

# **Web Science**

## Übung 1: Einrichtung VM / Datenmodelle

**Prof. Dr. Bogdan Franczyk**  
**Institut für Wirtschaftsinformatik – Professur für Informationsmanagement**

**Sommersemester 2017**

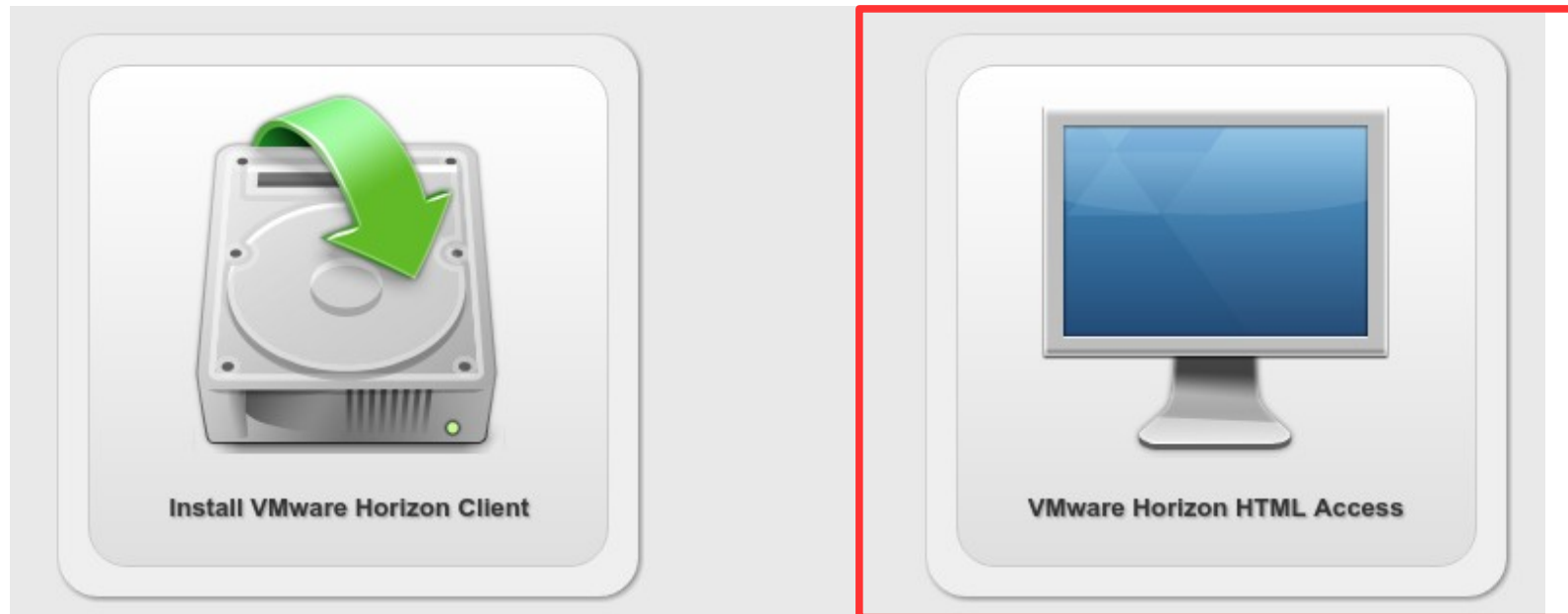
# Überblick für heute

- Zugang zur Arbeitsumgebung, Vertraut machen mit Tools
- Modellierungsbeispiel aus Vorlesung in verschiedenen Datenmodellen explorieren
- Vor- und Nachteile eruieren
- *Wie verdienen sich unsere Datensätze den vierten bzw. fünften Stern?*

## Zugang zu Ihrer Arbeitsumgebung

*Achtung, diese Anleitung ist für die externe Nutzung!  
Falls Sie der Übung in den Pools folgen wollen, überspringen Sie diesen Schritt!*

Link hinterlegt im Moodle: <https://view.wifa.uni-leipzig.de>



# Zugang zu Ihrer Arbeitsumgebung

- Login mit Zugangsdaten:  
z.B. *wir16abc*
- Wichtig! Ändern Sie die  
Domäne auf „DOM“!

vmware Horizon

Uni-Login

.....

