



Bibliotheken: Wir öffnen Daten

**Zum Stand der Entwicklung einer
offenen Dateninfrastruktur**

Bibliothekartag. Bremen, 04.06.2014.

Struktur

2

- I. Eröffnung: Ziele
- II. Elemente einer (offenen) Dateninfrastruktur
- III. Open Discovery?
- IV. Konkrete Schritte

I. Eröffnung: Ziele

(Discovery) Services, die...

4

- nützlich,
- einfach,
- stabil,
- kostengünstig und
- zukunftssicher sind.

Nutzer/innen, die...

5

- zahlreich und
- zufrieden sind.

Mitarbeiter/innen, die...

6

- motiviert,
- neugierig,
- innovativ,
- freundlich und
- hilfsbereit sind.

Bibliotheken, die...

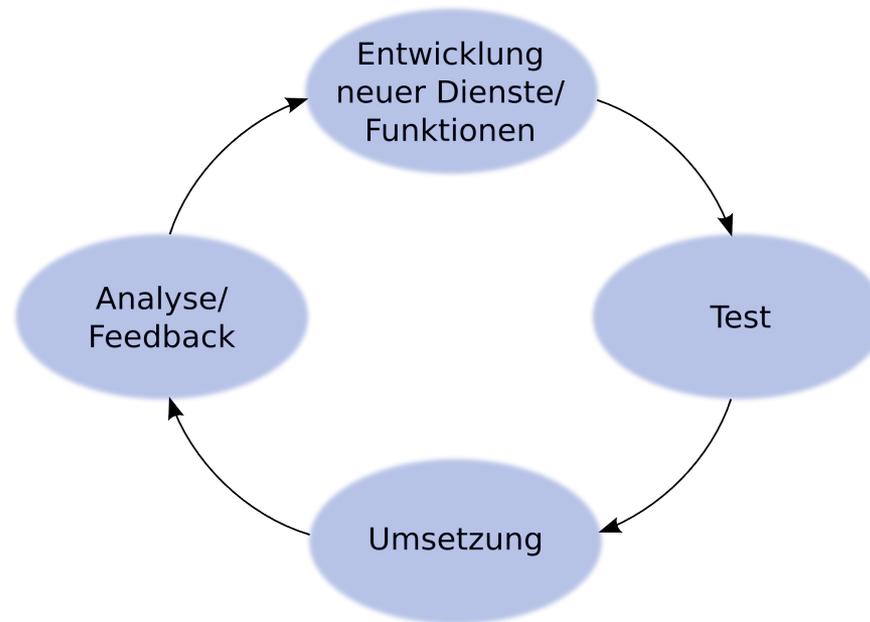
7

- relevant,
- vielbesucht,
- innovativ sind und
- fortbestehen.

Aber...

8

- Diese Ziele werden nie abschließend erreicht.
- An ihnen muss immer wieder gearbeitet werden:



- Dafür gibt es gute und schlechte Umgebungen.

Offene Umgebungen

9

Offene Umgebungen laden ein

- zum Experimentieren,
- zum Teilen und
- zur Zusammenarbeit.

Sie fördern langfristig:

- Effizienz,
- Transparenz,
- Innovation,
- Zukunftssicherheit.

“Offen”

