Overview Tuesday

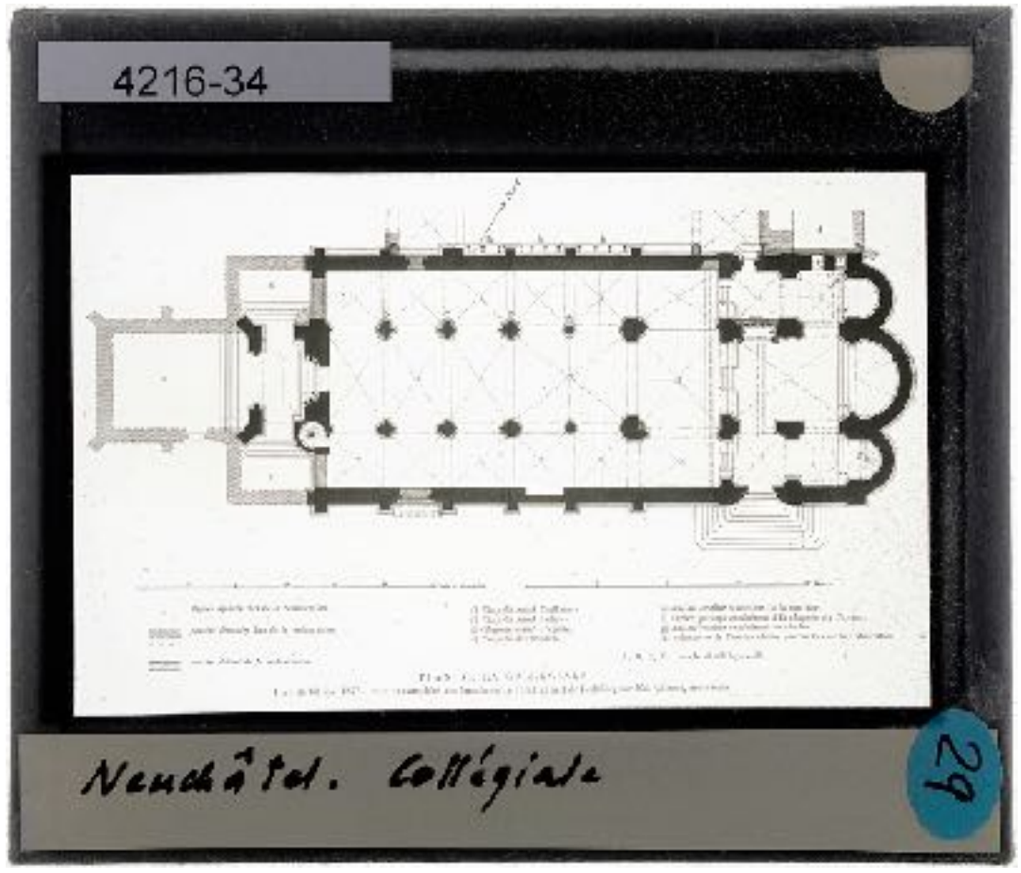
- Questions
- CIDOC-CRM
- Extensions
- Afternoon: Setup and modelling of own data

Data Model

- Write down the classes and properties for a historic slide archive


Archiv für Schweizerische Kunstgeschichte, Basel
Eigentum des kunstgeschichtlichen Seminars an der Universität Zürich


Carona : S. Maria. 10.
4207-25


4216-34
Neuchâtel. Collégiale
29


Eigentum des kunstgeschichtlichen Seminars an der Universität Zürich
Nr.

CIDOC-CRM

Introduction

A bit of history

- Until 1998 CIDOC existed as an Entity Relationship model, being derived from the technology of relational databases.

- Not being flexible enough, it meant supporting a highly complex system that is impossible to maintain.

- In 1996 CIDOC-CRM was born as a project to replace the E-R-model.

Relational Databases are not good at relations

Name	ID
Peter	1
Susanne	2
Hans	3
Julia	4

Relation	ID
Mother	1
Father	2
Daugther	3
Son	4

Name_ID	Relation_ID	Name_ID
1	2	3
2	1	3
4	3	1
3	4	1

A bit of history

- CRM is an object-oriented model, that allowed for new use-cases to be added on an ongoing basis.

- It was also conceived to be database technology agnostic.

- The primary objective of the CRM initiative was to allow exchange and sharing of information.

A bit of history

- Coming from a Computer Science background, automated reasoning was a big thing for the CRM creators. They define it as: „the ability to formally manipulate the data using logical rules in order to generate new information“.

- CRM clearly followed political topics: „In contributing to this resource of information institutions become important members of a revolutionary digital research community.“

- Or: „The value and relevance of data increases when it is communicated with its full meaning and context. This relevance is magnified when the knowledge of different institutions is combined to enable different perspectives to be preserved.“

Definition - general

- The CIDOC CRM is an ontology - a form of knowledge representation. An ontology represents the categorical knowledge within a domain, in this case the cultural heritage domain. The function of a domain ontology is to mediate the variability within a domain and provide a framework under which we can collaborate despite having different datasets. It is a language, not a statement of current scholarly convictions.

Definition - technology

- It is independent of any technical implementation framework. It is commonly employed using Resource Description Framework (RDF) databases, the lingua franca of linked data, but could also be used with other meta-models. Different technologies create a different set of constraints. The design of a knowledge representation system should not be based, or dependent upon, a particular technology. It should represent knowledge in a more generic form. Its only logical restriction is the kind of positive statements information systems can support so far.

Definition - no fields or values

- It does not mandate any fields or values. Unlike other standards that work by using an agreed set of fields and/or values the CRM supports variability. The reason why there are so many field/value based standards is because different cultural groups will naturally have different requirements. The CRM provides a semantic framework that describes more general entities (including events) and the relationships between them.

Definition - bottom up

- It is an empirically based ontology. Rather than being defined by a committee (top down), the CRM is based on empirical analysis of real practice and local knowledge (bottom up). The CRM develops as a result of understanding existing models of practice that have themselves developed over a considerable period of time; it represents nearly twenty years of international research. It is unlikely that a similar exercise would come up with a significantly different result. It is scientifically constituted and not influenced by the strength of opinion of a particular group or expert.

Definition - poly-hierarchical

- It is poly-hierarchical (not a flat linear structure) providing an optimal range of generalisation/specialisation above the point of individual institutional terminological descriptions. In such a framework context and semantics become important.

Definition - terminology alignment

- It does not concern itself with differences in terminology between institutions, it supports the ability to “plugin” local terminologies and provides an ontological framework under which these vocabularies (conceptual terminology) can be compared and linked.

Definition - automation

- It provides a framework for matching instances of people, places, things, events and periods using the information and context around these entities. It does not need to rely on primitive string matching techniques.

Definition - reasoning

- It has the ability to support rich computer-based reasoning. The ontology is based on the concept of object-oriented classes with carefully designed relationships that conform to rules of logic. The CRM provides the opportunity for a computer to infer new information by putting together fragments of information (semantically harmonised) from different sources and creating the conditions in which logical propositions can be concluded.

