

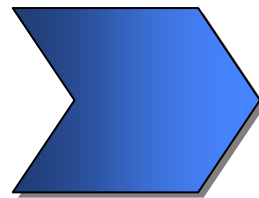
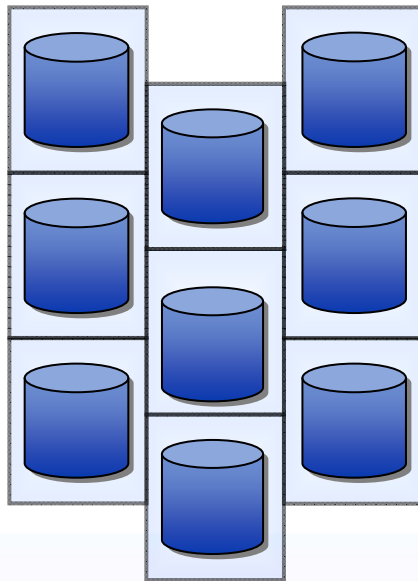
Jürgen Kett

# Linked Data Cloud: Die frühen Jahre

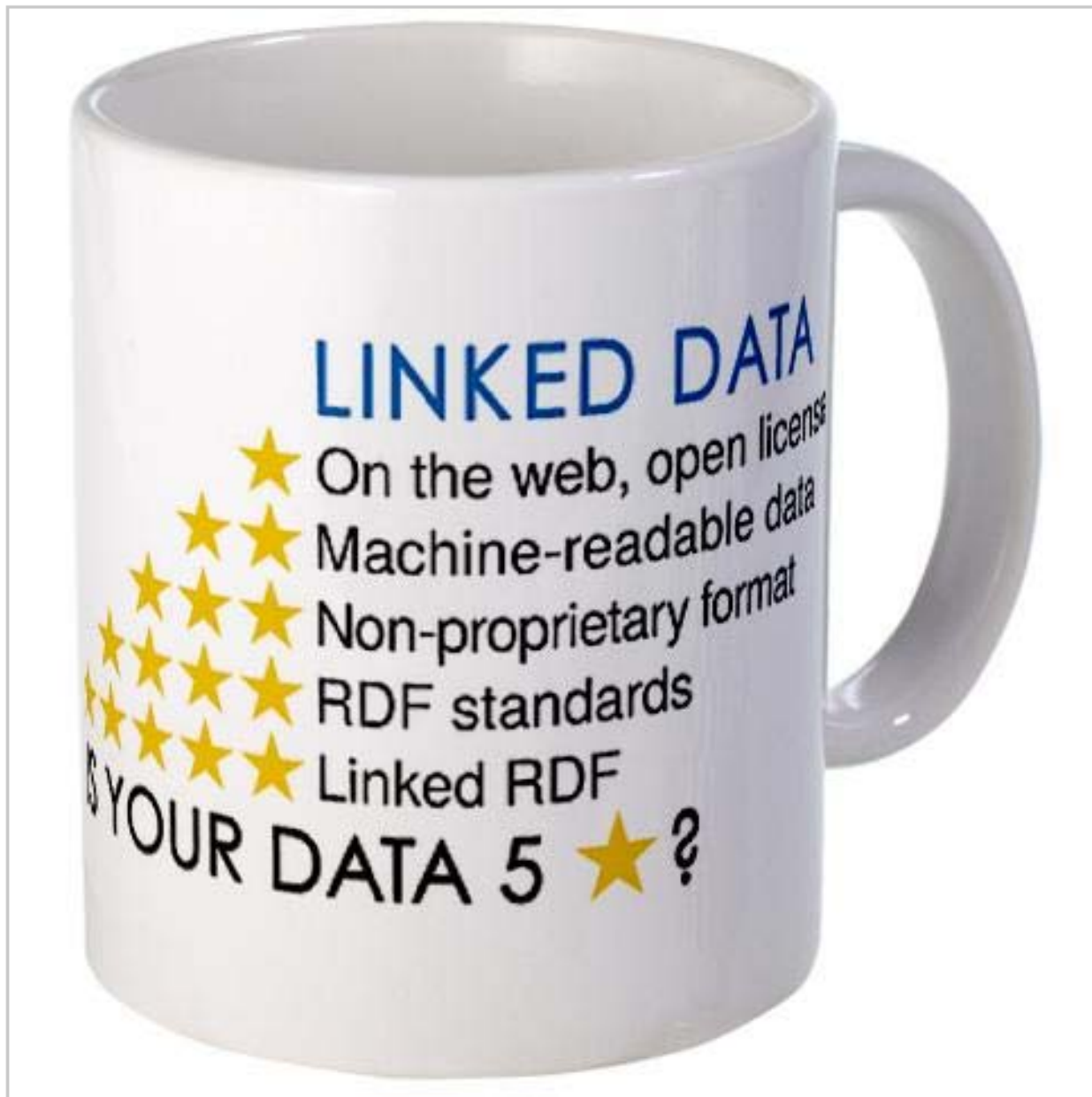


# Giant Global Graph (GGG)

Datensilos



Weltweite Datenbank  
aus vernetzten Daten



## Mug

**£10.00**

Qty:

[Add to Basket](#)

Availability: In Stock.

Product Number: 030-480759174



Like



Be the first of your fr



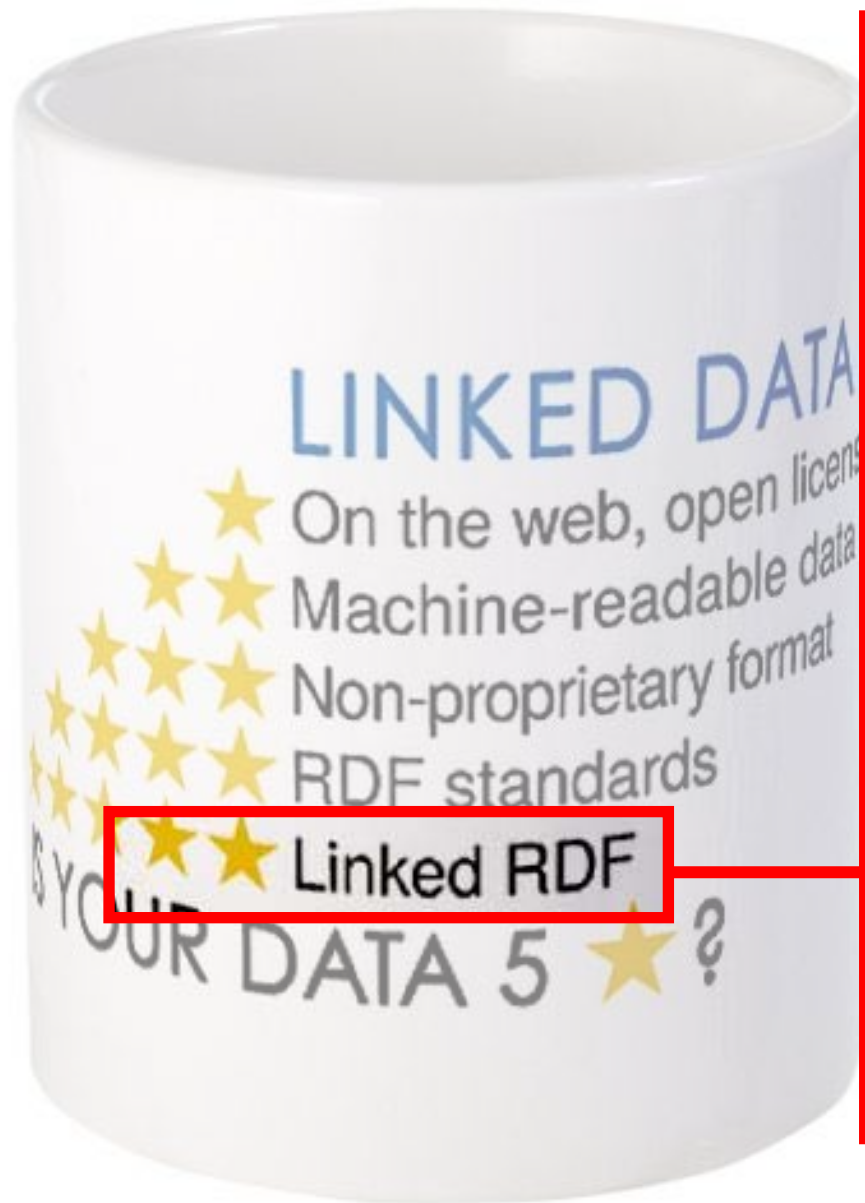
Share



### Product Information

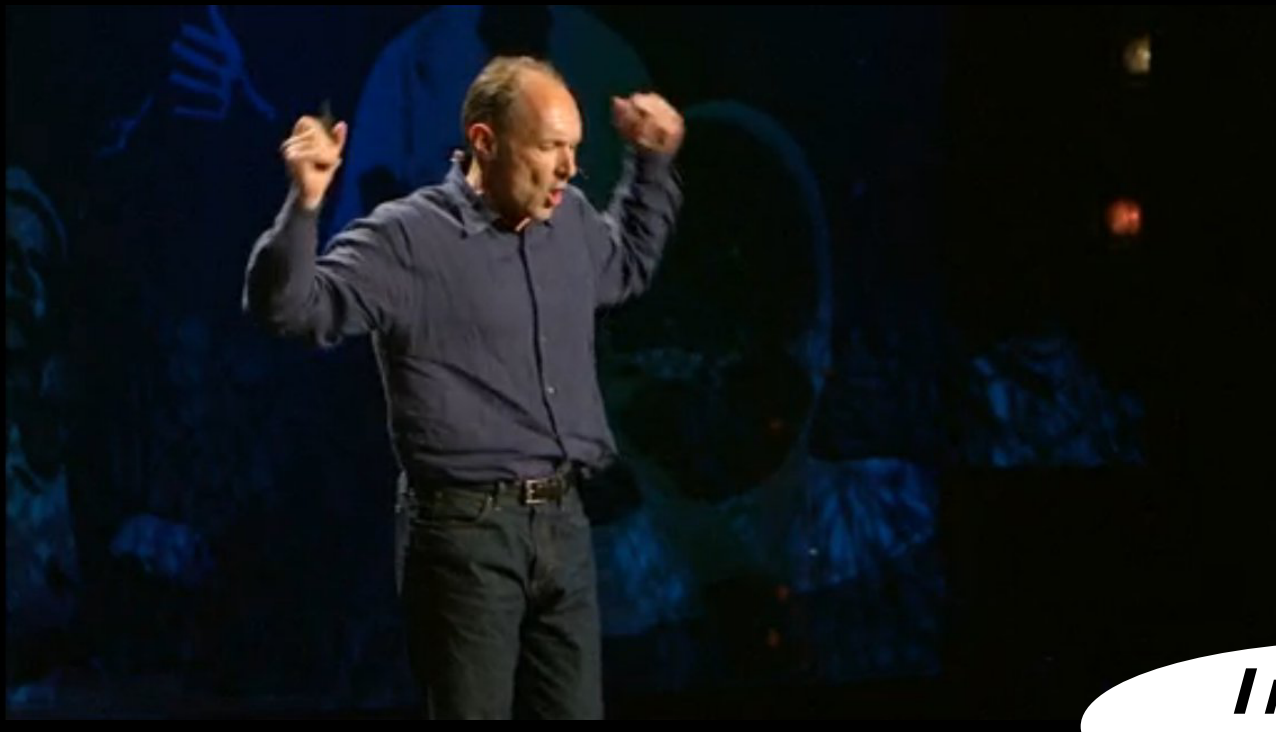
The perfect size for your favorite night brew. Large, easy-grip handle. A gift to someone special.

