express a relationship:

- a simple way. e.g.: an agent created a record;
- a more complex, event-based way. e.g.: a creator participated in a creative event which resulted in this product. Even if you do not know the agent, you have an event of creation and you can express it.

### Discussion about how to express the idea of duration when dealing with place and date.

CIDOC-CRM makes a difference between dates which are a human construction (e.g. the French revolution) and “normal” dates.

In CIDOC-CRM, they systematically distinguish the name, which is a human thing, and the fact: so

there are two classes to express the date (“Time Appellation”, “Period”).  
 In diplomatics, you have a “topographical” date (which is the place) and a “chronological” date. It is used in particular in the conventions according to whom we describe medieval documents.

**“Information object” and its subclasses in the ontology.**

Bogdan. The subclasses of “InformationObject” (“Set”, “CulturalObjectDescriptionContent”, “Function”, “Mandate” and “RecordContent”) do not seem to share so much as to be put under the same class named “InformationObject”.

Daniel. They are informations about the object and as such they share one common point.

Bogdan. Does not see a “mandate” as an “information object”. Why should “function” should be an “information object” and not the “record”?

Daniel. We do not talk of a particular instance of a function, we talk of a conceptual representation. “Function” is also a class of “InformationRealization”.

Jaana. What we model is the representation of the information, not the reality.

Bogdan. See for “Information object”, the definition in OAIS: “A data object (“either a physical object or a digital object”) together with its representation information”.

Jaana. Thoughts about the naming of the classes. We should not call the same things with different labels and we should take into account work done by other communities.

Terminology is an ontological problem.

Daniel. It is possible to map different terminologies.

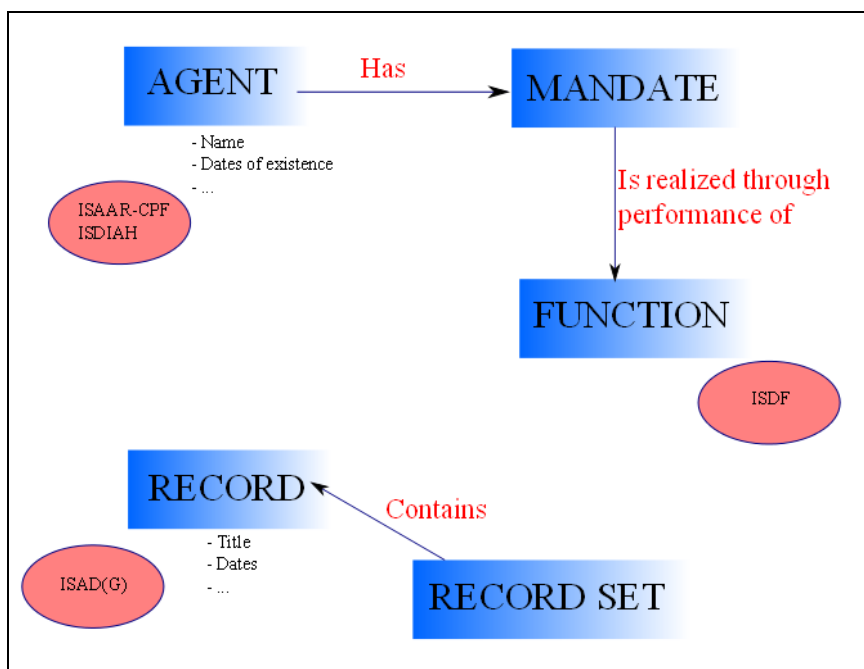
Gavan. Is there another conceptual object other than an information object?

In CIDOC-CRM, “symbolic object” is under “conceptual object” and comprises “information object”, as well as a subclass for “appellation”.

Particular history of the model.

**Afternoon discussion**

**Discussion based on a schema representing the five high level entities of the conceptual model.**



**Schema: five high level entities**

## Comments:

- “Agent” will serve also for archival institutions. It is possible to include properties describing archives managing informations and using elements taken from ISDIAH.

