Open Data in Cultural Heritage – OpenGLAM in Germany

1 Year in Digital Cultural Heritage:
The Walls I Ran into Most Often...

Stephan Bartholmei
s.bartholmei@dnb.de
Deutsche Digitale Bibliothek

and how to wear them down

DISCLAIMER: THIS IS A VERY PERSONAL TALK AND THE OPINIONS PRESENTED THEREIN ARE MY OW. THEY MAY NOT NECESSARILY BE THE OFFICIAL POSITIONS OF THE ORGANIZATIONS I WORK FOR.

Introduction:
- technical project coordinator for German Digital Library https://www.deutsche-digitale-bibliothek.de (short:„DDB")
- at German National Library http://www.dnb.de
- trained physicist, switched careers 1 year ago (so take my statements with a grain of salt)
- anecdote: held original letter by Albert Einstein 1st day at work => sense of belonging, right place for me
- transition from end user to provider of cultural data almost completed: ideal moment to talk about some possible improvements

Explanation of „strange" title:
- chinese folk myth: chinese ghosts can only walk in straight lines => defense mechanism: place otherwise useless „shadow walls" behind doors and portals, so that the ghosts run into that walls
- I ran into some walls during my first year, more about that later...
- possible metaphor for some obstacles that present themselves again and again in digital cultural heritage
- solution maybe not to wear nor even tear those walls down but to find our ways

There's a time shift ...

"Dr. Who" by aussiegal
http://commons.wikimedia.org/wiki/File:Dr_Who_(316350537).jpg

... there's a time shift, or a time lag of about 10 years: To put it bluntly, the Internet is perceived like it was still 2004:
- web portals are seen as state of the art
- either the mobile and the social revolution haven't registered with a lot of people
- APIs are few and far between

But things are about to change:
- Europeana Labs in beta http://labs.europeana.eu/
- DDB published her API in November 2013 http://api.deutsche-digitale-bibliothek.de/

Our data and services must go where our endusers already are.

© www.fotogenese.de (Axel Zucker) http://www.flickr.com/photos/29745442@N04/4720105077

Today, everybody moves in her/his „filter bubble" to prevent drowning in the flood of information.

- Except from those with an extra motivation nobody actively hunts for data & information anymore.
- Consequence: We must offer our data and provide our services where our endusers are:
    - physically: on mobile devices and other appliances
    - community-wise: social media!
- Even better: empower others to write the apps and build the websites that are taylored to the specific needs of the people.

One important way to do this is to provide a public API, I'm going to present two examples how the DDB API is used.

„DDB Katzen" – Twitterbot by Peter Mayr
- came up almost immediately after API publication
- copy cat of similar bot by DPLA http://twitter.com/HistoricalCats, more serious example: Europeanabot
- underlying concept of serenditpitous encounter much stronger tnan this play on an internet meme
- insightful tlak by Peter Mayr https://www.youtube.com/watch?v=l_i-_2YEpT0, about how to bring light to the dark corners of the deep web

We're hiding our treasures in the deep web.
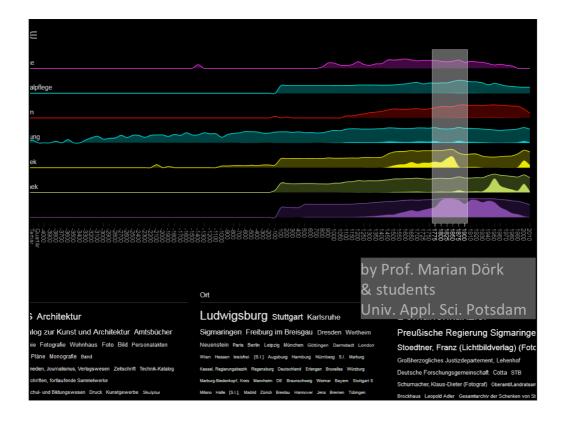
Riesending-Schachthöhle

Right now, we're hiding our treasures behind search slits, in catalogues and portals ... like Gollum.

Our users need expert knowledge, and special kind of attitude to go speluniing in our data.

They may get lost sometimes like the explorer who hurt himself in the depicted cave some weeks ago
http://de.wikipedia.org/wiki/Riesending-Schachth
%C3%B6hle#Unfall_des_H.C3.B6hlenforschers_Johann_Westhauser

We need different access methods so that different users can discover our treasures differently, but more easily.

One possible approach demonstrated in this screenshot of an interactive visualization of the DDB data as a whole.

