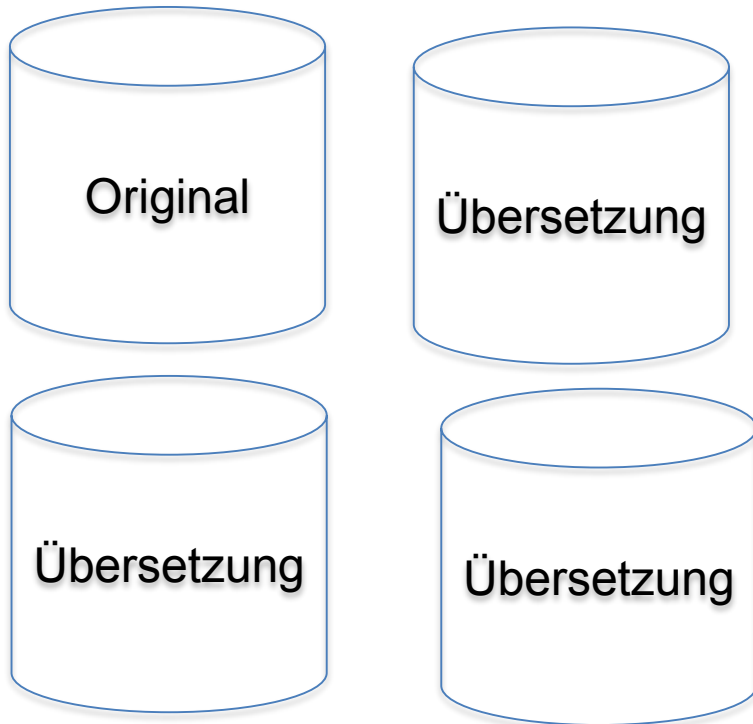


# “Empfehlungen für die Zukunft – Mehrwerte für Ihre Daten”

**Prof. Felix Sasaki**  
**(Deutsches Forschungszentrum für Künstliche Intelligenz)**



- Proprietäre Formate
- Keine Verknüpfung der Daten
- Keine Wiederverwendung außerhalb des Kontexts

## „5 Star“ Linked Data\*

1. Daten im Web
2. Maschinenlesbare Form
3. Standards nutzen
4. RDF als zentrales Format
5. Verknüpfung von Daten

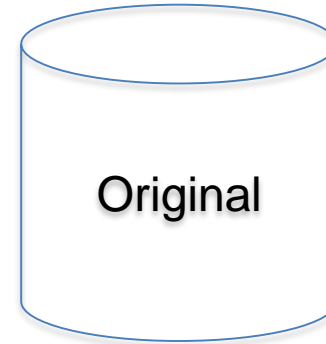


\* Vgl.  
<http://www.w3.org/DesignIssues/LinkedData.html>



„5 Star“ Linked Data:

1. **Daten im Web**
2. Maschinenlesbare Form
3. Standards nutzen
4. RDF als zentrales Format
5. Verknüpfung von Daten



Web =  
zusätzlicher  
Publikationskanal  
für Daten



<http://example.com/content/original/file1>



„5 Star“ Linked Data:

1. Daten im Web
- 2. Maschinenlesbare Form**
3. Standards nutzen
4. RDF als zentrales Format
5. Verknüpfung von Daten



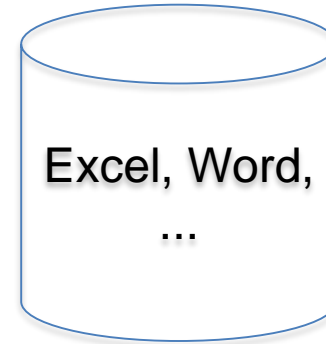
<http://example.com/content/original/file1>  
Filetyp: HTML, CSV, XML, ...