- Measures 3.75" tall, 3" diameter
- Dishwasher and microwave safe



1. Use **URIs** as names for things
2. Use **HTTP** URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using standards (**RDF**, **SPARQL**)
4. Include **links** to other URIs, so that they can discover more things

*([Tim Berners-Lee, 2007](#))*



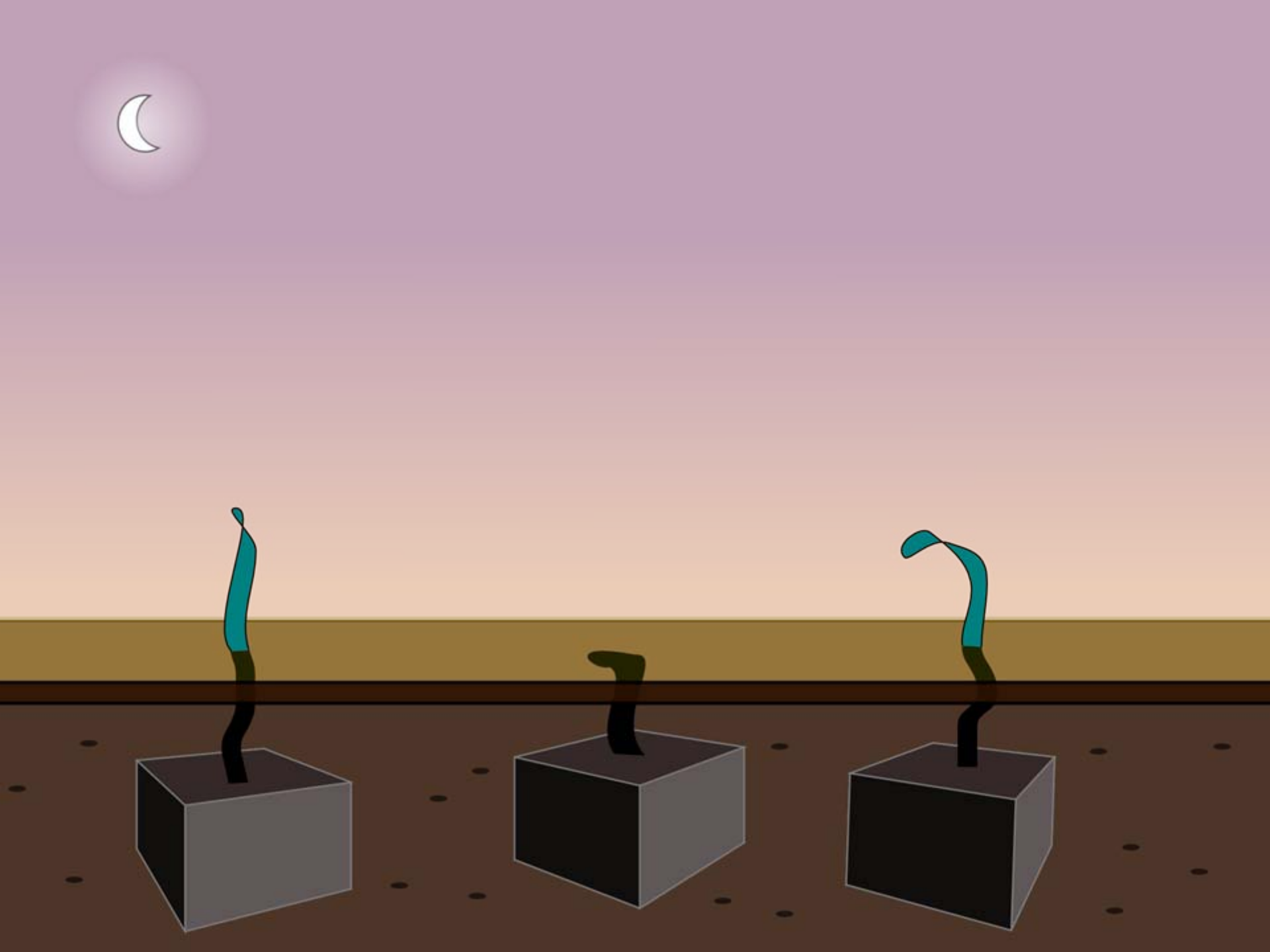
*Imagine  