DOM

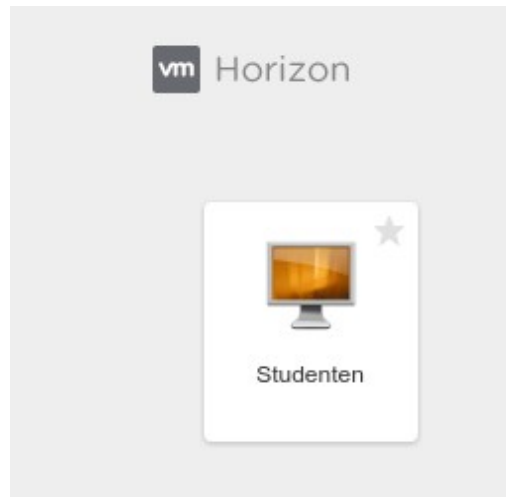
Login

Cancel

[Help with VMware Horizon](#)

## Zugang zu Ihrer Arbeitsumgebung

- Auswahl der BFR VM, auf welcher der Zugang zu allen benötigten Werkzeugen gegeben ist:



- Achten Sie darauf, Daten nur unter ihrem persönlichen Ordner zu speichern, sonst **verfallen** diese!

# Datenmodelle

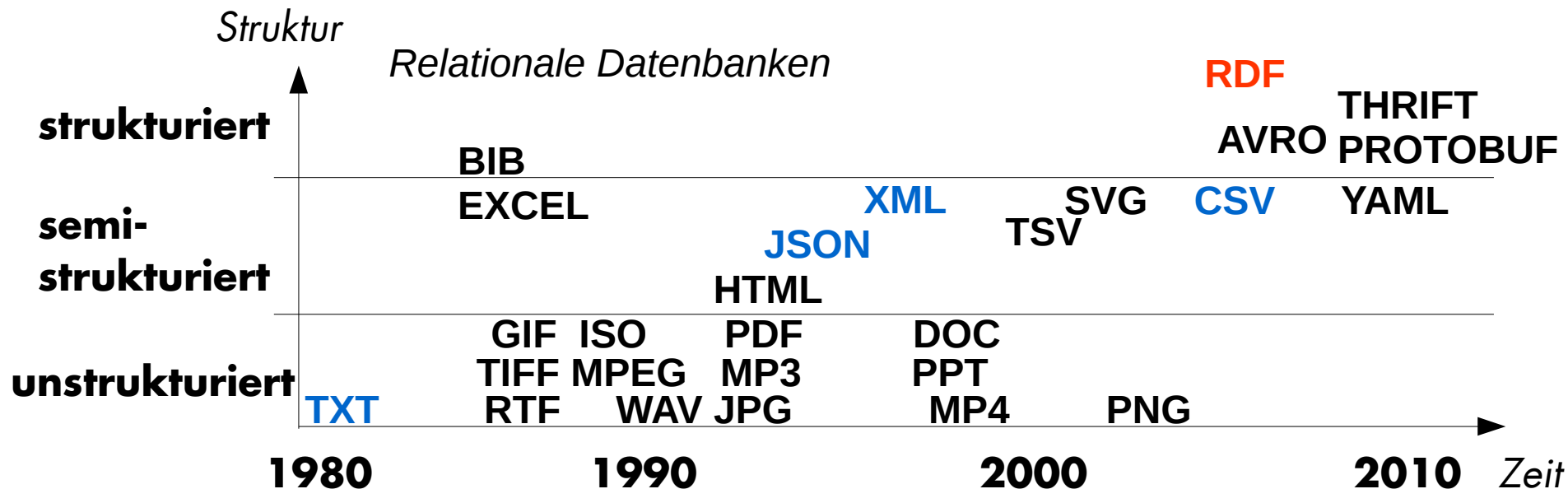
- Welche Datenmodelle gibt es?
- Was macht diese Datenmodelle relevant?
- Was sind deren Vorteile / Nachteile?
- Welche (Datenbank-) Softwaresysteme implementieren das Datenmodell?
- Welche Anfragesprachen existieren?