## „5 Star“ Linked Data:

1. Daten im Web
2. Maschinenlesbare Form
- 3. Standards\* nutzen**
4. RDF als zentrales Format
5. Verknüpfung von Daten

\*Standards = Kosten sparen durch

- ✓ Bessere Interoperabilität
- ✓ Nutzbarkeit von Open-Source Tools
- ✓ Standardisierte Technologien zumeist lizenzfrei („royalty-free“) verfügbar
- ✓ Offene Formate erleichtern Erweiterungen

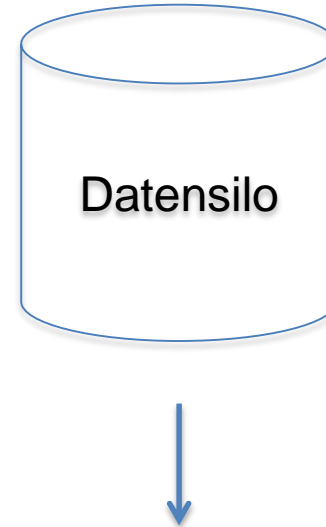


<http://example.com/content/original/file1>  
Filetyp: TBX, TMX, XLIFF (nicht verlinkt) >  
RDF (verlinkt)



„5 Star“ Linked Data:

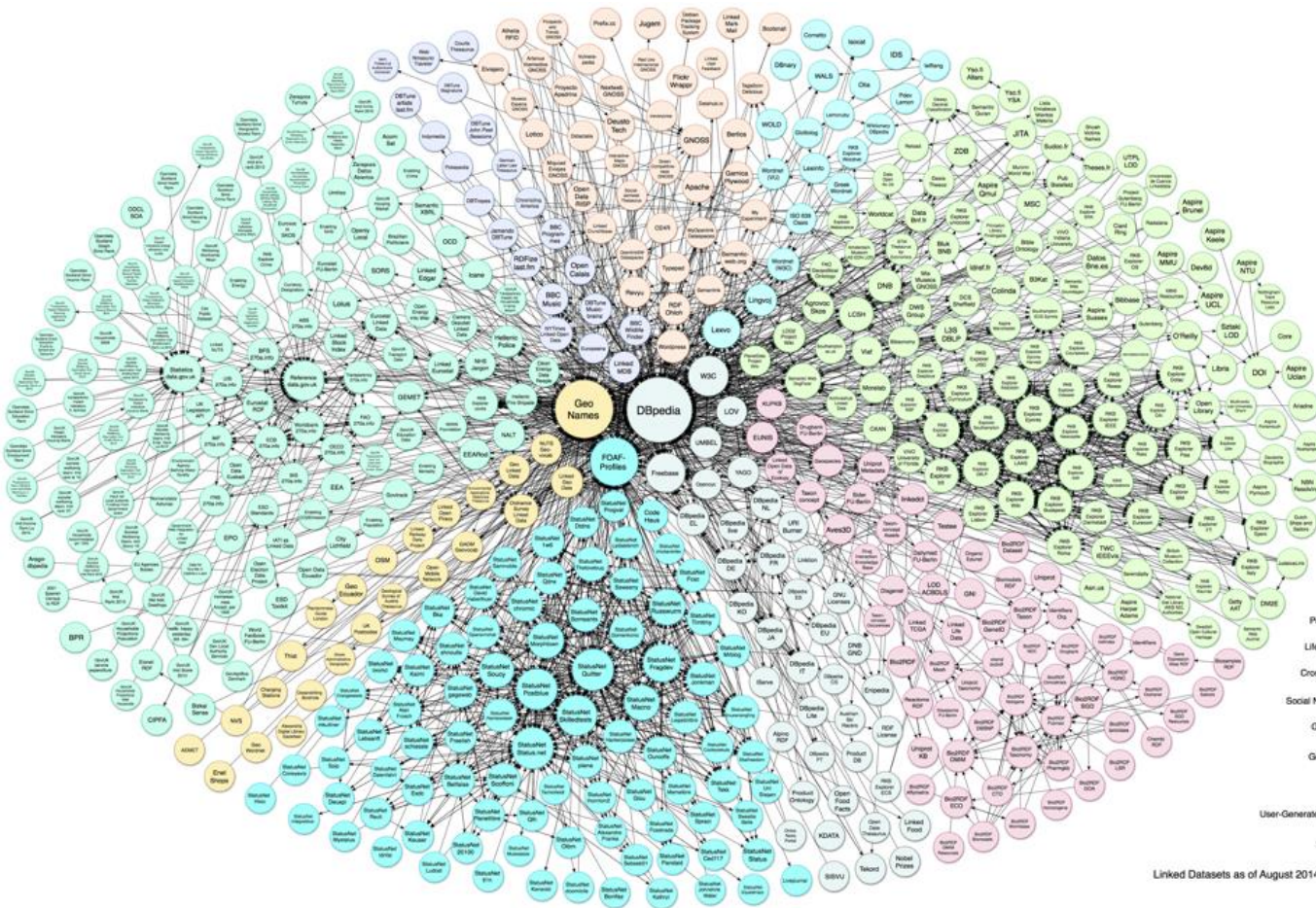
1. Daten im Web
2. Maschinenlesbare Form
3. Standards nutzen
4. **RDF\* als zentrales Format**
5. **Verknüpfung von Daten**



