

Daten und ihre Semantik für Menschen und Maschinen

NIE-INE
Ontologie-Workshop



31/03/2017

Agenda

- Das Was: Digitale Daten und Semantik
- Das Wie: Ein Semantisches Web
- Beispiele

Heterogene digitale Daten und Benutzer

- Schnell wachsende heterogene Datenmengen
- Zu kommunizieren mit einer wachsenden Anzahl unterschiedliche Benutzer (multidisziplinär)
- → Zunehmendes Bedürfnis für Interoperabilität von Computersystemen
- → Sich änderende Informationsbearbeitungsparadigma
- → Von menschlicher Interpretation zur **maschine-brauchbaren Semantik** → *machine reasoning*

Digitale Daten sind heterogen

- Viele unterschiedliche **natürliche Sprachen** mit eigenen Interpretationsmodellen von Wahrnehmungen
- Meist **un-/kaum strukturierte** Daten (freier Text)
- Unterschiedliche inter- aber auch intradisziplinäre **Terminologien**
- Unterschiedliche **Datenbankmodelle** (Schemata), oft ad hoc dargestellt, d.h. ohne genaue Repräsentation von Domainwissen

Was brauchen wir?

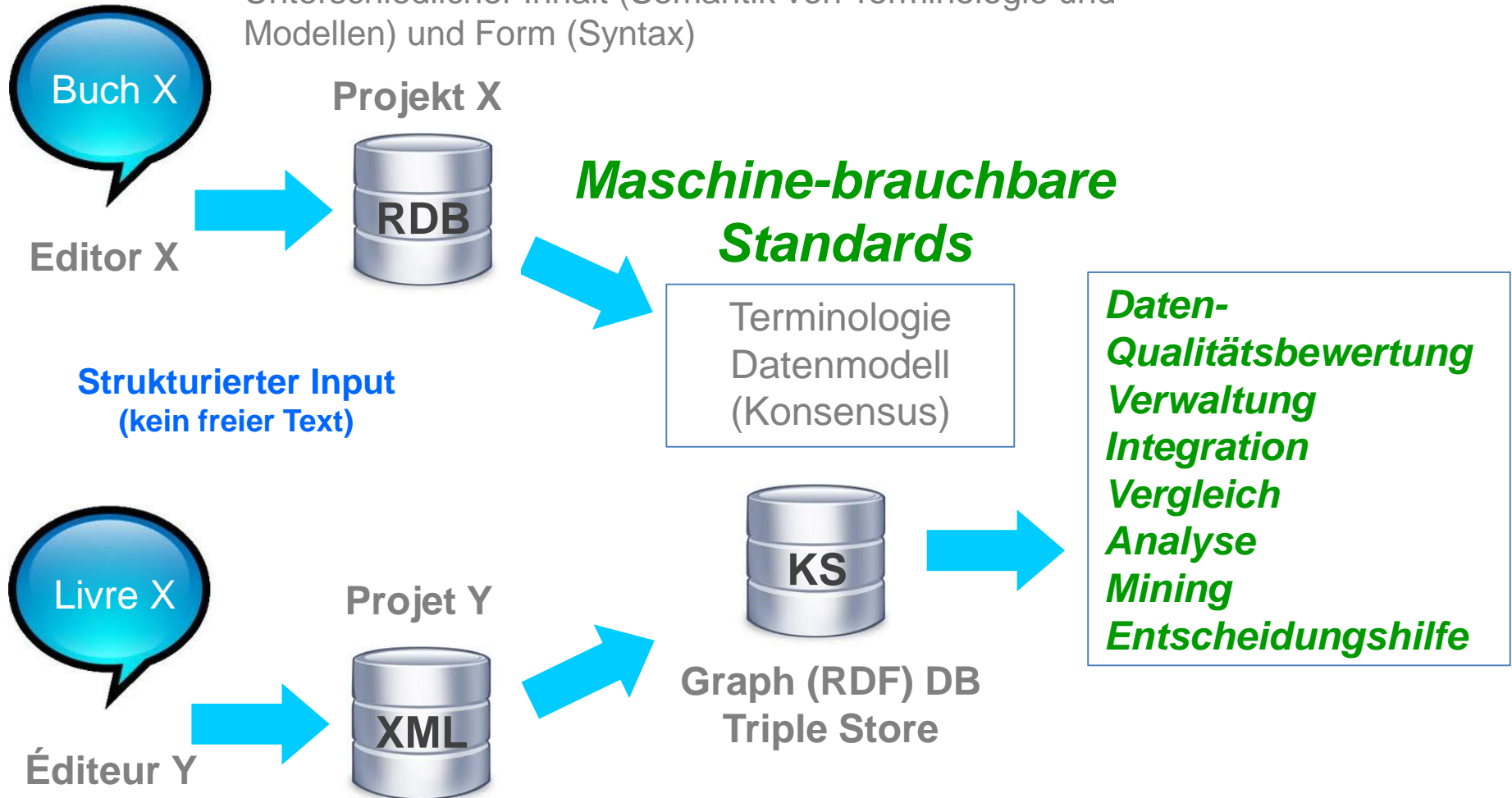
- Semantische Formalisierung von Daten unabhängig von natürlichen Sprachen
- Formal heisst: **explizit, unifiziert und maschinen-nutzbar**
- → Disambiguierte Semantik deklariert in formalen Vokabularien oder Ontologien
- → Konsensusmodell von Domainwissen

