

Mapping to the CIDOC CRM

Basic Overview

George Bruseker ICS-FORTH
CIDOC 2017
Tbilisi, Georgia
25/09/2017

Table of Contents

1. Pre-requisites for Mapping
 - Understanding, Materials, Tools
2. Mapping Method
 - Source Analysis, Target Analysis
3. Mapping Recipe
4. Mapping Example

1. MAPPING PRE-REQUISITES

What do I need to do a mapping?

Understanding

The Knowledge

- Know your source!
 - What the field intends to document not what its label says it is
 - What the data in the field actually is. Was it used for it?
- Know your target!
 - Study

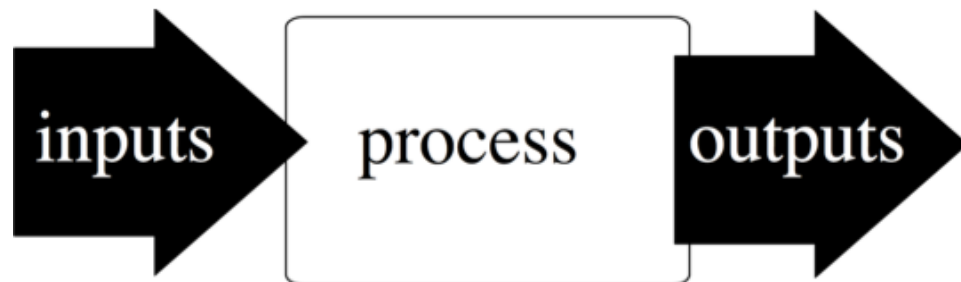


What do I need to do a mapping?

Materials

The Documents

- Source
 - Description of the Schema
 - Copy of the encoded Schema
 - Sample Data
- Target!
 - Description of the Ontology
 - Encoding of the Ontology
 - CRM @ <http://www.cidoc-crm.org/>



What do I need to do a mapping?

Tools

The Tools

- Planning
 - Pen and Paper
 - Time
- Executing
 - A Mapping software
 - Time
 - 3M @ <http://139.91.183.3/3M/Login>



2. MAPPING METHOD

Mapping Method: Analyzing the Source

Questions to ask

- What is the schema about?
i.e.: what is the subject of the data structure?
- What kind of statement does each field make about the subject?

Tip

- Think of the natural language sentences / propositions that the data structure encodes
- For the overall schema write 'There is an X' where X = subject
- For each field write 'X is called', 'was created in year' etc.

Crimea Conference

Historical Archives....

The whole schema says:

There is a document

The fields say:

- The document has a type
- The document has a title
- The document has a second title
- The document came into existence at a date
- The document was created by 1 or more people
- The document was published by one or more institutions
- The document is relevant to a particular subject

Field	Value
Type	Text
Title	Protocol of Proceedings of Crimea Conference
Subtitle	Declaration of Liberated Europe
Date	February 11, 1945
Creator	<ul style="list-style-type: none">• The Premier of the Union of Soviet Socialist Republics• The Prime Minister of the United Kingdom• The President of the United States of America
Publisher	State Department
Subject	Postwar division of Europe and Japan

