Datenmodellierung

Woche 12: Ontologien

Institut für Digital Humanities, Historisch-Kulturwissenschaftliche Informationsverarbeitung | Prof. Dr. Øyvind Eide

Literatur zur Vertiefung (optional)

- Busse, Johannes et al. 2014. “Was bedeutet eigentlich Ontologie? - Ein Begriff aus der Philosophie im Licht verschiedener Disziplinen.” *Informatik Spektrum* 37, S. 286–297. https://doi.org/10.1007/s00287-012-0619-2
- Eide, Øyvind und Christian-Emil Smith Ore. 2018. “Ontologies and data modeling.” S. 168–196 in: Flanders, Julia und Jannidis, Fotis (Hg.): *The Shape of Data in Digital Humanities: Modeling Texts and Text-based Resources,* Routledge, Taylor & Francis. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315552941-8/
- Rehbein, Malte. 2017. “Ontologien.” S. 162-176 in: Jannidis, Fotis, Kohle, Hubertus, Rehbein, Malte (Hg.): *Digital Humanities.* J.B. Metzler. https://doi.org/10.1007/978-3-476-05446-3_11
- https://chin-rcip.github.io/collections-model/en/resources/current/how-to/cidoc-crm-tutorial

Universität zu Köln

Ontologie: Definition

"An ontology is an explicit specification of a conceptualization. The term is borrowed from philosophy, where an ontology is a systematic account of Existence. For knowledge-based systems, what 'exists' is exactly that which can be represented."

Gruber, Thomas. 1993. "A Translation Approach to Portable Ontology Specifications." In: Knowledge Acquisition, 5(2): 199-220.

- Philosophie: Lehre vom Seienden
  - Teil von der Metaphysik
- Informatik: geteilte formalisierte Konzeptualisierung
  - Ein formales Begriffssystem zur Wissensrepräsentation für einen bestimmten Gegenstandsbereich (domain)
  - Hohes Abstraktionsniveau, semantisch explizit und präzise, maschinell verarbeitbar
  - ➢ Semantic Web
  - ➢ Künstliche Intelligenz

Ontologien: Elemente

Klassen (classes)

- Klassenhierarchie → Vererbung
  - Is-a-Beziehung: X is a Y
  - Oberklasse (superclass) / Unterklasse (subclass)
  - auch multiple Vererbung (multiple inheritance) möglich

Eigenschaften (properties)

- Eigenschaftenhierarchie → Vererbung
  - superproperty / subproperty
- Können verweisen auf Klassen, dann auch Beziehungen (relationships) genannt, oder auf Instanzen / literale Werte
- Quantifizierung
- Domain und range von Eigenschaften beachten, d.h. welche Klassen als Subjekt bzw. Objekt erlaubt

Instanzen (particulars)

- Ein Element, das zu einer Klasse gehört

Praktische Übung: Elemente einer Ontologie

Machen Sie sich eine Kopie des Dokuments (auch in lehre.idh verlinkt)

Füllen Sie dann Ihr Dokument aus (allein oder als Gruppe):

- Blatt 1 + 2
  - Identifizieren Sie zunächst, ob es sich jeweils um eine Klasse, Instanz oder Eigenschaft handelt
- Blatt 3
  - Bestimmen Sie eine Hierarchie zwischen den Klassen
- Blatt 4
  - Ordnen Sie zu, welche Instanz zu welcher Klasse gehört
- Blatt 5
  - Bestimmen Sie Domain und Range von Eigenschaften

Quelle: CC BY-ND: George Bruseker - Takin.solutions (2023)
https://tinyurl.com/mvadh8jy / https://tinyurl.com/2vspx8wx

Wozu Ontologien?

Beim **Entwerfen** von Ontologien:

- Besseres Verständnis einer Domäne erhalten
- Abstraktion, Strukturierung

Bei der **Verwendung** von Ontologien:

- Gemeinsames Verständnis
- Interoperabilität
  - Austausch von Daten
  - Integration von Daten aus unterschiedlichen Quellen
- Logische Schlussfolgerungen (reasoning)
  - Schlussfolgern neuer Informationen → ‚Wissen erzeugen'

Wozu Ontologien?

→ Integration durch Mapping auf eine gemeinsame Referenzontologie

"If one sets out to integrate data from ten different information systems one could in principle map each of them to each other. That would lead to **99** mappings and no clear centre of gravity to be used for searching. If one maps each of the information systems to CIDOC-CRM, one only has to make **ten** mappings. The data can then be exported using these mappings to a common system facilitating searches across the collections."

Øyvind Eide und Christian-Emil Smith Ore. 2018. Ontologies and data modeling. S. 168–196 in: Julia Flanders und Fotis Jannidis (Hg.): *The Shape of Data in Digital Humanities: Modeling Texts and Text-based Resources,* Routledge, Taylor & Francis. (S. 190)

Wozu Ontologien?

