# Requirements for Transdisciplinary Use of Data

Metadata Standards, Authorities, Policies

RDF, OWL, SPARQL

Unique and Persistent Identifiers

Ontologies

# Requirements for Transdisciplinary Use of Data

Metadata Standards, Authorities, Policies

# Metadata Standards

- Why should Metadata Standards be used?
  - Easy data exchange
  - Data is stored in the same structure
  - Help to avoid logical mistakes in documentation
  - Existing tools can be used
  - Provide persistent high quality documentation
  - Standards are persistent themselves

# Metadata Standards & Formats

- There are many standards for differing purposes and domains
  - Dublin Core (DC)
    - Initially used to describe resources (pages) on the web
  - Machine-Readable Cataloging (MARC)
    - Format for bibliographic data and related information
  - Lightweight Information Describing Objects (LIDO)
    - Harvesting format for museum data
  - Exchangeable Image File Format (EXIF)
    - Embedded Metadata in digital images

# Authorities

- Defined set of names or terms / a vocabulary for a domain of discourse
- Common vocabulary for consistent descriptions
- Example: Getty Vocabularies
  - Thesaurus of Geographic Names
  - Art and Architecture Thesaurus
  - Union List of Artist Names

# Getty Thesaurus of Geographic Names® Online
# Full Record Display



Click the icon to view the hierarchy.

**ID: 7004269** **Record Type: administrative**

**Sibiu (inhabited place)**

*Coordinates:*
Lat: 45 46 00 N *degrees minutes* Lat: 45.7667 *decimal degrees*
Long: 024 09 00 E *degrees minutes* Long: 24.1500 *decimal degrees*

**Note:** Located on Cibin river, on Turnu Roşu pass through the Transylvanian Alps; was site of Roman colony; town was founded by German colonists from Saxony in 12th cen.; destroyed by Tartars in 1241; suffered in Turkish wars, & went to Austria in 1699.

**Names:**

**Sibiu** (**preferred**,C,V)
**Hermannstadt** (C,O)
**Nagyszeben** (C,O)
**Hermannsdorf** (H,O) ..... name used by Saxon colonists in 12th cen.
**Cibinium** (H,O) ..... Roman, name refers to location on Cibin river

**Hierarchical Position:**

World (facet)
.... Europe (continent) (P)
..... Romania (nation) (P)
..... Sibiu (county) (P)
..... Sibiu (inhabited place) (P)

# Policies

- A policy is a guideline describing
  - What kind of data may be entered and what kind of data shall not be entered
  - How data shall be entered
  - Where data shall be entered
  - How data should look like
  - The semantics of the data
  - If data is valid and why certain data is invalid

# Example: Policy

- A policy could describe the following:
  - Stick to writing and formatting conventions
  - Use certain standards
  - Use certain authorities
  - Use fields for input of certain data
- Some example policies for museums
  - http://cco.vrafoundation.org/index.php/toolkit/cco_pdf_version/
  - http://www.d-nb.de/standardisierung/regelwerke/afs_dok_reg_entw.htm
  - http://www.collectionstrust.org.uk/index.cfm/collection-management/spectrum