10

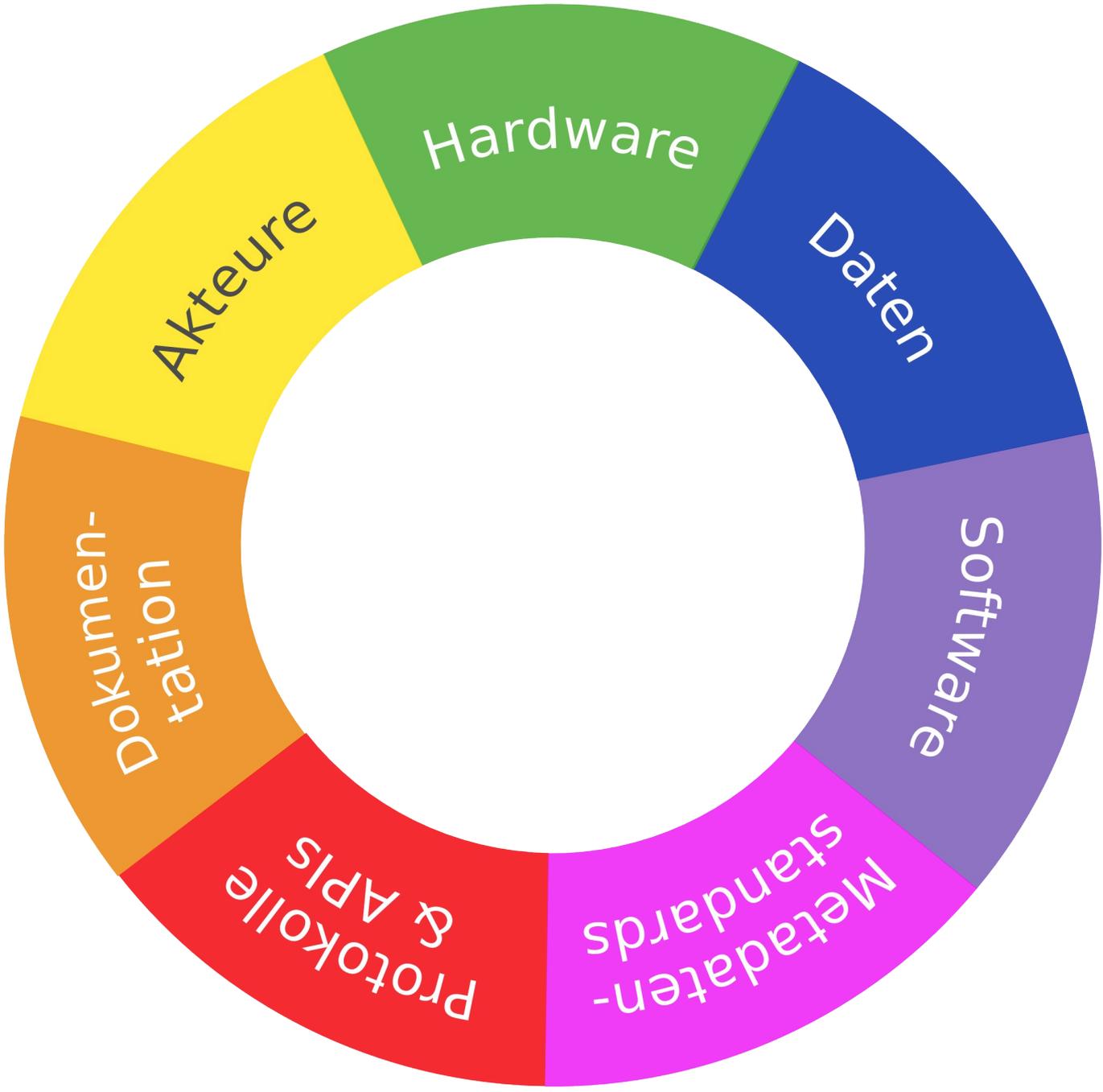
Etwas ist offen, wenn es selbst bzw. seine Spezifikation/Baupläne

(a) frei zugänglich sind,

(b) in öffentlich dokumentierten Formaten vorliegen und

(c) es jedem erlaubt ist, es zu benutzen, weiterzuverwenden und -zuverbreiten.