...!*

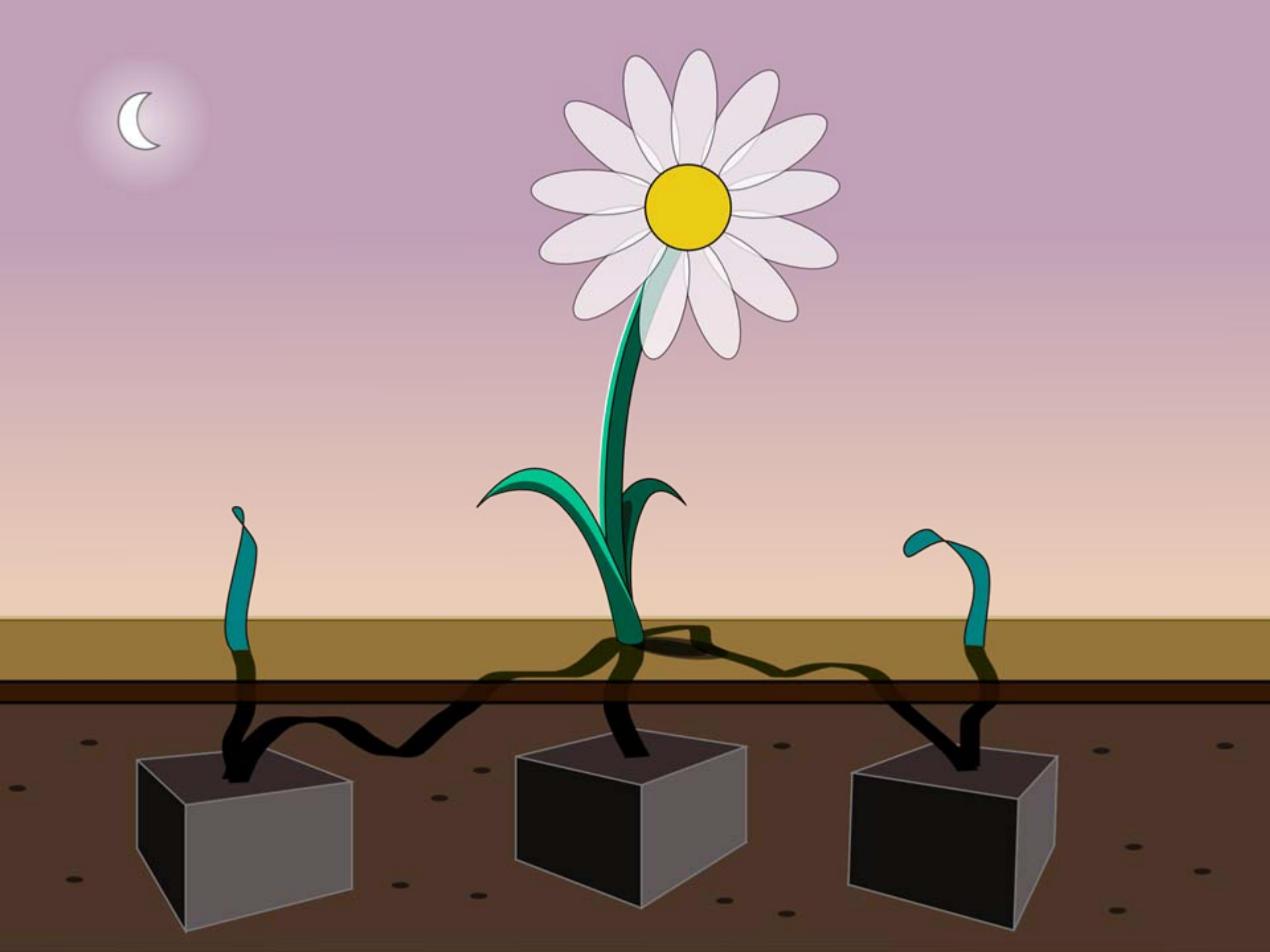


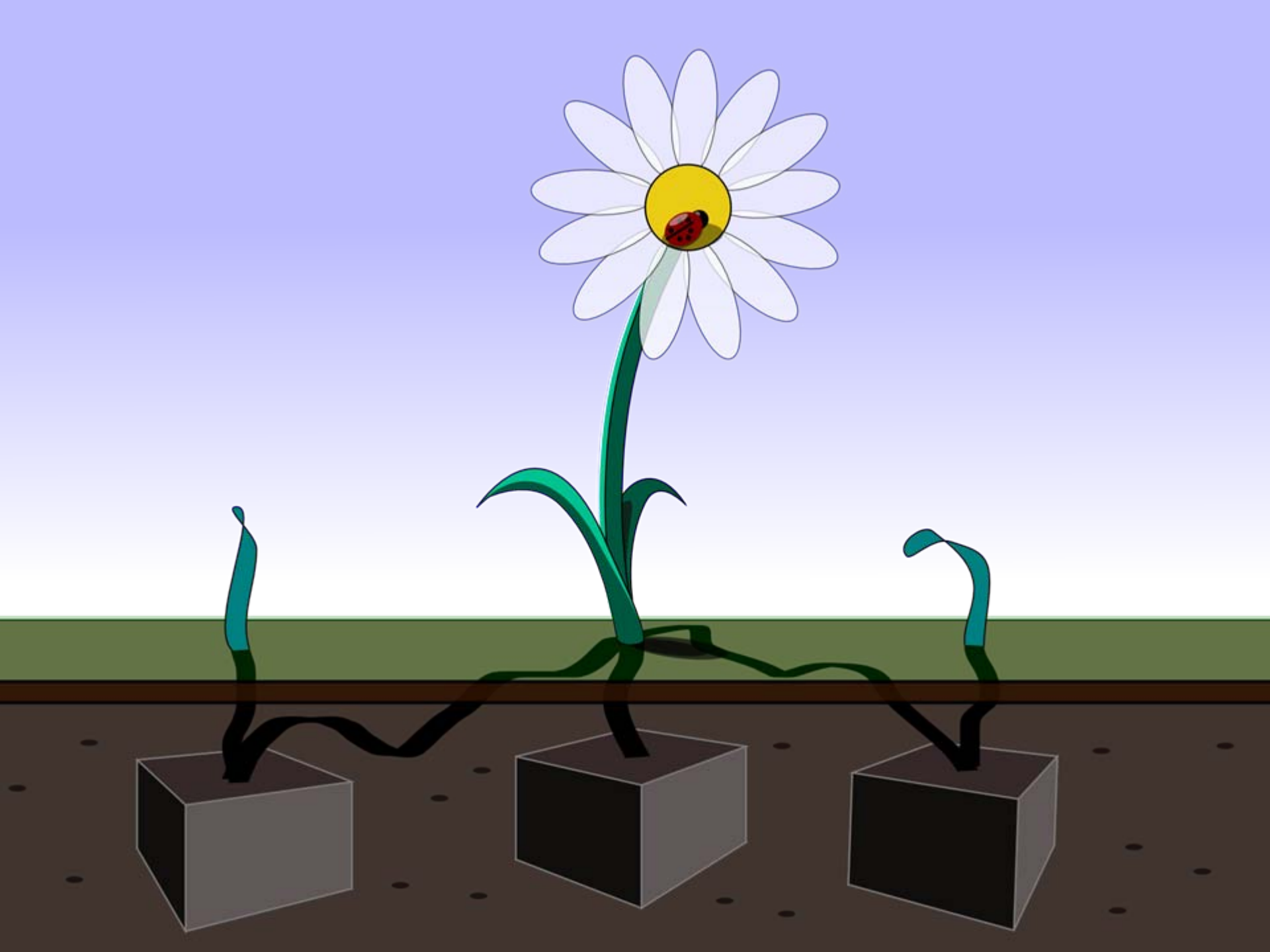
Berners-Lee  
TED-Talk 2009

[http://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html)





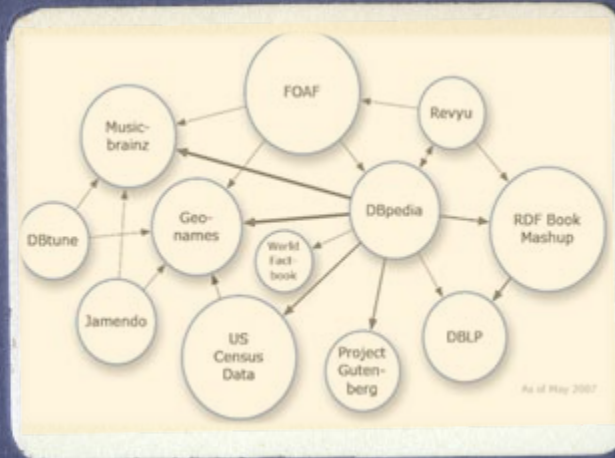






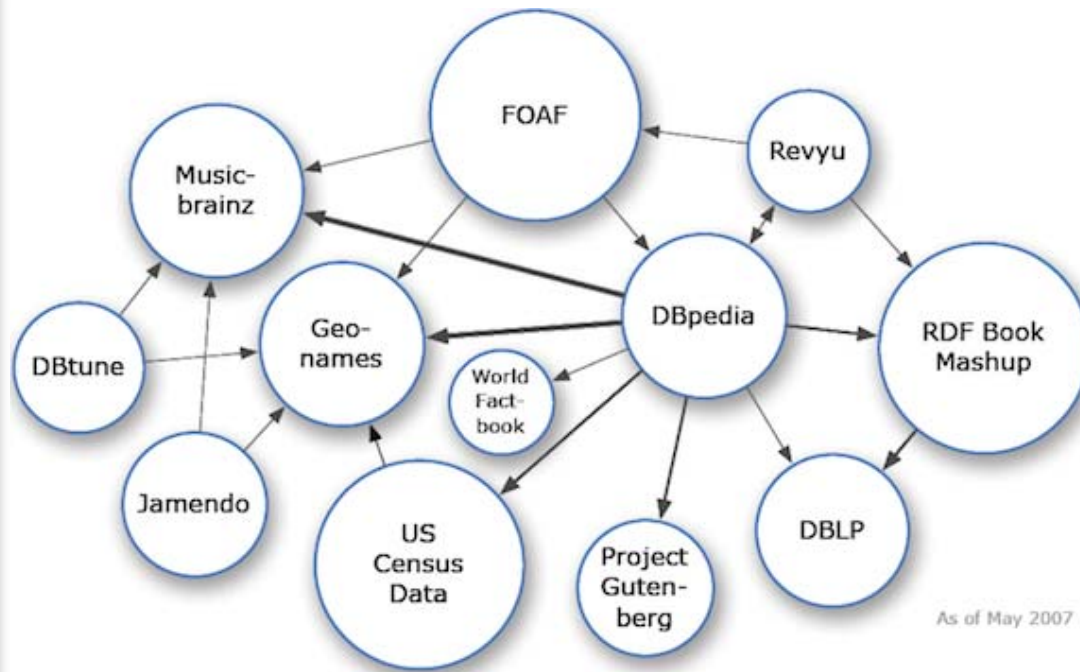
Soweit  
die  
Vision

...



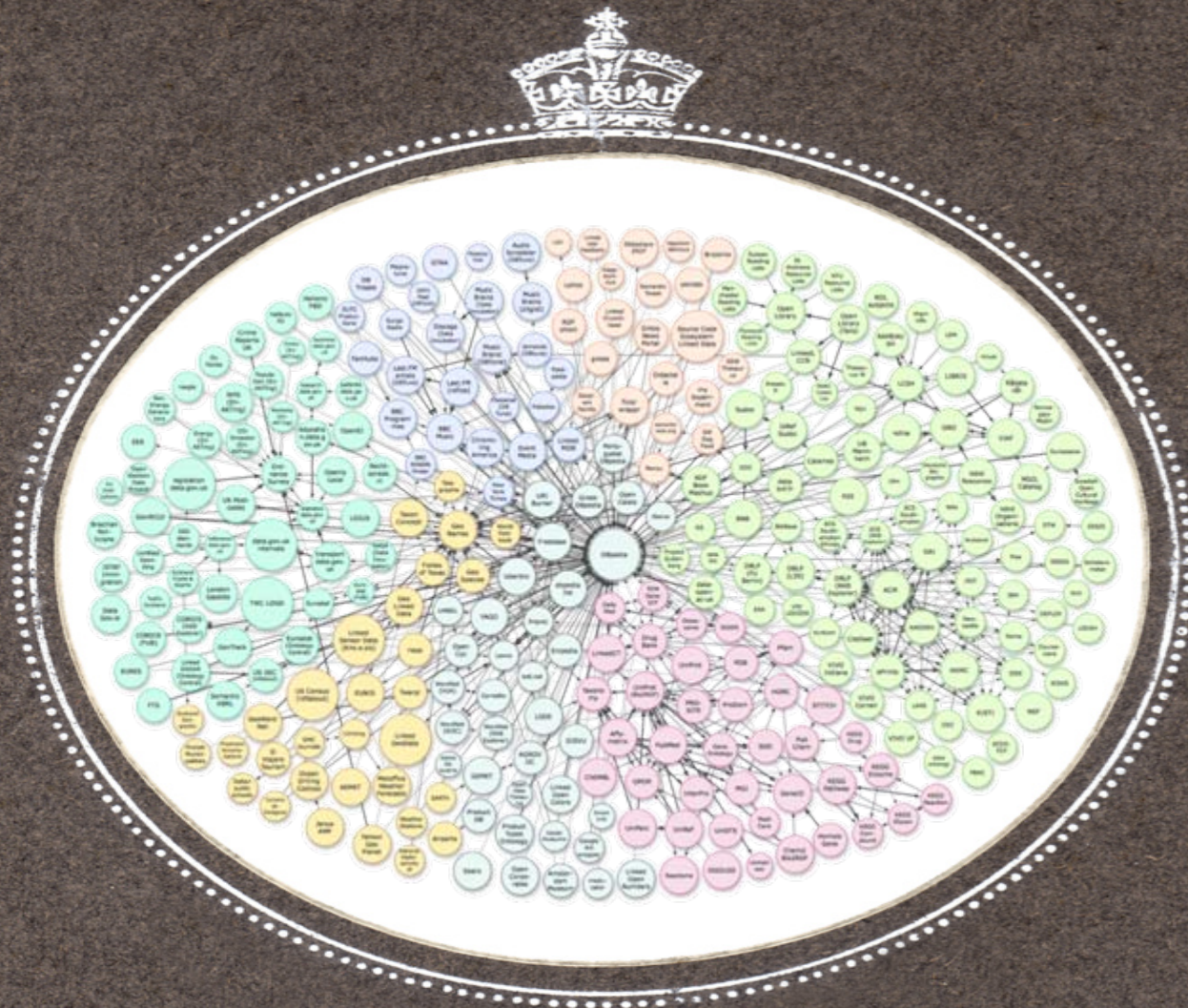
Und die  
Realität  
?