Resultat

- Standardisierte maschinen-lesbare Semantik
- Unifiziertes [transparentes] Domainwissen
- Datenqualitätsbewertung und -verbesserung (explizit)
 - → Ambiguitätsreduktion
 - → Flexiblere Datenverwaltung
 - → Semi-Automatische Datenintegration
 - → Datenvergleich, Datenanalyse, Data-Mining
 - → Entscheidungshilfe (Decision Support)

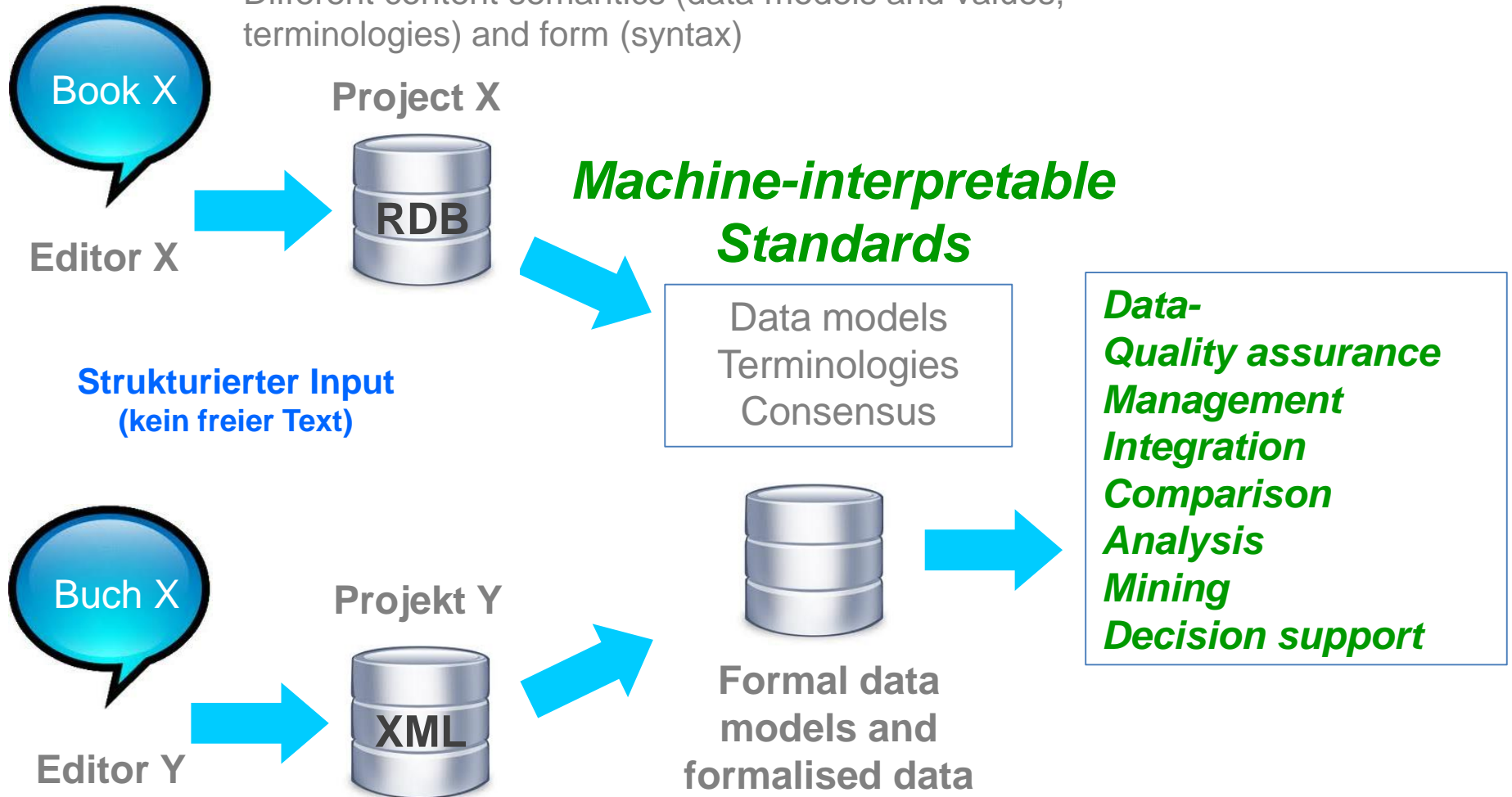
Daten über Editionen

Unterschiedlicher Inhalt (Semantik von Terminologie und Modellen) und Form (Syntax)