# Modellierungsbeispiel

- Woran misst sich Datenqualität?
  - Struktur
  - Konsistenz
  - Integrität

# Datenformate / - struktur

- Sind notwendig, um Daten digital darzustellen
- Oftmals durch Dateisuffix *impliziert*, z.B. bericht.pdf
- Struktur und Format stehen miteinander in Beziehung





# Klassifikation der Struktur von Datenformaten

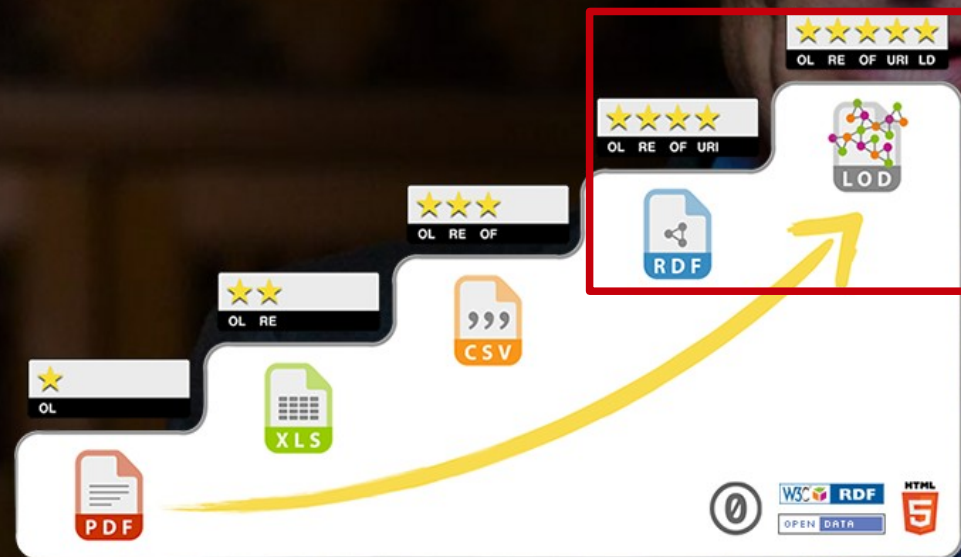
- **Strukturierte** Daten enthalten Metadaten und Daten und folgen einem Datenmodell, Syntax und Semantik sind für die Maschine dargestellt
  - Daten sind **sehr gut verwendbar**!
- **Semistrukturierte** Daten folgen einer vorgeschriebenen Syntax / Grammatik. Sie sind für die Maschine lesbar.
  - **Aufwand unterschiedlich**, für Rohdaten typisch!
- **Unstrukturierte** Daten folgen weder einem einheitlichen Schema noch einer Grammatik
  - der Aufwand, die Daten zu verwerten, ist **hoch**

# Wie klassifiziert man Datenstruktur?

→ <http://5stardata.info/de/>

## 5 ★ OFFENE DATEN

Tim Berners-Lee, der Erfinder des World Wide Webs und Initiator von Linked Data schlug ein 5-Sterne-Modell für Offene Daten (Open Data) vor. Hier geben wir Beispiele für jede Stufe und erklären die damit verbundenen Kosten und Nutzen.