Ein Blick ins Fotoalbum ...



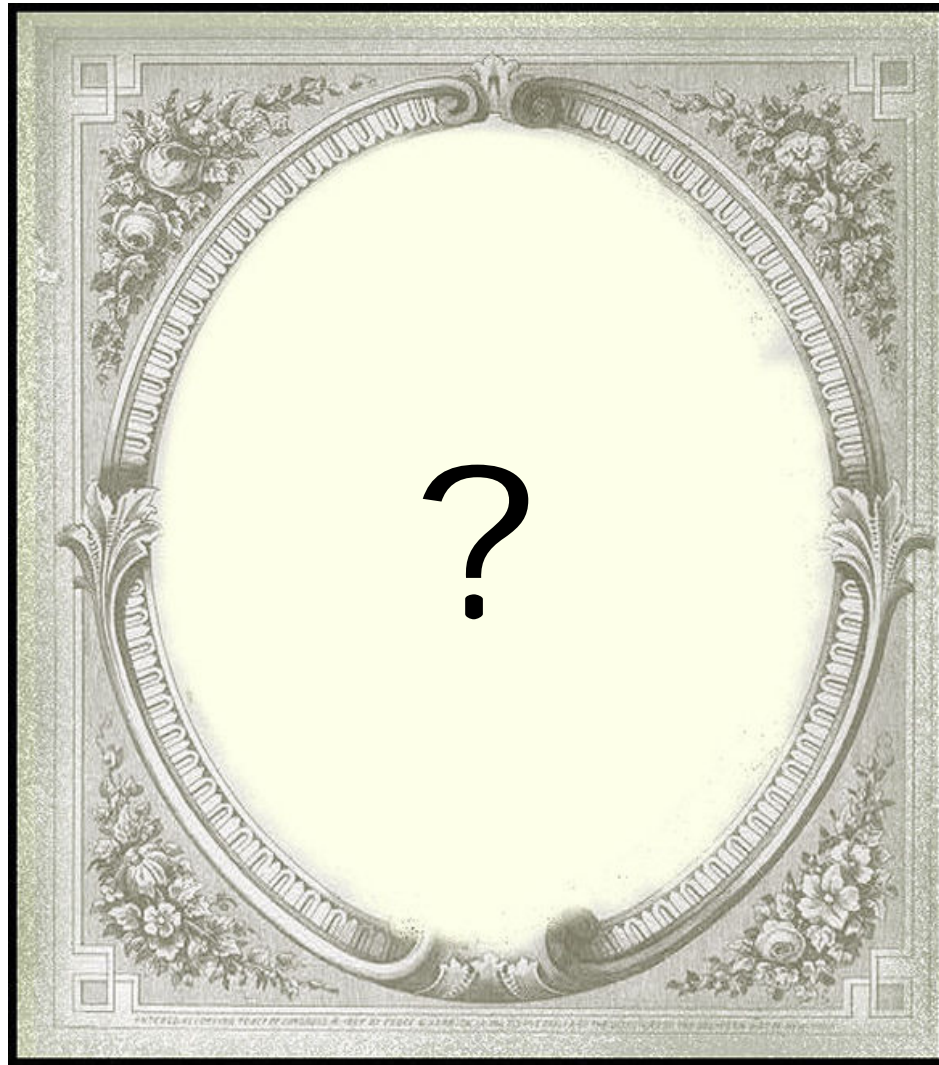
*LOD-Cloud Mai 2007*





*LOD-Cloud September 2011*





*LOD-Cloud 2020*

Bibliotheken und die LOD-Cloud:

# Liebe (fast) von Anbeginn





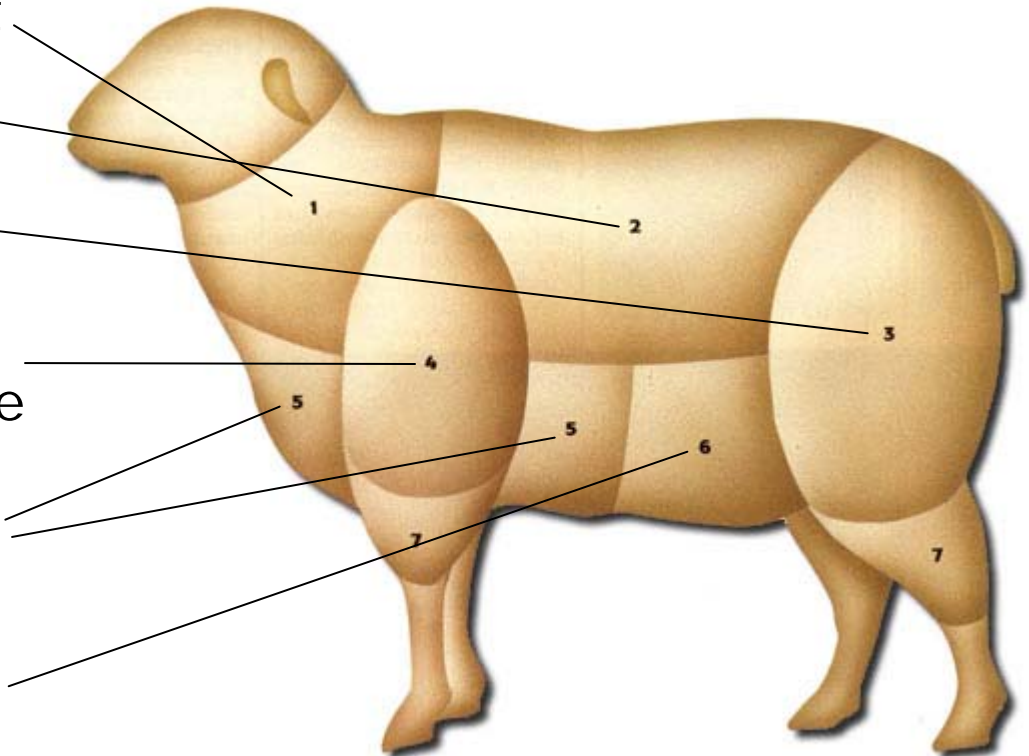
Librarian's  
Best

LOD  
Cloud



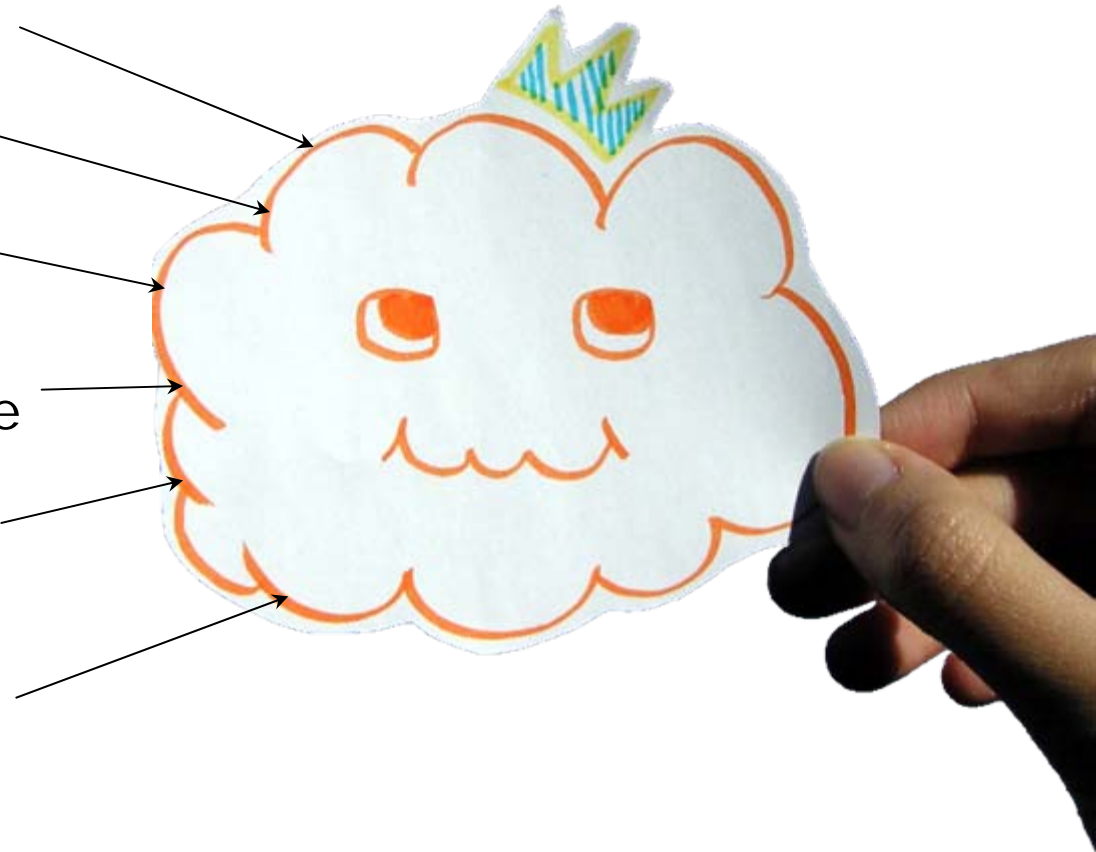
# Keine Selbstlosigkeit – Wir wollen etwas zurück ...