# Why use policies?

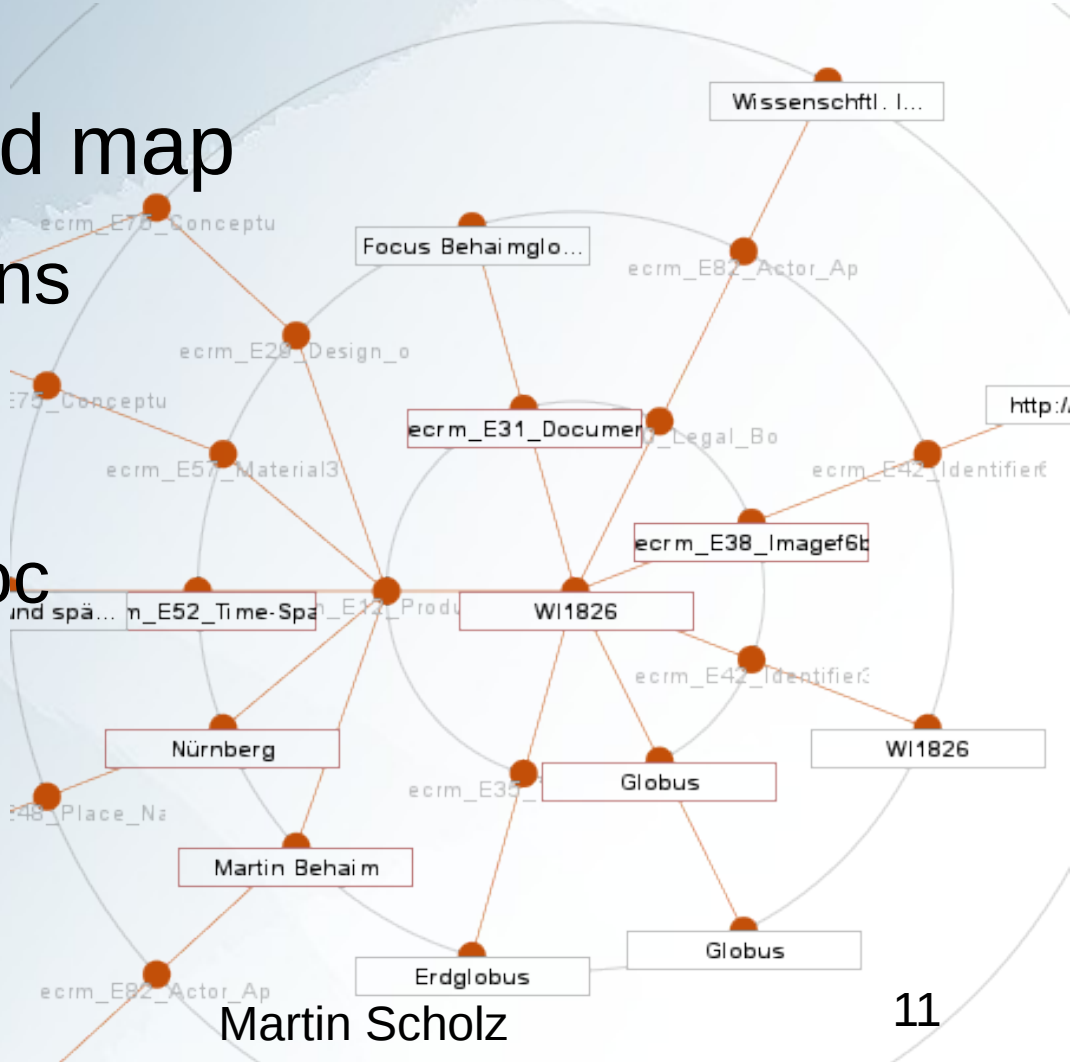
- They help you to
  - Be more consistent
  - Organize your data
  - Manage your data
  - Store the right things with the right meaning
- And therefore they help you to create high quality data for long term use

# Requirements for Transdisciplinary Use of Data

RDF, OWL, SPARQL

# Representing data: RDF

- RDF is a framework for representing metadata
- Think of RDF as a mind map
  - Connections/relations between „things“ (web resources)
  - Establish links adhoc
  - Classify things



# Representing data: OWL

- OWL = RDF + X !
- OWL adds structure and semantics
  - Express complex classification rules
  - Conditions for class membership

# What RDF & OWL provide

- Flexible data schema
- Designed for data exchange
- Open data formats and a lot of (free) tools
- Sound inference rules
  - consistency checks
  - find implicit knowledge
  - helps in drawing conclusions

# SPARQL – Searching & Querying RDF/OWL

- SPARQL is a query language for RDF/OWL databases
- Looks a bit like SQL
- Modes
  - Yes/no questions
  - Get all data describing an object
  - Build tabular respresentation (cf. field search)
  - Build network / mind map representation

```
SELECT * WHERE {
  GRAPH ?g {
    ?s ?p ?o .
    ?s a crm:E1_CRM_Entity .
  }
}
LIMIT 10
```

# Requirements for Transdisciplinary Use of Data

Unique and Persistent Identifiers

# Unique and Persistent Identifiers

- Why do we need identifiers?
  - To refer to things
- Why unique?
  - To be sure to talk about the same thing
- Why persistent?
  - Usable for a long time (till eternity?)
  - Meaning should not change over time

# Example: Identifiers

- Locally unique identifier
  - Inventory number of an object in a museum
  - Your room number in your hotel
- Globally unique identifier
  - Address
  - URL in the internet

# Unique and Persistent Identifiers

- Using unique and persistent identifiers enable
  - Talk about the same things in different disciplines
  - Easy aggregation of data from diverse sources