→ Integration durch Mapping auf eine gemeinsame Referenzontologie

“This common system can facilitate searching beyond what is possible in the source information systems. One reason for this is the well-defined hierarchical is-a system. In CIDOC-CRM, there is a set of classes including E12 Production, E65 Creation, and E66 Formation that are all subclasses of E63 Beginning of Existence. Each of these three classes is-a E63 Beginning of Existence. If source database A has a number of E12s, E65s, and E66s, and source database B has only E63s, the inheritance system means that information from these four classes can be integrated as if they all belonged to the same class E63, facilitating better information integration.”

Øyvind Eide und Christian-Emil Smith Ore. 2018. Ontologies and data modeling. S. 168–196 in: Julia Flanders und Fotis Jannidis (Hg.): *The Shape of Data in Digital Humanities: Modeling Texts and Text-based Resources,* Routledge, Taylor & Francis. (S. 190)

Wozu Ontologien?

→ Schlussfolgern neuer Informationen auf Grundlage der Wissensbasis

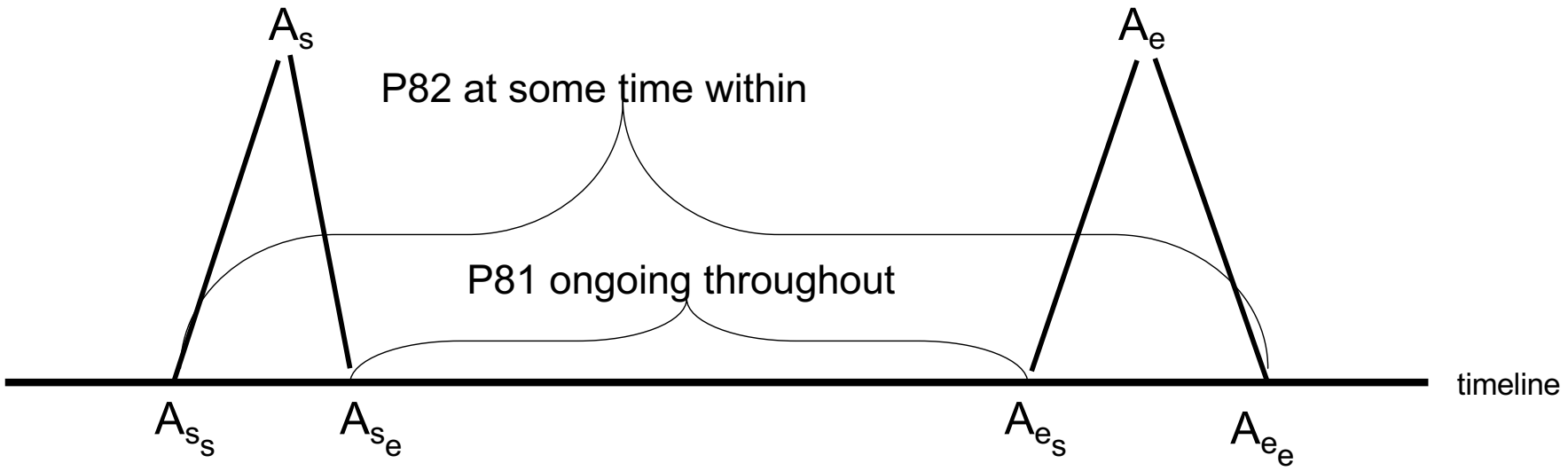
Sokrates ist ein Mensch.

Menschen sind sterblich.

Daraus folgt: Sokrates ist sterblich.

Modellierung von Zeitangaben

Interval on the timeline (time primitive) for a time span A



Interval for start point measurement

Interval for end point measurement

A_{s_s}: A must have started after

A_{s_e}: A must have started before

A_{e_s}: A must have ended after

A_{e_e}: A must have ended before

Validity conditions on intervals:

$A_{s_s} < A_{s_e} < A_{e_e}$

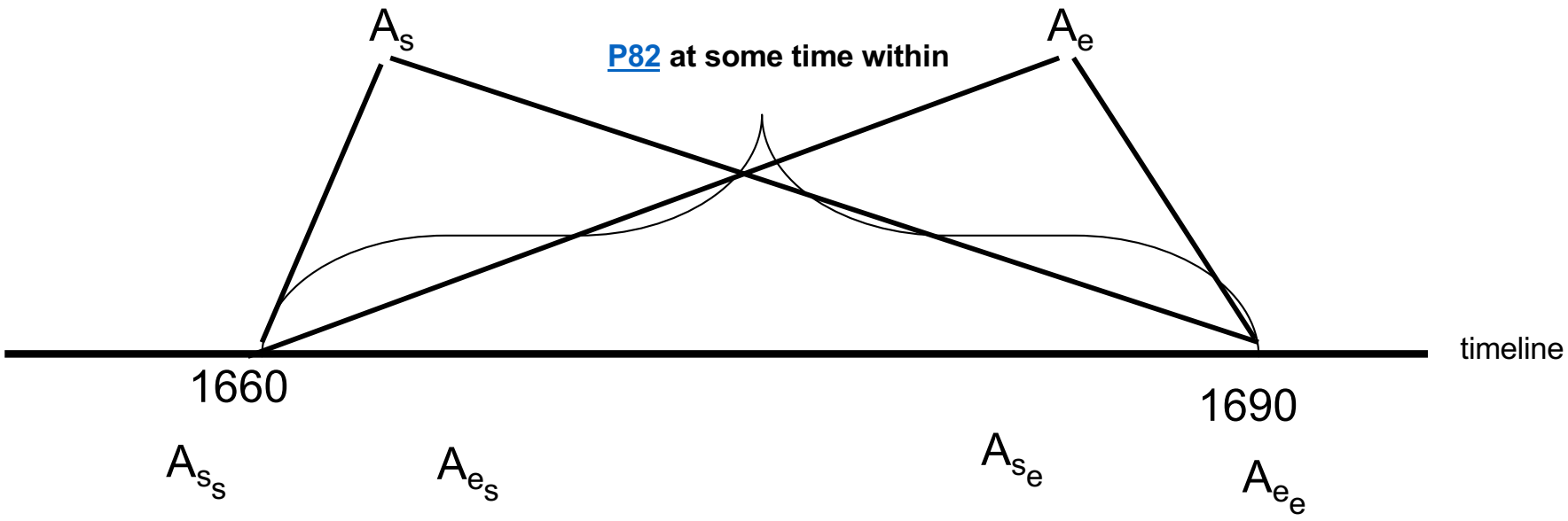
$A_{s_s} < A_{e_s} < A_{e_e}$

Universität zu Köln

Temporal entity A with undetermined start and end points

Document, dated 1660: “There are plans to build a church at place X”.
Document, dated 1690: “On place X a church has been built”.

Knowledge about the building activity A:



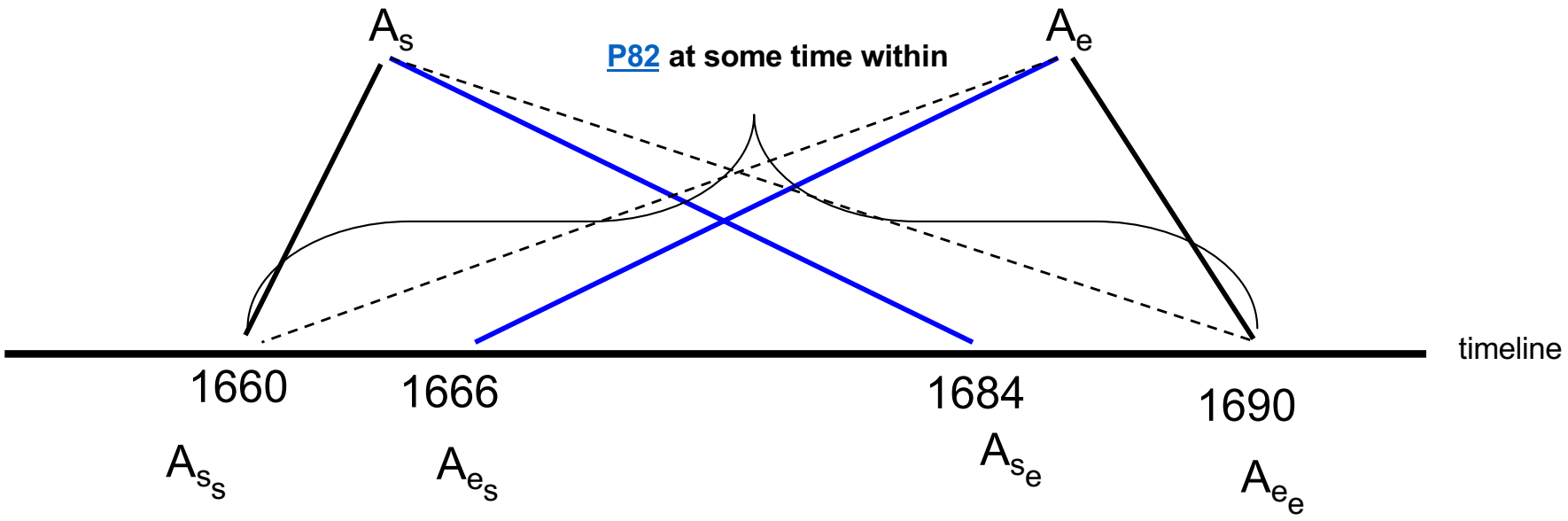
There is no information about the actual building period:

A_{s_s} not before 1660 A_{e_e} not after 1690

Temporal entity A with undetermined start and end points

Document, dated 1660: “There are plans to build a church at place X”.
Document, dated 1690: “On place X a church has been built”.
Document, dated 1711: “The construction of the church took at least 6 years

Knowledge about the building activity A:



There is some information about the actual building period:

A_{s_s} not before 1660

A_{e_e} not after 1690

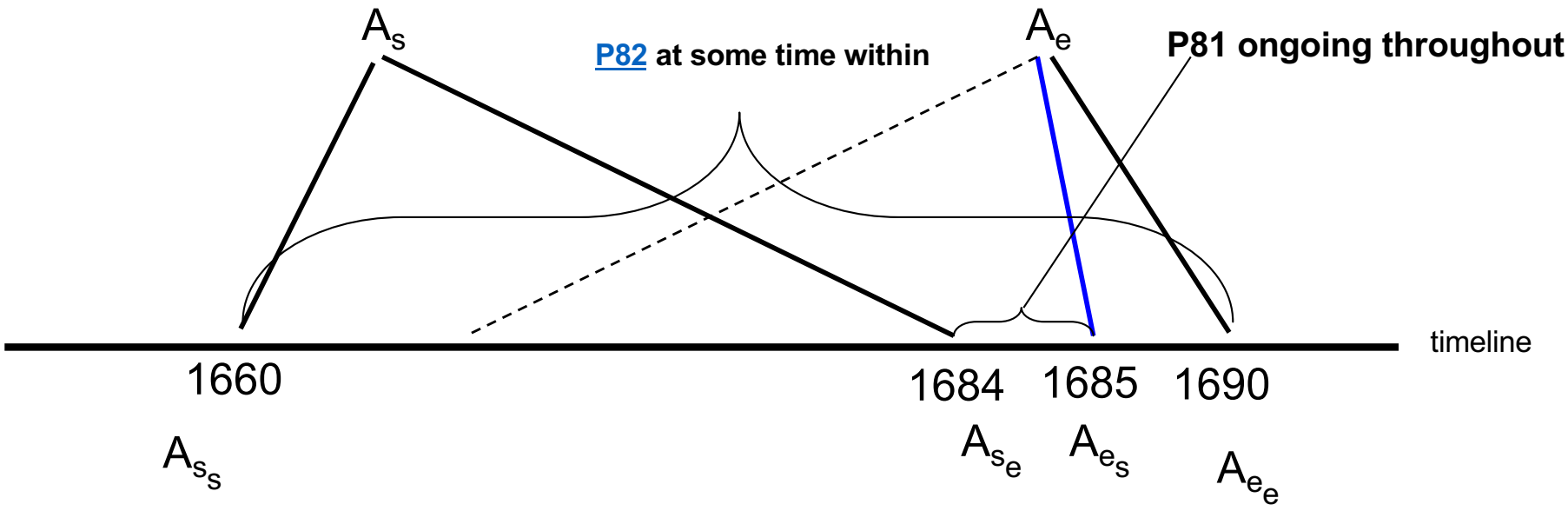
A_{e_e} not after 1684

A_{e_s} not before 1684

Temporal entity A with undetermined start and end points

Document, dated 1660: “There are plans to build a church at place X”.
Document, dated 1690: “On place X a church has been built”.
Document, dated 1711: “The construction of the church took at least 6 years
Report, date 1984: A coin from the reign of James II was found in the foundation

Knowledge about the building activity A:



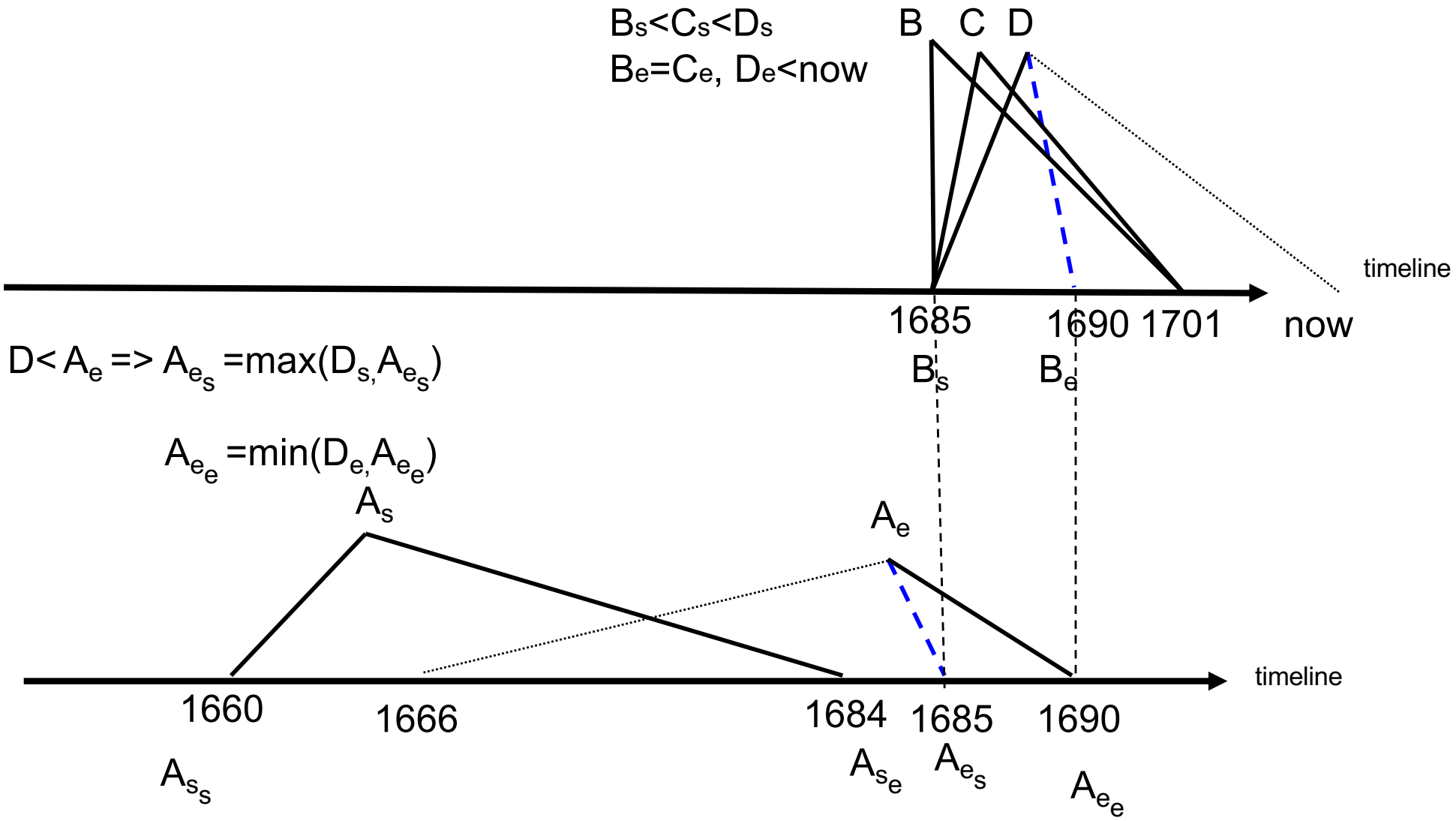
There is some information about the actual building period:

A_{s_s}	not before 1660	A_{e_e}	not after 1690
A_{e_e}	not after 1684	A_{e_s}	not before 1685

Relate activity D, coin disposal, with activity A, building of the church

A chain of events:

B: James II Regent Period (1685 -1701)
C: Production of Coins from James II
D: Disposal of the Coin in Church



Beschreibung von Ontologien

Um eine Ontologie und Aussagen (d.h. Daten) formal auszudrücken und maschinell zugänglich/verknüpfbar zu machen, nutzt man formale Beschreibungssprachen für Ontologien:

- First-order logic
- Resource Description Framework (RDF)
- RDF Schema (RDFS)
- Web Ontology Language (OWL)

Resource Description Framework (RDF)

https://www.w3.org/TR/rdf11-concepts/

https://www.w3.org/TR/rdf11-schema/

- Tripel aus drei Elementen: Subjekt Prädikat Objekt
- Jedes Element ist durch eine URI identifiziert
- Jedes Element kann wiederum typisiert werden (der Typ wird ebenfalls durch eine URI identifiziert)
- Definition eines Schemas mittels RDF Schema (RDFS) möglich
- Serialisierung als Terse RDF Triple Language (Turtle) oder in JSON- oder XML-basierter Syntax