- The most important kinds of computer-based reasoning the CRM can support are generalisations of relationships and deductions from highly indirect relations such as what parts have in common with their wholes, what wholes inherit from their parts and what is transferred across meetings and processes of derivation. These are not meant to replace scholarly conclusions but to comprehensively detect facts relevant to answer research questions. Besides others this ensures that highly specialized knowledge stays accessible to generic questions regardless the specificity of representation.

Mapping

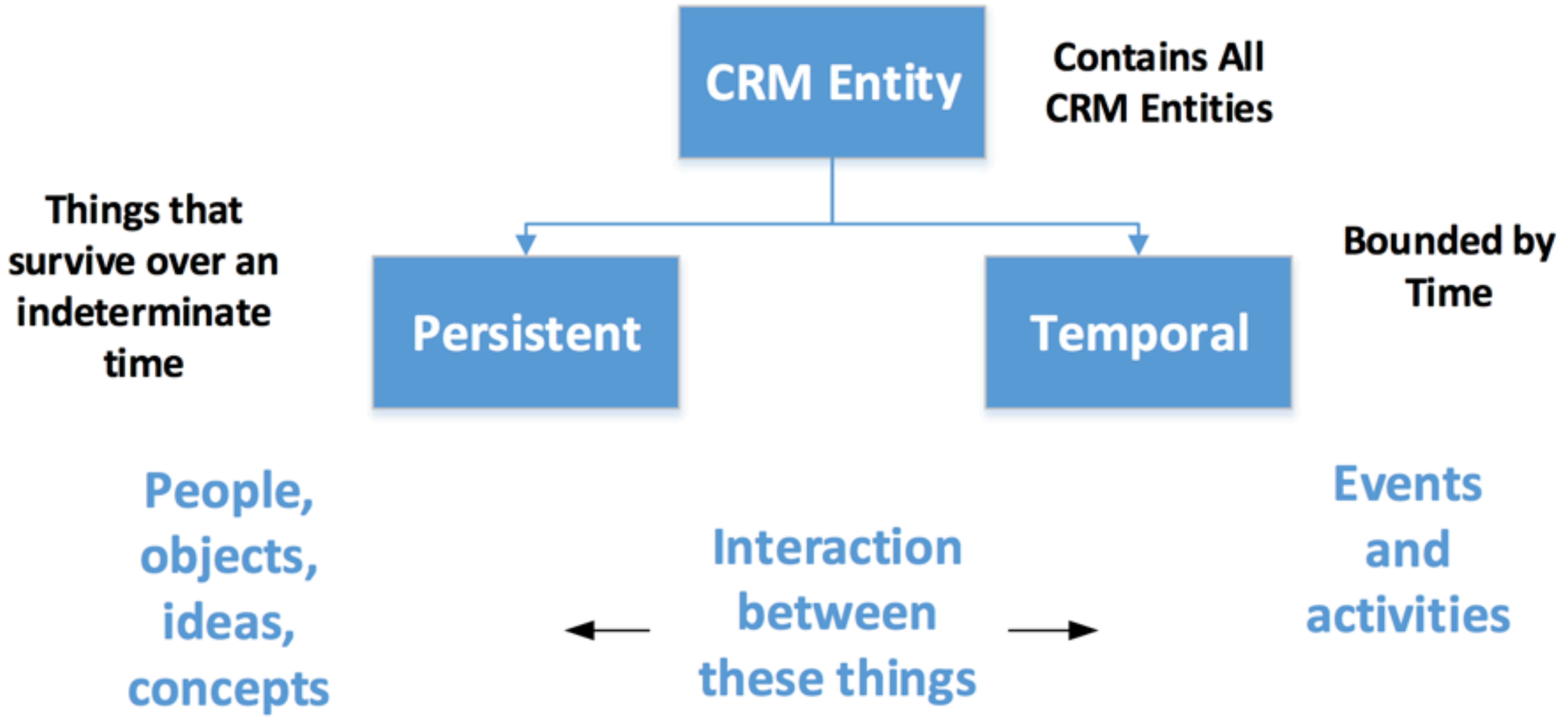
- CRM acknowledges that somebody has to do the mapping of the data. Because if the source data has only implicit semantics, then a human (read domain expert) is needed to express this information explicitly.