1. Größere Sichtbarkeit
2. Größere Reichweite
3. Leichter Zugang
4. Neue Dienste /  
Anwendungskontexte
5. Interoperabilität mit  
anderen Domänen
6. Mehr Arbeitsteilung/  
Nachnutzung



# Keine Selbstlosigkeit – Wir wollen etwas zurück ...

1. Größere Sichtbarkeit
2. Größere Reichweite
3. Leichter Zugang
4. Neue Dienste /  
Anwendungskontexte
5. Interoperabilität mit  
anderen Domänen
6. Mehr Arbeitsteilung/  
Nachnutzung

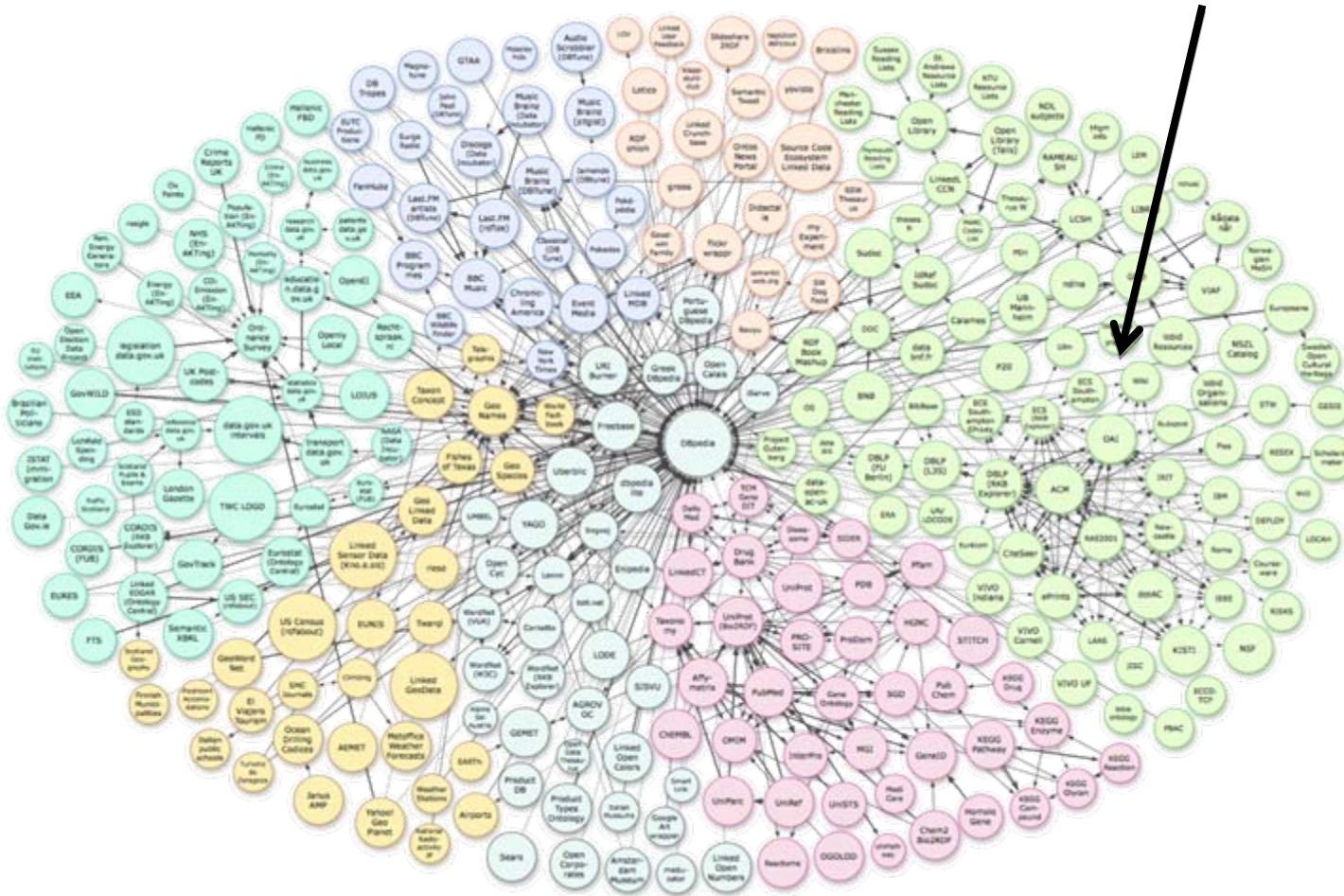


Wir wollen noch mehr...

Bibliografische Erschließung nicht als  
Auslauf- sondern als *Zukunftsmodell*

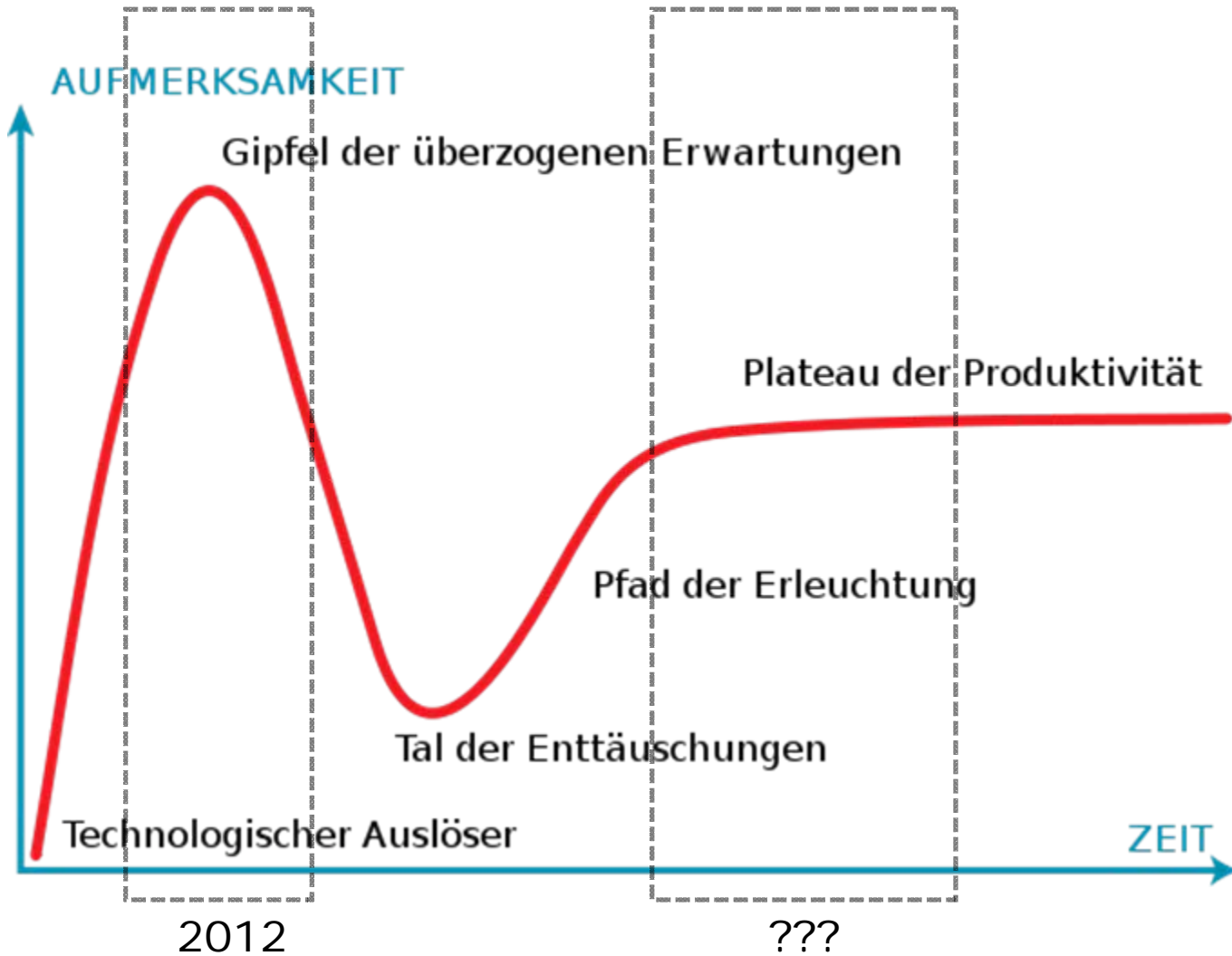
## Bibliothekare in der Wolke:

*„Wir waren nicht die Ersten, aber wir haben bald die dicksten Blasen.“*



Heilsbringer „Linked Data Cloud“:

# Erfolgsgeschichte oder Hype?





Hype	Erfolgsgeschichte
<p>Viele Datasets unbrauchbar: uninteressant, instabil Motto: Hauptsache dabei</p>	<p>Rasante Zunahme von LD-Datensets</p>
<p>Buzzword für Förderprojekte</p>	<p>Umdenken findet statt: die eigenen Daten als Teil einer weltweiten Datenbank begreifen.</p>
<p>RDF nicht immer praxistauglich, keine Lösung für alles</p>	<p>Kritik wird zum Teil angenommen</p>
<p>Infrastruktur steckt noch in den Kinderschuhen</p>	<p>F&amp;E-Projekte in diesem Bereich werden gefördert</p>
<p>überzogene Erwartungen (Reasoning)</p>	
<p>Relevante Anwendungen?</p>	
<p>Konzepte für die dauerhafte Organisation einer globalen Datenbank fehlen.</p>	

# Linked Data: Ein Hype.

- Linked Data steht noch am Anfang
- Zentrale Organisatorische, rechtliche und infrastrukturelle Fragen sind noch offen.
- Um den Vernetzungsgrad zu erhöhen und Stabilität zu erreichen, würden zusätzliche Infrastruktur, Werkzeuge und passgenaue Erschließungsprozesse benötigt.