RDF – Beispiel

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix bot: <http://beispielontologie.de/terminologie#> .
@prefix boa: <http://beispielontologie.de/instanzen#> .
bot:Ding                rdf:type                rdfs:Class .
bot:Person              rdfs:subClassOf         bot:Ding .
bot:Mann                rdfs:subClassOf         bot:Person .
bot:Frau                rdfs:subClassOf         bot:Person .
bot:Nobelpreis          rdfs:subClassOf         bot:Ding .
bot:istBruderVon        rdf:type                rdf:Property ;
                        rdfs:domain             bot:Mann ;
                        rdfs:range              bot:Person .
bot:Nobelpreisträger    rdfs:subClassOf bot:Person .
bot:gewinnt             rdf:type                rdf:Property ;
                        rdfs:domain             bot:Person ;
                        dfs:range               bot:Nobelpreis .
```

Quelle: Rehbein, Malte. 2017. “Ontologien.” In: Jannidis, Fotis, Kohle, Hubertus, Rehbein, Malte (Hg.): *Digital Humanities.* J.B. Metzler. https://doi.org/10.1007/978-3-476-05446-3_11

RDF – Beispiel

```
boa:Literaturnobelpreis      rdf:type          bot:Nobelpreis .

boa:Thomas_Mann              rdf:type          bot:Mann ;
                             bot:gewinnt       boa:Literaturnobelpreis .

boa:Heinrich_Mann            rdf:type                bot:Mann ;
                             bot:istBruderVon        boa:Thomas_Mann .
```

Quelle: Rehbein, Malte. 2017. “Ontologien.” In: Jannidis, Fotis, Kohle, Hubertus, Rehbein, Malte (Hg.): *Digital Humanities.* J.B. Metzler. https://doi.org/10.1007/978-3-476-05446-3_11

Web Ontology Language (OWL)

https://www.w3.org/TR/owl2-overview/

- Erweiterung von RDF(S)
- Ermöglicht weitergehende logische Kriterien zu definieren; erlaubt es z.B.....
  - Zwei Individuen als ‘identisch’ zu kennzeichnen: owl:sameAs
  - Kardinalität von Beziehungen festzulegen: owl:maxCardinality, owl:minCardinality
  - Schnittmenge zweier Klassen als leer definieren: owl:disjointWith
  - etc.
- Es können so weitere Details in der Ontologie explizit gemacht oder die logische Konsistenz überprüft werden

Ontologien: Weitere Begriffe

- Tripel (triple)
  - Gerichteter Graph
  - ➢ Linked Data
- Terminologisches Wissen vs. assertionales Wissen
- Triplestore: Datenbank zur Speicherung von Tripeln
  - Abfragesprache: SPARQL (SPARQL Protocol And RDF Query Language)
- Wissensbasis (knowledge base) oder -graph (knowledge graph)

- Closed World Assumption vs. Open World Assumption
  - ➢ Sind Aussagen, die nicht in der Ontologie/Wissensbasis enthalten sind, FALSE oder nicht zu beantworten?

Universität zu Köln

SPARQL
SPARQL Protocol And RDF Query Language

https://www.w3.org/TR/sparql11-query/

- Graphenbasierte Abfragesprache
- WHERE-Klauseln werden auch als Tripel formuliert
  - Erlaubt semantische Anfragen
- Suche ist über verschiedene SPARQL-Endpoints hinweg möglich (federated search)

