

# Linking Data and/with people

Sophie-Carolin Wagner, Christiane Fritze, Stefan Karner, Stefan Frühwirth, Georg Petz, Can Yilmaz, Christoph Steindl, Paul Sommersguter

# Überblick

- Strategische Ausrichtung der ÖNB
- Digitale Editionen an der ÖNB
- Library Labs an der ÖNB
- Crowdsourcing an der ÖNB

# ÖNB – Vision und Strategie

- September 2012

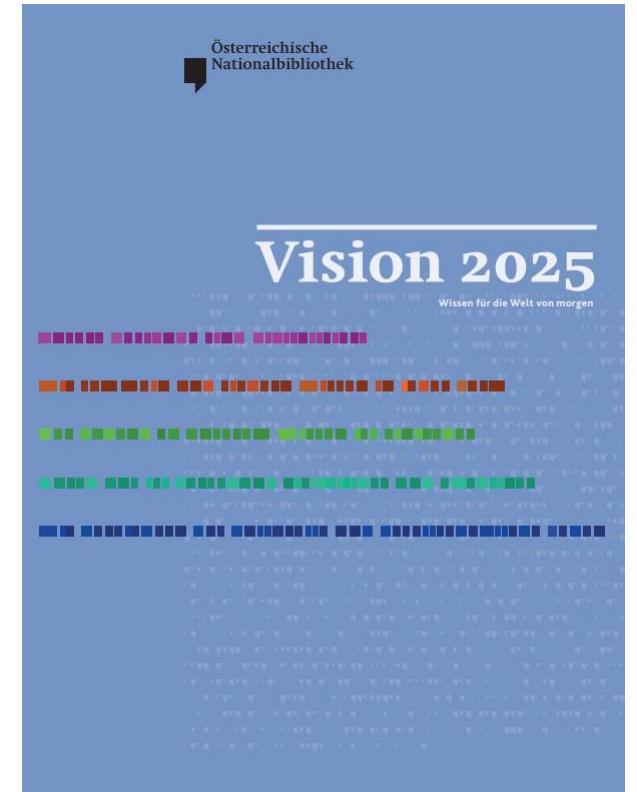
**Unsere Bestände sind digitalisiert**

**Wir sammeln und sichern Wissen in jeder Form**

**Der Zugang zu unserem Wissen ist einfacher**

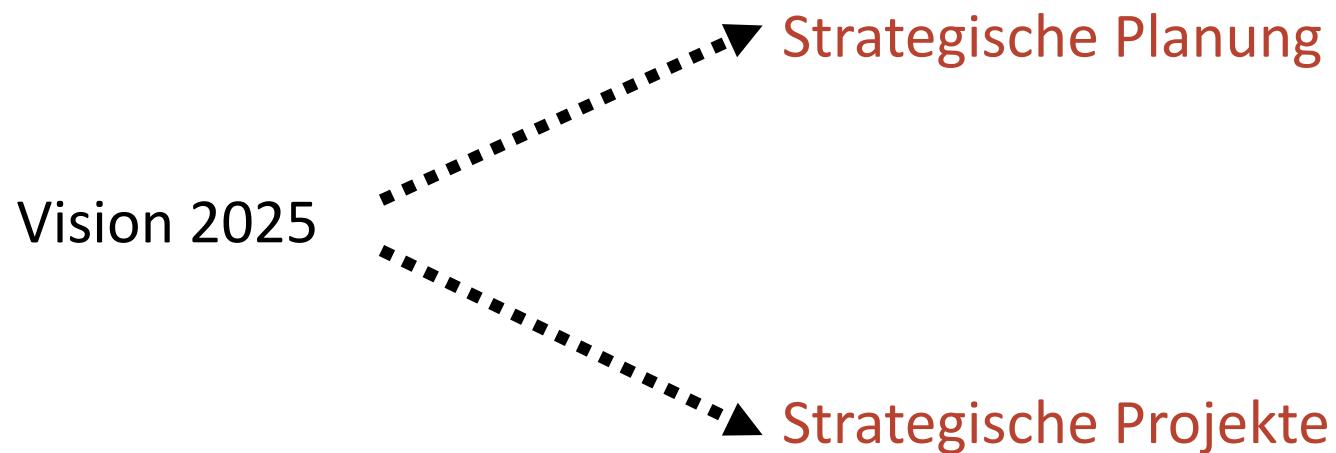
**Durch uns ist Forschung vielfältiger und effektiver**

