

Semantische Integration heterogener und unterschiedlichsprachiger Wissensorganisationssysteme

DFG-Projekt CrissCross und jenseits

Michael Preuss

Institut für Informationsmanagement
Fachhochschule Köln

4. Juli 2006



- 1 Möglichkeiten semantischer Integration klassischer Wissenorganisationssysteme
- 2 CrissCross
 - Cookbook
 - Probleme und offene Fragen
- 3 Verbesserung des Retrievals durch integrierte Vokabulare
 - Direkte Umsetzung
 - Nutzung von Strukturanreicherung und Strukturänderung
 - Retrievalagenten mittels einer Kernontologie
 - Schluss: Klassische Metadaten vs. Ontologien

MACS als prototypischer Versuch der Verknüpfung klassischer WOS

MACS

Das Projekt „Multilingual ACcess to Subjects“ versucht, drei Indexierungsvokabulare gleichgewichtig miteinander zu verbinden.

Paradigmen:

MACS als prototypischer Versuch der Verknüpfung klassischer WOS

MACS

Das Projekt „Multilingual ACcess to Subjects“ versucht, drei Indexierungsvokabulare gleichgewichtig miteinander zu verbinden.

Paradigmen:

- Volläquivalenz

MACS als prototypischer Versuch der Verknüpfung klassischer WOS

MACS

Das Projekt „Multilingual ACcess to Subjects“ versucht, drei Indexierungsvokabulare gleichgewichtig miteinander zu verbinden.

Paradigmen:

- Volläquivalenz
- Symmetrie der Verknüpfungsrelationen

MACS als prototypischer Versuch der Verknüpfung klassischer WOS

MACS

Das Projekt „Multilingual ACcess to Subjects“ versucht, drei Indexierungsvokabulare gleichgewichtig miteinander zu verbinden.

Paradigmen:

- Volläquivalenz
- Symmetrie der Verknüpfungsrelationen
- Bildung semantischer Cluster

MACS als prototypischer Versuch der Verknüpfung klassischer WOS

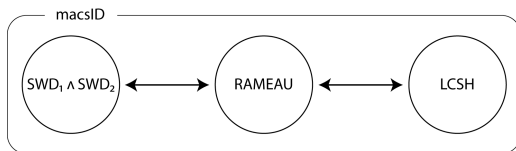
MACS

Das Projekt „Multilingual ACcess to Subjects“ versucht, drei Indexierungsvokabulare gleichgewichtig miteinander zu verbinden.

Paradigmen:

- Volläquivalenz
- Symmetrie der Verknüpfungsrelationen
- Bildung semantischer Cluster
- Abstraktion von der Struktur der Thesauri

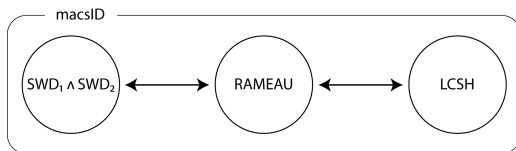
Paradigmen und Probleme von MACS



Probleme:

- Kein Einbezug von strukturellen Differenzen

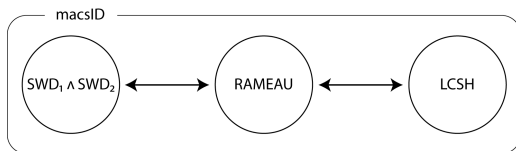
Paradigmen und Probleme von MACS



Probleme:

- Kein Einbezug von strukturellen Differenzen
- Komplette Abkoppelung von den Quelldateien

Paradigmen und Probleme von MACS



Probleme:

- Kein Einbezug von strukturellen Differenzen
- Komplette Abkoppelung von den Quelldateien
- 1-zu-n-Äquivalenzen brechen das Symmetrie-Prinzip

Vokabular 1

↔

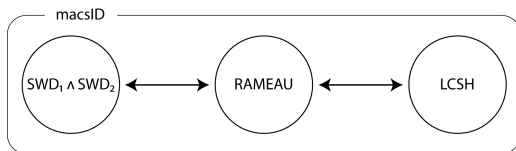
Vokabular 2

Objektorientierte Datenbank

↔

Objektorientierung \wedge Datenbank

Paradigmen und Probleme von MACS



Probleme:

- Kein Einbezug von strukturellen Differenzen
- Komplette Abkoppelung von den Quelldateien
- 1-zu-n-Äquivalenzen brechen das Symmetrie-Prinzip

Quellvokabular



Zielvokabular

Objektorientierte Datenbank



Objektorientierung \wedge Datenbank

CrissCross

Bei CrissCross geht es um die Verknüpfung von DDC und SWD und außerdem, unter Rückgriff auf die Daten von MACS, um die Anbindung der Library of Congress Subject Headings (LCSH) und des Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU).