May have any form, in principle, but the internet already has unique identifiers we can use: URIs

# Uniform Resource Identifier (URI)

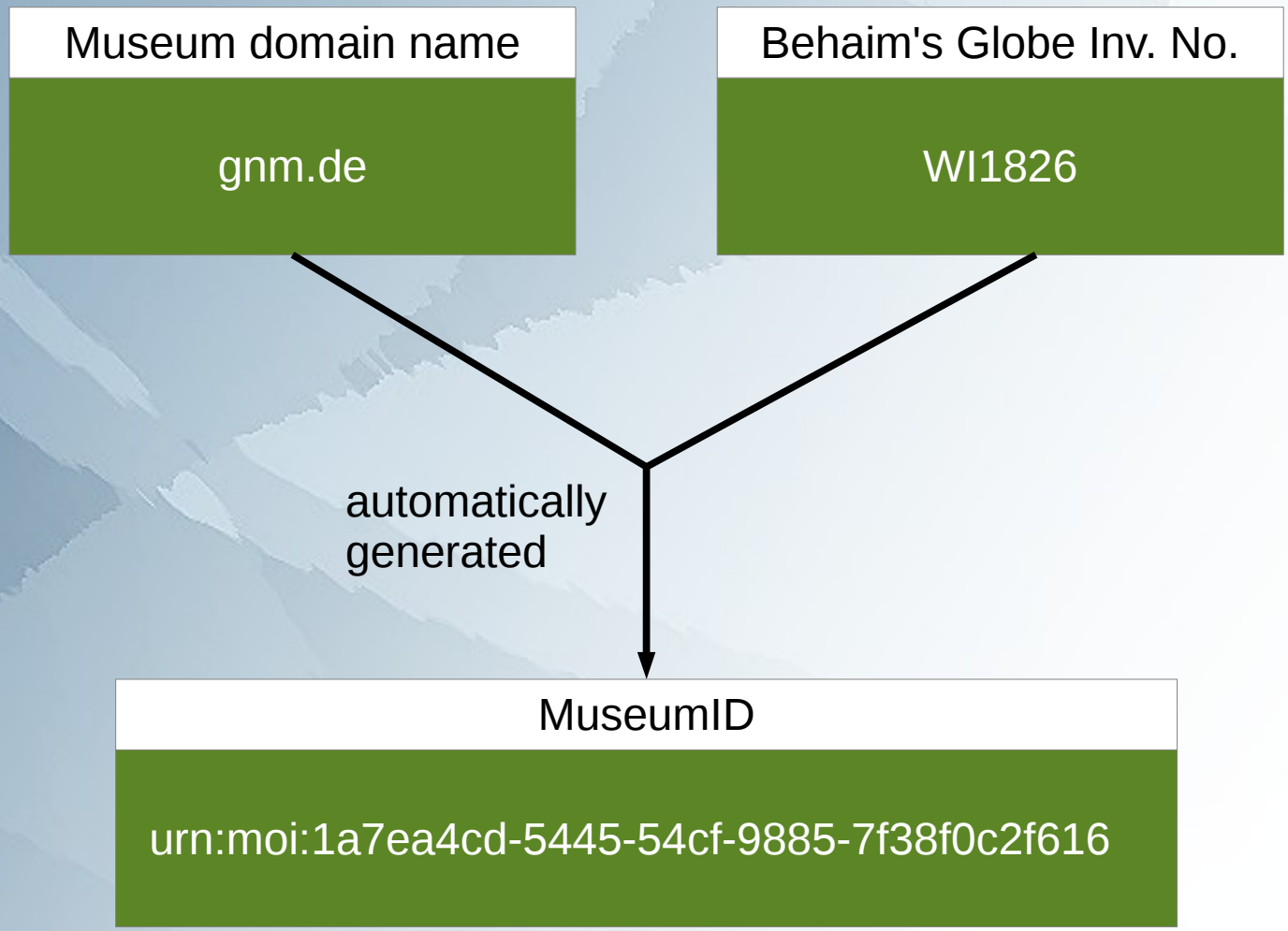
- URIs are used to denote things in RDF
- A set of identifier schemes
    - URL, e-mail address, URN, DOI, ...
- Most URI schemes use namespaces
    - low risk of collision
    - eg. domain name
- Example of a URI

http://erlangen-crm.org/110404/E1_CRM_Entity

# MuseumID

- A way of generating URIs for museum objects
  - A type of URN
  - Created from
    - Museum internet domain name
    - Inventory number
  - Easy to create and process for machines
    - No special chars, etc.
    - Human readable through a resolver
- http://museumid.net
- Talk on Tuesday 10:00