Daten über Editionen

Different content semantics (data models and values, terminologies) and form (syntax)



Agenda

- Digitale Daten und Semantik
- Semantisches Web
- Beispiele

Semantic Web - aka Web 3.0

- Initiative des W3C, ab 2001
- W3C: international standards organization
 - E.g. HTML, XML → *driving the web*
- Eigentlich schon Teil der ursprünglichen Vision eines Web (Tim Berners-Lee, 1990)
- Open Standards, Open Licenses
- Anwendung vom Web (neu) zur Semantik (alt)

Semantic Web Selbstbeschreibung

- Darstellung von Daten → Information (verknüpfte Daten) → Wissen (Theorie, Inferenz)
- Formale und maschinen-interpretierbare Semantik anstelle von natürlichen Sprachen
- → Verbesserte Interoperabilität von Computersystemen
- “Language of logic” (Tim Berners-Lee)
 - Basiert auf Mengenlehre, Modelltheorie, Prädikatenlogik erster Stufe
- → Global formal language

= was wir brauchen

Semantic Web Bausteine



Logik: Deduktion, Abduktion, Induktion	Folgerungsmaschine (machine reasoning)
Inferenz-Regeln	Formale Sprachen (N3 Syntax)
Abfrage	Grundstein SemWeb: wir verstehen einander
Formale Daten und Vokabulare oder Ontologien (Modellen)	Formale Sprachen: RDF, RDFS, OWL (Language of logic)
Identifikatoren	IRI: International Resource
	Grundstein WWW: wir sprechen über das Gleiche

Entwicklung eines Modells

- Menschliche Observation von [Realität]:
 - **Interpretation** durch Referenz oder Modell (Mustererkennung)
- Modell-/Interpretationstheorie:
 - **Spezifizieren von Semantik** (cf. Wörterbuch) in einer globalen formalen Sprache
 - Limitierte Interpretation von Ausdrücken
 - Minimalisierte Ambiguität
- **Explizit:**
 - Keine verborgenen Annahmen (klare beabsichtigte Bedeutung)

Mengenlehre

- **Klasse** (Sammlung, Kategorie):
Repräsentiert abstrakten Begriff, z.B. 'Person'
 - **Instanz** einer Klasse (Mitglied, Individuum):
Gleiche Semantik, z.B. 'Person_x', 'Person_y'
 - **Eigenschaft** (Relation, Prädikat):
Verknüpfen von Klassen und Mitgliedern, z.B. 'ist_ein'
- Expression: Person_x ist_eine Person
(Person_X ist ein Mitglied der Klasse "Person")

Semantic Web Expression

Einfacher Satz mit 3 Elementen

(RDF) TRIPLE

Person_x ist_eine Person.

Subject **Predicate** **Object**

Editor_x ist_ein Editor.

Person_x hat_Rolle Editor_x.