- How should we treat a “compound record”? It should be treated as a subclass of record, and not as a record set.

e.g. A text document with some photographs, a charter and its seal, etc. They really are a record, and not a record set.

- What about topical subjects?

In some of the national models which were compared, they could be found as high-level entities, even if the choices (definitions, scopes, formalism) were different.

Topical subjects can be named events, named places, named periods, topic controlled vocabularies, etc.

Essential distinction. A record can be issued in a place and can deal about a place, etc. So a “period” can be viewed as a topic.

Jaana. There is a distinction in CIDOC-CRM between “information from” (evidence of) and “information about”.

Florence. “About” can be expressed through a property, not through a class.

Agents can be topics too.

Libraries make the distinction between jurisdictional places (administrative areas, with a geographic extent) (which also are corporate bodies) and geographical features.

Places without a name.

Discussion about the notion of “evidence”. We do not want to mean any intention of interpretation from the archivist. To soften this term, we could say “the record is a trace of the performing of a function”, and not “the record is an evidence of”.

## **Breakout sessions.**

Group A: Gavan, Bill, Florence.

Investigate:

- date and place;
- event, action, transaction and relation.

Group B: Javier, Vitor, Alice, Bogdan, Stefano.

Investigate:

- diplomatics (form and structure in relation to medium and content);
- library conceptual models and their relevance to our work;
- digital records and migration of data formats;
- copies of records;
- compound records;
- what is a record.

Group C: Jaana, Victoria.

Investigate:

- topical subjects.

## **Results sharing.**

## Discussion about topical subjects.

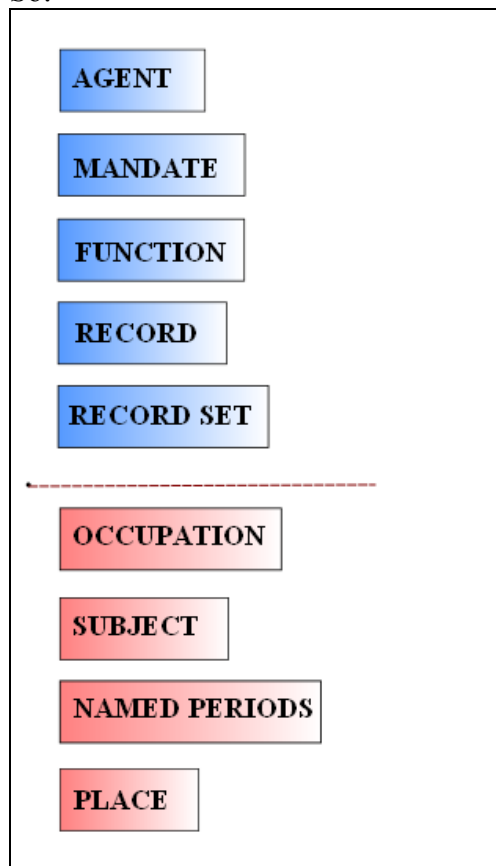
These things are related to records.

Any of our five entities can be treated as a subject: a record can be the subject of another record.

Library of Congress subject headings. These terms have also their history.

Occupation, subject, date and period. Things that can be named, identified and documented outside of records.

So:



**Schema: high level entities and topics**

Some topics can be combined. In the Dolce ontology, there is a “space/time region”.