E and P labels

- Entity Types/Classes = E

- Relationships/Properties = P

Persistent and Temporary Things



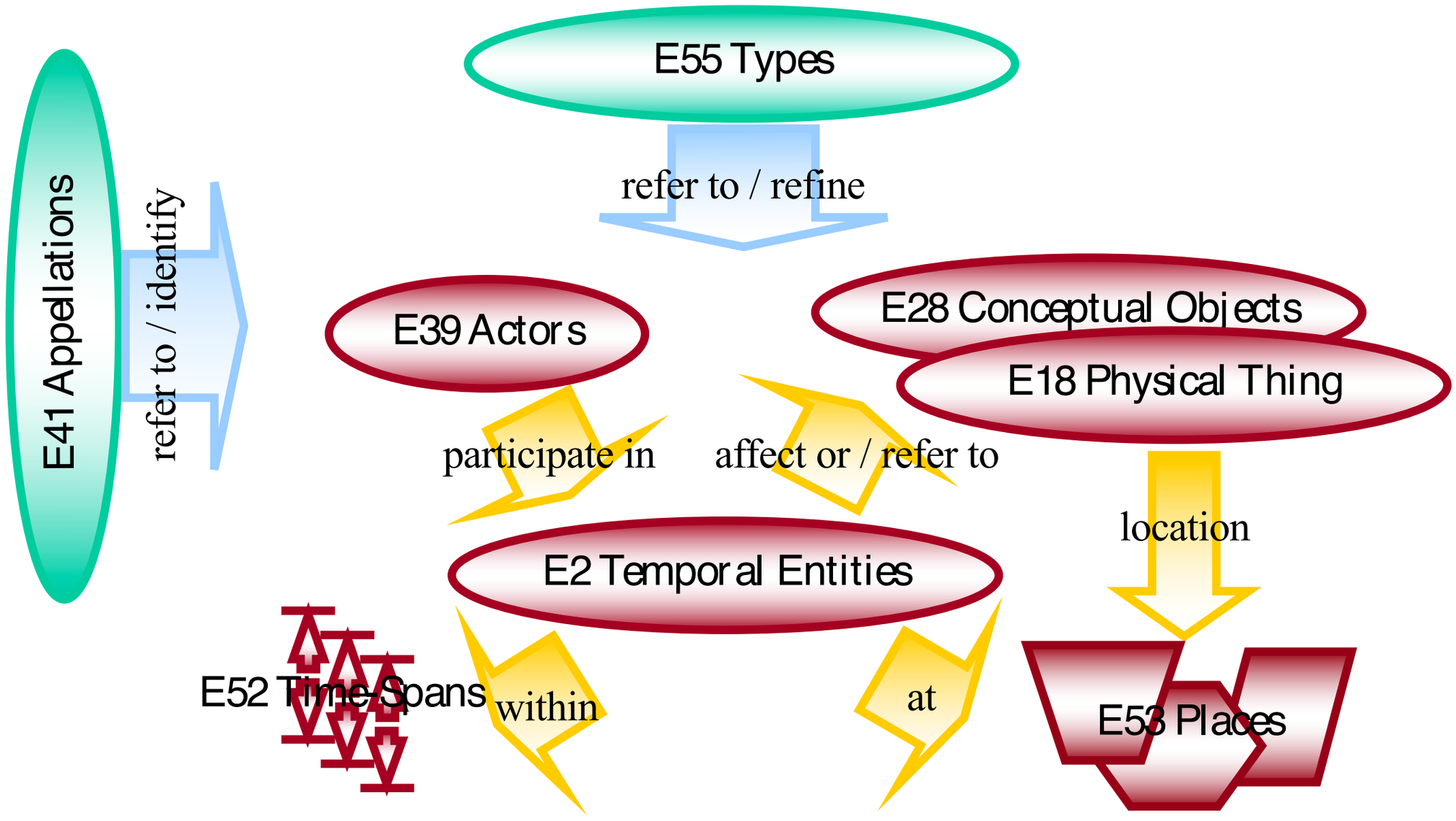
Persistent and Temporary Things

Identity 

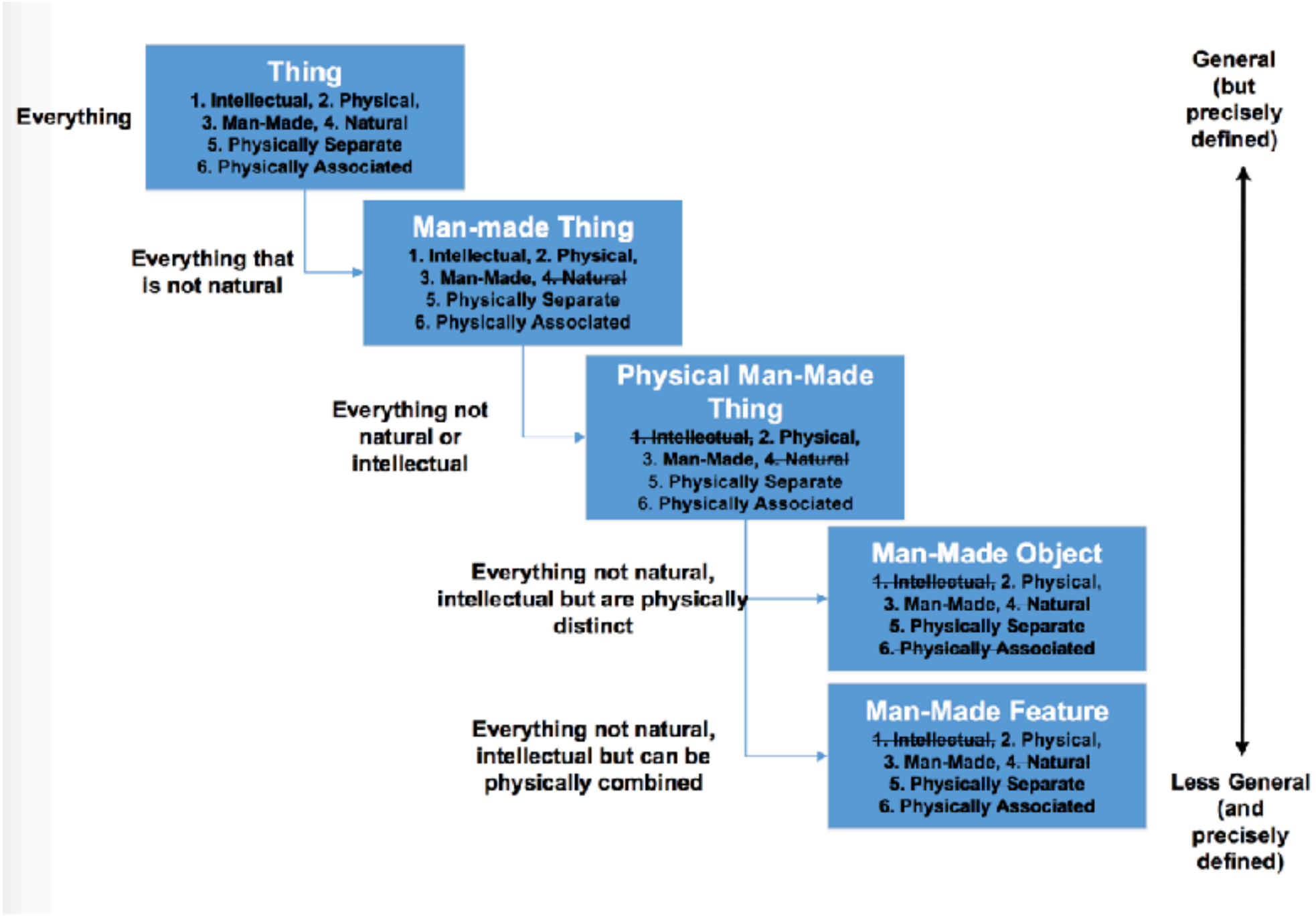
Physical life 

The CIDOC CRM

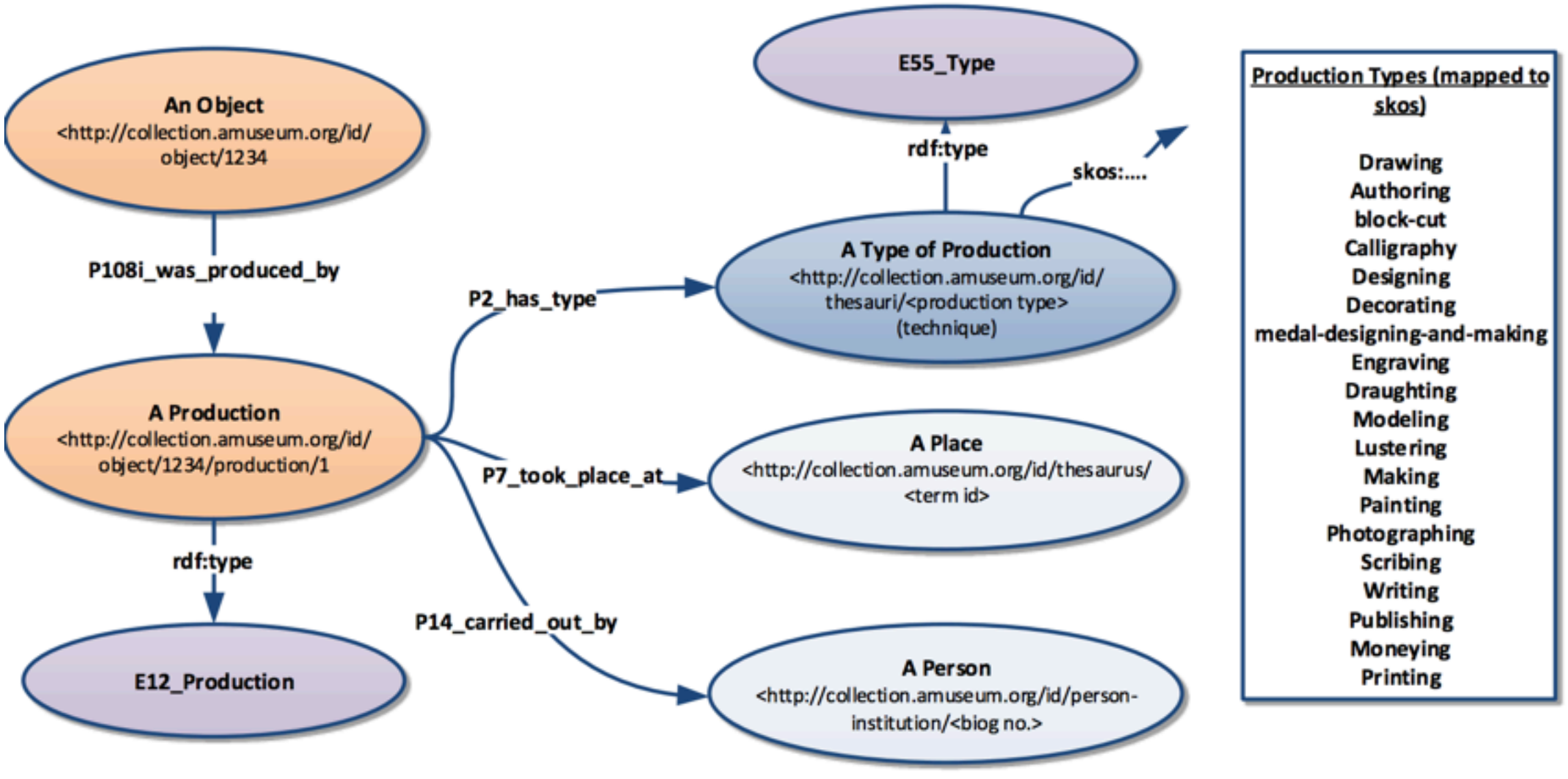
Top-level classes useful for integration