\* RDF = „Resource Description Framework“, Technologie und Sammelbegriff für Linked Data Technologien



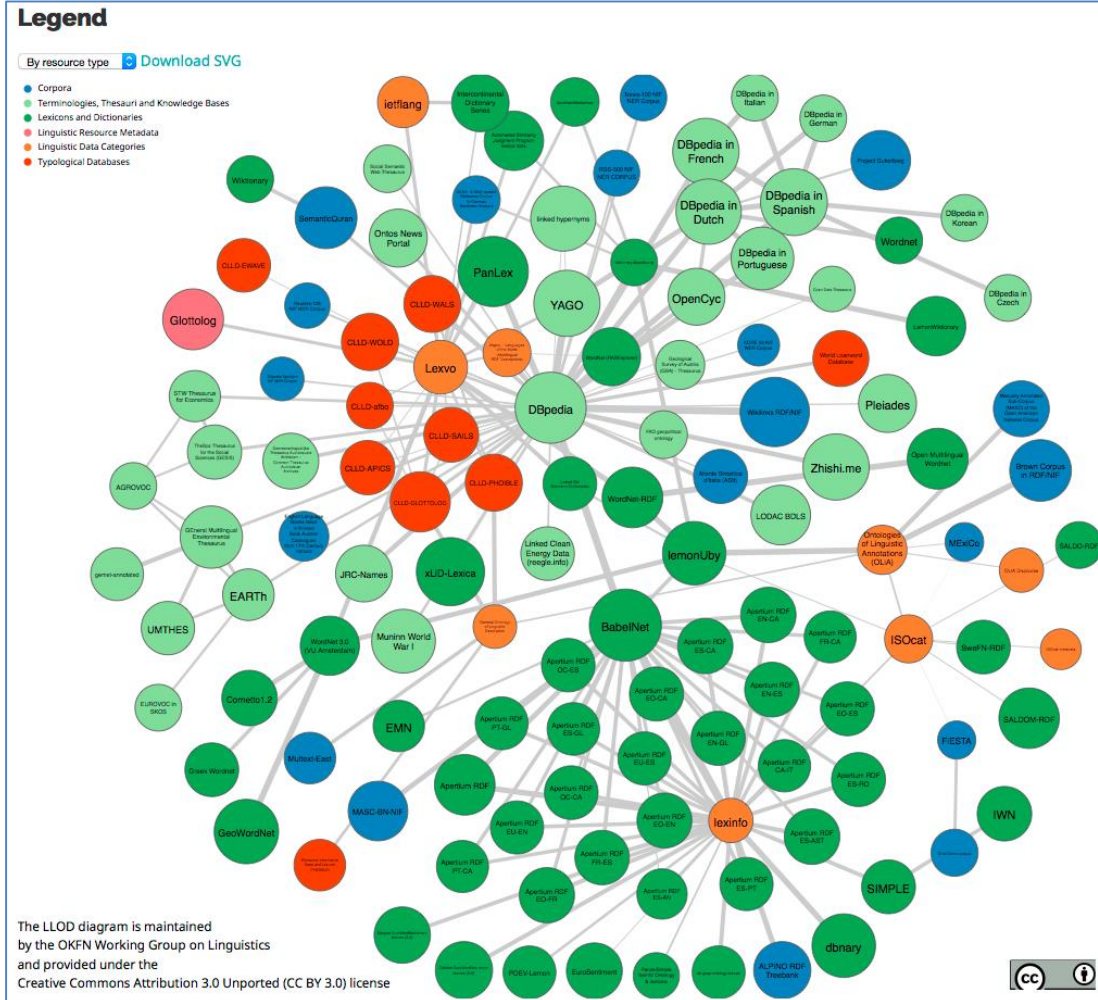
- Die Linked Data Cloud und die Linguistic Linked Data Cloud (vgl. nächste Folien)
- Kein Ersatz für TBX, TMX, XLIFF etc.; ein zusätzlicher Publikationskanal