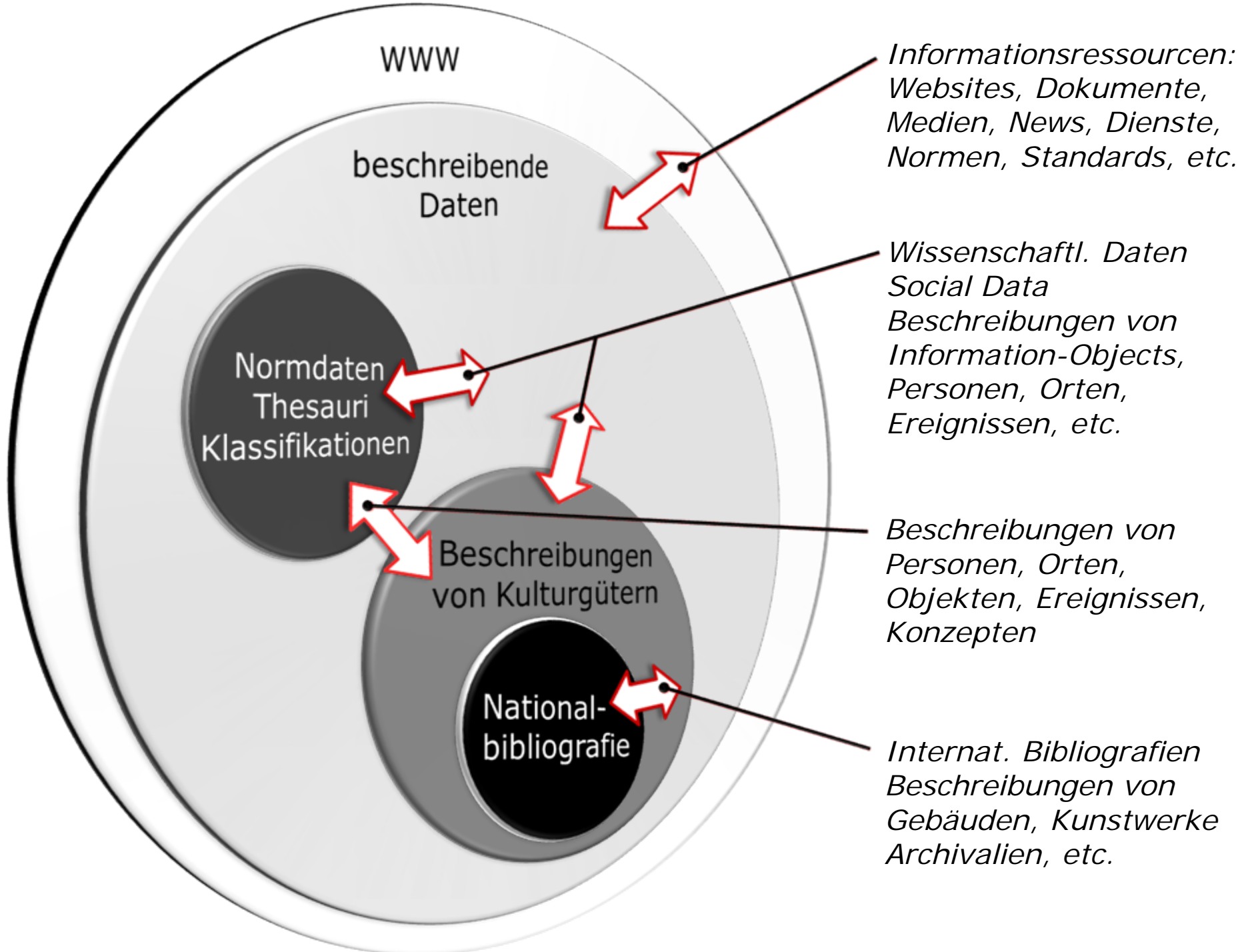
# Linked Data: Die Zukunft!

- Vernetzte Daten im WWW sind die Zukunft!
- **Linked Data ist eine Sichtweise:** die eigenen Daten als Teil einer weltweiten Datenbank begreifen.
- Veröffentlichung der Daten im Web als Linked Data ist ein idealer Startpunkt.
  
- aber eine Frage bleibt ...

Bibliografische Erschließung und Dienste  
im Zeitalter des Webs:

# Auslauf- oder Zukunftsmodell?

Linked Data ist (nur) eine Sichtweise  
und hilfreich, um Erschließung neu  
zu definieren.





**Hofladen:** Für spezialisierten Kunden  
+ Kontrolle über die Vermarktung und Nutzung  
+ Direkter Kundenkontakt



- Schwerer zu finden und zu erreichen
- weniger Auswahl / spezielle Produktpalette



# Großhandel/Supermarkt:

- + Großer Kundenkreis
- + Leicht erreichbar / Abdeckung in der Fläche
- + Integration in Produktionsketten und höherwertige Produkte
- Verlust an Sichtbarkeit und Transparenz?



# Rollenverschiebung bei bibliothekarischen Dienstleistungen

## Mehr Backend / weniger Frontend

bibliografische Recherche =  
Recherche im Web

Qualitätsdebatte:

Zukunftsfähige Erschließung heißt  
Massenproduktion für den  
Großhandel?

# Erschließung: Prozess in einem globalen „Datenökosystem“

Daten zu  
Personen/  
Körperschaften

Daten zu  
Orten/Geografika

Daten zu  
Ereignissen

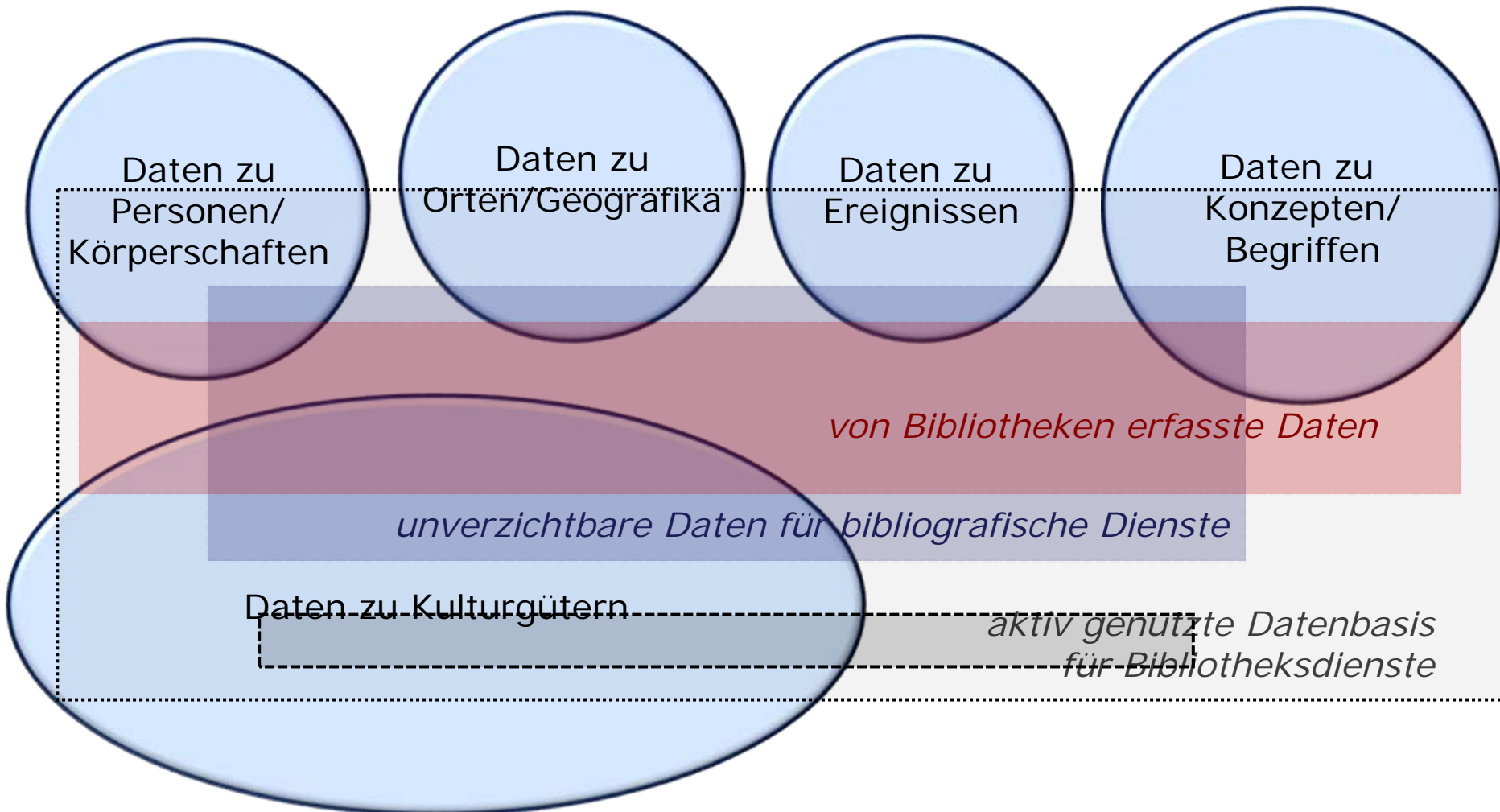
Daten zu  
Konzepten/  
Begriffen

*von Bibliotheken erfasste Daten*

*unverzichtbare Daten für bibliografische Dienste*

Daten zu Kulturgütern

*aktiv genutzte Datenbasis  
für Bibliotheksdienste*





Bestimmende Qualitätsmerkmale:

Persistenz, Transparenz,  
Nachnutzbarkeit, Interoperabilität,  
Offenheit, Vernetzungsgrad

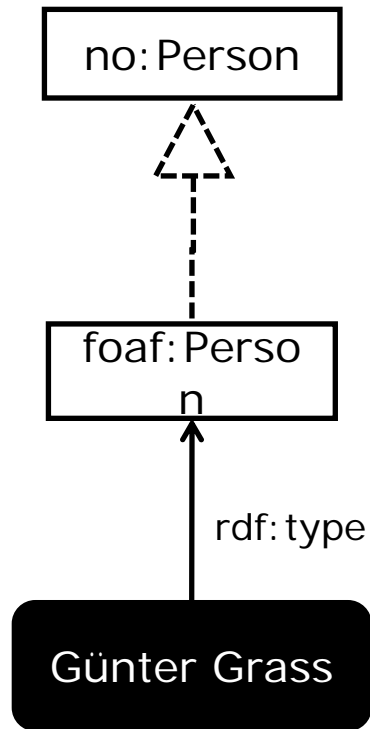
# Herausforderung Verlässlichkeit

- Überprüfung externer Datenquellen (langfristige Perspektive, Qualität, Fachkompetenz)
- Transparenz
  - Sichtbarkeit der Urheberschaft von Daten
  - Änderungshistorie
- Persistenz von Daten und Ressourcen
- Qualitätssicherungsprozesse
- Kriterien für vertrauenswürdige Fremddaten-Anbieter