Poly-hierarchical



Roles


An Object
<http://collection.amuseum.org/id/object/1234
P108i_was_produced_by
A Production
<http://collection.amuseum.org/id/object/1234/production/1
rdf:type
E12_Production
P2_has_type
P7_took_place_at
P14_carried_out_by
E55_Type
rdf:type
skos:....
A Type of Production
<http://collection.amuseum.org/id/thesauri/<production type>
(technique)
A Place
<http://collection.amuseum.org/id/thesaurus/<term id>
A Person
<http://collection.amuseum.org/id/person-institution/<biog no.>
Production Types (mapped to skos)
Drawing
Authoring
block-cut
Calligraphy
Designing
Decorating
medal-designing-and-making
Engraving
Draughting
Modeling
Lustering
Making
Painting
Photographing
Scribing
Writing
Publishing
Moneying
Printing

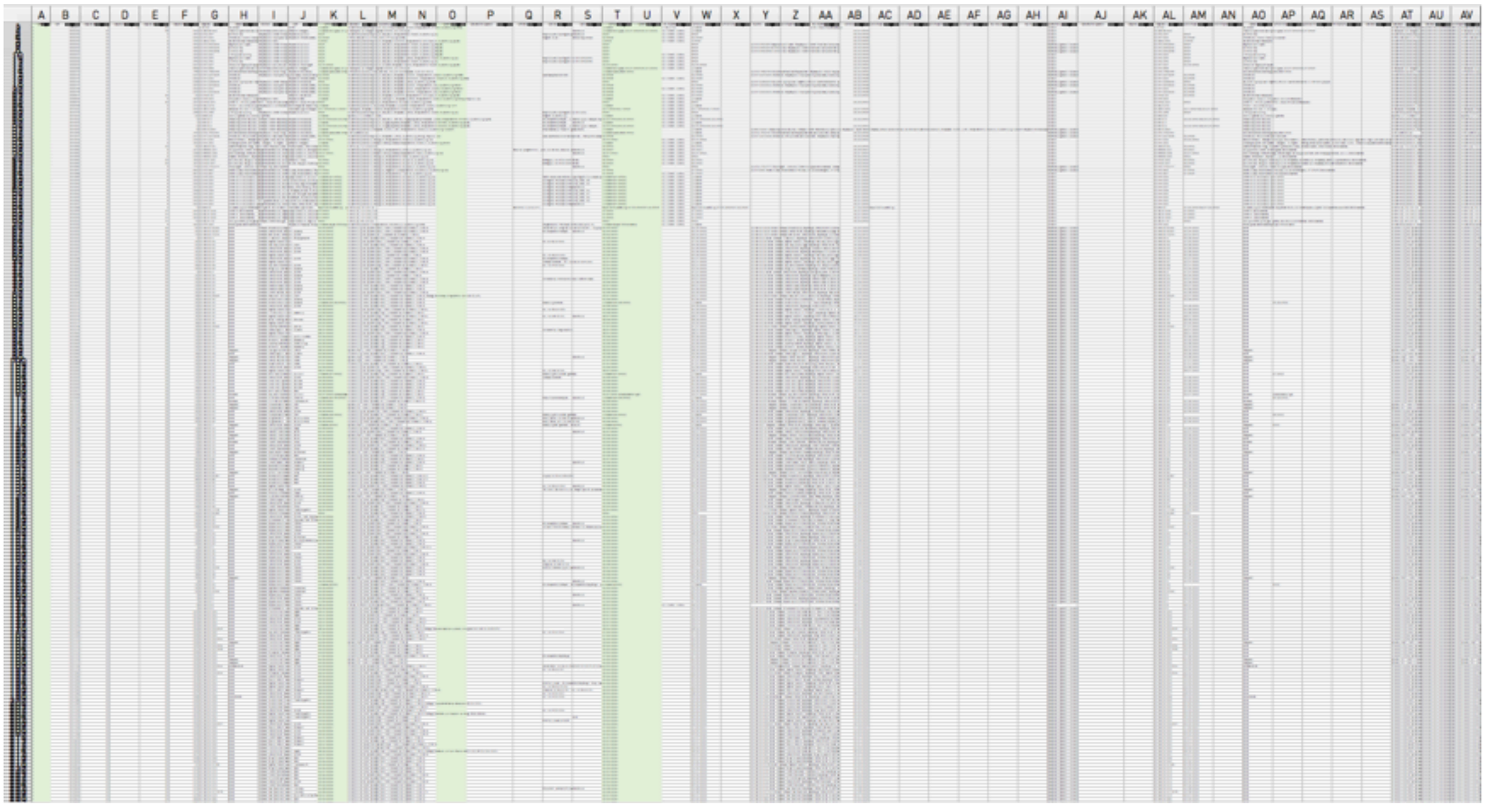
What data are we working with

- Manual, meaning human, data collection
- Data collection sometimes spread over several decades
- Data has had previous migrations and additional imports from external sources
- People entering data know more than the data shows

