

Praktische Erfahrungen aus der Linked-Data-Publikation bei der ZBW

Joachim Neubert

IT-Entwicklung

Deutsche Zentralbibliothek für Wirtschaftswissenschaften
Leibniz-Informationszentrum Wirtschaft

Semantic Web in Bibliotheken (SWIB09)
Köln, 25.11.2009





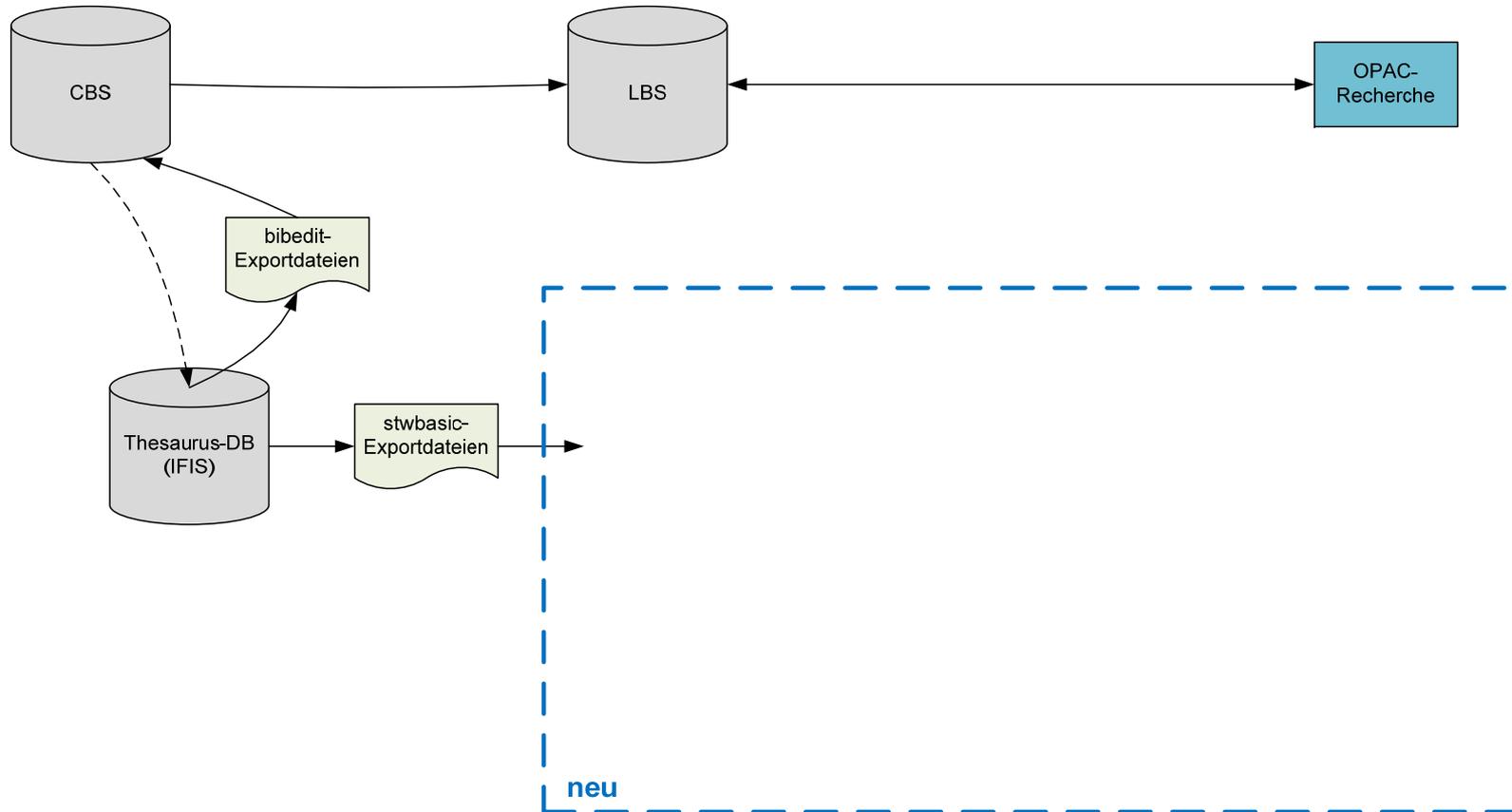
- 1) Womit anfangen?
- 2) Die Daten
 - a. RDF-Vokabulare
 - b. Datenkonvertierung
 - c. RDF(a)-Validierung
- 3) Das Web
 - a. Linked Data Grundlagen – URIs
 - b. URI-Konzept STW und Pressearchiv
 - c. Webserver-(Apache-)Konfiguration
 - d. Linked-Data-Validierung
- 4) Entwicklungsprozess und Community-Feedback

Womit anfangen?