Paradigmen:

CrissCross

Bei CrissCross geht es um die Verknüpfung von DDC und SWD und außerdem, unter Rückgriff auf die Daten von MACS, um die Anbindung der Library of Congress Subject Headings (LCSH) und des Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU).

Paradigmen:

- DDC als Pivot/Backbone

CrissCross

Bei CrissCross geht es um die Verknüpfung von DDC und SWD und außerdem, unter Rückgriff auf die Daten von MACS, um die Anbindung der Library of Congress Subject Headings (LCSH) und des Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU).

Paradigmen:

- DDC als Pivot/Backbone
- Retrievalorientierung

CrissCross

Bei CrissCross geht es um die Verknüpfung von DDC und SWD und außerdem, unter Rückgriff auf die Daten von MACS, um die Anbindung der Library of Congress Subject Headings (LCSH) und des Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU).

Paradigmen:

- DDC als Pivot/Backbone
- Retrievalorientierung
- Aufbau lokalisierterer Strukturen

CrissCross

Bei CrissCross geht es um die Verknüpfung von DDC und SWD und außerdem, unter Rückgriff auf die Daten von MACS, um die Anbindung der Library of Congress Subject Headings (LCSH) und des Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU).

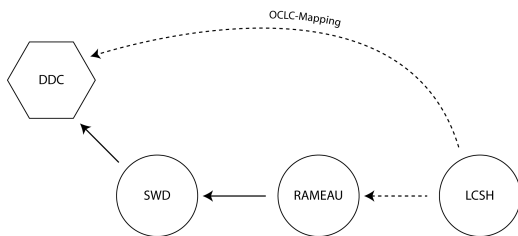
Paradigmen:

- DDC als Pivot/Backbone
- Retrievalorientierung
- Aufbau lokalisierterer Strukturen

Interoperabilität über die Themenfelder der DDC-Klassen:

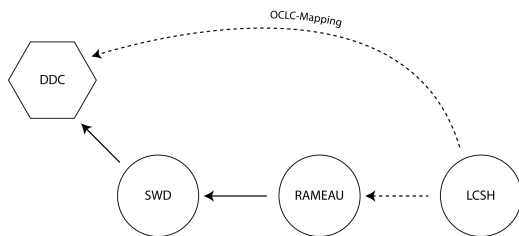
$$x \mapsto DDC \mapsto (y_1, z_1); \dots; (y_n, z_n)$$

CrissCross-Cookbook



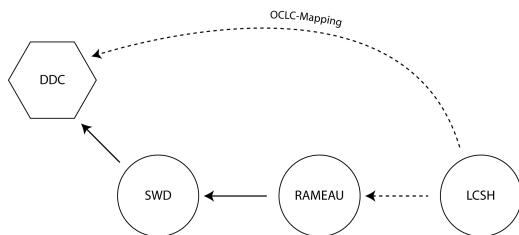
- Erweiterung des MAB-Formats zur Verankerung der Verknüpfungen

CrissCross-Cookbook



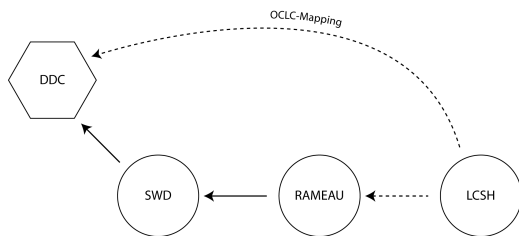
- Erweiterung des MAB-Formats zur Verankerung der Verknüpfungen
- Struktur von Feld 816