(RDF) GRAPH

Eine Menge Triples

Es gibt auch andere Triples



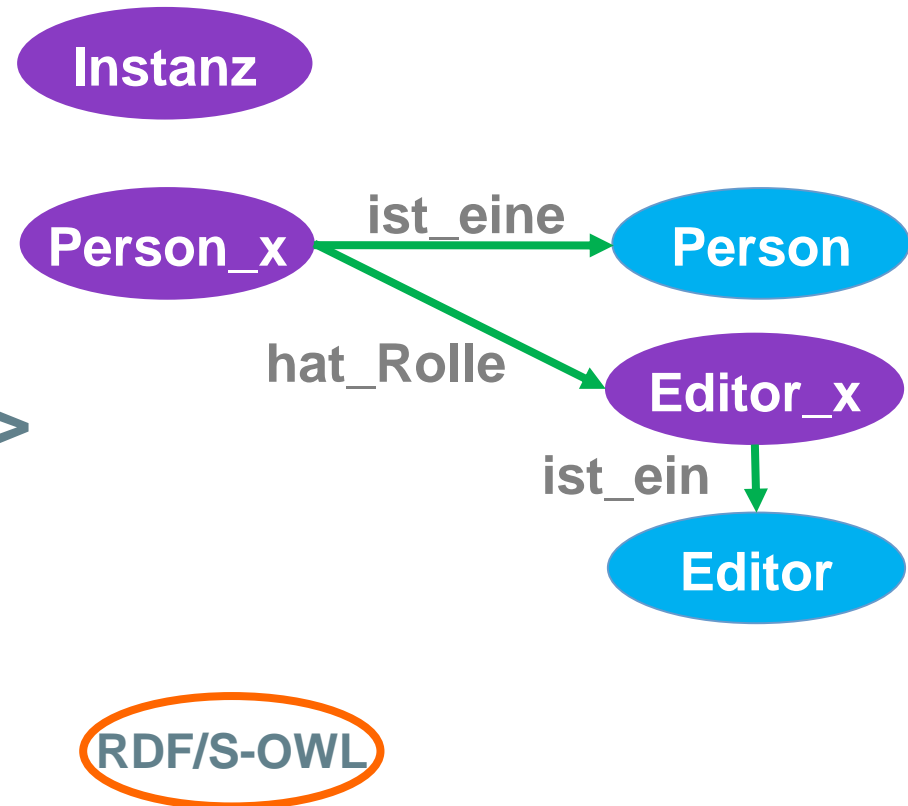
Nur 8,4%

Modellierung und Datenformalisierung

Vokabular oder Ontologie



Formale Daten



Konsensusmodell von Domainwissen. Notwendig und genügend, um Daten auszudrücken und zu formalisieren

Ontologie Definition

W3C

Conceptualization of a domain
enabling knowledge sharing¹

Representation of terms and their interrelationships²

¹ <http://www.w3.org/2005/Incubator/w3pm/XGR-w3pm-20091008/#A>

² <http://www.w3.org/TR/2004/REC-owl-features-20040210/#s1>

Ontologien modellieren: Quellen

- **Implizites Fachwissen** von Domain-Experten: heute
- **Existierende Modelle**, e.g.
 - E. Pierazzo: *Digital Scholarly Editing: Theories, Models and Methods*. Ashgate Publishing Company, 2015 - ISBN: 9781472412119/-26/-33
 - > Chapter 2: Modelling (Digital) Texts
- **Wissenschaftliche Artikel**, e.g.
 - F. Ciotti, F. Tomasi. *Formal Ontologies, Linked Data, and TEI Semantics*. <https://jtei.revues.org/1480>, 2014
- **Bücher**: E. Pierazzo, cf. supra

Ontologien modellieren: Quellen

- **Existierende RDF/S-OWL Ontologien:**
 - Open source
 - Fundamentelle: W3C: RDF/S – OWL Ontologien
 - Basische, z.B. Friend of a friend (FOAF)
 - Domain: z.B. Dublin Core (DC),
- Wachsende Anzahl von Ontologien in akademischem und industriellem Umfeld

Grundlegende Ontologien

- **Friend of a friend (FOAF):**
 - Beschreibt Akteure wie eine Person und ihre Eigenschaften, z.B.
 - Klassen: Agent, Person, Organization, Group name, address
 - Eigenschaften: familyName, givenName, name, member
 - <http://xmlns.com/foaf/0.1/>

Domain-Ontologien

- **Dublin Core (DC):**
 - Beschreibt Ressourcen, Metadaten, z.B.
 - Klassen: BibliographicResource, MediaType
 - Eigenschaften: creator, publisher, title, language, subject
 - Standard ISO15836
 - <http://purl.org/dc/elements/1.1/>
 - Extension: <http://purl.org/dc/terms/>

Domain-Ontologien

- **FRBR:**
 - Functional Requirements for Bibliographic Records
 - <http://metadataregistry.org/schema/show/id/94.html>
 - Klassen: z.B. Publication Expression, Publication Event
- **SPAR-Serie:**
 - Semantic Publishing and Referencing
 - Document Components Ontology:
<http://www.sparontologies.net/ontologies/doco>
 - FRBR-aligned Bibliographic Ontology
<http://www.sparontologies.net/ontologies/fabio>

Agenda

- Digitale Daten und Semantik
- Semantic Web
- Beispiele

Beispiele

- **Generell: DBPedia:**
 - “RDF-izing” Wikipedia
 - Large, Multilingual, Semantic Knowledge Graph
 - <http://wiki.dbpedia.org/>
- **Europeana:**
 - Liberating Cultural Heritage for Use in Research
 - <http://labs.europeana.eu/api/linked-open-data-data-structure>
 - Europeana Data Model (EDM)
 - Plattform mit Triplestore und Abfragesystem (SPARQL endpoint)

Beispiele

- **Knora, Digital Humanities Lab, UniBas:**
 - Knowledge Organization, Representation, and Annotation
 - Software framework for storing, sharing, and working with primary sources and data in the Humanities
 - Applikationsorientierte Semantik mit gewisse Bedürfnisse
 - Grundlage von NIE-INE
 - NIE-Projektontologien werden mit den Knora Ontologien verknüpft, z.B. der Knora Base Ontology



Danke