Laying the foundation for sustainable content aggregation, collaboration and innovation

Lessons from the German Digital Library

SADI Workshop
Pretoria/Johannesburg
December 3rd, 2014

Stephan Bartholmei
s.bartholmei@dnb.de
+49 (0) 69 1525-1783
A personal note ...
The Deutsche Digitale Bibliothek in a nutshell

- joint venture of the 16 states & the federal government
- linking up the digital inventories of Germany’s cultural heritage institutions
  - >10 Millionen digital objects (images, text, video & audio) from 160 data providers from 6 domains
    - film & audio archives
    - archives
    - monuments
    - research
    - museums
    - libraries
- normalised, enriched and linked metadata
- available via portal and API
- national aggregator for Europeana
- collaborative network

https://www.deutsche-digitale-bibliothek.de
In the beginning ...
... there was a letter, written in 2005,

from Jacques Chirac to European Commission’s president Manuel Barroso ...
Lessons learned

A top-down approach usually leads to extremely unrealistic expectations – especially, if started by politicians.

Initial funding is excellent, sustainable operation is difficult.

Top-down approaches favor Big Design UpFront – slow, tend to produce outdated results.

Resist the temptations of „All you can do, I can do better.“

Be agile.

The concept of „digital libraries“ is very much in the flow, we have to make it up, as we go along.
As important as presentation is ...
... DO NOT neglect the infrastructure!
but beware of your greatest enemy ...

die eierlegende Wollmilchsau (egg laying & milk giving sheep-pig)
In the cultural sector, Germany doesn’t look like that.
Baden-Württemberg  Bavaria  Berlin  Brandenburg  Bremen  Hamburg  Hesse

Lower Saxony  Mecklenburg-Vorpommern  North Rhine-Westphalia  Rhineland-Palatinate  Saxony-Anhalt  Thuringia

16 states – 16 legislations
There's no federal digitization programme (yet).

BUT: There is federal funding for digitization projects at single institutions via the German Research Foundation.
2008-2009: concept phase

- Spring 2008: initial concept presented by joint federal-state expert group
- Spring 2009: feasibility study by Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS)
- October 2009: states’ prime ministers and
- December 2009: federal cabinet pass

DDB’s founding documents:
1. common outline of the federal government, the
Competence Network
Deutsche Digitale Bibliothek
Organisational Structure

- board of trustees („Kuratorium“)
- general assembly of the competence network
- managing board (H. Parzinger, R. Griebel, R. Kretzschmar)
- CEO (F. Frischmuth)
- admin. office
- coordinator
- service desk
- technical operator
- working groups
- domain desk
- development partners
- museums:
  - Frankfurt a. M.
  - Institut für Museumsforschung
    Staatliche Museen zu Berlin
November 28, 2012: Beta-Launch
Lessons learned

beta launch

• A lot of ideas from the concept papers and founding documents didn’t make it into the prototype.

• We chose the wrong data model (CIDOC-CRM) => switch to EDM
  • working implementations: (n)one
  • domain experts were unfamiliar
  • no interoperability with Europeana

• We had to re-implement our frontend. Outsourcing core tasks = bad idea. => technical operator with frontend development expertise

• Basic concept (data clearing, ingest) worked!

• Huge benefit: mutual improvement of domains’ metadata culture through collaboration in network’s working groups.

• We need a central service desk!
From the beta launch to normal operations...
How do the data get into the DDB?
The Service Desk
from data...
... to digital objects
preliminaries

registration

content survey

cooperation agreement

first contact
data clearing

preliminaries
formalities & Co.

quality assurance
technical mapping
conceptual mapping
test data
content survey
cooperation agreement
registration
first contact

< dcterm s:language ns3:resource="http://id.loc.gov/vocabulary/iso639-2/ger"
xm lns:dctem rms="http://purl.org/dc/terms/"/>
< dcterm s:LinguisticSystem ns3:about="http://id.loc.gov/vocabulary/iso639-2/ger"
xm lns:dctem rms="http://purl.org/dc/terms/"> <ns3:valu e>ger</ns3:value> 
</ dcterm s:LinguisticSystem>

< datafield t ag="041" ind1=" " ind2=" ">
< subfield c ode="a">ger</subfield>
</ datafield>
real processes are quite complex
data workflow (simplified)

delivery  splitting  clean-up  transformation  enrichment
person entity pages

- contextualization
- entity pages as nodes for the semantic web
- linkage to authority files
  - dates of birth & death, etc.
  - further links (external)
- filled by web service „Entity Facts“
  - basiert auf Gemeinsamer Normdatei (GND)
- web of related objects (internal)
How does the DDB work?