CrissCross-Cookbook



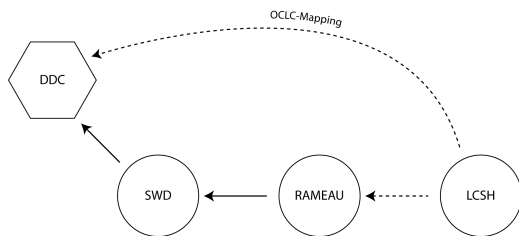
- Erweiterung des MAB-Formats zur Verankerung der Verknüpfungen
- Struktur von Feld 816
- Anpassung des DDC-Webclients MelvilClass

CrissCross-Cookbook



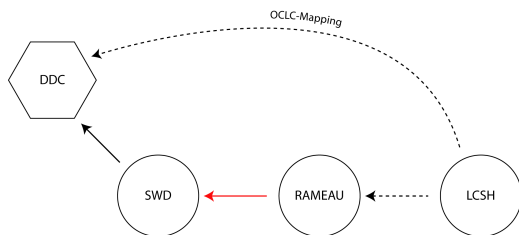
- Erweiterung des MAB-Formats zur Verankerung der Verknüpfungen
- Struktur von Feld 816
- Anpassung des DDC-Webclients MelvilClass
- Determiniertheit der Mappings als Gütemaß

CrissCross-Cookbook



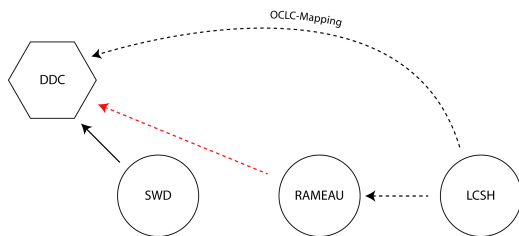
- Erweiterung des MAB-Formats zur Verankerung der Verknüpfungen
- Struktur von Feld 816
- Anpassung des DDC-Webclients MelvilClass
- Determiniertheit der Mappings als Gütemaß
- Anbindung von MACS-Tupeln entweder explizit oder abgeleitet

CrissCross-Cookbook



- Erweiterung des MAB-Formats zur Verankerung der Verknüpfungen
- Struktur von Feld 816
- Anpassung des DDC-Webclients MelvilClass
- Determiniertheit der Mappings als Gütemaß
- Anbindung von MACS-Tupeln entweder **explizit** oder abgeleitet

CrissCross-Cookbook



- Erweiterung des MAB-Formats zur Verankerung der Verknüpfungen
- Struktur von Feld 816
- Anpassung des DDC-Webclients MelvilClass
- Determiniertheit der Mappings als Gütemaß
- Anbindung von MACS-Tupeln entweder explizit oder **abgeleitet**

MelvilClass / Klassen: 658.31124 - Opera

Datei Bearbeiten Ansicht Lesezeichen Widgets Extras Hilfe

Neuer Tab

http://melvil.dcb.de/melvilclass/classes?id=1805470

Google

Michael Preuss
FH Köln - CrissCross

MelvilClass
The dot in 025.431

Klassen Suche     Hilfe Logout

Suche:

Klassen

Notation	Bezeichnung
	<u>DDC</u>
<u>600</u>	<u>Technik, Medizin, angewandte Wissenschaften</u>
<u>650</u>	<u>Management & Öffentlichkeitsarbeit</u>
<u>658</u>	<u>Allgemeines Management</u>
<u>658.3</u>	<u>Personalmanagement (Human Resource Management)</u>
<u>658.31</u>	<u>Bereiche des Personalmanagements</u>
<u>658.311</u>	<u>Anwerbung und Auswahl des Personals</u>
<u>658.3112</u>	<u>Auswahl</u>
658.31124	Auswahlgespräche

Anmerkungen, Hinweise und Verweisungen

Klassifiziere Vorstellungsgespräche aus Sicht des Arbeitssuchenden, umfassende Werke über Vorstellungsgespräche in [650144](#)

Registereinträge

- Auswahlgespräche — Personalauswahl
- Bewerbungsgespräche — Personalauswahl
- Einstellungsgespräche — Personalauswahl
- Interviews — Personalauswahl
- Vorstellungsgespräche — Personalauswahl