# Herausforderung Interoperabilität / Nachnutzbarkeit

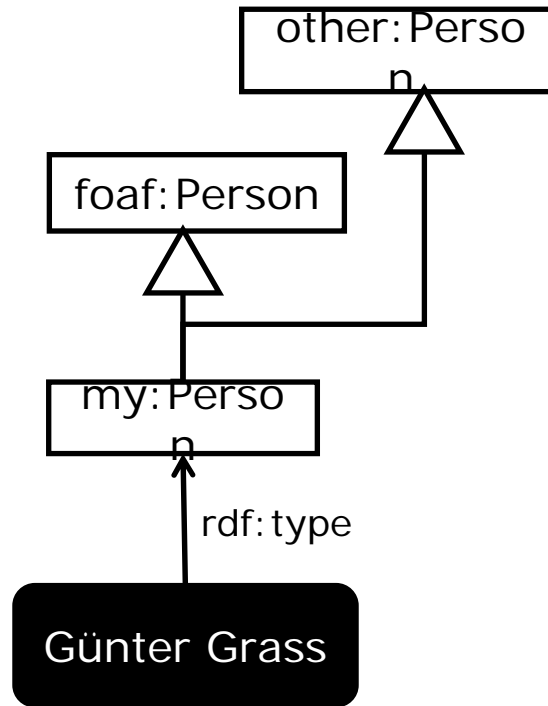
- Regelwerksgerechtigkeit bis ins letzte Detail kein alleiniges Kriterium für hohe Qualität
- ... dennoch braucht es mehr Regeln und Best-Practices
  - KIM / AG der Verbundsysteme / BibFrame
- Anwendungsfallgetriebene Modellierung vs. Übertragbarkeit
- Debatte „ontology alignment vs. direct re-use“: Unabhängigkeit und Präzision vs. Pflegeaufwand und Einfachheit

### direct re-use



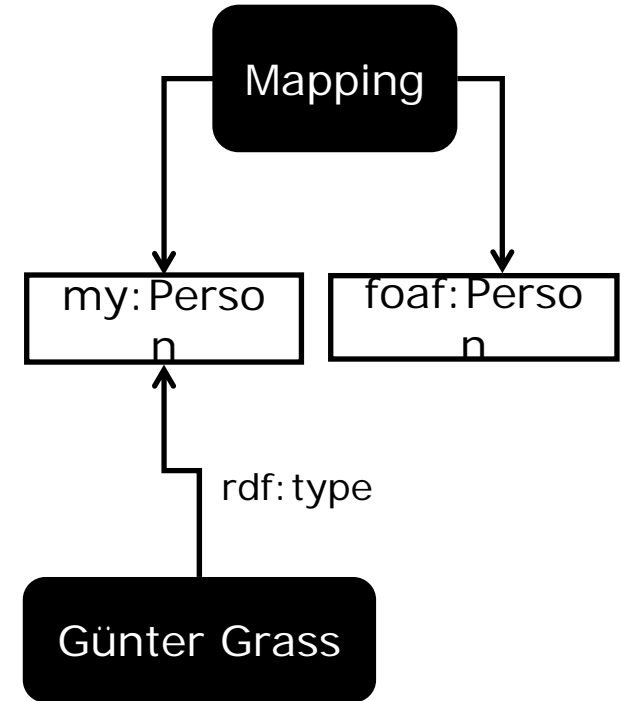
- + kein Pflegeaufwand
- + Mapping auch ohne RDFS & Co möglich
- Semantisch ggf. nicht 100% treffend
- Abhängigkeit
- Risiko unabgestimmter Änderungen

### class re-use über Vererbung



- + Unabhängigkeit und Präzision
- + Mehrfachvererbung möglich
- Pflegeaufwand
- Interpretation der Ontologie notwendig

### class mapping



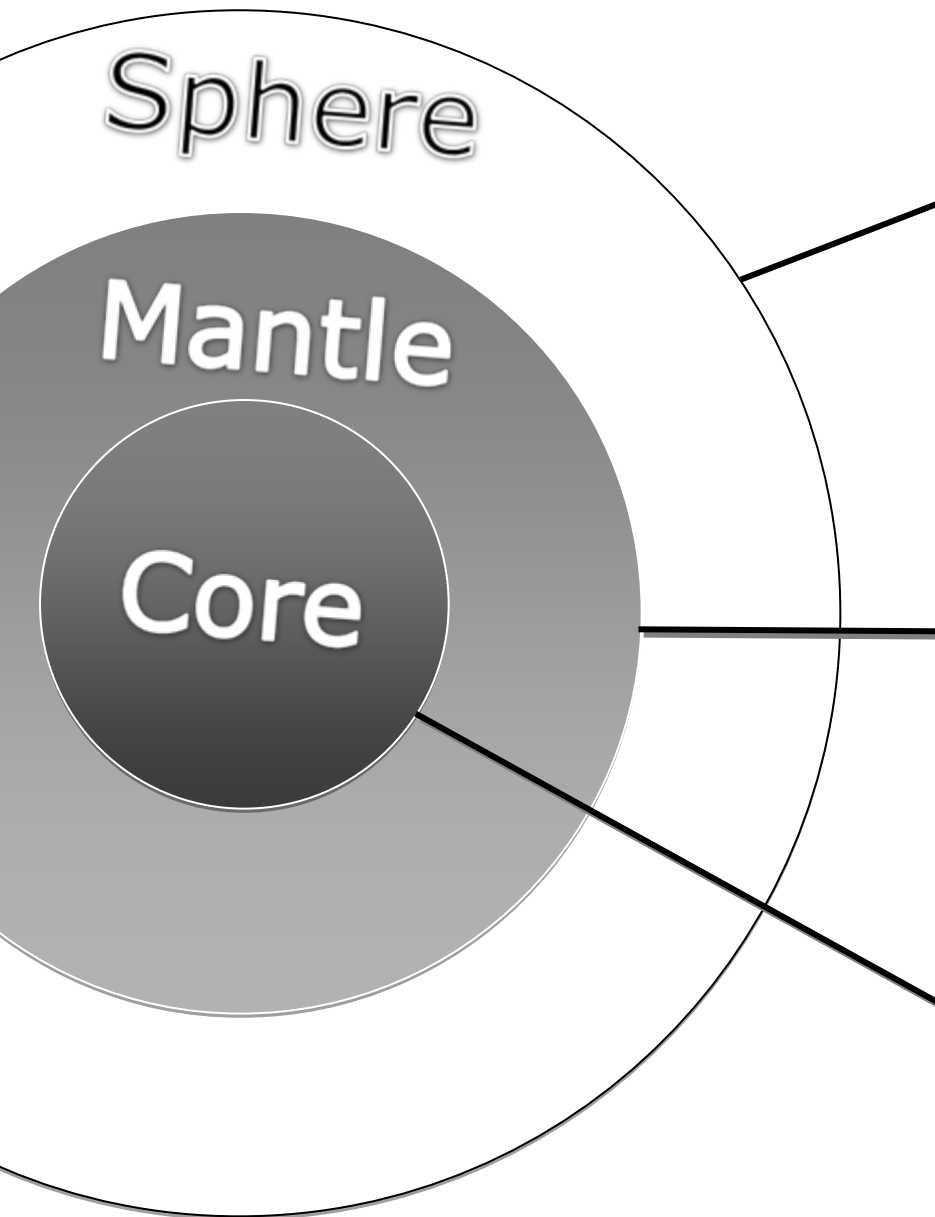
- + präzisere Modellierung von Klassenrelationen möglich
- + geringe Kopplung zwischen den Ontologien
- zusätzliche Mapping-Ontologie notwendig

# Herausforderung verteilte Datenhaltung

- Höhere Komplexität durch steigende Abhängigkeit:
    - Synchronisation zwischen Datentöpfen
    - Stabilität / dauerhafte Verfügbarkeit vs. Aktualität
    - Umgang mit Massenänderungen / Angriffen / Fehlern
    - Abonnieren relevanter Änderungen
- Neue Infrastruktur erforderlich

# Herausforderung: Datenvernetzung

- Retrospektive Vernetzung:
  - Bestehende Datenbestände vernetzen/verknüpfen
  - Dubletten erkennen und zusammenführen
- Künftige Erschließung
  - Keine Dubletten entstehen lassen / Keine Doppelarbeit mehr
  - Erschließung muss leichter werden nicht komplexer: Unterstützung in Erschließungswerkzeugen
- Vernetzung mit Dritten
  - Domänenübergreifende Recherche
  - Anreicherung durch Fach-Communities
  - Anlegen von Interessensprofilen
  - Inhaltliche Vernetzung schon beim Entstehen der Publikation



**Externe Datensets /  
Nutzer-generierte Daten**

Verlagsdaten, soziale Netze,  
Nutzerfeedback, Wikipedia, Nachrichten,  
etc.

---

**Qualitätsgeprüfte Automatismen**

- Automatische Verfahren zur Vermittlung  
zwischen Kern und Sphäre.  
- Gepflegt und verwaltet durch KWEs  
z.B.: *culturegraph.org*

---

**Stabil und verlässlich**

- Persistente Knoten mit einer strikten,  
transparenten Policy: keine Löschungen,  
Versionierung  
- Gepflegt und verwaltet durch Kultur- und  
Wissenschaftseinrichtungen (KWE)  
Basierend auf etablierten Standards