- by Prof. Dr. Marian Dörk and his students from the University of Applied Sciences Potsdam
- shown: number of objects itemized by cultural sector over time
- focus on the libraries (color: yellow): drop of number of digitized works after 1875 apparent
- „black hole of copyright": a work, written in 1875 by an author at age 20, who lived for 90 years will enter the public domain in 2015
- rise in modern days due to works that are published digitally and needn't be digitized (they are not necessarily in the public domain)

(Copy)rights issue is the biggest obstacle of all.

While we – as a future OpenGLAM group – won't be able to change the law single-handedly, there are two things we can do:
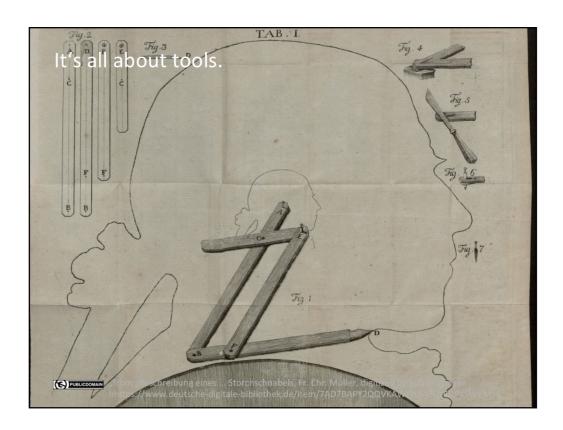- educate the cultural institutions, especially the smaller ones
- empower them to …

Let's be a little bit bolder.

© Deutsche Fotothek, „Dampfer Lützow, Deckspiele, Kissenkampf"
https://www.deutsche-digitale-bibliothek.de/item/USDLN3KIOONMYNLOJ66WTQ2Z6ZOB5TYL

... be a little bit bolder, especially in legal matters

<some controversial remarks>

Say, you're a small cultural institution, e.g. a one-person-library in a museum and have some objects you'd like to publish digitally under an open license. Than there are other problems ...

First you need to digitze your objects, e.g. scan Berlin's first telephone book.
- access to the big digitization facilities is nearly impossible
- there are initiatives like digiS (http://www.servicestelle-digitalisierung.de/) and its counterpart in Brandenburg (http://www.fh-potsdam.de/informieren/ organisation/fachbereiche/informationswissenschaften/fachbereich/ koordinierungsstelle-brandenburg-digital/) to overcome that hurdle
- to make your digitized objects findable and consumable you need
    - provide metadata, publish them in portals like DDB, Europeana
- metadata standards are not your day-to-day business, tools available are too complicated or
- not availabe at all, quote „In 25 years I haven't seen a better tool for conceptual mappings than an excel sheet."
- vendors of collection management software don't support standards
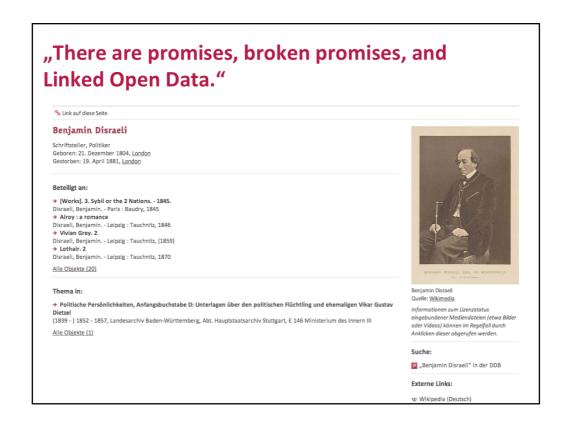
... and infrastructure

Somebody has to host the data.

Lots of development going on, Europeana and DDB haven't found their final form and business model – will look quite differntly in a few years.

Let me close with a personal rant of mine ...
Maybe you know the famous quote about lies from 19th century British prime minister Benjamin Disraeli: „There are three kinds of lies: lies, damn lies and statistics."

When I started about one year ago, I read enthusiastically about ontogies, RDF, triple stores, SPARQL queries ... only to find out, that for many of these concepts there are no robust implementations out there. One cURL statement may be enough to bring down every triple store around.

... so I was quite disillusioned about Linked Open Data.

- But, as of latest, not so much anymore ... I'm tentatively optimistic , that finally concepts, technology and acceptance may come together.
- similar to „Internet of things"
- example: DDB person entity pages,
    - Benjamin Disraeli: https://www.deutsche-digitale-bibliothek.de/entity/ 118526014
    - fed by LOD service at DNB, that pulls life data & links from GND („Gemeinsame Normdatei" = Integrated Authority File), portrait pictures from Wikipedia
- CHANGE IS IN THE AIR, another sign for that is presented by Helene Hahn, who will talk about the phenomenal success of Germany's first hackathon with open culture data

Thanks for having me.