SWD-Schlagwörter

|s|Strukturiertes Interview

W3C HTML 4.01  W3C CSS  ddc.deutsch 

MelvilClass / Klassen: 794.122 - Opera

Datei Bearbeiten Ansicht Lesezeichen Widgets Extras Hilfe

Neuer Tab

http://melvil.ddb.de/melvilclass/classes?id=601282

Michael Preuss
FH Köln - CrissCross

MelvilClass
The dot in 025.431

Klassen Suche     Hilfe Logout

Suche:

Klassen

Notation	Bezeichnung
	<u>DDC</u>
<u>700</u>	<u>Künste und Unterhaltung</u>
<u>790</u>	<u>Sport, Spiele, Unterhaltung</u>
<u>794</u>	<u>Unterhaltungsspiele für drinnen</u>
<u>794.1</u>	<u>Schach</u>
<u>794.12</u>	<u>Strategie und Taktik</u>
794.122	Eröffnungen

Anmerkungen, Hinweise und Verweisungen

Registereinträge

Eröffnungen (Schach)

SWD-Schlagwörter

- |s|Abtauschvariante
- |s|Alapin-Eröffnung
- |s|Albins Gegengambit
- |s|Aljechin-Verteidigung
- |s|Altindische Verteidigung
- |s|Angenommenes Damengambit
- |s|Antiindische Eröffnung
- |s|Basman-Sale-Variante
- |s|Benoni-Verteidigung
- |s|Bird-Eröffnung
- |s|Blackmar-Diemer-Gambit
- |s|Blumenfeld-Gambit
- |s|Bogoljubov-indische Verteidigung
- |s|Bogoljubovs Verteidigung
- |s|Budapester Gambit
- |s|Caro-Kann-Verteidigung
- |s|Colle-System

WinIBW 2.000 - [ZENTRAKATALOG DNB - Hauptbestand Vollanzeige]

Datei Bearbeiten Ansicht Optionen Macro ZDB-Funktionen SWD/PND-Funktionen Maj Fenster 2

Review Index Löschen Neuaufnahme Bearbeiten Sichern Hilfe

CrisAnmeldenMxNew CrisX CrisNeu CrisFuegeSWDein CrisMxAn CrisSet CrisSetX CrisShowCK CrisShowTitle CrisSucheSW CrisShowCkpl

SET: S8 [21] TTL: 9 PPN:948822201

Eingabe: 1250:14-10-96 Änderung: 1220:20-06-06 11:17:37 Status: 1220:20-06-06

005 Tsx
012 /xxw
021 4427224-8
800 s|Strukturiertes Interview
808 |x|pr
808 |a|Lex. Soz.
808 |c|Nichtpermutierendes USW, falls die auf einen Untersuchungsgegenstand angewandte Methode ausgedrückt wird. Nicht als Forms Schlagwort zugelassen; als solches verwende Interview
810 5.1b;9.2b
816 158.39#2#
816 658.31124#1#
850 s|Interview

F1 für Regelwerksinformationen 00:10 NUM EID

Implikationen des Mappings

- Mapping als „Framing“ der semantischen Dimensionen des Schlagworts

Implikationen des Mappings

- Mapping als „Framing“ der semantischen Dimensionen des Schlagworts
- Res(t)itutierung des Begriffs in Kontexten (vgl. Relative Index der DDC), trotz stabiler Extension

Implikationen des Mappings

- Mapping als „Framing“ der semantischen Dimensionen des Schlagworts
- Res(t)itutierung des Begriffs in Kontexten (vgl. Relative Index der DDC), trotz stabiler Extension
- Funktionale vs. ontologische Relationen

Implikationen des Mappings

- Mapping als „Framing“ der semantischen Dimensionen des Schlagworts
- Res(t)itutierung des Begriffs in Kontexten (vgl. Relative Index der DDC), trotz stabiler Extension
- Funktionale vs. ontologische Relationen

Beispiel: s Ropes Course

302.3	Soziale Interaktion in Gruppen	funktional
371.384	Unterricht im Freien	funktional
616.8916	Psycho- und Aktivitätstherapien	funktional
796.5	Aktivitäten im Freien	ontologisch (isA)

Probleme und offene Fragen

- Begrenzungen von MAB: Kein Mapping von Nicht-Deskriptoren, ...