## Discussion about events, and about how to express relationships.

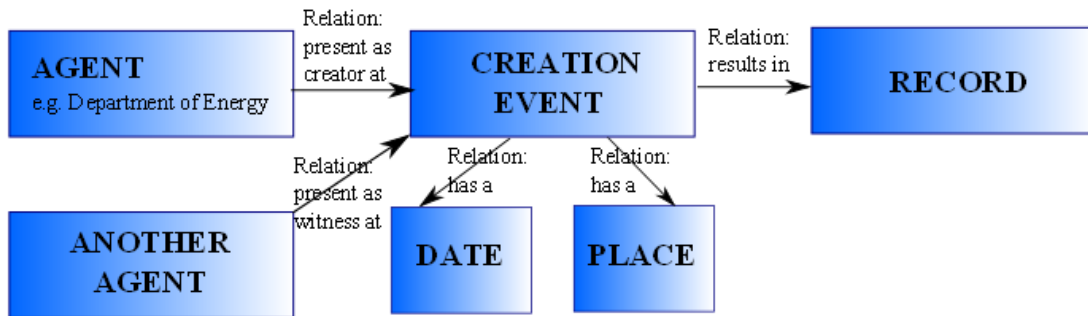
There are several ways to formally express a relationship between two classes:

- A simple way;
- An event-based model, such as in CIDOC-CRM;
- An alternative could be a relationship-based model, with a “relation” entity. The Australian model has such an entity. In this model, the relations are only associative events.

**First model of expressing a relationship.**



**Second model of expressing a relationship (cf CIDOC-CRM).**

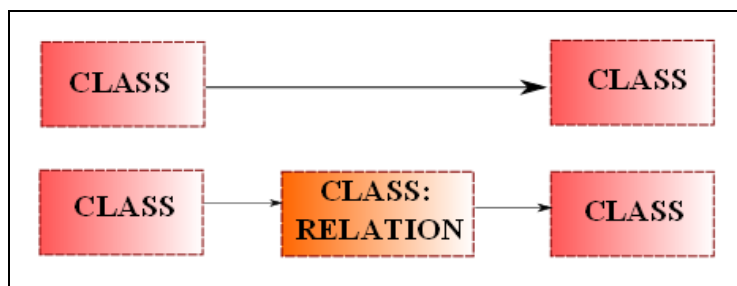


**Third model of expressing a relationship (cf AGRKMS).**



**Schema: expressing relationships**

The event-based model from CIDOC-CRM seems particularly complex to handle. We could combine two ways to express our model: we could use the first and third options (a simple one and a more complex one, based on a relationship entity):



**Schema: comparison of the elements'articulation in a simple model and in a relationship-based model (“From this formalism to this other formalism”)**

Comments:

The main fact about the third model is that it is possible to express a lot more of things, because if relationships are treated as an entity, they can be characterized through a range of properties.



Remark. Sometimes, we only have to establish a relation between two entities, but we can say nothing more about it other than it simply exists, so we could use the “relation” entity without any attributes.

Some thoughts about existing controlled lists of property values.

Library of Congress list of agent roles (author, book producer, bookseller...). A list of this kind exists in the BnF (National Library of France).

It can be applied to the “archival bond”.

i.e. “A record is related to another record” could be formally expressed with:



We could have event-based relationships.

We could also describe relationships that agents do not cause.

Bogdan. It is important, when choosing the model, to know if it is sized to the type of queries we will do. For example, what can we do with an event-based model? What queries do we do?

Bogdan. From a terminological point of view, an entity should exist in itself. A relation can not. So, from a strict theoretical point of view, it could be impossible to treat the relationships as an entity.