# Example: MuseumID

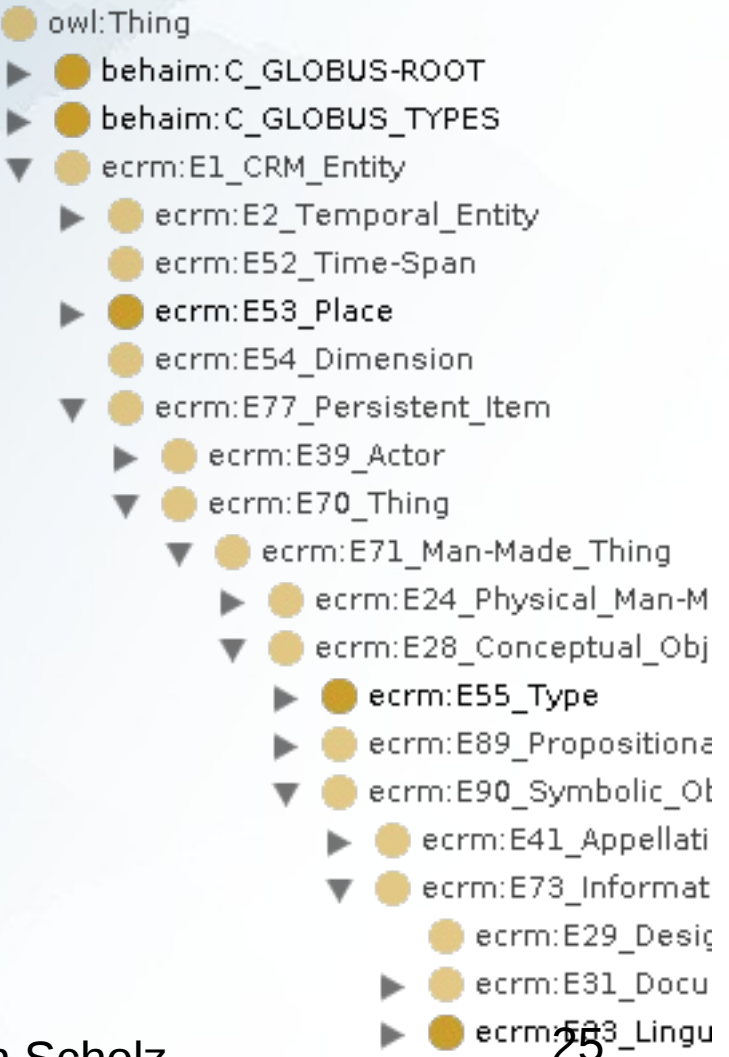


# Requirements for Transdisciplinary Use of Data

Ontologies

# Organizing knowledge with ontologies

- Define the terminology you use to talk about your domain
  - What do you want to talk about?
    - Concepts
    - Relations
  - What do you want to distinguish?
- Combine and reuse ontologies
  - For easier data exchange
  - Reduce costs



# What's in an Ontology

- An ontology consists of a set of concepts and defined relations between instances of these concepts
- An ontology is used to describe distinct parts of the real world with logical constructs and the relations between these parts
- Often also contains identifiers for the objects of discourse, the „things“

# Example: Ontologies

- Examples for concepts:
  - Material non-living concepts like „painting“, „frame“, …
  - Material living concepts like „artist“, „animal“, ...
  - Immaterial concepts like „plan“
- Examples for relations:
  - Relation between the concept of humans and the concept of artists
  - Relation between artists and paintings
  - Mother-child relationship between humans

# Using ontologies: Storing data

- Data is stored in propositions about instances of concepts (individuals)
- Proposition: Subject Predicate Object
- Example:
  - „The selfportrait“ is a painting.
  - „The selfportrait“ is damaged.
  - „Albrecht Dürer“ painted „the selfportrait“.

    Here „the selfportrait“ is a specific instance.