Probleme und offene Fragen

- Begrenzungen von MAB: Kein Mapping von Nicht-Deskriptoren, ...
- Unterdeterminiertheit des Kontextes (Beispiel: s Muskel)

Probleme und offene Fragen

- Begrenzungen von MAB: Kein Mapping von Nicht-Deskriptoren, ...
- Unterdeterminiertheit des Kontextes (Beispiel: s Muskel)
- Kontraintuitive Typisierung der Schlagwörter in der SWD (Beispiel: s Indianerin, aber: g Indianer)

Probleme und offene Fragen

- Begrenzungen von MAB: Kein Mapping von Nicht-Deskriptoren, ...
- Unterdeterminiertheit des Kontextes (Beispiel: s Muskel)
- Kontraintuitive Typisierung der Schlagwörter in der SWD (Beispiel: s Indianerin, aber: g Indianer)
- Nichtidentitätsproblematik (die Bedeutung eines Terms hat sich verändert; ein Term hat seine Bedeutung geändert)

Retrieval und integrierte Vokabulare

Wie wird das Mapping unter Wahrung des Mehrwerts in Retrievalszszenarien integriert?

Retrieval und integrierte Vokabulare

Wie wird das Mapping unter Wahrung des Mehrwerts in Retrievalszenerarien integriert?

Direkte Umsetzung des Mappings

Retrieval und integrierte Vokabulare

Wie wird das Mapping unter Wahrung des Mehrwerts in Retrievalszszenarien integriert?

Direkte Umsetzung des Mappings

- durch Transaktionsmoduln

Retrieval und integrierte Vokabulare

Wie wird das Mapping unter Wahrung des Mehrwerts in Retrievalszenarios integriert?

Direkte Umsetzung des Mappings

- durch Transaktionsmoduln
- mit aktivem Zugriff auf die verknüpften Vokabulare (URI [info:lccn], OAI-PMH)

Retrieval und integrierte Vokabulare

Wie wird das Mapping unter Wahrung des Mehrwerts in Retrievalszszenarien integriert?

Direkte Umsetzung des Mappings

- durch Transaktionsmoduln
- mit aktivem Zugriff auf die verknüpften Vokabulare (URI [info:lccn], OAI-PMH)
- unter Ausnutzung der Gütemaße

Anreicherung von Thesaurusrelationen

Ausnutzung der Strukturanreicherung durch das Mapping und darüber hinausgehende Förderung der Strukturänderung der beteiligten Dateien.

Anreicherung von Thesaurusrelationen

Ausnutzung der Strukturanreicherung durch das Mapping und darüber hinausgehende Förderung der Strukturänderung der beteiligten Dateien.

Ausdifferenzierung der OB/UB-Relationen in

Anreicherung von Thesaurusrelationen

Ausnutzung der Strukturanreicherung durch das Mapping und darüber hinausgehende Förderung der Strukturänderung der beteiligten Dateien.

Ausdifferenzierung der OB/UB-Relationen in

- generische Relationen und Instanzrelationen

Anreicherung von Thesaurusrelationen

Ausnutzung der Strukturanreicherung durch das Mapping und darüber hinausgehende Förderung der Strukturänderung der beteiligten Dateien.

Ausdifferenzierung der OB/UB-Relationen in

- generische Relationen und Instanzrelationen
- Umwandlung der partitiven Relationen zu VB-Relationen

Anreicherung von Thesaurusrelationen

Ausnutzung der Strukturanreicherung durch das Mapping und darüber hinausgehende Förderung der Strukturänderung der beteiligten Dateien.

Ausdifferenzierung der OB/UB-Relationen in

- generische Relationen und Instanzrelationen
- Umwandlung der partitiven Relationen zu VB-Relationen
- Erben von generischen (isA-)Relationen aus der gemappten Klassifikation

Anreicherung von Thesaurusrelationen

Ausnutzung der Strukturanreicherung durch das Mapping und darüber hinausgehende Förderung der Strukturänderung der beteiligten Dateien.