```
SELECT ?personA ?personB
WHERE {
        ?personA        rdf:type                bot:Person ;
                        bot:gewinnt             bot:Nobelpreis .
        ?personB        bot:istBruderVon        ?personA .
```

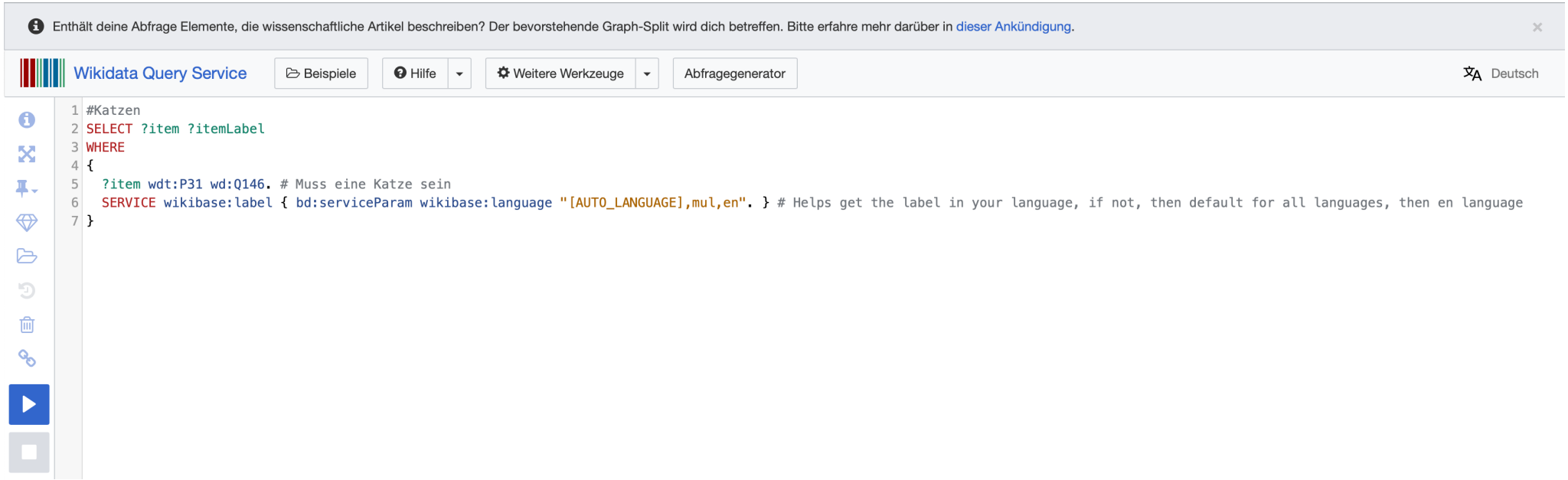
Quelle: Rehbein, Malte. 2017. “Ontologien.” In: Jannidis, Fotis, Kohle, Hubertus, Rehbein, Malte (Hg.): Digital Humanities. J.B. Metzler. https://doi.org/10.1007/978-3-476-05446-3_11

Praktische Übung: SPARQL-Abfrage von Wikidata

Senden Sie Abfragen an Wikidata mit dem Wikidata Query Service: https://query.wikidata.org/

- Unter ‘Examples’ können Sie Beispielabfragen laden (und diese verändern)
- Wenn Sie über den Wikidata-Klassen und -Eigenschaften hovern, sehen Sie was die Nummern bedeuten

In Wikidata können Sie sich ansehen, welche Statements überhaupt existieren (z.B. für diese Katze: https://www.wikidata.org/wiki/Q851190)



Klassifikation von Ontologien

- Ontologien können von sehr klein und spezifisch bis sehr generell und umfassend sein
- Durch ontologische Annahmen: Simplifizierung
- Begriff Ontologie wird teils synonym verwendet für Klassifikation/Katalogisierung

- Lightweight Ontology: z.B. kontrollierte Vokabulare, Thesauri
- Dazwischen verortet: z.B. Taxonomien
- Heavyweight Ontology

- Top-Level Ontology / Upper Ontology
- Domain Ontology
- Task Ontology
- Application Ontology

Ontologien und Datenmodellierung

- Teilweise schwammige Definitionen, wenn Begriffe aus der Philosophie übertragen wurde
  - Zur Abgrenzung des Begriffs ist Unterscheidung nach Zweck sinnvoll, z.B.
    - Klassifikation/Katalogisierung → Thesaurus;
    - Zusammenhänge zwischen den Entitäten → Ontologie
- Doppelte Semantik
  - Einerseits Formalisierung notwendig für maschinelle Verarbeitung der Ontologie
  - Andererseits natürliche Sprache notwendig für Beschreibung und Beispiele, damit die Ontologie korrekt angewendet wird
    - Nur ein Label ist in der Regel nicht ausreichend zum Verständnis
- Modell und Daten sind im Triplestore enthalten und können auf dieselbe Art abgefragt werden
- Jede Ontologie ist auch ein (abstraktes) Datenmodell, aber nicht jedes Datenmodell ist eine Ontologie
- Vergleich: relationale Datenbank vs. Ontologie
  - Datenbankschema: Fokus auf Daten
  - Ontologie: Fokus auf Bedeutung

Universität zu Köln

Beispiel: IFLA Library Reference Model (LRM)
(früher: FRBR)

- Ontologie für bibliographische Metadaten
- Entitäten u.a.: Unterscheidung von... (Produkte intellektueller/künstlerischer menschlicher Tätigkeit)
  - Work
  - Expression
  - Manifestation
  - Item
- LRMoo: Harmonisierung des LRM und CIDOC CRM

Universität zu Köln

Beispiel: IFLA Library Reference Model (LRM)
(früher: FRBR)

Work

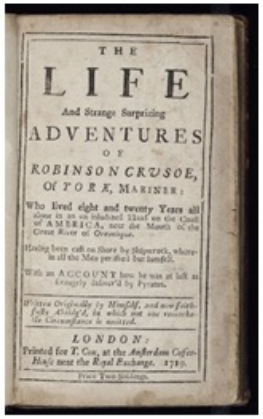
Robinson Crusoe

Expression
text version

"Cox-Abridgment" or
"Amsterdam Coffee-House edition"
(first English abridgement)

Ludwig Vischer's German version
(first translation into German)

Manifestation
edition


THE
LIFE
And Strange Surprizing
ADVENTURES
OF
ROBINSON CRUSOE,
OF YORK, MARINER:
LONDON:
Printed for T. Cox, at the Amsterdam Coffee-House near the Royal Exchange. 1719.

Piracy edition
(London: T. Cox , 1719)

Was Leben
und die
ganz ungemeine Begebenheiten
des Weltberuffenen Engelländers/
Mr. ROBINSON
CRUSOE,
welcher durch Sturm und Schiffbruch (worin
alle seine Reise-Gefährten elendiglich ertrunken) auff
der Americanischen Küste/worin an dem grossen Fluß
Oroonoko, auff eine unbewohnte Insul getrieben, allwo
und zwantzig Jahr lang darauff gelebet, und gulest