# CIDOC Conceptual Reference Model (CRM)

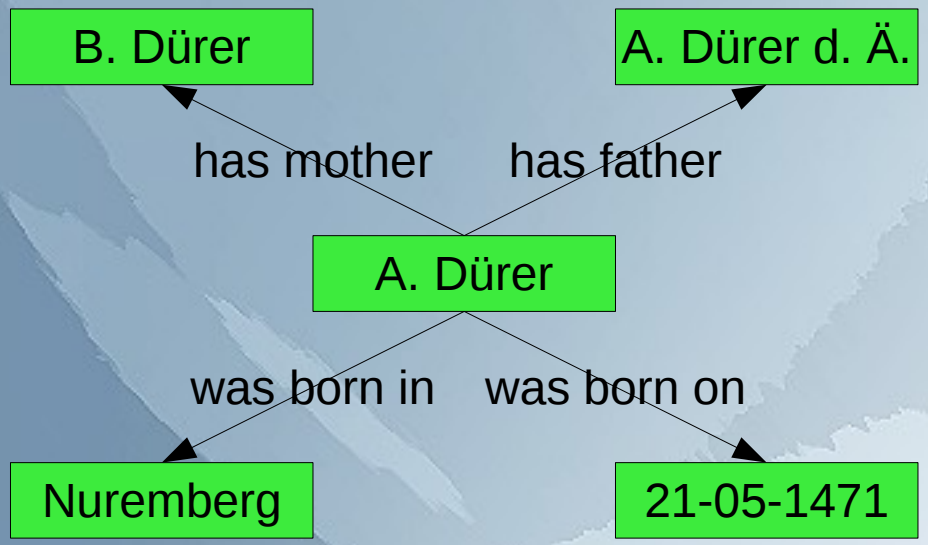
- Ontology initially created for the domain of cultural heritage
- Defines 86 concepts (entities) and 137 relations (properties)
- Manifested as ISO 21127
- Top level ontology: suitable for all cultural heritage domains
- http://cidoc-crm.org/

# CIDOC Conceptual Reference Model (CRM)

- Features of the CRM:
  - Event-oriented approach
  - Support for authorities
  - Distinction between the individual and it‘s name
  - CRM can be used as a common terminological framework across all different scientific domains in cultural heritage