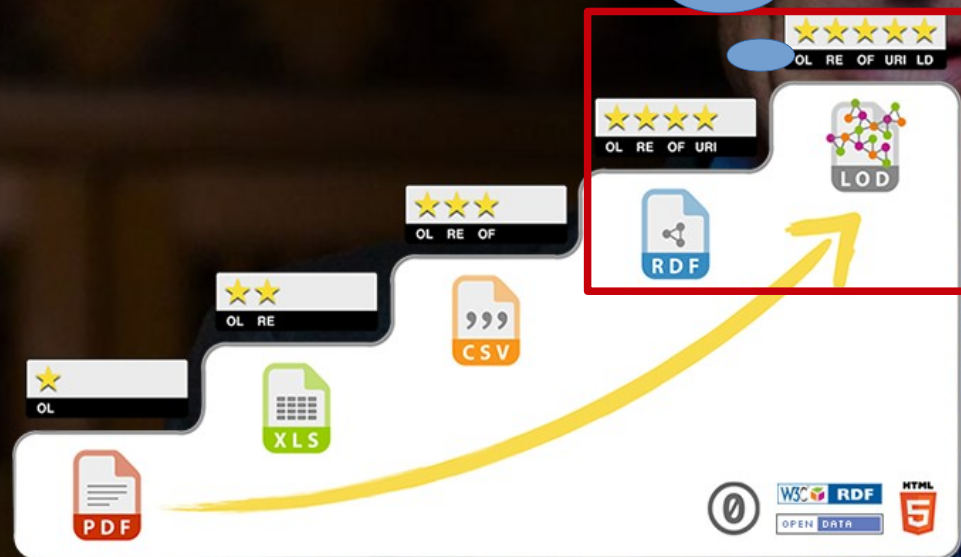


# Wie klassifiziert man Datenstruktur?

→ <http://5stardata.info/de/>

Wie verdienen sich unsere Datensätze den vierten bzw. fünften Stern?

Tim Berners-Lee, ... und Initiator von Linked Data schlug ein **5-Sterne-Modell** für **Offene Daten** (Open Data) vor. Hier geben wir Beispiele für jede Stufe und erklären die ... den Kosten und Nutzen.



## Modellierungsbeispiel (1)

- Es existieren verschiedene Himmelskörper, *Sterne*, Planeten, Trabanten und Asteroiden.
- Jeder dieser Gebilde besitzt einen Namen, Radius.
- Planeten besitzen einen Tag-Nacht Zyklus.
- Sterne haben eine geschätzte Restlebensdauer.
- Himmelskörper können andere umkreisen, d.h. ihr Satellit sein.
  - Closed World assumption / Open World assumption?

## Modellierungsbeispiel (2)

- Folgende Daten sollen eingefügt werden:
  - Sonne (radius:  $6957e5$ , lebenszeit:  $6e9$  jahre)
    - Merkur (radius:  $2439.7$ , tagnachtzyklus:  $1407.6$  h)
    - Venus (radius:  $6051.9$ , tagnachtzyklus:  $5832.45$  h)
    - Erde (radius:  $6372.8$ , tagnachtzyklus:  $34$  h)
      - Mond (radius:  $1737.1$ )
    - Mars (radius:  $3402.5$ , tagnachtzyklus:  $24.616666666$  h)
      - Phobos (radius:  $11.1$ )
      - Deimos (radius:  $6.2$ )

## Modellierungsbeispiel (3)

- Es soll die Dauer hinzugefügt werden, die beim Umkreisen eines Satelliten benötigt wird
- Wie könnte ein Anfrage auf alle Monde (Satelliten von Planeten) erfolgen?
  - Was ist, wenn Daten da sind, aber keine Klarheit über das gesamte Schema, die Miniwelt, besteht?

## Take away

- Eine standardisierte Möglichkeit, Metadaten zu formulieren und die Semantik hinter den Metadaten einer Maschine zugänglich zu machen, ermöglicht eine tiefere maschinelle Verständlichkeit des Datensates, d.h. den vierten Stern im Lee'schen Klassifikationsschema
- Die Verknüpfung der Datensätze mit anderen in diesem Kontext ermöglicht eine wiederum bessere maschinelle Datenverständlichkeit (z.B. DBPedia)
- Kodierung des Schemas (Metadaten) Standardisieren!
- Fragestellungen: Was sind Daten, was Metadaten?  
Wie kodiert eine Impl. eines Datenmod. impliziert Wissen?