- Überschaubares Gebiet und Datenset aussuchen
- Lizenzierungs- und Attributierungsfragen für die Daten ansprechen und eine Klärung innerhalb der Institution anschieben
- Datenspeicherung und –pflege sowie bibliothekarische Geschäftsgänge möglichst lassen, wie sie sind
- Quer-Abhängigkeiten zu bestehendem Software-Systemen minimieren
- Einfache Datenschnittstellen nutzen
 - z.B. simple, textbasierte Dateischnittstelle

Systemübersicht STW (Anbindung an Produktionssysteme)





- für STW SKOS als Thesaurus-Datenmodell,
für Pressearchiv OAI-ORE (nested aggregations)
- zusätzliche Properties aus anderen RDF-Vokabularen
nach Bedarf, z.B.

```
rdf:seeAlso  
dcterms:creator, dcterms:issued  
owl:versionInfo  
cc:licence
```

- darüber hinaus benötigte Klassen und Properties als
Subclasses und Subproperties von vorhandenen
abgeleitet, z.B.

```
zbwext:Descriptor, zbwext:Thsys rdfs:subClassOf skos:Concept .  
zbwext:ppn rdfs:subPropertyOf dcterms:identifier .*
```

* da der GBV inzwischen die Einführung eines PPN-Property plant (vgl.
<http://www.gbv.de/wikis/cls/GBV-Ontology>), wird sich ein "Eigenbau" hier erübrigen



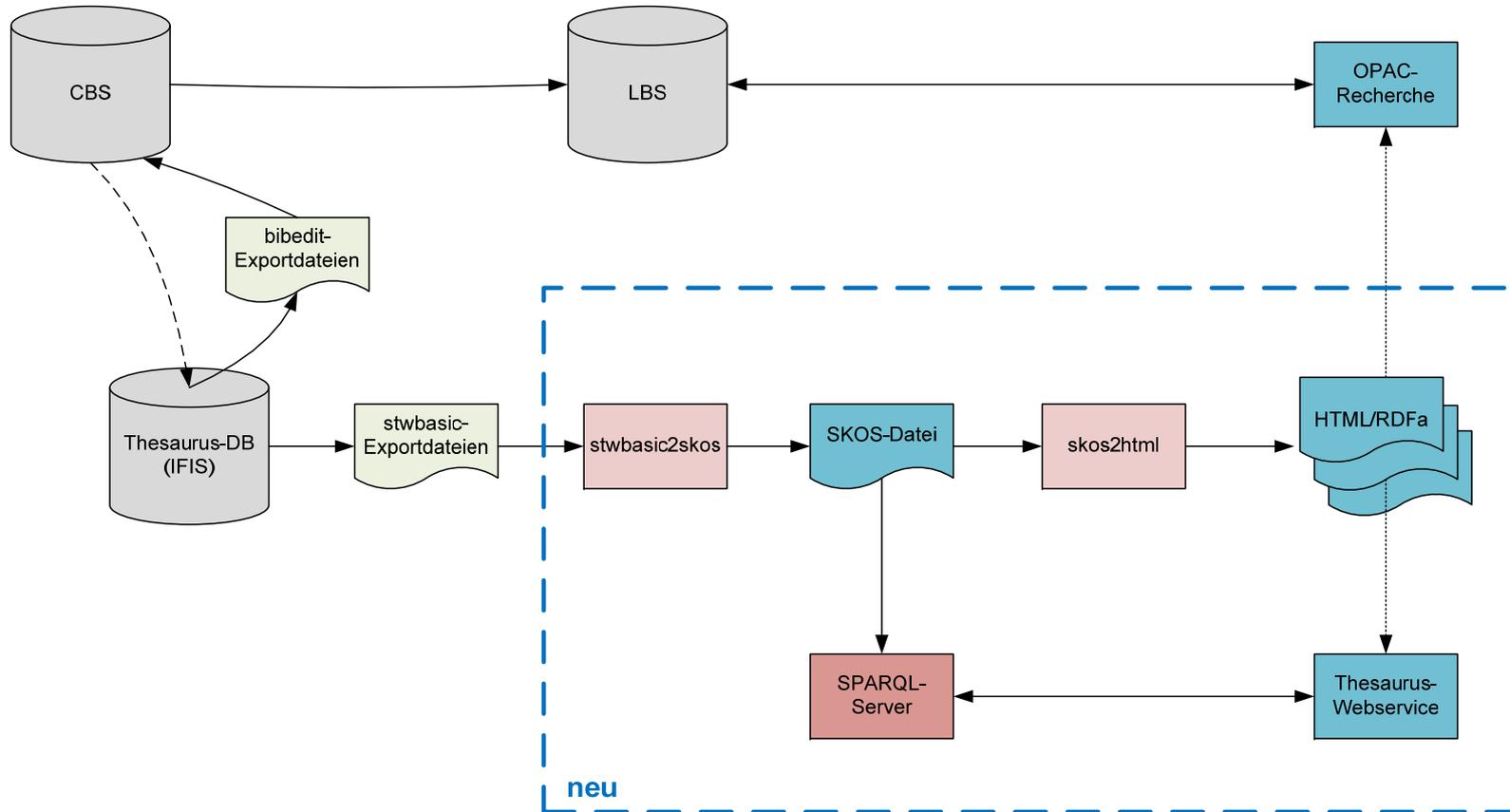
beim STW benutzt:

- Perl mit RDF::Redland Modulen (<http://download.librdf.org/> - CPAN-Fassung veraltet)
- Parsen, Aufbau eines RDF-Modells und Serialisierung per Programm
- SKOS-Datei in RDF/XML (standardisiertes Austauschformat) und Turtle (für Menschen lesbar)
- kein Versuch, RDF/XML direkt zu lesen, zu schreiben oder durch XSLT zu manipulieren
- Ausgabe mit HTML::Template und eingebettetem, "handgeschriebenem" RDFa => Webseite aus statischen Files

Mögliche Alternativen:

- direktes Schreiben einer RDFa-Datei mit Hash URIs (für einfache Anwendungsfälle)
- dynamisch aus Datenbank generierte Seiten (für große Sites)
- ...

Systemübersicht STW als Linked Data



RDF(a)-Überprüfung und Validierung



- Visualisierung von eingebetteten Triples möglich mit <http://www.w3.org/2006/07/SWD/RDFa/impl/js/>
- aber nicht unbedingt korrekt (Entwicklungsstand 2007!)
- generierte Daten nach jeder Änderung der Templates validieren
 - valides XHTML
 - logisch korrekte RDF-Tripel
- komfortable Lösung: einbettete Buttons
 - <http://validator.w3.org/check?uri=referer>
 - (<http://www.w3.org/2007/08/pyRdfa/extract?uri=referer>)
http://zbw.eu/beta/tools/rdfa_extract_turtle_from_referer

[Demo]



- 1) Womit anfangen?
- 2) Die Daten
 - a. RDF-Vokabulare
 - b. Datenkonvertierung
 - c. RDF(a)-Validierung
- 3) **Das Web**
 - a. Linked Data Grundlagen – URIs
 - b. URI-Konzept STW und Pressearchiv
 - c. Webserver-(Apache-)Konfiguration
 - d. Linked-Data-Validierung
- 4) Entwicklungsprozess und Community-Feedback

Semantic Web heißt vor allem: Web



- Web- und Webserver-Know-How akquirieren
- Bedeutung von persistenten Identifizieren (URIs) in der Institution bewußt machen
- Persistenz von URIs kann m.E. nur garantiert werden, wenn
 - die verwendete Domain-Adresse stabil gehalten werden kann
 - Zugriff auf die Konfiguration des Webservers für diese Domain-Adresse besteht (insbesondere reverse proxy und rewrite Funktionen)
- frühzeitig Kontakt zu Verantwortlichen für den produktiven Webserver aufbauen und klären, ob sie das unterstützen
- ggf. Alternativen (z.B. purl.org) erwägen



Unterscheidung zwischen

- der Sache selber (das Konzept "Finanzkrise", der Staat "Frankreich" etc.), und
 - der Web- oder RDF-Datei, die diese Sache beschreibt
- erfordert unterschiedliche URIs.