► Transdisciplinary Approach

# Event Oriented Documentation



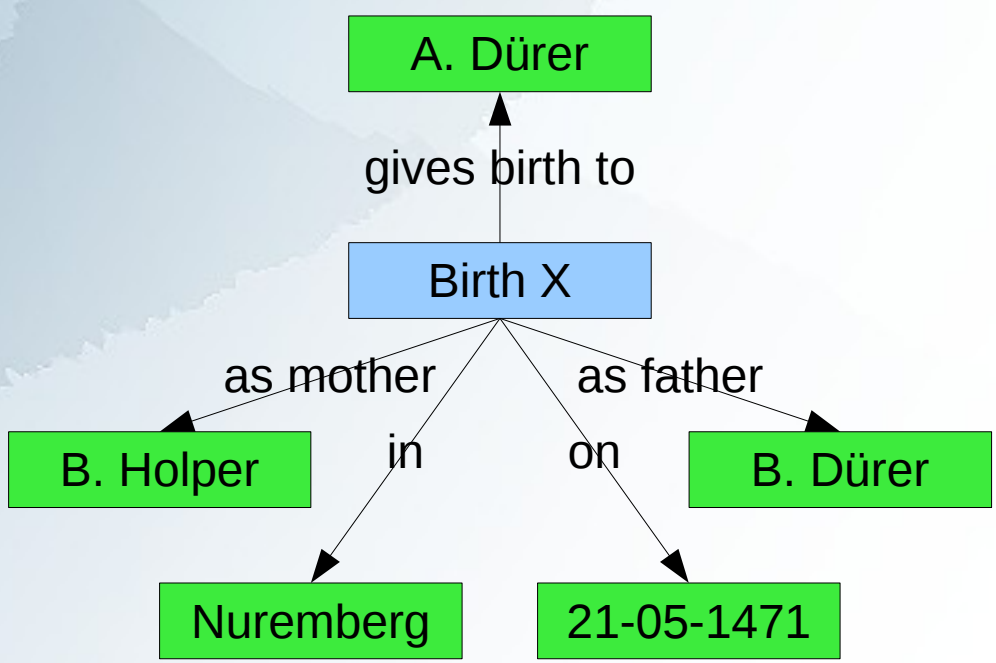

**Propositions:**

**A. Dürer has mother B. Dürer**

**A. Dürer has father A. Dürer d. Ä.**

**A. Dürer was born in Nuremberg**

**A. Dürer was born on 21-05-1471**




**Propositions:**

**Birth X gives birth to A. Dürer**

**Birth X as mother B. Dürer**

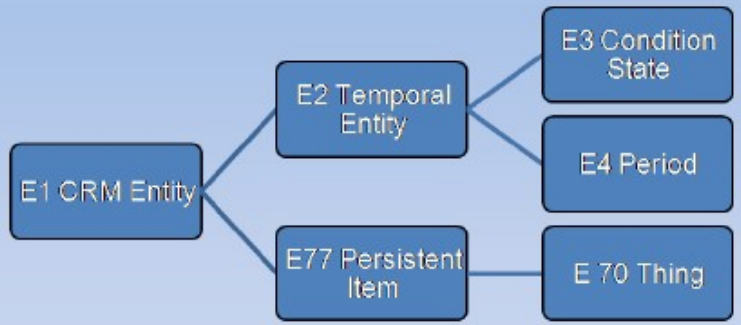
**Birth X as father A. Dürer d. Ä.**

**Birth X [took place] in Nuremberg**

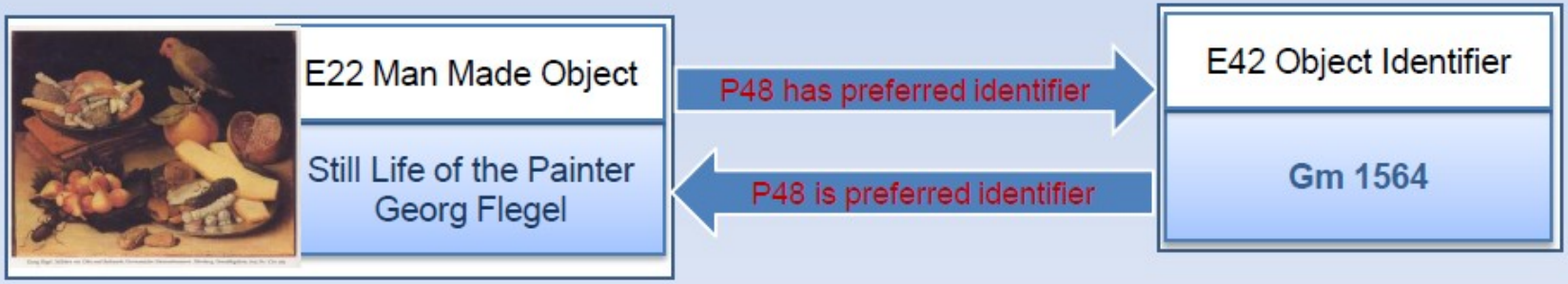
**Birth X [took place] on 21-05-1471**

# What is the CIDOC-CRM?

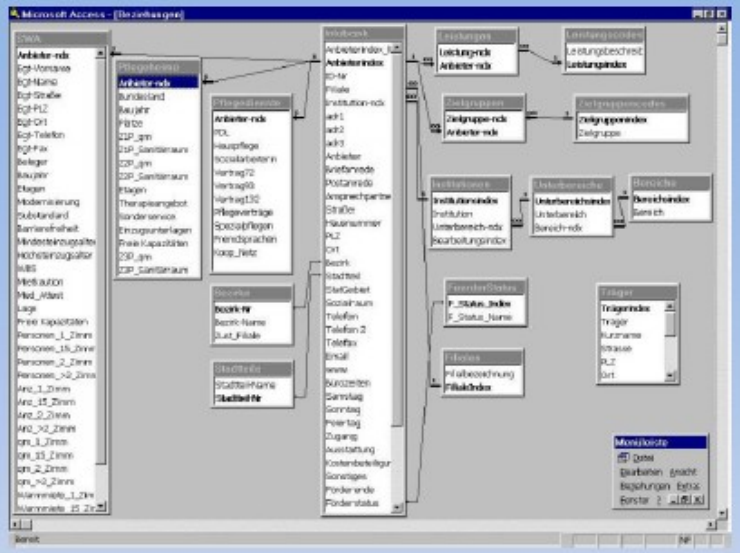
- It is an empirically developed ontology of the museum community
- It is organized in an object oriented class model




- It is based on linguistic “Subject - Predicate - Object” relations (Properties)



# The CIDOC-CRM is not:



- The CIDOC-CRM is **not** a readymade data model for the implementation of databases.


- But the CIDOC-CRM could be used as a guide to best practice in conceptual modeling