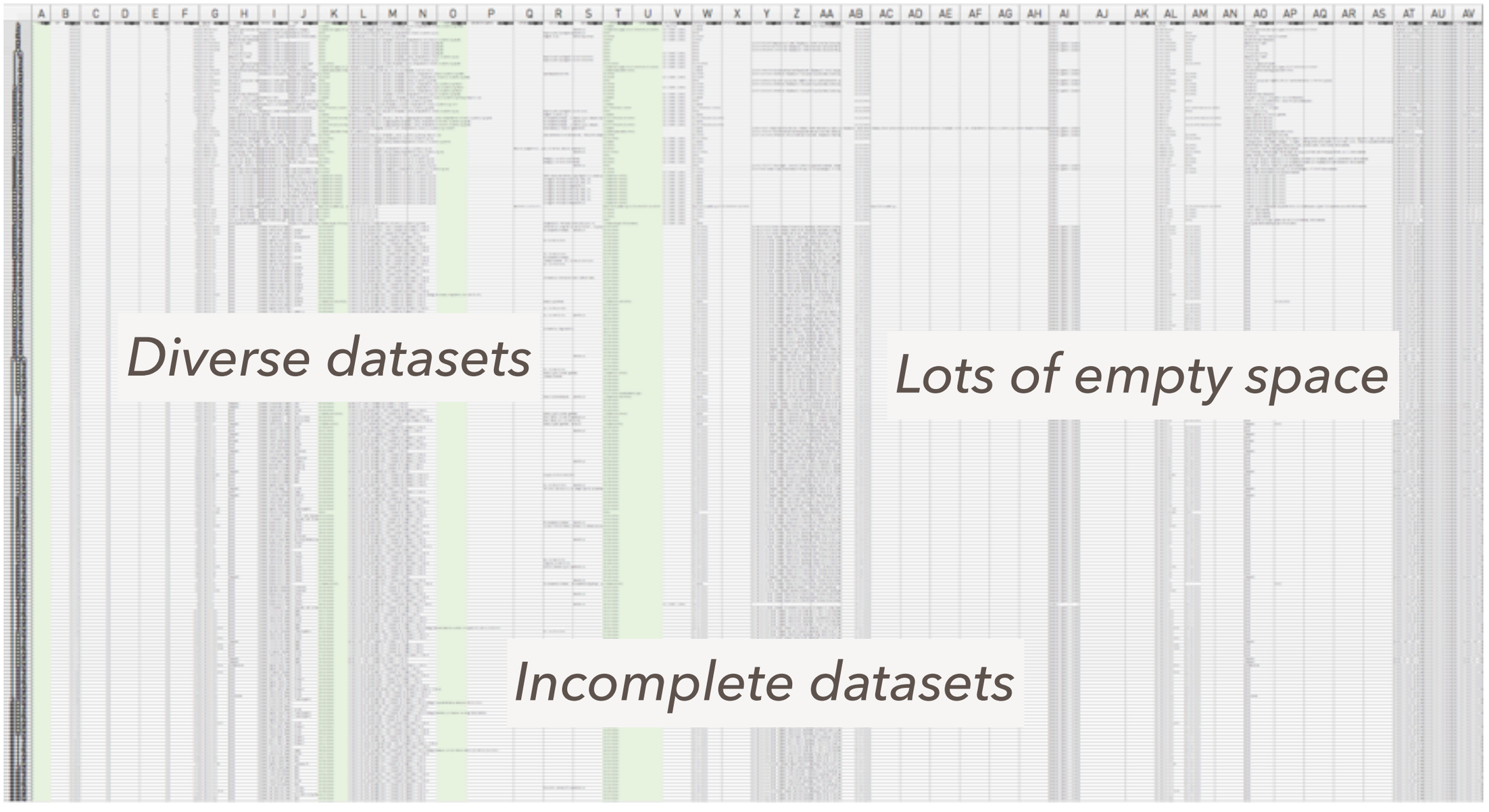
What data are we working with

- Manual, meaning human, data collection
  - **Different backgrounds, mistakes, bad days**
- Data collection sometimes spread over several decades
  - **Different computer literacy, changing leadership**
- Data has had previous migrations and additional imports from external sources
  - **Already messy data is becoming even messier**
- People entering data know more than the data shows
  - **Limitations due to data formats, user interface, lack of support**

Here is how it looks like



Here is how it looks like



Here is how it looks like

	A	B	C	D	E	F	G	H	I	J
998	997	Aesch / Maur/ZH			Aesch / Maur/ZH		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
999	998	Aesch bei Maur/ZH			Aesch bei Maur/ZH		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1000	999	Aesch/BL			Aesch/BL		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1001	1000	Aesch/LU			Aesch/LU		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1002	1001	Aesch/ZH			Aesch/ZH		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1003	1002	Aeschengraben / St. Jakobstrasse, Basel			Basel, Aeschengraben / St. Jakobstrasse		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1004	1003	Aeschengraben 9, Basel			Basel, Aeschengraben 9		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1005	1004	Aeschengraben, Basel			Basel, Aeschengraben		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1006	1005	Aeschengraben/Nauenstrasse, Basel			Basel, Aeschengraben/Nauenstrasse		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1007	1006	Aeschenplatz/St. Albananlage, Basel			Basel, Aeschenplatz/St. Albananlage		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1008	1007	Äussere Baselstrasse 170, Riehen/BS			Riehen/BS, Äussere Baselstrasse 170		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1009	1008	Äussere Ringstrasse 36, Thun/BE			Thun/BE, Äussere Ringstrasse 36		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1010	1009	aussere Seefeldstrasse, Zürich			zürich, aussere Seefeldstrasse		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1011	1010	Äusserer Damm, Oldenburg			Oldenburg, Äusserer Damm		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1012	1011	Affoltern am Albis/ZH			Affoltern am Albis/ZH		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1013	1012	Affoltern im Emmental/BE			Affoltern im Emmental/BE		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03

Here is how it looks like

	A	B	C	D	E	F	G	H	I	J
998	997	Aesch / Maur/ZH			Aesch / Maur/ZH		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
999	998	Aesch bei Maur/ZH			Aesch bei Maur/ZH		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1000	999	Aesch/BL			Aesch/BL		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1001	1000	Aesch/LU			Aesch/LU		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1002	1001	Aesch/ZH			Aesch/ZH		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1003	1002	Aeschengraben / St. Jakobstrasse, Basel			Basel, Aeschengraben / St. Jakobstrasse		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1004	1003	Aeschengraben 9, Basel			Basel, Aeschengraben 9		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1005	1004	Aeschengraben, Basel			Basel, Aeschengraben		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1006	1005	Aeschengraben/Nauenstrasse, Basel			Basel, Aeschengraben/Nauenstrasse		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1007	1006	Aeschenplatz/St. Albananlage, Basel			Basel, Aeschenplatz/St. Albananlage		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1008	1007	Äussere Baselstrasse 170, Riehen/BS			Riehen/BS, Äussere Baselstrasse 170		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1009	1008	Äussere Ringstrasse 36, Thun/BE			Thun/BE, Äussere Ringstrasse 36		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1010	1009	aussere Seefeldstrasse, Zürich			zürich, aussere Seefeldstrasse		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1011	1010	Äusserer Damm, Oldenburg			Oldenburg, Äusserer Damm		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1012	1011	Affoltern am Albis/ZH			Affoltern am Albis/ZH		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03
1013	1012	Affoltern im Emmental/BE			Affoltern im Emmental/BE		Alex Winiger	02.08.2017 15:31:03	Alex Winiger	02.08.2017 15:31:03