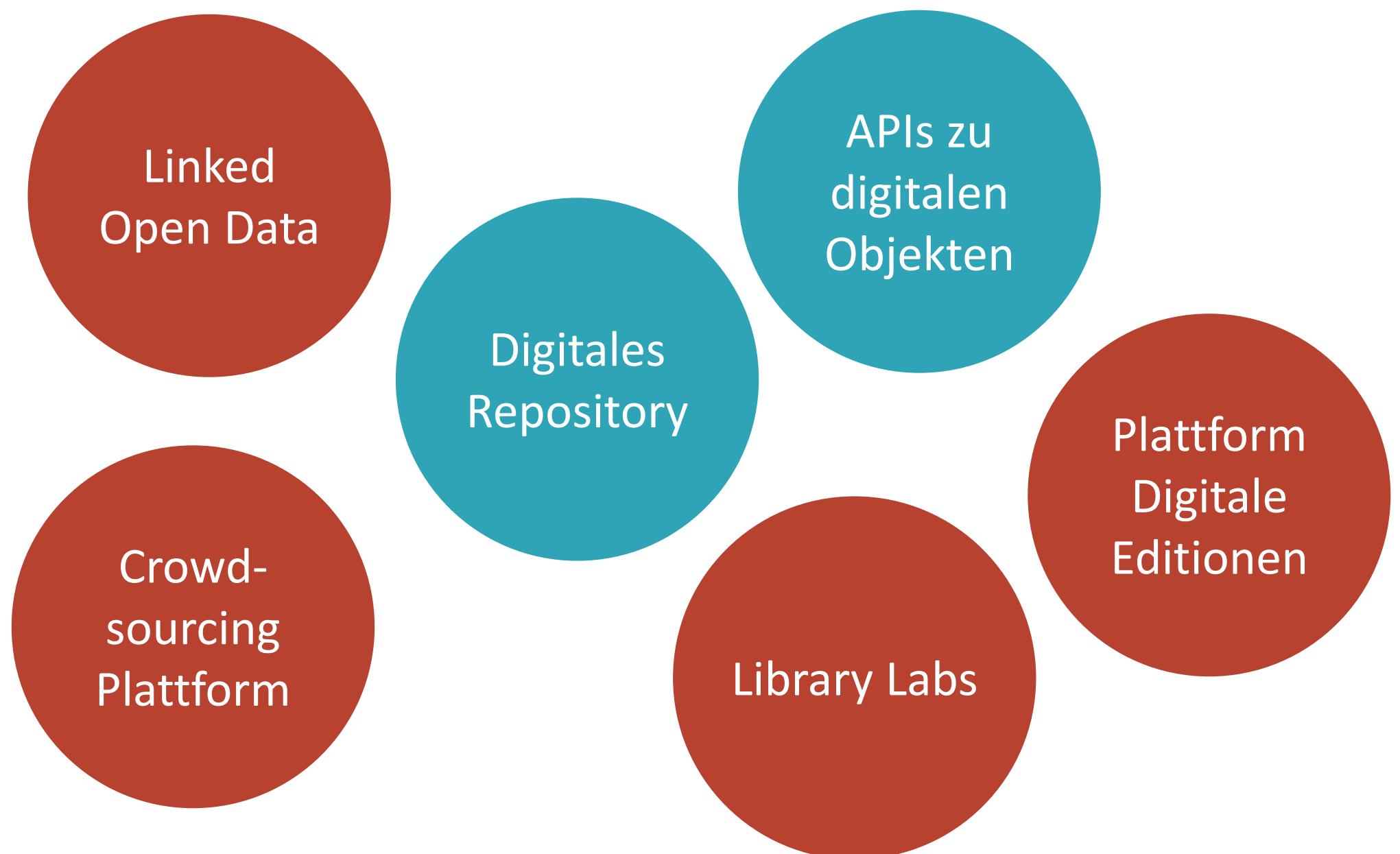
**Wir bereichern das kulturelle und gesellschaftliche Leben**