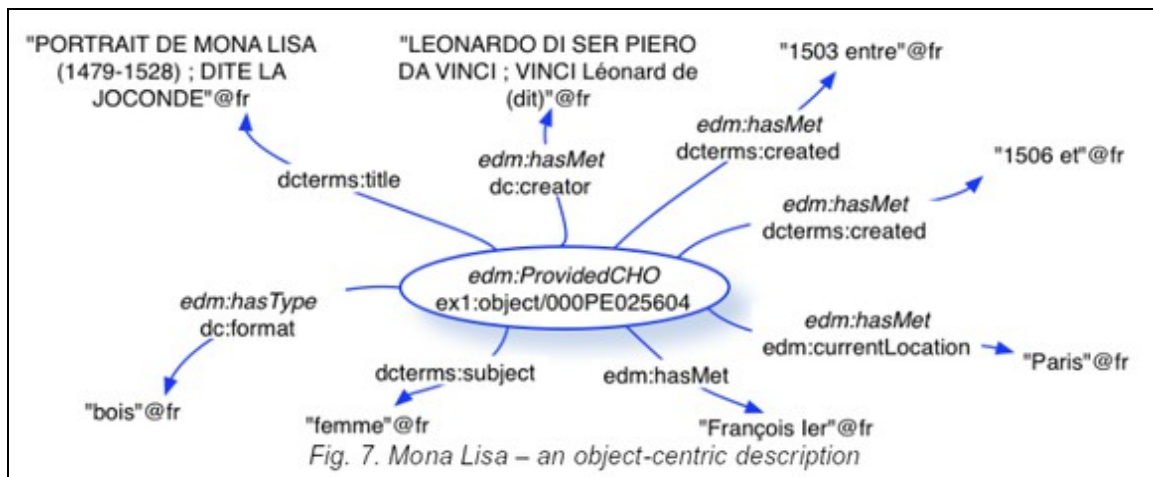
Daniel shows some other model references.

- Europeana data model primer.

<http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5>

There are two approaches: an object-centric approach and an event-centric approach.

We have a look at some examples for Mona Lisa. Below are some copies from this document:



(schema from page 13)