durch Gott-Räuber wunderbahrer Weise da-
von befreyet worden.
Die Zweyte Hamburgische Auffage.
Dis Authoris Bedancken über
die ungleich-grösere Anzahl derer Heyden/
auff den Christen auf dem Erdboden, und einem unver-
muthlichen Vorschlag, zu der Christenthums
ungemeinen Erweiterung:
Ingleichen einem accurateen Abriß obbemeld-
ter Insul vermehret.
HAMBURG, gedruckt und verlegt durch Joh. Thomas von
Wierings Erben bey der Börse, im güldenen A,B,C. 1721.

2nd edition
(Hamburg: Wierings Erben, 1721)

Das Leben
und die ganz ungemeine
Begebenheiten
Des
ROBINSON
CRUSOE,
Welcher unter andern auf der Americani-
schen Küste durch Sturm Schiffbruch erlitten,
und bey dem Ausfluß des grossen Stroms Oroonoko an
eine unbewohnte Insul verschlagen worden, auf welcher
er über acht und zwantzig Jahr biß zu seiner wunder-
baren Befreyung gelebet hat.
Von ihm selbst beschrieben, und um seiner
Fürtrefflichkeit willen aus dem Englischen
ins Teutsche übersetzt.
Die fünffte Auflage
Mit vielen Kupffern nebst einer accuraten Land-Charte.
Der erste Theil.
Franckfurth und Leipzig,
1720.

Piracy edition
(Frankfurt/ Leipzig 1720)

Item
copy

British Library
G.13243

Hamburg Staats u. Uni-Bibliothek
Scrin A/211

Weimar Anna Amalia Bibliothek
309306 - A

HAB Wolfenbüttel
Lq.250:1

CIDOC Conceptual Reference Model (CRM)

The primary role of the CIDOC CRM is to enable the exchange and integration of information from heterogeneous sources for the reconstruction and interpretation of the past at a human scale, based on all kinds of material evidence, including texts, audio-visual material and oral tradition. It starts from, but is not limited to, the needs of museum documentation and research based on museum holdings. It aims at providing the semantic definitions and clarifications needed to transform disparate, localised information sources into a coherent global resource, be it within a larger institution, in intranets or on the Internet, and to make it available for scholarly interpretation and scientific evaluation. These goals determine the constructs and level of detail of the CIDOC CRM.

https://cidoc-crm.org/sites/default/files/cidoc_crm_version_7.3.1_1.pdf (S. 9)

- Kontext/Anwendungsbereich: kulturelles Erbe, Museumsdokumentation
- Ziele:
  - Austausch und Integration von Informationen
  - Definition der zugrunde liegenden Semantik – semantische Interoperabilität
  - Aber: keine Vorgaben zu Implementierung

CIDOC Conceptual Reference Model (CRM)

- Entwickelt durch Arbeitsgruppen des International Committee for Documentation (CIDOC) des International Council of Museums (ICOM): zunächst durch CIDOC Documentation Standards Working Group, heute durch CIDOC CRM Special Interest Group (SIG)
  - Mitgliedschaft in der CIDOC CRM SIG durch Institutionen
    - ➢ Liste der Mitglieder
- Kontinuierliche Überarbeitung → Versionen; aktuellste stabile Version: Version 7.3.1.
- ISO-Norm seit 2006, erneuert 2014
- Soll als grundlegende, erweiterbare Basis-Ontologie dienen (wird daher auch CRMbase genannt)
  - ➢ Erweiterungen: u.a. CRMgeo, CRMarcheo, CRMinf

Hausaufgabe
zum 13. Juli 2026

Bitte schauen Sie sich die offizielle Dokumentation von CIDOC CRM 7.3.2. an, um sich ein bisschen damit vertraut zu machen:

- https://cidoc-crm.org/versions-of-the-cidoc-crm

➢ Scrollen Sie durch die einleitenden Kapitel und Diagramme (S. 9–48), überfliegen/lesen Sie bei Interesse

➢ Schauen Sie sich einzelne Klassen und Eigenschaften genauer an, die interessant klingen (Sie finden Sie diese z.B. über das Inhaltsverzeichnis)