---



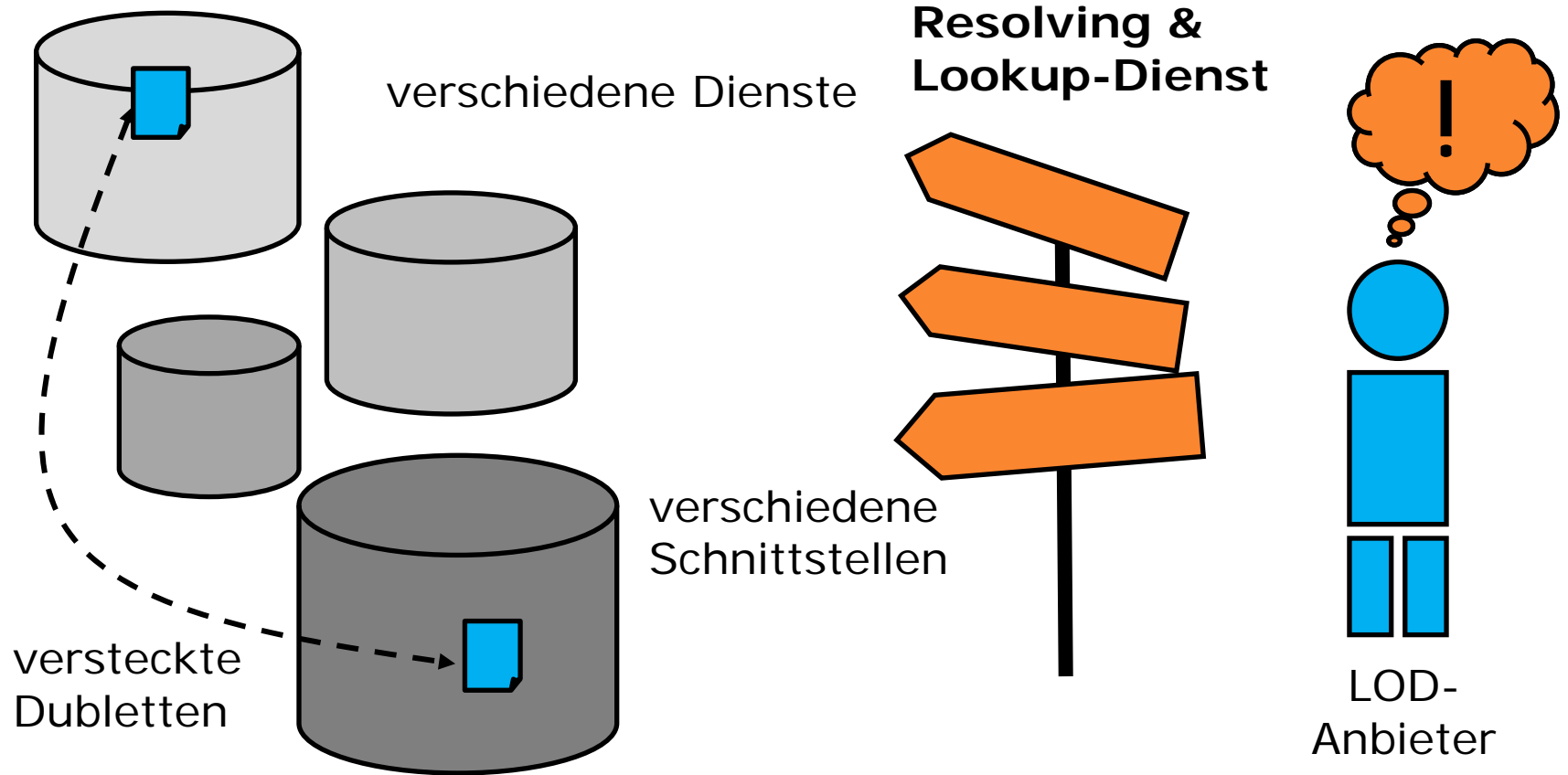
# Culturegraph.org

## Offene Plattform zur Datenvernetzung

- persistente Identifizierung von Entitäten
- Verwaltung und Veröffentlichung von Mappings
- Berechnung von Ähnlichkeiten und Verknüpfungen
- Offene Daten / Freie Software

The screenshot shows the Culturegraph.org website interface. At the top, the word "culturegraph" is displayed in white on a blue background. Below this, there is a dark grey navigation bar with three links: "Home", "Über das Projekt", and "Impressum". The main content area is light blue and contains a search form with the text "Please enter an URN." above a text input field. To the right of the input field is a black "GO" button. Below the input field, the word "EXAMPLES:" is written in blue, followed by three example URNs: "urn:nbn:de:eki:DNB991052625", "urn:nbn:de:eki:ZDB515966-0", and "urn:nbn:de:eki:HBZHT015344773".

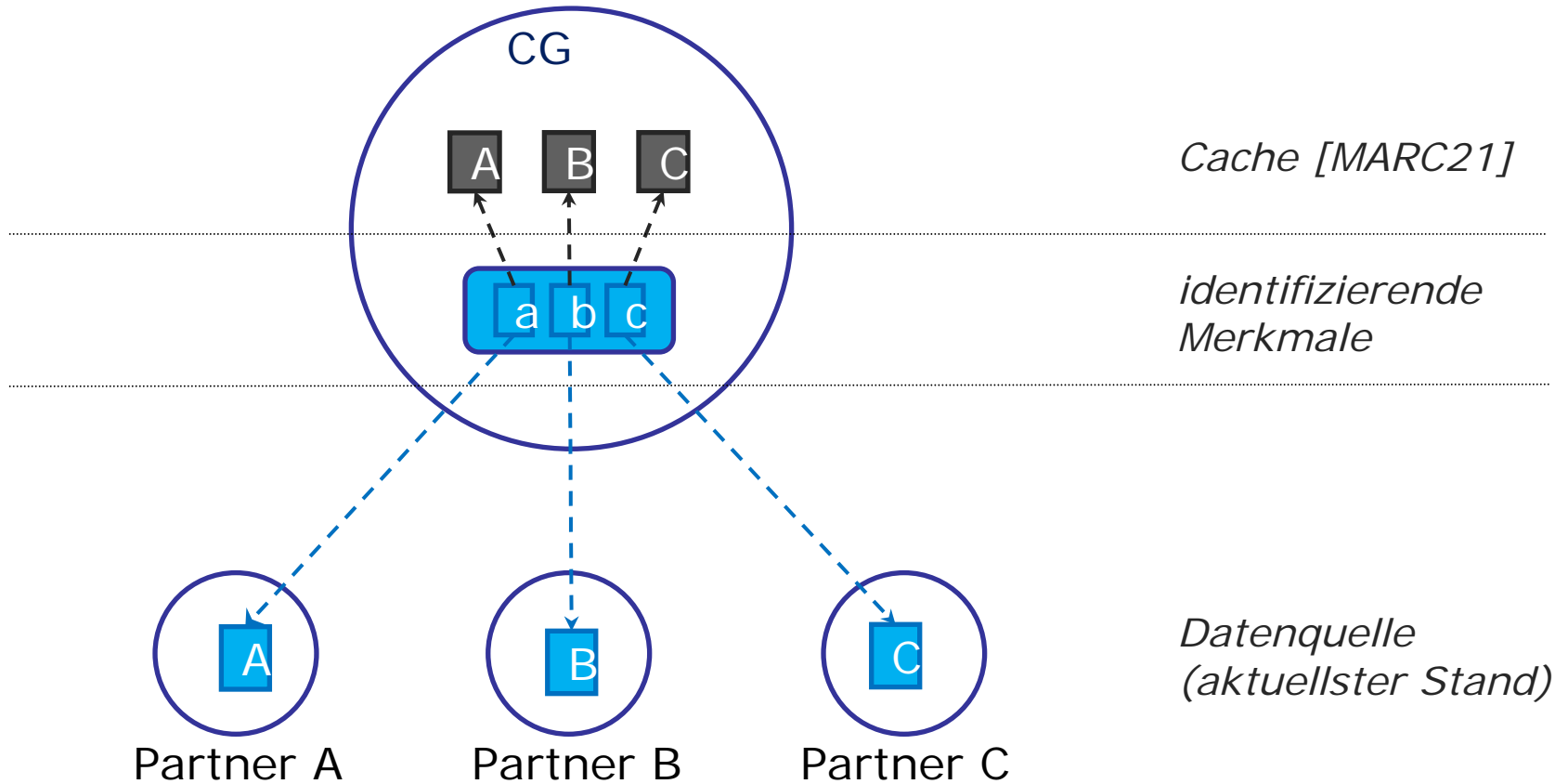
# culturegraph als Wegweiser in der LOD-Cloud



# Erreichter Stand

- Projekt erfolgreich abgeschlossen
- Skalierbare, performante Infrastruktur wurde aufgebaut
- Aktueller Datenbestand:
  - Monografien seit 1945
  - aktuell ca. 88 Mio. Datensätze im System
  - Verarbeitung (Import/Analyse/Clustering) aller Daten in ca. 2 Std.
  - Siehe <https://wiki1.hbz-nrw.de/display/SEM/Datenlieferungen+der+Verbundene>
- Beta-Dienst online: <http://culturegraph.org>

# Culturegraph: Datenmodell



# Weiteres Vorgehen

- Datenlieferung:
  - Bereinigung fehlerhafter EKI-Verwendung
  - Regelmäßige Datenlieferungen in MARC21
  
- Folgeprojekte:
  - Retrospektive Vernetzung der AGV:
    - Intensive Nutzung der Plattform
  - Künftige Erschließung:
    - Integration von CG in den Erschließungsprozess
  - Vernetzung mit Dritten:
    - Erweiterung um weitere Fremddaten

Fachlicher Teil

Dienstleistungen



Applikationen



Anforderungen  
Servicekatalog  
Policy  
Web-Standards

# Idee AGV- Serviceplattform:

synergetische Arbeitsteilung

Gemeinsame Dienstleistungen  
aus einem Guss /  
geschlossenes Auftreten

Technischer Teil

API



Gemeinsam gepflegte Infrastruktur  
und Daten als Basis:

✓Erweiterbare und offene  
Plattform

✓Einbindung Dritter über  
offene Schnittstellen

✓Verteilte serviceorientierte  
Architektur

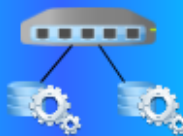
✓Integration ins Web /  
Linked Open Data

✓Echtzeitsynchronisation über  
moderne Webstandards

Services



Nachrichten-  
infrastruktur



Technische Infrastruktur



# Fazit

- Vernetzte Daten im WWW sind die Zukunft!
- Wir stehen vor noch einigen Herausforderungen
  - Ontologien: Best-Practices (national und international)
  - Alt und Neu zur selben Zeit (ein klarer Schnitt wird nicht kommen)
  - Datenvernetzung / zwischen Bibliotheken und Domänen-  
übergreifend
  - Anpassung von Erschließungsprozessen / -werkzeugen
  - Nachweis des Mehrwerts unserer Daten
- **Größtes Risiko:** Fehlende Kommunikation, zu wenig Kooperation!



Aber deshalb sind wir ja hier 😊

Vielen Dank!

# Credits (Pictures)

- Slide 1: Matt Burtch. <http://www.flickr.com/photos/cortexbomb/4552278669>
- Slide 9: Ard Hesselink. <http://www.flickr.com/photos/docman/3750030>
- Slide 10: Glen Dahlman. <http://www.flickr.com/photos/gldahl43/4934478836>
- Slide 11: lisabee73. <http://www.flickr.com/photos/24882261@N04/4187681413>
- Slide 14: Jack Hynes. <http://www.flickr.com/photos/jackhynes/462872830>
- Slide 16: Karen Ka Ying Wong. <http://www.flickr.com/photos/kky/704056791>
- Slide 27: frankenschulz. <http://www.flickr.com/photos/frankenschulz/5047918350>
- Slide 28: Winfried. <http://www.flickr.com/photos/w-tommerdich/6883785180/>