II. Elemente einer (offenen) Dateninfrastruktur



Hardware

Daten

Software

Metadaten-standards

Protokolle & APIs

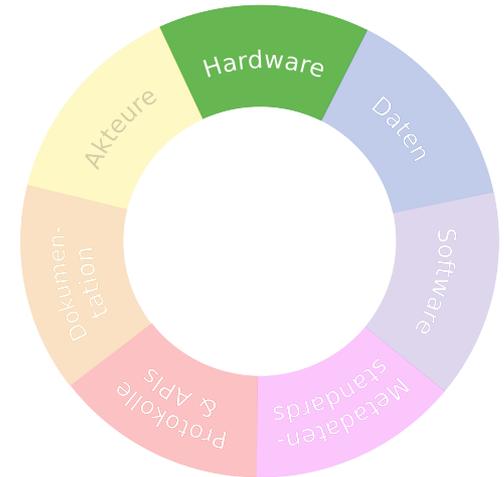
Dokumen-tation

Akteure

Hardware

13

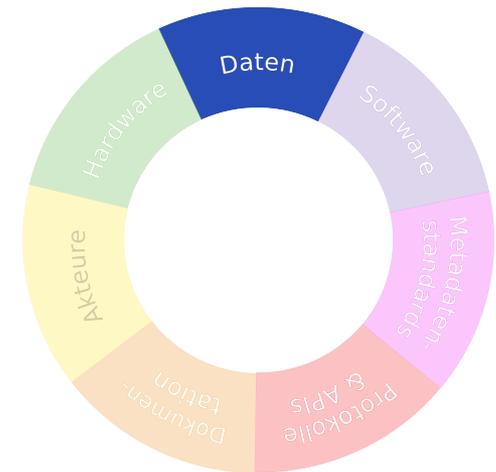
- Physische Netze
- Server
- Clients



Daten

14

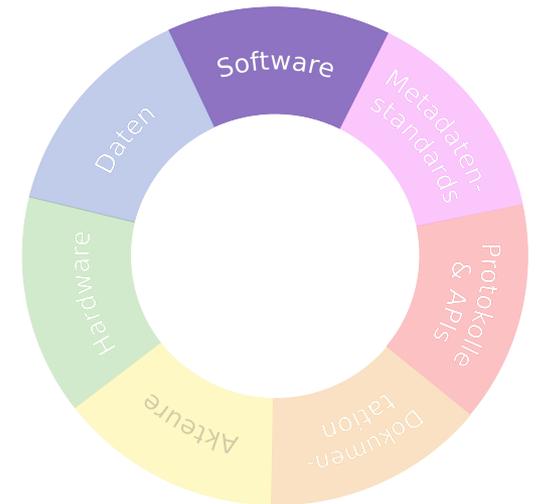
- Titeldaten (Monographien & Periodika vs. Zeitschriftenartikel)
- Kontrollierte Vokabulare
- Bestandsdaten
- Verfügbarkeitsdaten
- Bibliographien
- Statistische Daten (Ausleihdaten, Zugriffsstatistiken etc.)
- Forschungsdaten



Software

15

- Bibliothekssysteme
- Discovery-Lösungen
- Verbundsysteme
- Repositories
- Software zur Datentransformation

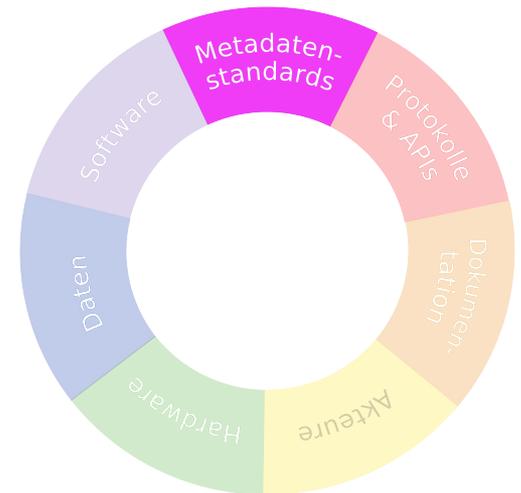


(Meta)datenstandards

16

Spezifikationen von

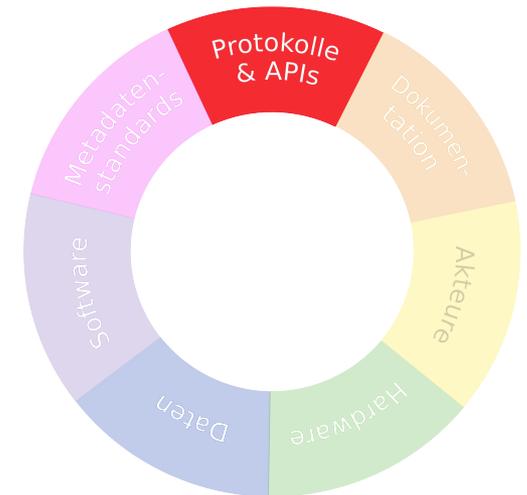
- Datenmodellen, Vokabularen/
Schemas,
- Erfassungsregeln,
- Serialisierungen,
- Encodings.