# ÖNB – Vision und Strategie

- Jänner 2017





# Strategiethemen zu Digital Humanities

---

Unsere Bestände sind digitalisiert

---

Wir sammeln und sichern Wissen in jeder Form

---

Der Zugang zu unserem Wissen ist einfacher

---

Durch uns ist Forschung vielfältiger und effektiver

---

Wir bereichern das kulturelle und gesellschaftliche Leben



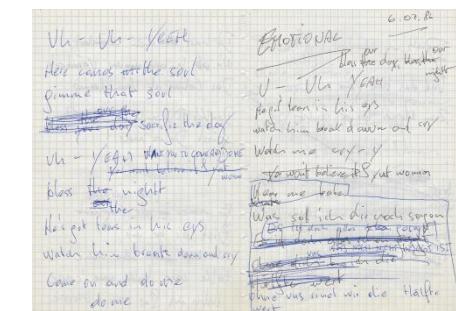
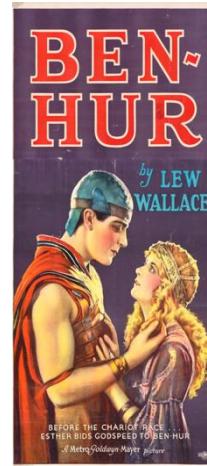


# NACHHALTIGE INFRASTRUKTUR FÜR DIGITALE EDITIONEN

# Bestände an der ÖNB

- Österreichische Nationalbibliothek verfügt über vielfältige, herausragende Bestände
- Quellen für Editionsprojekte
  - Literaturarchiv
  - Sammlung von Handschriften und Alten Drucken
  - Musiksammlung
  - Bildarchiv und Grafiksammlung
  - ...

# Vielfalt der ÖNB-Bestände



# Verortung Digitaler Editionen

- **Frage:** Wohin gehört eine Digitale Edition?
  - Dorthin, wo das Material erforscht wird – also in die Universitäten und Forschungsstätten?
  - Dorthin, wo die Quellen liegen – also die Gedächtnisinstitutionen wie Archive, Museen und Bibliotheken?
- **Antwort:** Ins Netz.
  - Für den Nutzer ist die Antwort bezüglich der Verortung egal, solange der Zugang dauerhaft gewährt wird.

# Ziele

- Zukünftige Editionsprojekte der ÖNB kommen in gemeinsame, nachhaltige Infrastruktur
- Aufbau & Erweiterung der Kompetenzen für Digitale Editionen in der ÖNB
- Nachhaltige Infrastruktur für Digitale Editionen erhöht die Sichtbarkeit der einzelnen Editionsprojekte

# Umsetzung

- Eine technische Basis für alle Digitalen Editionen an der ÖNB
  - Für eigene Projekte und Kooperationen mit Projekten, die mit Beständen der ÖNB arbeiten
  - Komponenten für Erstellung und Präsentation Digitaler Editionen
  - Einschulungen
  - Edition zum „Mitnehmen“
  - Vernetzung mit nationalen & internationalen Partnern



# Welche Technologien nutzen wir?

- Allgemeine Infrastruktur im Backend
  - GAMS (ZIM-ACDH)
- Generische Komponenten
  - TEI-Basiskodierung
  - Automatische Annotation via Normdateien
  - Synoptische Darstellung von Faksimile und Transkription
- Editionsspezifische Komponenten
  - Verschränkung von Themenkommentaren und Navigation
  - Bezüge zu Bibliotheksverzeichnissen