Die Verbindung zwischen beiden URIs wird ausgedrückt durch

- 303-Redirect von dem URI der Sache auf den URI der Datei
- Hash-URIs (z.B. <http://example.com/xyz#id>, vgl. Cool URIs for the Semantic Web <http://www.w3.org/TR/cooluris/>)



Mappe "Ernst Abbe"

<http://zbw.eu/beta/pm20/person/00012>

Informationen über diese Mappe

<http://zbw.eu/beta/pm20/person/00012/about>

und zwar, je nach Sprach- und Format-Anforderung (per Content Negotiation)

<http://zbw.eu/beta/pm20/person/00012/about.de.html>

<http://zbw.eu/beta/pm20/person/00012/about.en.html>

<http://zbw.eu/beta/pm20/person/00012/about.de.xhtml>

<http://zbw.eu/beta/pm20/person/00012/about.en.xhtml>

<http://zbw.eu/beta/pm20/person/00012/about.de.rdf>

<http://zbw.eu/beta/pm20/person/00012/about.en.rdf>

<http://zbw.eu/beta/pm20/person/00012/about.de.ttl>

<http://zbw.eu/beta/pm20/person/00012/about.en.ttl>

Alle Repräsentation haben den gleichen Informationsgehalt!

Alle HTML-Links führen explizit auf eine HTML-Datei derselben Sprache, damit die benutzer-gewählte Sprache beibehalten wird

Content Negotiation in Apache-Konfiguration



Global (irgendwo in /etc/httpd/conf.d/*) gesetzt:

```
Option +MultiViews
```

In .htaccess-Datei (im Wurzelverzeichnis der Applikation)

```
# deliver en, when no or languages other than de are requested
```

```
LanguagePriority en de
```

```
ForceLanguagePriority Fallback
```

```
# ensure files served with appropriate content types
```

```
# (if not present in main apache config)
```

```
AddType application/xhtml+xml .xhtml
```

```
AddType text/turtle .ttl
```

```
AddType application/rdf+xml .rdf
```

```
# Rewrite engine setup
```

```
RewriteEngine On
```

```
RewriteBase /beta/pm20
```

```
# rely on content negotiation for the correct file extensions
```

```
# 303 redirect for /beta/pm20
```

```
RewriteRule ^/?$ http://zbw.eu/beta/pm20/about [R=303,L]
```

```
# 303 redirect for URIs below /beta/pm20/person, ...
```

```
RewriteRule ^((person|company)(/[0-9]+)*)$ http://zbw.eu/beta/pm20/$1/about [R=303,L]
```



Validierung von Linked Data 303-Redirect und Content Negotiation

- Vapour Linked Data Validator
<http://validator.linkeddata.org/vapour>

Genereller RDF-Validator

- RDF:Alerts
<http://swse.deri.org/SWSEAlerts/>

Hinweise der Pedantic Web Group nutzen:

- "Frequently Observed Problems on the Web of Data"
<http://pedantic-web.org/fops.html>

Entwicklungsprozess und Community-Feedback



- funktionaler Prototyp (mit wenigen Datensätzen)
- build/test/deploy möglichst automatisiert (z.B. mit ant)
 - Entwicklungszyklen werden vielfach durchlaufen
 - Schritt zum Regelbetrieb und Deployment neuer Versionen wird einfacher
- frühzeitig ins Web stellen
- auf Listen um Feedback bitten, z.B.
 - <http://groups.google.com/group/pedantic-web>
 - <http://groups.google.com/group/bibcode>
 - public-lod@w3.org
 - public-esw-thes@w3.org
 - public-rdfa@w3.org
 - semantic-web@w3.org



- How to Publish Linked Data on the Web
<http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>
- Linked Data Tutorial – NG
<http://ld2sd.deri.org/lod-ng-tutorial/>
- RDFa for HTML Authors
<http://www.w3.org/MarkUp/2009/rdfa-for-html-authors>
- ESW-Wiki – Wiki der W3C Semantic Web Community
<http://esw.w3.org/topic/FrontPage>
- Übersicht SemanticWebTools
<http://esw.w3.org/topic/SemanticWebTools>



Vielen Dank!

Fragen?

Links:

<http://zbw.eu/stw>

<http://zbw.eu/beta/stw-ws>

<http://zbw.eu/beta/sparql>

<http://zbw.eu/beta/pm20>

Über die Publizierung des STW als Linked Data:

http://ceur-ws.org/Vol-538/ldow2009_paper7.pdf

Joachim Neubert

Deutsche Zentralbibliothek für

Wirtschaftswissenschaften (ZBW)

Leibniz-Informationszentrum Wirtschaft

j.neubert@zbw.eu