Mapping Method: Understanding the Target

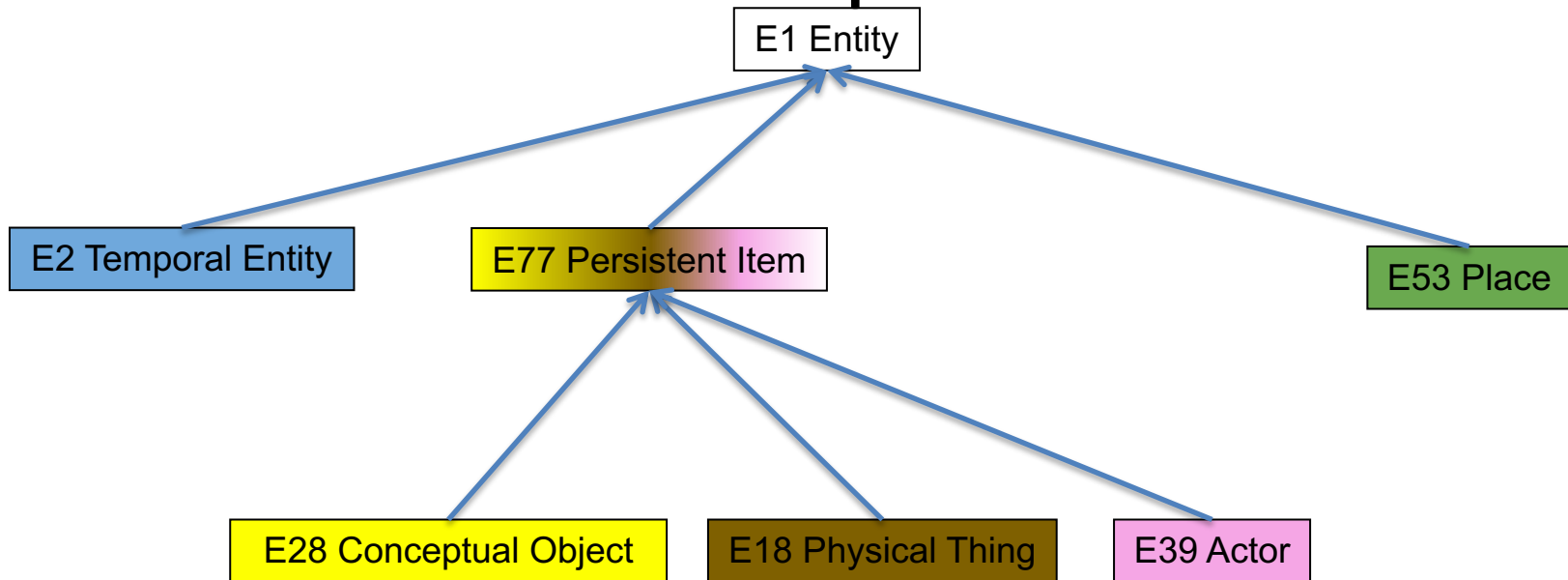
Read / Question / Understand the Top Level Classes

For each ask:

- What kind of things does it allow me to talk about?
- What does it allow me to say about that kind of thing?
- Think about the nature of the object that the source is talking about and what it says about it
- Are the target class and its relations adequate to express this?
- Do I want/need to say more?



Mapping Method: CIDOC CRM Top Level Classes



- To anything I can give a name or a type
- Some things are consistent in identity through time (objects) while some change (temporal entities) but have an identity overall
- Of objects we can distinguish: conceptual things (that are not limited to one instances), physical things which are unique, and actors which have the unique property of agency in the world
- Places define a geometric location bound to some object in time

Mapping Method: Using the Target

- Formal Ontologies are arranged hierarchically.
- The highest classes are the most abstract and define - **through their relations** - the highest levels of discourse within a domain. These are your starting point for understanding/mapping.
- Everything that can be said about a superclass can also be said of a subclass. Everything that can be said about a super relation can be said of its subrelation
- Once you find the branch of the hierarchy where your concept fits... find how low you can go!

E1 Entity

E2 Temporal Entity

E4 Period

E5 Event



E1 What you can say about anything

E1 CRM Entity	
Scope	This class comprises all things in the universe of discourse of the CIDOC Conceptual Reference Model.
Relations	P1 is identified by (identifies): E41 Appellation P2 has type (is type of): E55 Type P3 has note: E62 String

For the Crimean document example, let's assume: **Whole thing = E1**

This allows us to say:

- title = **E1 CRM Entity** *p1 is identified by* **E41 Appellation**
- type = **E1 CRM Entity** *p2 has type* **E55 Type**

But we obviously need to say more. So, we descend the hierarchy.

E2 Expressing things about time

E2 Temporal Entity

Scope	This class comprises all phenomena, such as the instances of E4 Periods, E5 Events and states, which happen over a limited extent in time. In some contexts, these are also called perdurants.
Relations	P4 has time-span (is time-span of): E52 Time-Span ...

For the Crimean document example, we must say:
Whole thing != E2 Temporal Entity

The document as such *is not an event*; it has no time. But it results from an event.

We know there are dates associated to it so it implies that we will have to find a way to connect the document to an event.