# Hybridition Tagebücher

## Andreas Okopenko

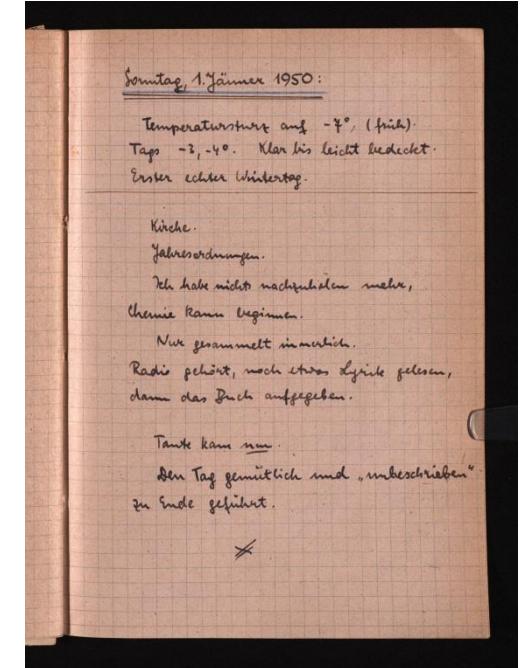
- FWF-Projekt in Kooperation mit der Universität Wien (2015-2018)
- Zeitraum: 1950-1955
- Umfang, ca. 4300 Seiten
  - Bisher ca. 3700 Seiten transkribiert und nach TEI kodiert



```

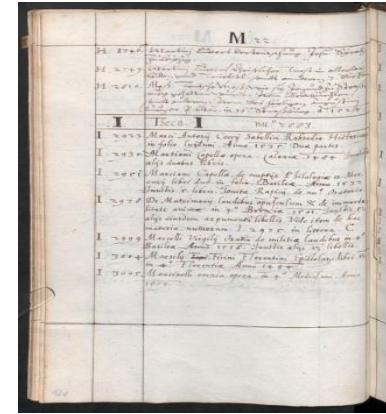
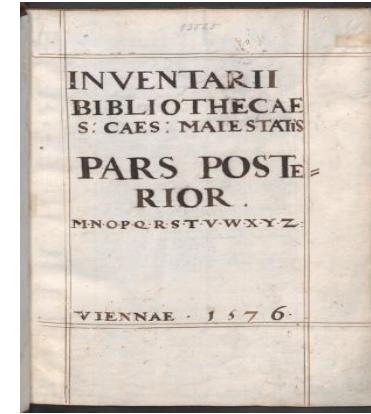
<TEI>
<fileDesc>
  <titleInfo>
    <title>Gewalttaten in Graz Wien und Voitsberg</title>
    <statement>Ein Kaffetessen wird gesucht, viele Famili- und Spaziergäste eingeschlagen, Autos umgestoßen</statement>
    <statement>Angriffe auf die Arbeitnehmenden</statement>
    <statement>Krawalle und Steinwürfe auch in Innsbruck</statement>
    <statement>Ordnungsstörungen</statement>
  </titleInfo>
  <textCreation>Keine aktiven Beamten über 65 Jahre</textCreation>
  <textCreation>Wiener Portier gewinnt im Sportloto 286.000 !</textCreation>
  <textCreation>Über 25 Millionen Schilling Gesamteinatz</textCreation>
  <textCreation>Doppelkette 1949 Hietzinger Poststraße</textCreation>
</fileDesc>
<text>
  <!-- Text content of the newspaper clipping -->
</text>
</TEI>

```



# Digitale Edition des Blotiuskatalogs

- Erster Bibliothekskatalog von Hugo Blotius: „Inventarii bibliothecae S. C. Maiestatis pars posterior (M-Z)“
- Zeitraum: 1576
- Umfang: 270 Blatt
  - Zurzeit: laufende Transkription nach TEI
  - [http://data.onb.ac.at/rec/  
AL00163360](http://data.onb.ac.at/rec/AL00163360)



# Aktueller Projektstand

- Prototypentwicklung für Projekte Okopenko und Blotius
- Laufende Transkriptionen für Blotius und Okopenko
- Beratung in technischen Belangen für zukünftige Projekte



# LIBRARY LABS

# British Library Labs

## Experiment with our digital collections



Search

Home

BL Labs Awards

BL Labs Events

Digital collections

- Datasets
- Images
- Books / Text
- Music
- Maps
- Sounds
- Multimedia

Resources and tools

Digital Scholarship blog

Contact BL Labs

About BL Labs

BL Digital Scholarship

British Library website

BL Labs Competition

## British Library Labs

British Library Labs is an initiative funded by the [Andrew W. Mellon Foundation](#) which supports and inspires the public use of the British Library's [digital collections and data](#) in exciting and innovative ways. We actively encourage scholars and innovators to work with us, thereby supporting the British Library's mission to guarantee that the wealth and diversity of the Library's intellectual digital heritage is available for the research, creativity and fulfilment of everyone.

### Engage and collaborate with British Library Labs - Enter our Annual Awards!

- Learn more and enter the British Library Labs [Awards](#) which recognises outstanding work that has been created using the British Library's [digital collections](#), **deadline is midnight Tuesday 11th October, 2017**.



['16x16 Colourful Faces from the British Library Collection'](#), created by [Mario Klingemann](#), a [winner](#) of the Labs Awards. The images are collated from the Library's [Flickr Commons 1 million collection](#).

# Join us and explore the KB's digital treasure trove

The KB Lab hosts all experimental tools and data sets based on the KB's digitised collection.