Ausdifferenzierung der OB/UB-Relationen in

- generische Relationen und Instanzrelationen
- Umwandlung der partitiven Relationen zu VB-Relationen
- Erben von generischen (isA-)Relationen aus der gemappten Klassifikation
- Automatische Einbeziehung lexikalischer Relationen aus bestehenden Strukturen (z.B. WordNet/GermaNet)

Typisierung und Vermehrung der VB-Relationen

- Im Hinblick auf Retrieval, nicht Erstellung und Pflege des Thesaurus

Typisierung und Vermehrung der VB-Relationen

- Im Hinblick auf Retrieval, nicht Erstellung und Pflege des Thesaurus
- Maß für semantische Distanz über Anzahl der besuchten Knoten

Typisierung und Vermehrung der VB-Relationen

- Im Hinblick auf Retrieval, nicht Erstellung und Pflege des Thesaurus
- Maß für semantische Distanz über Anzahl der besuchten Knoten

RT (Plain) for undifferentiated associative relationships

Meaning connection

Meaning overlap

Distinguished from

Antonym

Conjoined terms

Causal (taken broadly – many inter-facet relationships might fit here)

Dependency/requires

Uses

Product

Patient

and possibly Spatial and Temporal connections

Partitive (taken broadly)

Constituent parts

Aggregate group

Property/attribute

Tudhope et. al., „Augmenting Thesaurus Relationships: Possibilities for Retrieval“, in: *JoDI* 1/8 (2001)

Ontologien

Zwei unterschiedliche Operationalisierungen von Ontologien:

Ontologien

Zwei unterschiedliche Operationalisierungen von Ontologien:

- 1 Sprachunabhängige Systemstrukturierung auf konzeptueller Ebene (Metaphysik)

Ontologien

Zwei unterschiedliche Operationalisierungen von Ontologien:

- 1 Sprachunabhängige Systemstrukturierung auf konzeptueller Ebene (Metaphysik)
- 2 Spezifizierung eines Repräsentationsvokabulars und Regeln zur Verknüpfung

Ontologien

Zwei unterschiedliche Operationalisierungen von Ontologien:

- 1 Sprachunabhängige Systemstrukturierung auf konzeptueller Ebene (Metaphysik)
- 2 Spezifizierung eines Repräsentationsvokabulars und Regeln zur Verknüpfung

Versuch einer Synthese (Guarino, 1998):

Ontologisches *commitment* eines formalen Vokabulars zu einer bestimmten Konzeptualisierung der Welt

Einsatz einer Kernontologie

- Kompilationsphase vs. Konsolidierungsphase der Informationsintegration (nach Doerr)

Einsatz einer Kernontologie

- Kompilationsphase vs. Konsolidierungsphase der Informationsintegration (nach Doerr)
- Kompilationsphase: keine Erzeugung abgeleiteten Wissens, keine Auflösung von alternativen oder kontradiktorischen Informationen

Einsatz einer Kernontologie

- Kompilationsphase vs. Konsolidierungsphase der Informationsintegration (nach Doerr)
- Kompilationsphase: keine Erzeugung abgeleiteten Wissens, keine Auflösung von alternativen oder kontradiktorischen Informationen
- Konsolidierungsphase: „Ansichten“ über die beschriebene Realität werden gebildet

CIDOC Conceptual Reference Model (CRM)

- Top-Level und Domänenontologie für den Bereich des kulturellen Erbes

CIDOC Conceptual Reference Model (CRM)

- Top-Level und Domänenontologie für den Bereich des kulturellen Erbes
- Modell einer integrierenden, weitgehend auf die Kompilationsphase beschränkten Ontologie

CIDOC Conceptual Reference Model (CRM)

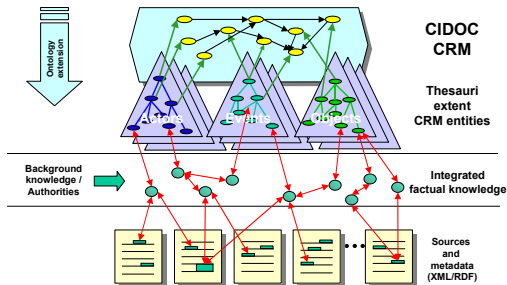
- Top-Level und Domänenontologie für den Bereich des kulturellen Erbes
- Modell einer integrierenden, weitgehend auf die Kompilationsphase beschränkten Ontologie
- Möglichkeiten des Einsatzes:

CIDOC Conceptual Reference Model (CRM)

- Top-Level und Domänenontologie für den Bereich des kulturellen Erbes
- Modell einer integrierenden, weitgehend auf die Kompilationsphase beschränkten Ontologie
- Möglichkeiten des Einsatzes:
 - ▶ Titeldaten (über P1 is identified by + E32 Authority Document + [Metadaten des Normdatensatzes])

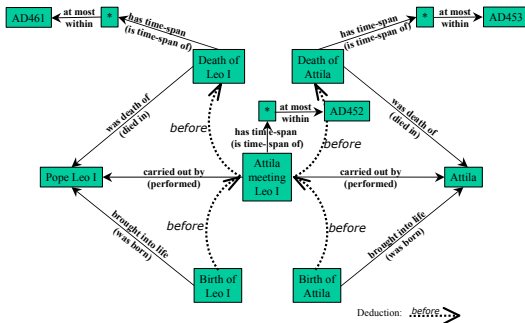
CIDOC Conceptual Reference Model (CRM)

- Top-Level und Domänenontologie für den Bereich des kulturellen Erbes
- Modell einer integrierenden, weitgehend auf die Kompilationsphase beschränkten Ontologie
- Möglichkeiten des Einsatzes:
 - ▶ Titeldaten (über P1 is identified by + E32 Authority Document + [Metadaten des Normdatensatzes])
 - ▶ Normdaten (über E55 Type oder über Mapping des Normdatenmodells)



Pid	Property Name	Domain	Range
P92	brought into existence (was brought into existence by)	E63 Beginning of Existence	E77 Existence
P94	- has created (was created by)	E65 Conceptual Creation	E28 Conceptual Object
P95	- has formed (was formed by)	E66 Formation	E74 Group
P98	- brought into life (was born)	E67 Birth	E21 Person
P108	- has produced (was produced by)	E12 Production	E24 Physical Man-Made Stuff
P93	took out of existence (was taken out of existence by)	E64 End of Existence	E77 Existence
P13	- destroyed (was destroyed by)	E6 Destruction	E19 Physical Object
P99	- dissolved (was dissolved by)	E68 Dissolution	E74 Group
P100	- was death of (died in)	E69 Death	E21 Person

Pid	Property Name	Domain	Range
P92	brought into existence (was brought into existence by)	E63 Beginning of Existence	E77 Existence
P94	- has created (was created by)	E65 Conceptual Creation	E28 Conceptual Object
P95	- has formed (was formed by)	E66 Formation	E74 Group
P98	- brought into life (was born)	E67 Birth	E21 Person
P108	- has produced (was produced by)	E12 Production	E24 Physical Man-Made Stuff
P93	took out of existence (was taken out of existence by)	E64 End of Existence	E77 Existence
P13	- destroyed (was destroyed by)	E6 Destruction	E19 Physical Object
P99	- dissolved (was dissolved by)	E68 Dissolution	E74 Group
P100	- was death of (died in)	E69 Death	E21 Person



Schluss: Klassische Metadaten vs. Ontologien

- Ressourcenzentriertheit vs. Eigenschaftszentriertheit

Schluss: Klassische Metadaten vs. Ontologien

- Ressourcenzentriertheit vs. Eigenschaftszentriertheit
- Präskriptivität vs. Deskriptivität

Schluss: Klassische Metadaten vs. Ontologien

- Ressourcenzentriertheit vs. Eigenschaftszentriertheit
- Präskriptivität vs. Deskriptivität
- Erstellt von Menschen vs. von maschinellen Tools

Schluss: Klassische Metadaten vs. Ontologien

- Ressourcenzentriertheit vs. Eigenschaftszentriertheit
- Präskriptivität vs. Deskriptivität
- Erstellt von Menschen vs. von maschinellen Tools
- Underdetermination vs. Überdetermination