I still need more expressive power to describe the document as such, but I know I will have to find it another branch of the ontology.

E77 talking about objects, what lasts

E77 Persistent Item

Scope	This class comprises items that have a persistent identity, sometimes known as “endurants” in philosophy. They can be repeatedly recognized within the duration of their existence by identity criteria rather than by continuity or observation.
-------	---

Relations	-
-----------	---

For the Crimean document example, we can say:

Whole thing = E77 Persistent Item

- Whole thing = **E1 CRM Entity** but also **E77 Persistent Item**
- **E77 Persistent Item** is more expressive

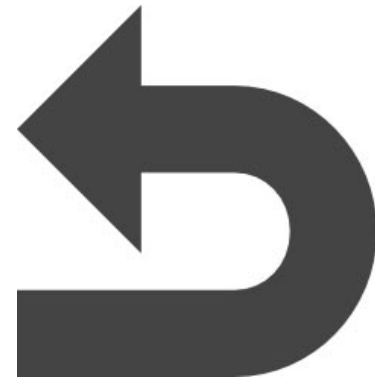
Finding a more particular class will give me more expressive power and allow me to express the other fields from the source.

How do I know when I have the right class?

Check Intension! Is the intension (scope) of the class I'm looking at in line with what I'm trying to describe?

Check relations! What does this class do? Does the thing it does cover the kinds of things I want to be able to talk about? If not, what's missing? Could it be elsewhere?

Go a little further! Does the class work but it feels a bit too generic? Maybe you can go further. Try a step further down until the intension and/or properties don't seem to fit.



3. MAPPING RECIPE

The Basic Mapping Recipe

1. Determine **for the whole or part of** a data structure, **what class describes it** in the target ontology. This is your Subject.
2. Determine **for each field** in the whole/part, **what class describes it** in the target ontology. This is your Object.
3. **Having understood the intended meaning of the field, select the relation or relations** that will allow you to link Subject and Object. This is your Verb.
4. Repeat

Tip!

- Mapping is NOT matching terms for terms
 - le.: I do not look for names of fields in the source and names of classes in the target and try to make their equivalence
- Mapping is NOT tagging
 - le.: we do need to find the right classes for mapping, but the semantics come in choosing the correct relation / series of relations to translate the data
- Mapping IS translating a data structure into formal propositions
 - le: my data will translate out into a triple structure that offers a pidgin version of natural language

4. MAPPING EXAMPLE

Crimea Conference

Historical Archives....