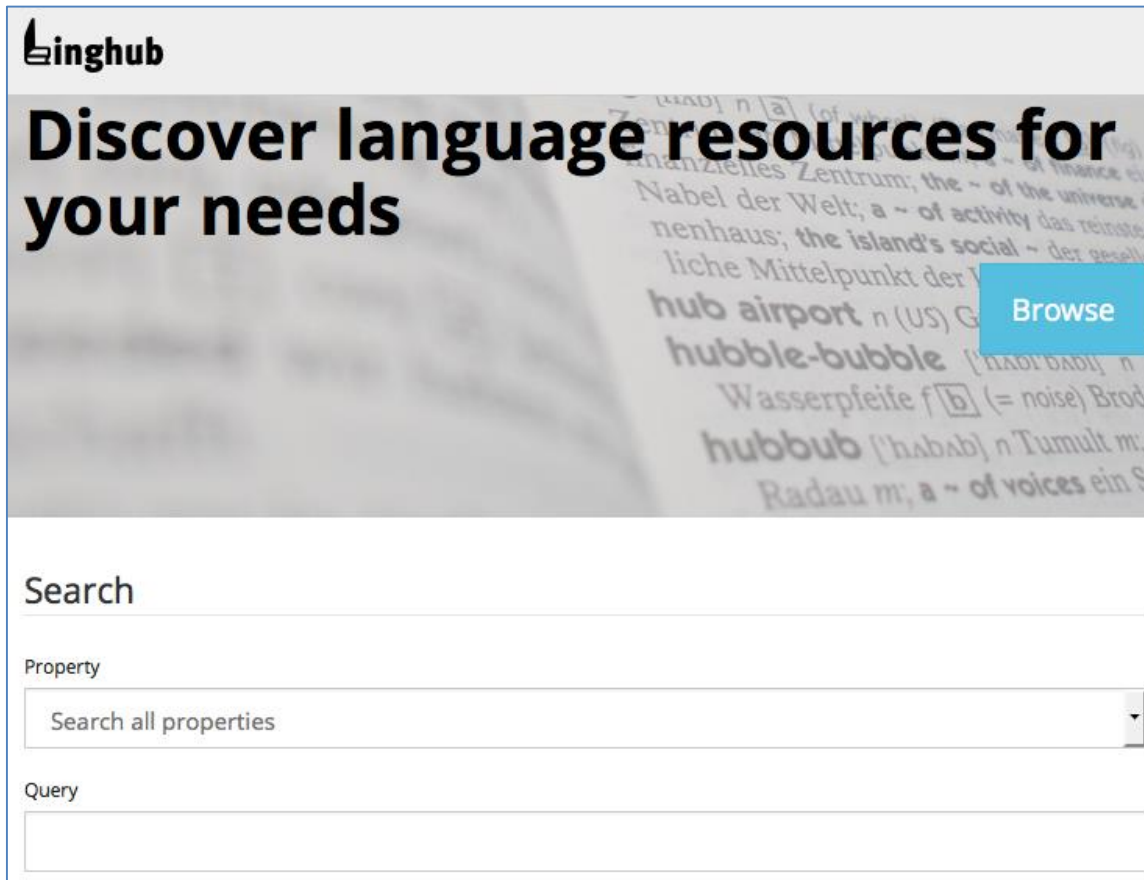
- Verlinkte Daten im Web
- Dezentral, wie das Web selber ☺
- Success stories in verschiedenen Domänen, z.B. Medizin & Bibliotheken
- Relevant für linguistische Daten – die Linguistic Linked Data Cloud