Typos

Vague statements

Conflicting entity types

Overloaded data fields

Data Reconciliation

	A	B	C	D	E	F	G	H
22	Aesch	Aesch	BL	Aesch/BL		https://www.wikidata.org/wiki/Q68367		
23	Aesch (Kanton Luzern)	Aesch	LU	Aesch/LU			Aesch (Kanton Luzern)	http://viaf.org/viaf/239214984
24	Aesch	Aesch	ZH	Aesch/ZH			Aesch	http://viaf.org/viaf/309920087
25	Aesch	Aesch	Maur	Aesch/Maur			Aesch	http://viaf.org/viaf/309920087
26	Aesch bei Birmensdorf	Aesch bei Birmensdorf	ZH	Aesch bei Birmensdorf/ZH			Aesch bei Birmensdorf	http://viaf.org/viaf/249378701
27	Aesch bei Maur	Aesch bei Maur	ZH	Aesch bei Maur/ZH				
28	Aesch-Neftenbach (Swit	Aesch bei Neftenbach		Aesch bei Neftenbach			Aesch-Neftenbach (Switzerland)	http://viaf.org/viaf/149253217
29	Aeugst am Albis	Aeugst	ZH	Aeugst/ZH			Aeugst am Albis	http://viaf.org/viaf/245841727
30	Aeugstertal	Aeugstertal	ZH	Aeugstertal/ZH			Aeugstertal	http://viaf.org/viaf/235220475
31	Affoltern am Albis	Affoltern am Albis	ZH	Affoltern am Albis/ZH			Affoltern am Albis	http://viaf.org/viaf/241446768
32	Affoltern im Emmental	Affoltern im Emmental	BE	Affoltern im Emmental/BE			Affoltern im Emmental	http://viaf.org/viaf/246284319
33	Afghanistan	Afghanistan		Afghanistan			Afghanistan	http://viaf.org/viaf/138292133
34	Afikim (Israel)	Afikim	IL	Afikim/IL			Afikim (Israel)	http://viaf.org/viaf/133774489
35	Afrika	Afrika		Afrika		https://www.wikidata.org/wiki/Q19834170		

Extracting other entities

Identifying correct entities

Special characters require UTF-8

Returning unclear places

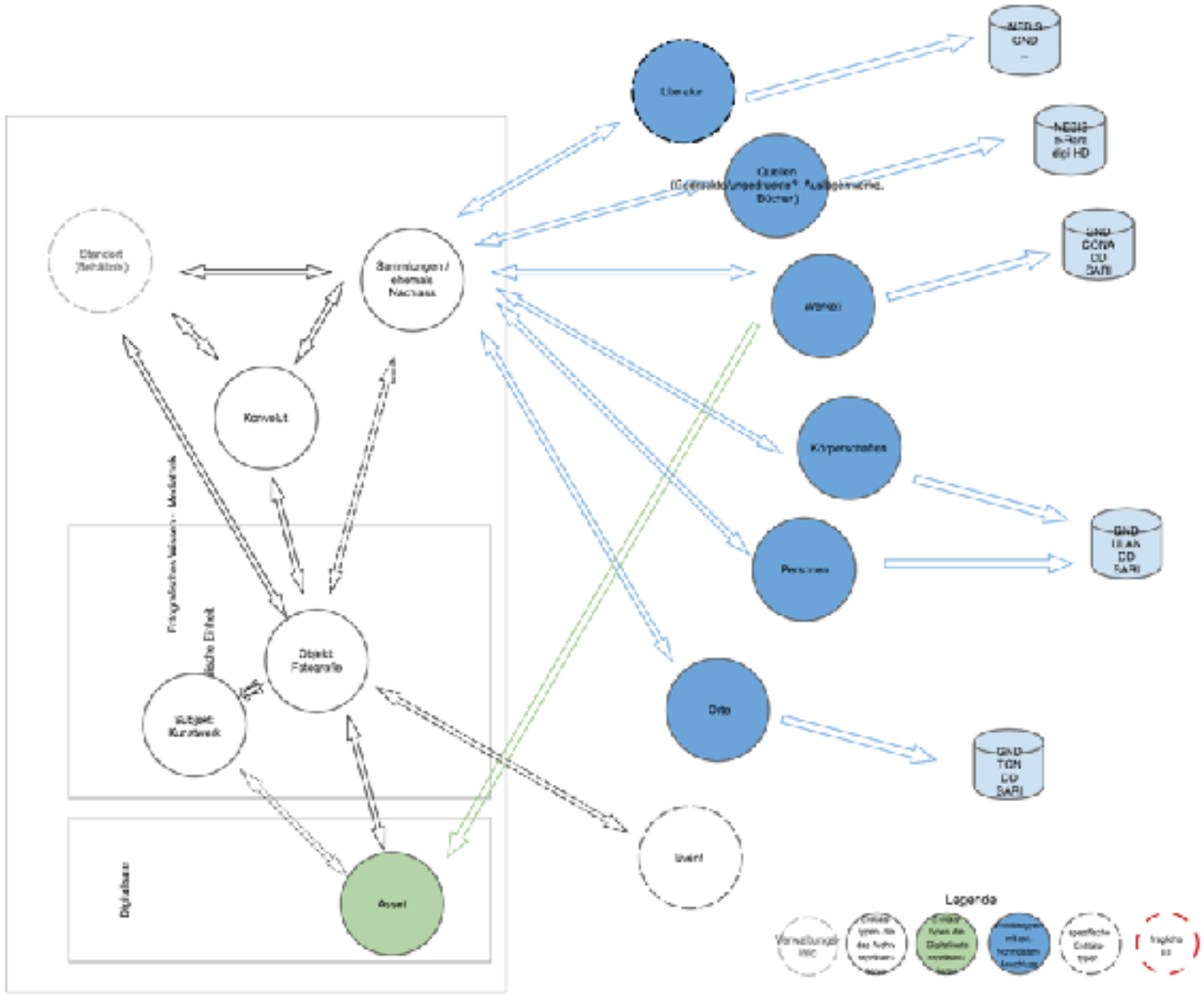
Your data model

- What kind of classes did you write down? And what kind of properties would you define?

- How did these definitions came about?