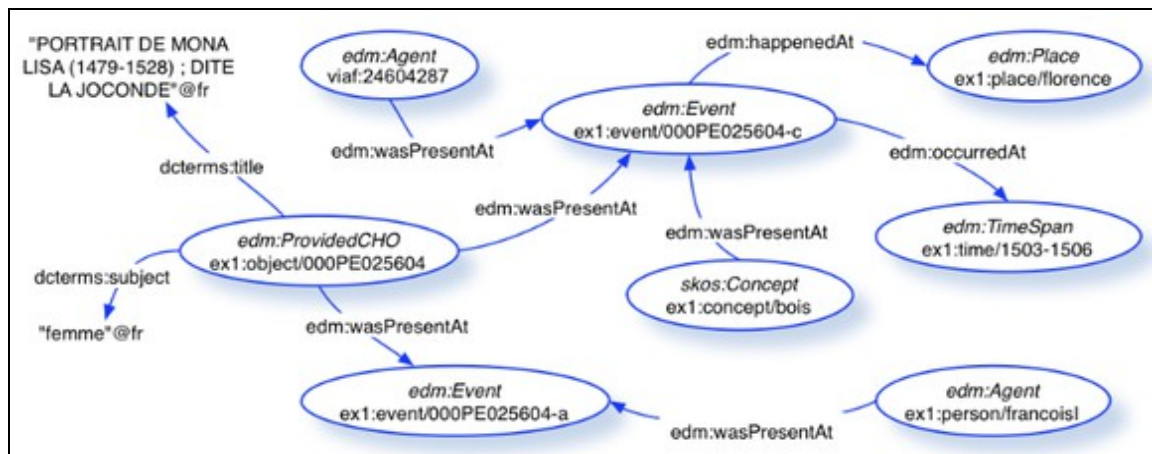


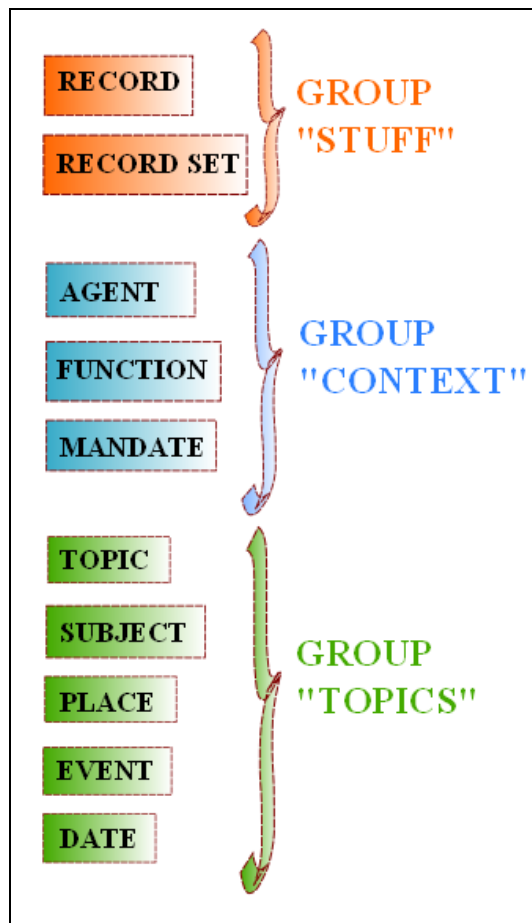
Fig. 10. Mona Lisa – an event-centric description

(schema from page 16)

The role of the agent can be defined through the relation.  
 Example: "Leonardo Da Vinci" edm:wasPresentAt "CreationEvent".

October 17<sup>th</sup>  
Morning discussion

**Introduction: links between the ontology and the conceptual model.**



**Schema from Salvatore: three groups of things to describe**

Daniel. One of the most important things we are trying to do is to pull out these things from ISAD(G) (records, levels of description, function, form, mandate, etc.) and deal with them independently, to create a finding aid, and also to be able to obtain something more: possibility to isolate a function, to identify the agencies that were responsible for its performing, and then access to the records that were created.

We are trying to think in terms of data. We should be able to combine them as we want. So, the data model should allow us to continue doing what we are used to do, and to do a lot of other things.

What are the links between the ontology and the conceptual model?

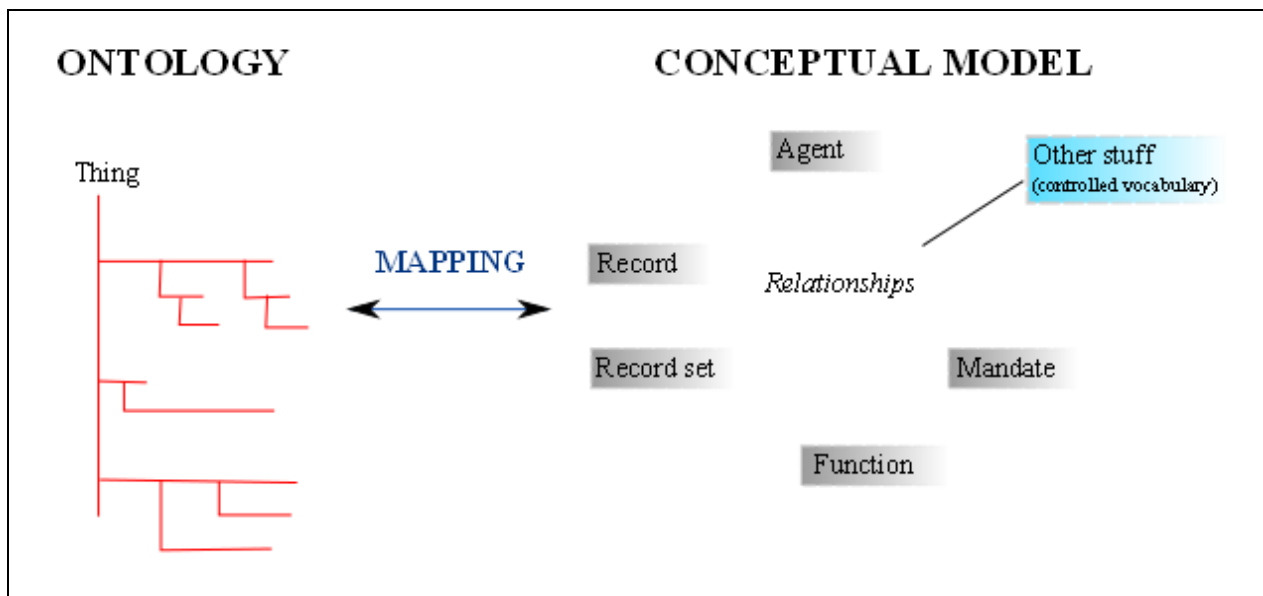
How does the conceptual model relate to the ontology?