Linked Datasets as of August 2014



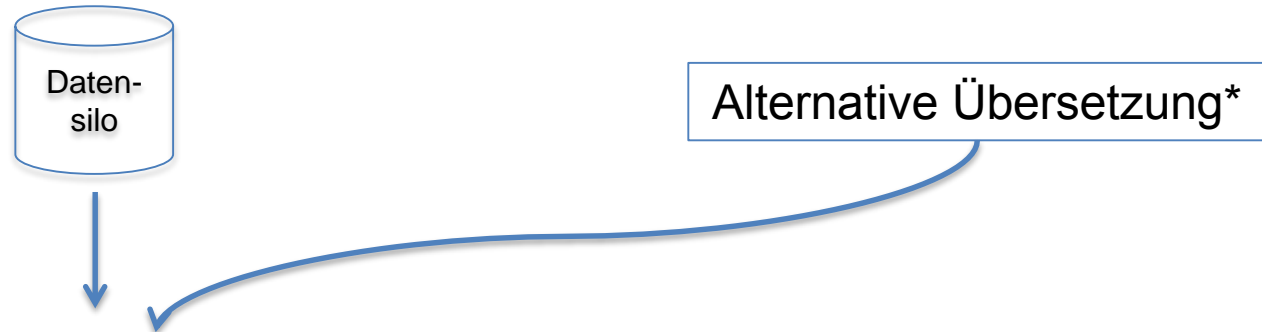


- Vgl. <http://linguistic-lod.org/lod-cloud>
- Linguistische Ressourcen als Linked Data
  - (Trainings)korpora, Terminologien, Lexika, ...
- Metadaten zu linguistischen Ressourcen
- Entdecken Sie was andere bereit stellen -> dezentral = **Mehrwert durch den Netzwerkeffekt** 😊



The screenshot shows the LingHub website interface. At the top left is the LingHub logo. Below it, the main heading reads "Discover language resources for your needs". A blue "Browse" button is visible on the right side of the main content area. Below the heading, there is a search section with a "Search" label, a "Property" dropdown menu (currently showing "Search all properties"), and a "Query" input field.

- Success Story „LingHub“  
<http://linghub.lider-project.eu/>
- Kein Registrieren von Sprachressourcen, sondern Harvesting verteilter Ressourcen anhand von Metadaten – der Netzwerkeffekt 😊
- Verknüpfung verbreiteter Metadatenschemata
  - META-SHARE, DataID, CLARIN, DCAT, ...



<http://example.com/content/uebersetzung/file1/segment1>

- Übersetzungssegmente
  - Als Linked Data repräsentieren
  - D.h. = im Web **identifizieren**
- Dadurch möglich: alternative Übersetzungen verlinken
  - Bewahrung von Provenienz zur Qualitätssicherung



Anreicherung mit weiteren Quellen



<http://example.com/content/uebersetzung/file1/segment1/term1>

- Worte, Terme und Phrasen
  - Als Linked Data repräsentieren
  - D.h. = im Web **identifizieren**
- Dadurch möglich: Verlinkungen mit weiteren Datenquellen, um Übersetzungsalternativen zu generieren oder Inhalte weiter anzureichern
  - Prominenten Quellen: BabelNet, DBpedia, EuroVoc, ...