**institute portal**
- lokale Recherche
- full view

**institution repository**
- metadata
- digital objects

**ASC**
- transformation
- ingest

**mapping rules**

**Cortex („Backend“)**
- Metadaten
- Derivate

**DDB portal („Frontend“)**
- search engine
- preview

**API**

**configuration**
(facets, search fields, ...)

references
A closer look at a DDB search...

Frontend (portal)

Backend (data pool)

search index

Repository

API
Application Programming Interface
content

- digital objects (derivatives)
  - digital first
  - no relevance criteria regarding content
  - formal criteria
    - metadata in standard formats
    - persistent identifiers, stable links
    - objects must be freely accessible
    - sustainable infrastructure for data supply
- accompanying metadata
  - default licensing CC0 (cooperation agreement)
  - exceptions: e.g. long descriptions

http://www.deutsche-digitale-bibliothek.de/item/26JVK5E5WTHCKZR4DARVHDD7QFSGMEPL
Digression: Content Re-Use
The people use the internet differently than 10 years ago. How do we react?
„As a museum we are used to having the people, either visitors or researchers, come to us.“
Our users decide, where, when and in which context they are reached by our data and services.
Portal ➔ Platform ➔ Ecosystem

standardization
open interfaces
community
tools & infrastructure
We hide our treasures in the „Deep Web“.
API usage example (1)

Twitterbot @ddbkatzen
by Peter Mayr (Vortragsvideo https://www.youtube.com/watch?v=l_i-_2YEpT0)
source code: https://github.com/hatorikibble/twitter-ddbkatzen
There’s plenty of room on a platform ...
API usage example (2)

Archivportal-D
domain specific frontend for the DDB [https://www.archivportal-d.de/](https://www.archivportal-d.de/)
API usage example (3)

map-based search in the data of the Hessian monument preservation agency
web application:  
source code:  

https://github.com/Deutsche-Digitale-Bibliothek/ddb-mapshowcase
API usage example (4)

interactive visualization of DDB’s data
by Prof. M. Dörk (FH Potsdam)  
http://infovis.fh-potsdam.de/ddb/
CODING DA VINCI

Der Kultur-Hackathon
26./27. April – 5./6. Juli 2014

http://codingdavinci.de/
16 INSTITUTIONS
Participating Cultural Heritage Institutions

BERLINISCHE GALERIE MUSEUM FÜR MODERNE KUNST

ETHNOLÓGICO MUSEO DE BERLÍN

LANDESARCHIV BADEN-WÜRTTEMBERG

STAATSBIBLIOTHEK ZU BERLIN

MAX-PLANCK-INSTITUT FÜR WISSENSCHAFTSGESCHICHTE

STIFTUNG STADTMUSEUM BERLIN

GÉRÓF ECKERT INSTITUT

DEUTSCHE NATIONALBIBLIOTHEK

HERZOG AUGUST BIBLIOTHEK WOLLENBÜTTEL

MUSEUM FÜR NATURKUNDE BERLIN

TEXTGRID BERLIN

ZLB
26 open data sets
Botanical Garden Berlin

Natural History Museum

city museum Berlin

26 Datensets — http://codingdavinci.de/daten/

Berlinische Galerie

Steinheim Institut, Essen

Georg-Eckert-Institut, Braunschweig
150 participants
27 project ideas
Startup Weekend
26.-27. April, Berlin
Data & Hacking

Der Kultur-Hackathon

10 week „sprint“
17 teams

presentations & award ceremony
06. July, Berlin
17 presentations
5 winners
Willkommen bei EthnoBand!

Mit EthnoBand können sie die Instrumente aus der Sammlung des Ethnologischen Museums selbst spielen! Ob mit Freunden, Familie oder alleine - einfach ein Instrument aussuchen und über die virtuellen Anschlagflächen spielen.
March 31, 2014: Presentation of DDB 1.0
Lessons learned
2013/2014

• Strategic change of focus: supply driven -> target audience oriented
• We have to decide what our main product is (ongoing).
• We have to keep developing from portal to platform.
• Our processes are too expensive.
• Google offers better search on our data than we do.
• 2/3 of our users come via Google => SEO, schema.org
• Nobody (exc. machines) is that much into metadata. Context is king.
• Content is king. In 2015 we’ll focus on ingesting new collections.
• After 10 years of broken promises, Linked Open Data is finally coming.
• We need permanent innovation. (And more money...)
Thanks for having me. When will I see the South-African Digital Library online?

Stephan Bartholmei
s.bartholmei@dnb.de
+49 (0) 69 1525-1783
Skype: stephan.bartholmei