The terms can sometimes be loosely used.

An ontology starts from a very basic thing, then has a rather elaborate structure. Its framework is broader than in the conceptual model.

Things from the conceptual model should sometimes find a rather unexpected place in the ontology, but everything that participates in the statements of the model gets represented in the ontology.

The WP3 group will be in charge of expressing all the elements from the conceptual model in the ontology.



**Schema: ontology and conceptual model**

### **Discussion about the breakout session on records.**

Report from Bogdan. To give a definition of a record, we started from ISAD(G)'s definition and added other aspects.

→ “Recorded information, in any documentary form or medium, created or received and set aside by an agent, in course of conducting the affairs, as an evidence or an information resource.”

“Set aside”, taken from InterPares. “Maintained” was the original term, but it includes a notion of preservation that is not always present. Records can be kept without a bother to manage them.

“Business” was removed and replaced by “affairs”, because for non English-speaking countries, it has a notion of economic value.

“Simple document” and “Compound document”.

→ “A record made of a document with multiple parts, that can not logically be separated”.

Example. A database with multiple data from different provenances.

Daniel: a simple document can have one part. Should we say a “discreet document”?

Its authenticity, its intellectual integrity is removed if a part is removed. It undermines its “recordness”.

The “logically” notion should be removed from this definition.

A compound document is made of multiple documents.

Examples. An email with 3 attachments. An office document with a photograph embedded in it and a spreadsheet. A letter with enclosed photographs.

If you remove a part, you describe something that is no longer what it was.

Use of the term “document” when talking about a compound record?

In a “simple” document, you describe one documentary form.

In a compound record, you can provide a documentary form description to all the components. Each element has a documentary form of its own.

Frequently we do not describe at the item level. But we will have to cover it in the model.

→ A compound record is a record with its own existence and its components are records that have a documentary form, an identity and an integrity of their own.

### **Discussion about levels of description and record/record set.**

Comparison between:

1) Notion of record set (gathering, aggregation, assemblage)

2) and the arborescence as expressed in EAD and ISAD(G):

EAD

<archdesc>

<c01>

<c02>

<c02>

<c01>

<c02>

<c02>

ISAD(G)

Fonds

Series

File

Item

#### Comments:

The arborescence from ISAD(G) and EAD is something constructed based on professional rules and practices about a hierarchy which can be different from one country to another.

“Set”, as we already said, is a mathematical notion: the members of the set share one or more properties.

But sometimes we have a miscellaneous set, where the members share no property.

At the fonds level, the description units can share a provenance property.

At a series level, the description units can share a function property. Bogdan: it can also be a provenance notion in a sense. In some countries, fonds are arranged in respect to the ideal/expected structure of the creator.

Florence. A set may be composed of one record only. Possibility of fonds with accruals.

There is an ontological distinction between record and record set.

“File as an item” vs “file as a set”.

Do not confuse levels of description and units of description.

Florence. Archivists, when dealing with a file composed of only one record, would not systematically treat it as a file, more as an item...

Bill. The arborescence deals with three layers of sources: administration, function and things in themselves. cf. T. Cook.

Florence. We could create an event about a set creation event, an arrangement event, whose result would be a set. You may not know who did it or why was the intention, but you know an event took

place.

A set which arrived as such / a set which was set up after.

Bogdan. Should we add a third class, which would top these two?