Protokolle & APIs

17

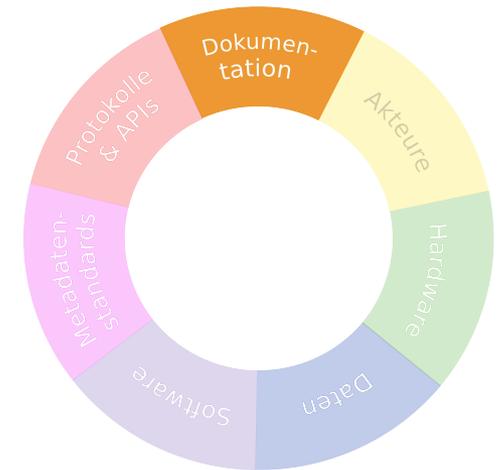
- Spezifikationen von Protokollen für
 - Datenabfrage,
 - Lesen/Schreiben,
 - Datensynchronisierung.
- Konkrete Schnittstellen/APIs
 - Zugriff
 - Dokumentation



Dokumentation

18

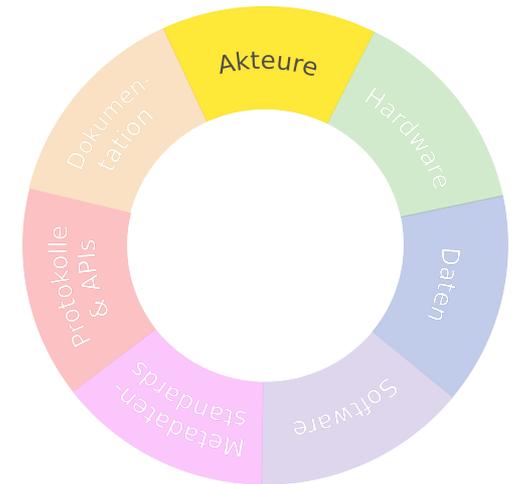
- Tutorials
- Mailinglisten und -archive
- Ticketsysteme
- Metadatenmappings
- Anwendungsprofile
- Projektanträge, -pläne, -berichte, Lessons Learned
- Konfigurationsdateien für Datentransformation...



Akteure

19

- Bibliotheken & Verbundzentralen
- Non-Profit-Organisationen
- For-Profit-Unternehmen
- Verlage
- Universitäten und Wissenschaftler
- Förderinstitutionen
- Sonstige



Status Quo: Beispiele

20

	offen	nicht offen
Hardware	-	<ul style="list-style-type: none"> • Physikalische Netze & Server • Tendenz: Zentralisierung
Daten	<ul style="list-style-type: none"> • hbz, b3kat, HeBIS • BNB, DNB, BnF, BNE • GND, VIAF, LCSH • DBLP 	<ul style="list-style-type: none"> • Journalartikelmetadaten • ViFa-/FID-Daten • Fachbibliographien • Discovery Indizes
Dokumentation	Mailinglisten(archive), Mappings	•Projektanträge und -berichte
Protokolle & APIs	<ul style="list-style-type: none"> • Protokolle meist offen (W3C-, NISO-Standards) • lobid-API, Nature-API 	•SLNP
Metadaten-standards	•Spezifikationen von Vokabularen, Serialisierungen & Encodings meist offen	•AACR, RDA
Software	<ul style="list-style-type: none"> • Koha, Evergreen, vufind • OPUS, DSpace, E-Prints, Fedora • Refine, Metafactory, Catmandu 	Alma, WMS, EDS, Summon, Primo, CONTENTdm, DigiTool

III. Open Discovery?

Derzeitige Discovery-Ansätze

22

Effizient?

