

Daten und Daten über Daten

Ein Zentrales Problem des Internet ist die Repräsentation der Semantik der Ressourcen.

- **XML**: Markup zur Erfassung von Struktur
- **Namespaces**: Definition eines Vokabulars zur Erfassung von Struktur
- **Dublin Core**: Standardvokabular zur Beschreibung von beliebigen Ressourcen (15 Elemente)

Problem: In welcher Form (Modell) sollen diese Bausteine im WEB kombiniert werden, um die Sacherschließung von Ressourcen (= Daten) im Information Retrieval auszunützen.

Gesucht ist also ein Standardmodell für Metadaten, d.h. Daten über Daten.

Beachte: Diese Definition bedeutet, dass Metadaten Daten sind, das Präfix *Meta-* wird nur durch den Kontext bestimmt und kann im Prinzip beliebig geschachtelt werden.

Einfachstes Beispiel:

In einer relationalen Tabelle sind die Attributnamen (= Beschreibungselemente der Daten) Metadaten.

In einem Data Dictionary (Codebook) sind diese Metadaten Daten. Die Beschreibungselemente dieser Metadaten sind Meta-Metadaten)

.....

Semantic Web

*Definition: The **Semantic Web** is the abstract representation of **data** on the World Wide Web, based on the RDF standards and other standards to be defined. It is being developed by the W3C, in collaboration with a large number of researchers and industrial partners.*

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." -- Tim Berners-Lee, James Hendler, Ora Lassila, [The Semantic Web](#), Scientific American, May 2001

Ein Beispiel zur Erläuterung des Grundprinzips von RDF

(nach K. Ahmed et al.: *Professional XML Metadata*, Wrox Press Ltd., 2001; ISBN 1861004516)

Unterscheide drei Repräsentationen zur Erfassung von Struktur

a) Beschreibung in HTML

```
<HTML>
  <HEAD>
    <TITLE> Glücklich ist wer Pensionist</TITLE>
    <META NAME= "DC.Title" CONTENT="Glücklich ist
      wer Pensionist">
    <META NAME= "DC.Creator" CONTENT="Norbert
      Stanzel">
    <META NAME= "DC.Subject" CONTENT=
      "Pensionsreform in Österreich">
  </HEAD>
  <BODY>
    Vorweg: Eine Pensionsreform.....
  </BODY>
</HTML>
```

b) Beschreibung als XML Dokument (Nur Metadaten)

```
<document type="News Comment"
  xmlns:dc="http://purl.org/dc/elements/1.1">
  <dc.Title> Glücklich ist wer Pensionist</dc.Title>
  <dc.Creator>Norbert Stanzel</dc.Creator>
  <dc.Subject>Pensionsreform in Österreich</dc.Subject>
</document>
```

c) Darstellung der Metadaten als relationale Tabelle

c1) Mit individuellen (lokalen)Tabellenköpfen (Attributnamen)

Titel	Author	Inhalt
Glücklich ist wer Pensioniost	Norbert Stanzel	Pensionsreform in Österreich

c2) Mit standardisierten Tabellenköpfen durch URLs

http://purl.org/dc/elements/1.1/Title	http://purl.org/dc/elements/1.1/Creator	http://purl.org/dc/elements/1.1/Subject
Glücklich ist wer Pensioniost	Norbert Stanzel	Pensionsreform in Österreich

Einheitliches Format aller Darstellungen:

Metadaten werden als **Name/Value-Pairs** repräsentiert

Name	Value
http://purl.org/dc/elements/1.1/Title	Glücklich ist wer Pensioniost
http://purl.org/dc/elements/1.1/Creator	Norbert Stanzel
http://purl.org/dc/elements/1.1/Subject	Pensionsreform in Österreich

Beachte Unterschied zwischen HTML und den anderen beiden Darstellungen:

HTML zeigt uns das Dokument und keine Metadaten, die anderen beiden Repräsentationen geben uns explizit die Metadaten als Daten.

In XML können wir leicht die Referenz zum Objekt (= Ressource) einfügen:

```
<document type="News Comment"
  url=" http://kurier.at/archiv/....."
  xmlns:dc="http://purl.org/dc/elements/1.1">
  <dc.Title> Glücklich ist wer Pensionist</dc.Title>
  <dc.Creator>Norbert Stanzel</dc.Creator>
  <dc.Subject>Pensionsreform in Österreich</dc.Subject>
</document>
```

Dies entspricht einer relationalen Triple-Struktur

Dokument/Name/Value

Document	Name	Value
http://kurier.at/archiv/.....	http://purl.org/dc/elements/1.1/Title	Glücklich ist wer Pensionist
http://kurier.at/archiv/.....	http://purl.org/dc/elements/1.1/Creator	Norbert Stanzel
http://kurier.at/archiv/.....	http://purl.org/dc/elements/1.1/Subject	Pensionsreform in Österreich
http://kurier.at/archiv/.....	Type	News Comment

Wir bezeichnen solche Tripel allgemein als

Resource-property-value Triples

Diese Tripel können offensichtlich alle Arten von Information darstellen

Etwas umständlich, aber strukturell einfach!!

Was wir benötigen ist auf Basis von XML ein formales Modell und eine Syntax, die diese Struktur repräsentieren kann.

Dadurch eröffnet sich die Möglichkeit derart repräsentierte Semantik über das WEB in standardisierter Form zu transportieren.

RDF = RESOURCE DESCRIPTION FRAMEWORK ist ein derartiges Modell