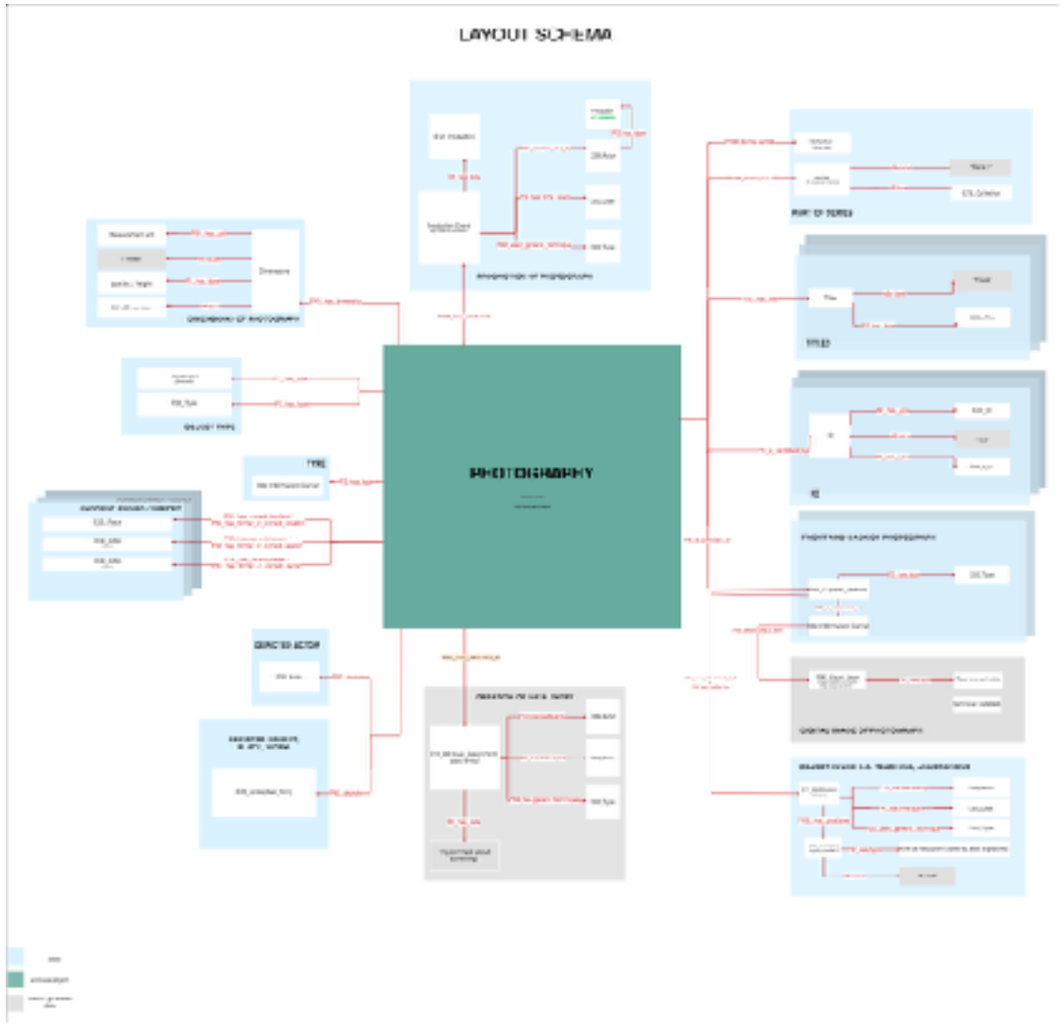
Result of First iteration

Clear focus on the workflow



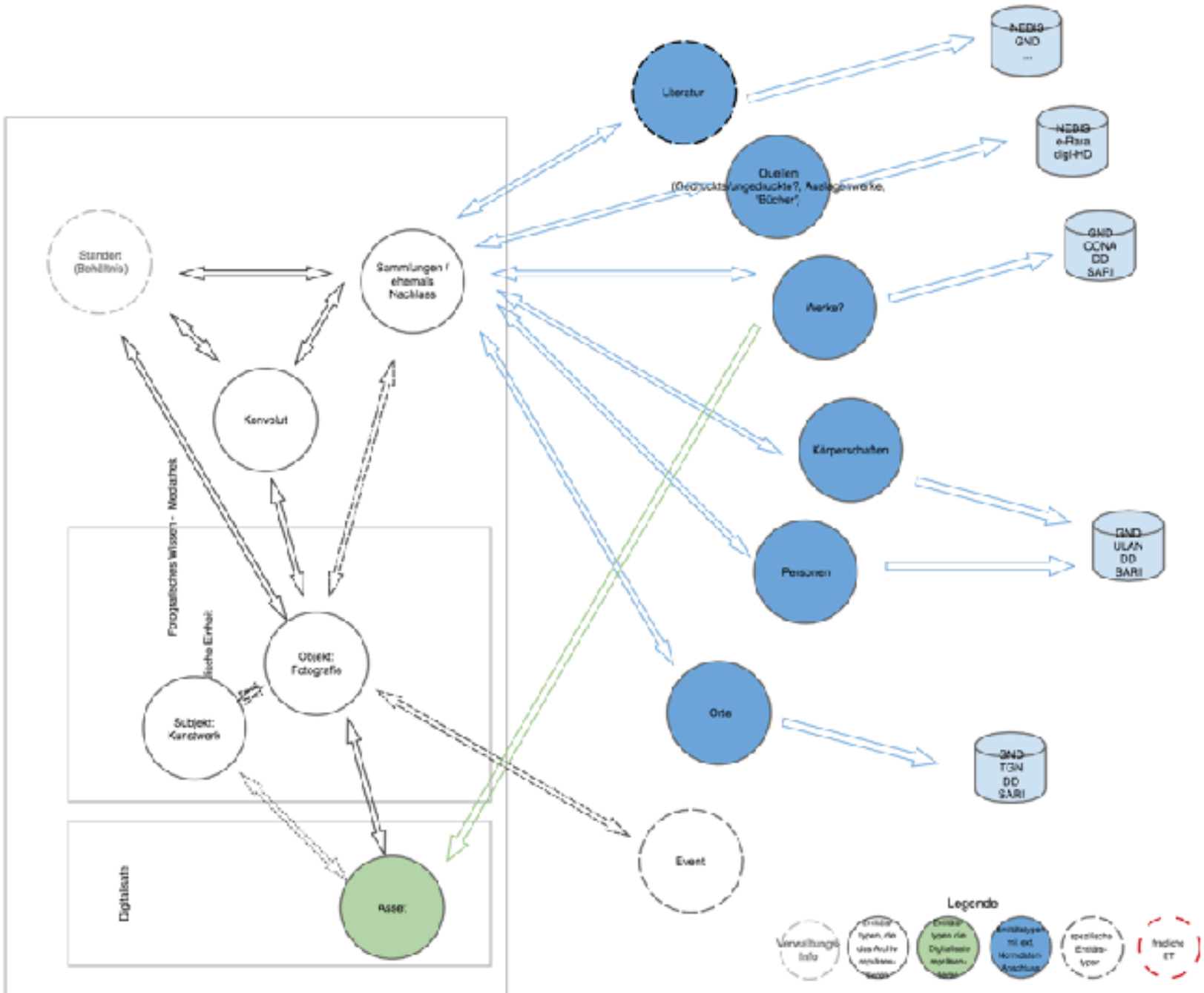
Current iteration

Reusing standards

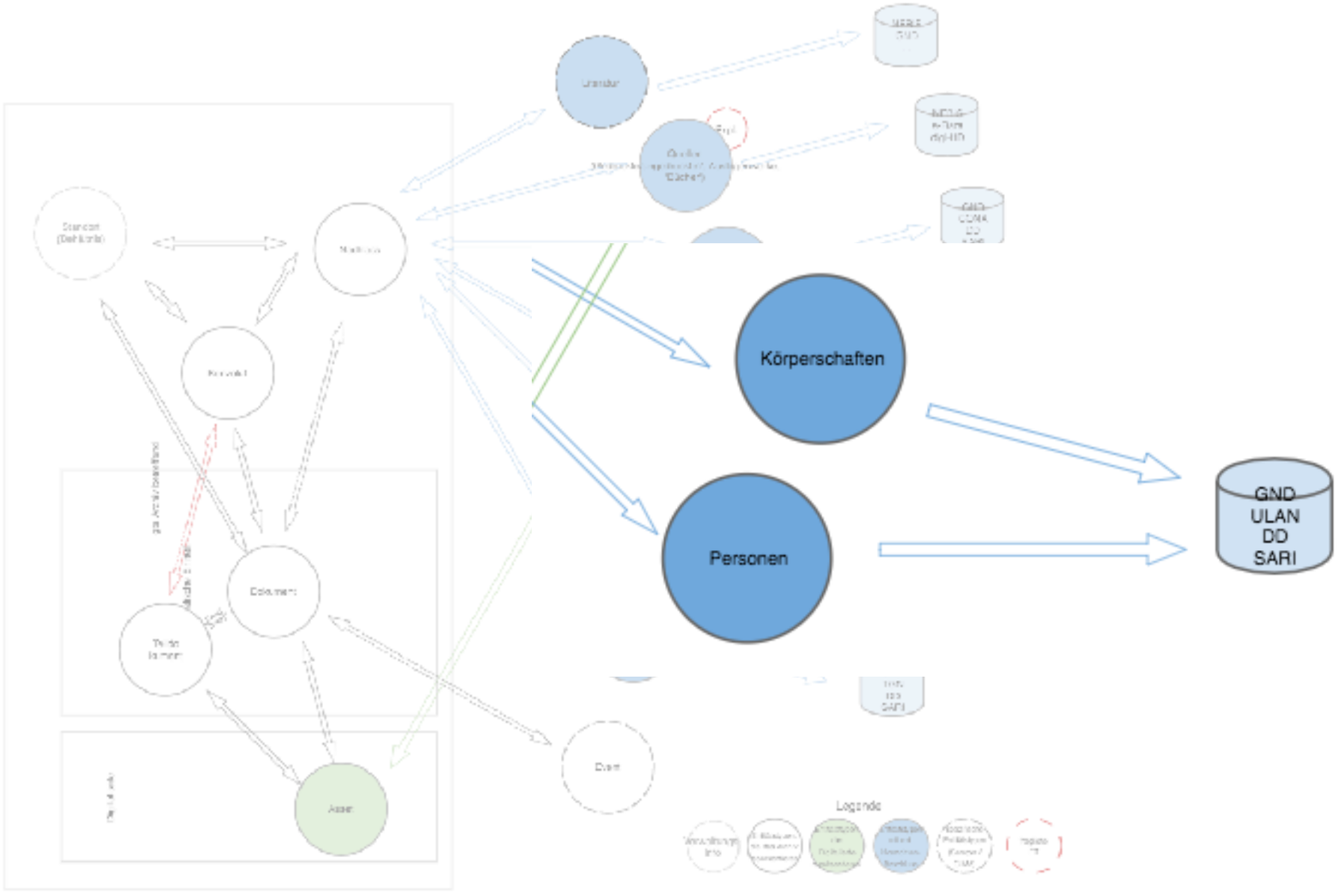


...

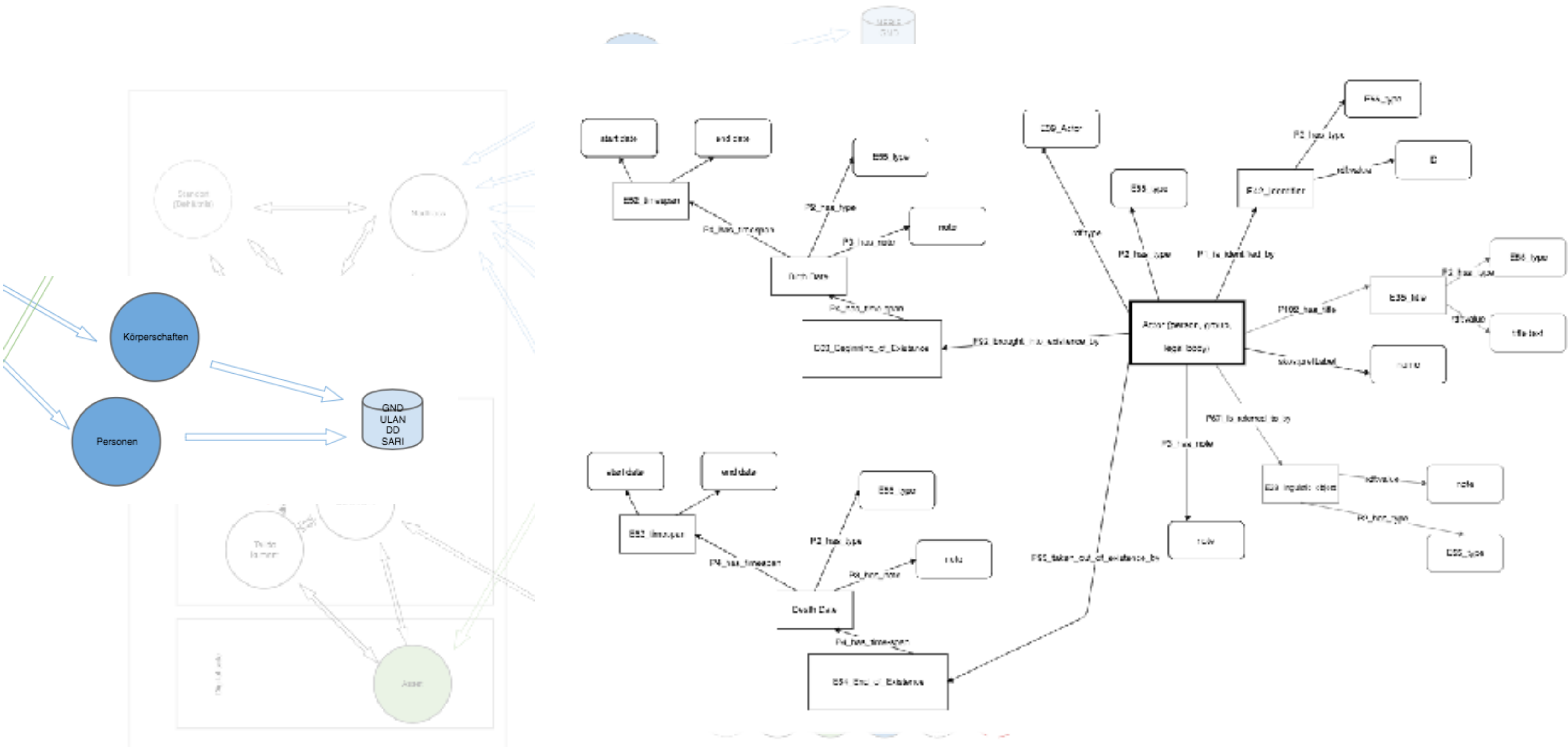
Example


Literatur
Quellen
(Geschäftsungedruckte?, Nachlassakten, 'Bücher')
Sammlungen / ehemals Nachlass
Konvolut
Fragestellung/Wissen - Medium/Inhalt
Objekt: Fotografie
Subjekt: Kartenwerk
Digitalisat
Asset
Werke?
Körperschaften
Personen
Orte
Event
Legende

Example



Example


Körperschaften
Personen
GND
ULAN
DD
SARI
start date
end date
E55_type
note
Birth Date
Death Date
E39_Actor
Actor (person, group,
legal body)
E35_title
title text
name
E64_End_of_Existence