Zukunftssicher?

Nützlich?

Kostengünstig?

Innovativ?

Stabil?

Einfach?

Transparent?

Kooperativ?

Nachnutzbar?

Probleme derzeitiger Ansätze

23

- Daten und Software nicht offen/nachnutzbar
- Kostenintensiv
- Intransparent (z. B. Ranking)
- Zugriff auf Daten lizenzierter Inhalte nicht gesichert
- ...

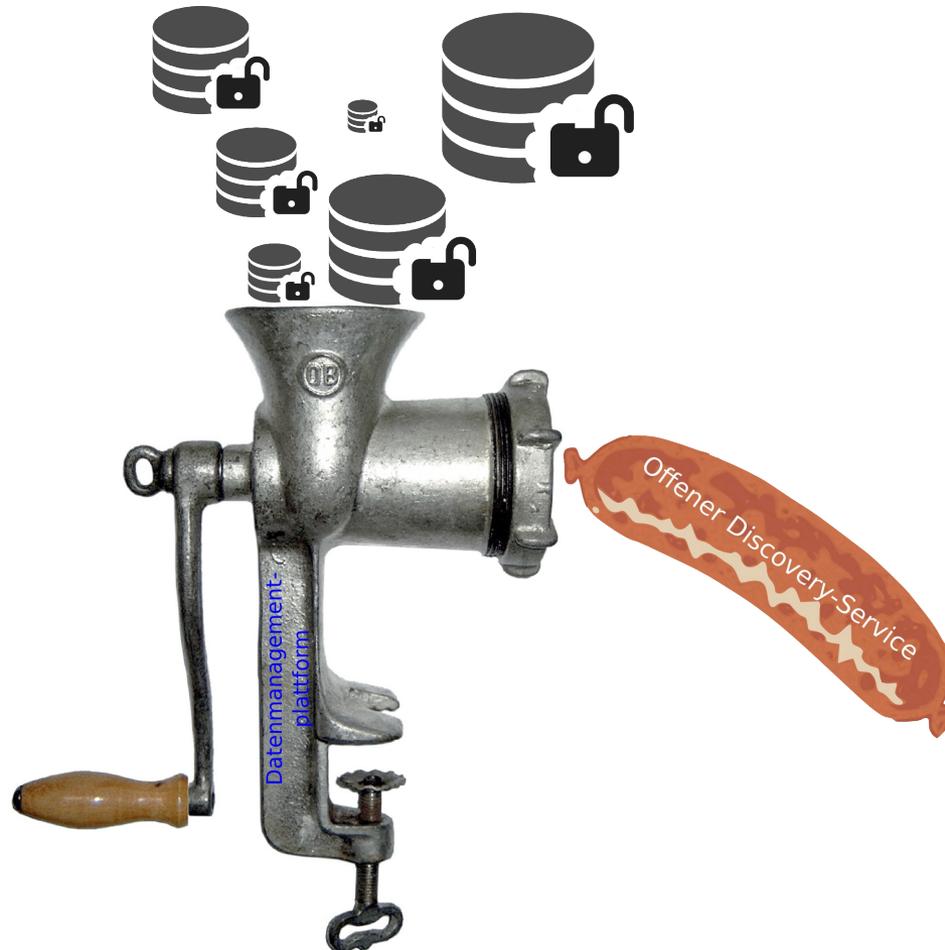
Lösungsansätze

24

- a) *Offene Daten*: Rechtliche Hürden beseitigen, um allen (Bibliotheken, Wissenschaftlern etc.) den Aufbau eigener Suchindizes zu ermöglichen.
- b) *Linked Data*: Die Publikation strukturierter Metadaten im Web (RDF+schema.org) vorantreiben, so dass Internetsuchmaschinen sie indexieren können.

a) Offene Daten einsammeln

25



b) Linked Data sammeln lassen

26



Google-Suche

Auf gut Glück!

schema.org & Google

27

Erste Erfahrungen & Experimente:

- Duke University Libraries, *Schema.org and Google for Local Discovery*
- *schema.org-basierter union catalog* (proof of concept) von Dan Scott

IV. Konkrete Schritte

1. “Eigene” Daten freigeben

29

- Fachbibliographien
- Virtuelle Fachbibliotheken / SSG / Fachinformationsdienste
- Kataloge von Spezialbibliotheken
- Bibliographien von WissenschaftlerInnen
- Hochschulschriftenserver
- (Egal, ob PDF, Word-Datei, XML, MAB.)