Subject

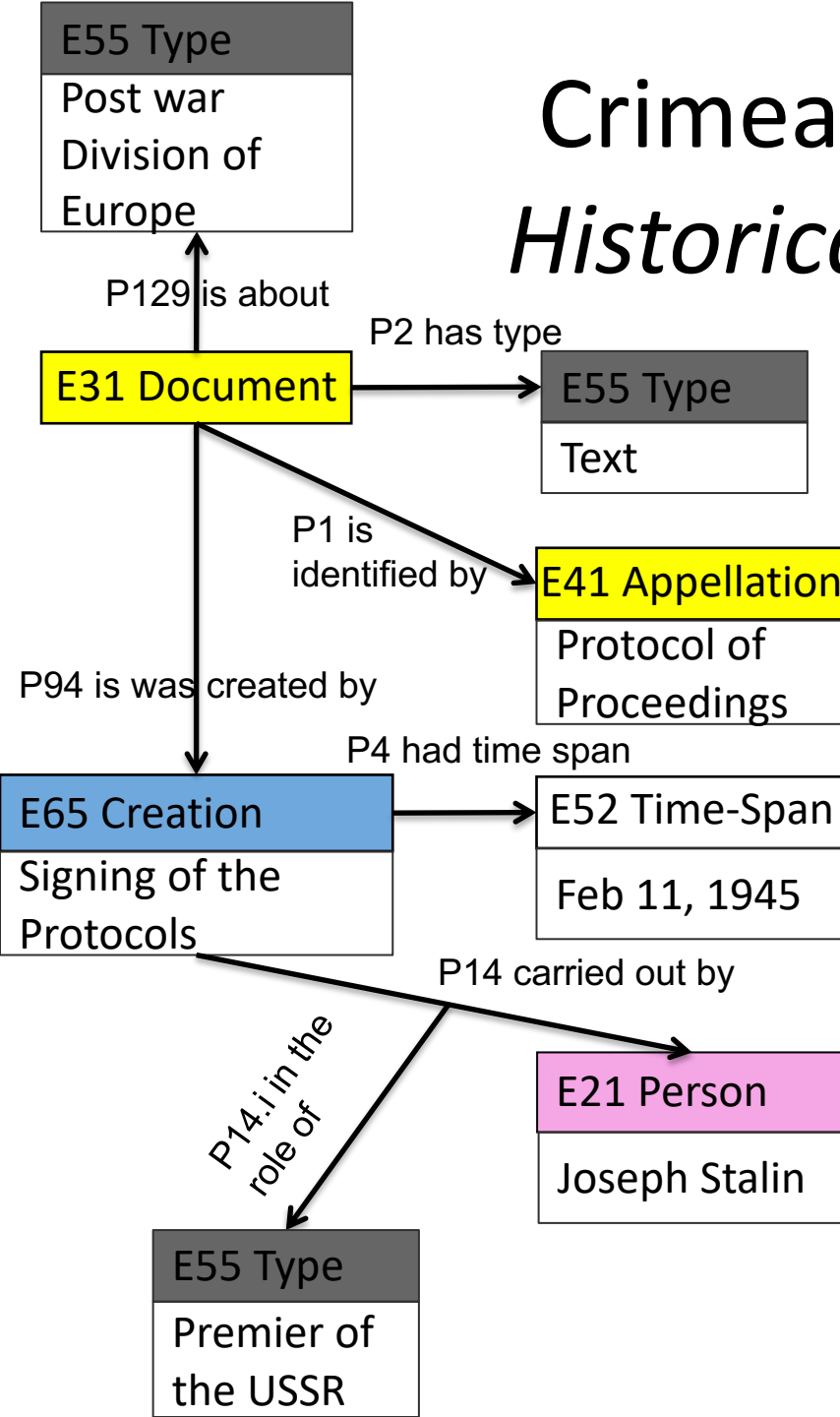
E31 Document

Object

Field	Value	Object
Type	Text	E55 Type
Title	Protocol of Proceedings of Crimea Conference	E41 Appellation
Subtitle	Declaration of Liberated Europe	E41 Appellation
Date	February 11, 1945	E52 Time-Span
Creator	<ul style="list-style-type: none">• The Premier of the Union of Soviet Socialist Republics• The Prime Minister of the United Kingdom• The President of the United States of America	E21 Person
Publisher	State Department	E40 Legal Body
Subject	Postwar division of Europe and Japan	E55 Type

Crimea Conference

Historical Archives....



Field	Value
Type	Text
Title	Protocol of Proceedings of Crimea Conference
Subtitle	Declaration of Liberated Europe
Date	February 11, 1945
Creator	<ul style="list-style-type: none"> The Premier of the Union of Soviet Socialist Republics The Prime Minister of the United Kingdom The President of the United States of America
Publisher	State Department
Subject	Postwar division of Europe and Japan

Crimea Conference

Historical Archives....mapped

Field	Value	CRM Translation = E31 Document
Type	Text	P2 has type E55 Type
Title	Protocol of Proceedings of Crimea Conference	p1 is identified by E41 Appellation
Subtitle	Declaration of Liberated Europe	p1 is identified by E41 Appellation
Date	February 11, 1945	P94i was created by E65 Creation p4 has time-span E52 Time Span
Creator	<ul style="list-style-type: none"> The Premier of the Union of Soviet Socialist Republics etc. 	P94i was created by E65 Creation p14 was carried out by E39 Actor
Publisher	State Department	P148i is component of E31 Document P94i was created by E65 Creation p14 was carried out by E39 Actor
Subject	Postwar division of Europe and Japan	P129 is about E55 Type

Time for mapping!

Questions:

George Bruseker

bruseker@ics.forth.gr

ICS-FORTH

<http://www.ics.forth.gr/>

CRM SIG

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

END