Conclusion from Daniel. We could abstract ourselves from a strict terminology of “fonds”, “sous-fonds”, etc. and produce something which should allow every archivist to type the levels or record sets according to his practices.

Cf mail from Daniel on October 24<sup>th</sup>:

“I think our descriptive alternative to merely typing the "levels" or Records Sets will prove to assist the community in understanding what they are doing when describing, and also help users by providing useful information on how to understand and use the description in finding and understanding records.”

### Afternoon discussion

#### Discussion on properties to assign to record and record set.

Set	Property	Record	Comments
Yes	Level of description	Item (implied)	
Yes	Identifier (multiple) : persistent identifier / URL	Yes	Mandatory
Yes	Title/appellation/name (descriptive)	Yes	Some records may not have a title. e.g. a letter Highly recommended.
Yes	Single date – Date range – Date set (typed)	Yes	A type is needed. Has a creation/registration, etc. date.
Yes	Scope/content note geog., time, subject, form	Yes	
Yes	Access restrictions	Yes	Authority, dates?
Yes	Use (rights) conditions	Yes	
Yes	Extent	Yes	
	Documentary form:		
No	language	Yes	
No	script	Yes	

#### - About “scope content”

What do we understand with this notion? It is defined in ISAD(G) with three types of information: type, scope and interest of the document.

Do we keep informations related to the document form and to the message in the same place?

Florence. Diplomatics: a content is seen as the pure message of the document and is abstract. It is the subject put in a narrative form of description. There is a verb in it. So it is different from a topic. e.g. Charlemagne allows an abbey to raise a tax on a bridge...

Does it include a documentary form?

Scope: in the French practice, the things you put in this field are different depending from the level you are describing.

You can not have a documentary form for a fonds. To be replaced by “type”?

We should allow a possibility to be as accurate as possible and also allow economic ways of describing.

There are very common documentary forms: “a will is a will”, etc.

#### - About rights and accessibility.

Stefano. Are “access restrictions” really a property to record? Is there not a relation to the rules that a custodian is setting to a record?

Bogdan. Is it not a management decision? A special section for management?

Remark: in the model used for the French “Archival standard for exchanging metadata” (SEDA), descriptive metadata were separated from management metadata, where all the informations about the access were placed.

Florence. This field will be difficult to formalize. It can be very complex, with processes.

They can be associated with dates.

“This authority placed this record on restricted conditions until this date.”

Can we provide a controlled list of values?

Controlled vocabulary: “open”, “restricted”, and dates. The question of “Why did we apply these conditions?” could be included in a note.

This type of information can be closely related to national policies.

There is a lot of expected publicity from accessibility rules: open data policy, transparency laws, according to governmental positions.

Victoria. What about records that are open to some people and closed to others?

Gavan. What about things that were created with no intention of publication? Private papers rules.

Is this not outside our scope?

#### - Are we describing the set, or the records in the set?

We have to be careful about what we are describing when talking about a record set's properties. Do we describe the properties of the set or the properties inferred from the records which constitute the set?

For example, the name assigned to a set describes the set, but the dates are actually the dates of the records in the set (the date of the creation of the set is different). The language of the set is actually the language of the records in the set.

Properties referring to the set: level of description, identifier, title, scope.

Properties referring to the records in the set: the rest?

There are “rules” for properties in an ontology. Cf. “A semantic web rule language” (SWRL), submission by a W3C member.

Conclusion: one of the next steps will be to define each property, with its cardinality (one to one, one to many, etc.) and eventual predetermined values (data value, constraints).

#### **Next meeting and work planning.**

Focus on the conceptual model to be released for comments. Keep it basic and underdetailed.

February 1<sup>st</sup>: deadline for a draft.

March 1<sup>st</sup> : release and call for comments (conceptual model, glossary and first input of ontology).  
60 days left for comments.

Between May and first two weeks of June: next meeting in Romania. 4 full days.