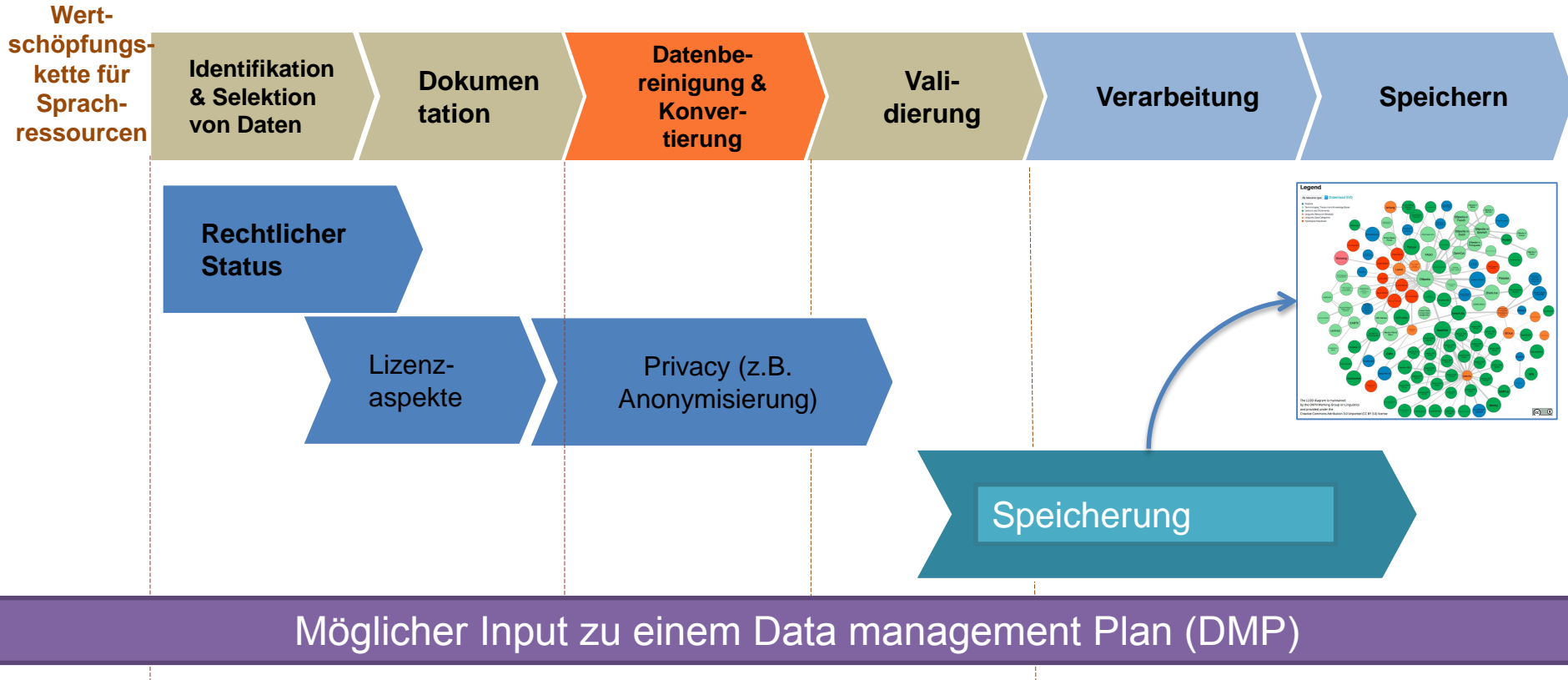


Lizenzsicherung: Nur der Besitzer  
einer Lizenz darf das Segment verarbeiten!

<http://example.com/content/uebersetzung/file1/segment1>

- Lizenzen der Daten auf kleinster Ebene sichern

# Datenmanagement von Sprachressourcen – die Rolle von Linguistic Linked Data



- Existierende Workflows müssen nicht geändert werden
- Linguistic Linked Data = ein zusätzlicher Publikationsworkflow
- Open Source Tooling und Beratung zur Nutzung von Linguistic Linked Data steht bereit 😊
- Ist Linguistic Linked Data einmal im Workflow integriert entsteht der Mehrwert von selbst – durch andere!
  - Die Nachnutzung Ihrer Daten > Der Netzwerkeffekt

- Die Zusammenfassung von übersetzungsrelevanten Daten im Web kann Mehrwerte schaffen
  - Daten: Originaldokumente, übersetzte Dokumente, Übersetzungsmemories, Terminologiedatenbanken, Lexika, ...
- Wer „5 Star Linked Data“ Prinzipien verfolgt kann diese Mehrwerte maximal ausnutzen
  1. Daten im Web
  2. Maschinenlesbare Form
  3. Standards nutzen
  4. RDF als zentrales Format
  5. Verknüpfung von Daten
- Existierende Daten und Workflows sollten nicht aufgegeben werden – Linked Data ist nur ein zusätzliches Publikationsformat
- Vorteile gibt es auf verschiedenen Ebenen
  - Sprachressourcen entdecken durch Linked Data Metadaten -> Die Linguistic Linked Data Cloud
  - Dritten erlauben die Sprachressourcen zu verbessern
  - Übersetzungen mit weiteren Informationen anreichern
- Linked Data basiert auf Standards
  - Etablierte Standards im Übersetzungsbereich: XLIFF, TBX, TMX, ...
  - Etablierte Standards im Bereich Linked Data. DCAT, LEMON, Provenance Ontology, .....
- Linked Data bietet auch Mechanismen um Rechte an Daten sicher zu stellen – nicht alles muss offen sein!





- LIDER <http://lider-project.eu/>
  - Support der Community um Linguistic Linked Data
- FALCON <http://falcon-project.eu/>
  - Linguistic Linked Data im Lokalisierungsprozess
- FREME <http://freme-project.eu/>
  - e-Services für die Anreicherung von digitalen (Übersetzungs-)workflows mit (Linguistic) Linked Data
- Verschiedene W3C Gruppen zu technischen Aspekten um Linguistic Linked Data und Standards für Übersetzung im Web
  - Vgl. Details unter <http://www.multilingualweb.eu/projects/>

**Präsentation für ELRC Workshop:**

# **“Empfehlungen für die Zukunft – Mehrwerte für Ihre Daten”**

**Author: Felix Sasaki**  
**Deutsches Forschungszentrum für Künstliche Intelligenz**