2. LOD veröffentlichen

30

- Persistente URLs
- RDF-Daten mittels JSON-LD oder RDFa einbetten
- Nutzung von schema.org
- Hilfreich: Empfehlungen der W3C Schema Bib Extend Community Group

3. (Linked) Open Data einfordern

31

- Bei der Lizenzierung von Inhalten (Journals, eBooks etc.) die freie Nutzbarkeit der zugehörigen Metadaten in den Vertrag aufnehmen.
- Entsprechende Musterklauseln entwerfen.
- Zusätzlich von Inhalteanbietern die Publikation strukturierter Metadaten im Web in Verbindung mit den eigentlichen (kostenpflichtigen) Inhalten verlangen.

4. Index aufbauen

32

- Daten einsammeln & aktuell halten
- Daten normalisieren
- Daten indexieren
- Alternativ: Google indexieren lassen und Custom Search aufsetzen
- Recherche und mehr ermöglichen (fachspezifisch oder -übergreifend)
- Freie Software nutzen, Ansatz und Erfahrungen dokumentieren, Skripte teilen

Danke.

Fragen?

**Gerne jetzt oder auch später an
pohl@hbz-nrw.de oder @acka47.**



Diese Folien stehen – soweit nicht anders angegeben – unter der Creative Commons Public Domain Dedication (CC0):
<http://creativecommons.org/publicdomain/zero/1.0/>.

Weiterführende Informationen

34

Aery, Shawn (2014): Schema.org and Google for Local Discovery: Some Key Takeaways.
URL:
<http://blogs.library.duke.edu/bitstreams/2014/03/27/schema-org-and-google-for-local-discovery-some-key-takeaways/>

Gruppe „Lizenzen“ der DINI AG KIM (2011): Empfehlungen zur Öffnung
bibliothekarischer Daten: <https://wiki.dnb.de/x/zA21Ag>

Christoph, Pascal / Pohl, Adrian (2014): Dezentral, offen, vernetzt – Überlegungen zum
Aufbau eines LOD-basierten FID-Fachinformationssystems. In: Bibliothek Forschung und
Praxis, Band 38, Heft 1, Seiten 114–123. ISSN (Online) 1865-7648, ISSN (Print) 0341-
4183, DOI: 10.1515/bfp-2014-0005. Preprint unter <https://wiki1.hbz-nrw.de/x/EYOf>.

Libraries Empowerment Manifesto: <http://etherpad.lobid.org/p/LEM>

Pohl, Adrian (2013): Discovery silos vs. The open web. In: Open Bibliography and Open
Bibliographic Data [Blog], 23.6.2013. URL:
<http://openbiblio.net/2013/06/23/discovery-silos-vs-the-open-web/>

Credits

35

„Discovery-Wurst“ (Folie 25) erstellt unter Verwendung von

- *Configuration, data* icon von Visual Pharm (https://www.iconfinder.com/icons/175662/configuration_data_icon),
- *Lock, open* icon von Stephen Hutchings, https://www.iconfinder.com/icons/216292/lock_open_icon, lizenziert unter CC-SA 3.0,
- *Fleischwolf*, <https://de.wikipedia.org/wiki/Datei:Fleischwolf.ganz.jpg>, Foto von de:user:Kku, modifiziert von de:user:Rainer Zenz, lizenziert unter CC BY-SA 3.0,
- *Sausage* von mcol, <http://openclipart.org/detail/6165/sausage-by-mcol>, Public Domain.

„RDF-Ernte“ (Folie 26) erstellt unter Nutzung von

weizen weizenfeld weizenähren ähren getreide korn von spinheike, lizenziert unter CC0.