4.878 lines of code



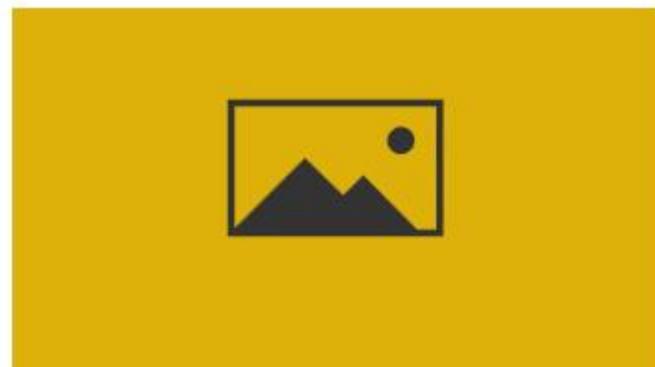
40.330 MB files



11 events

## Datasets

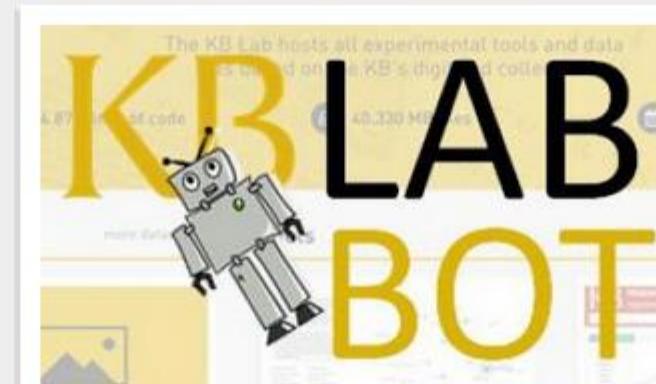
more datasets



KBK-1M

## Tools

more tools



KB Lab Bot



Newspaper ngram viewer

# LABO // LE BLOG

vendredi 16 janvier 2015

## La BnF expérimente le crowdsourcing



0 commentaires

Depuis le 24 novembre 2014, la BnF a lancé une expérimentation de production participative (*crowdsourcing*) suite à mise en ligne de la plateforme Correct.

Cette plateforme expérimentale est le fruit d'un projet de recherche réunissant 9 partenaires : Orange (pilote du projet), BnF, Jamespot, Urbilog, I2S, ISEP, INSA de Lyon, Université Lyon 1,

Rechercher dans ce blog

Suivez-nous...

- Sur notre fil Twitter
- Sur le site BnF

Suivre le flux RSS/Labo BnF<sup>21</sup>



This page has been archived. It is no longer updated.

---

[Articles & Databases](#)[Research Divisions](#)

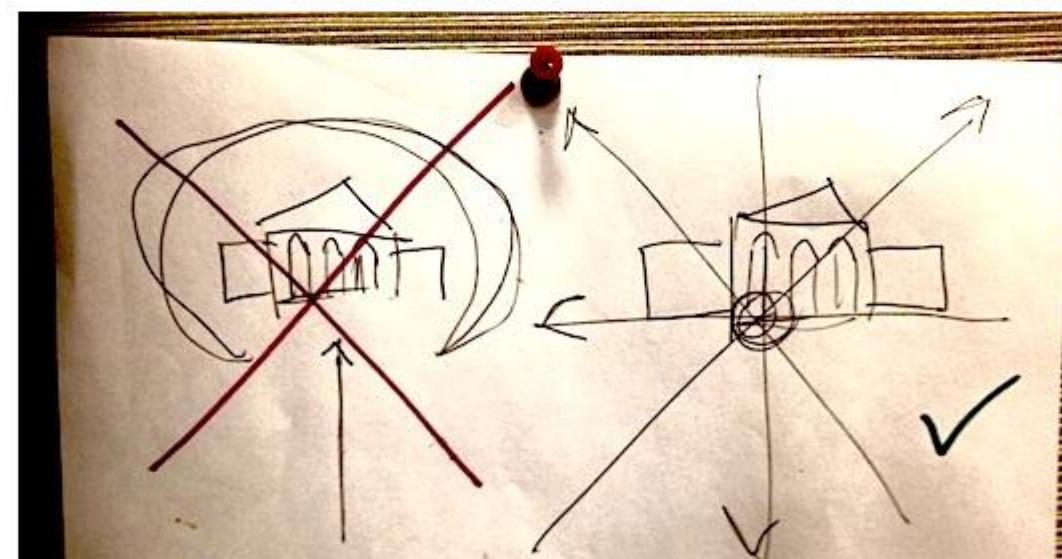
---

[NYPL Labs](#)

---

[NYPL Recommendations](#)

## NYPL Labs

 [Like 95](#) [RECOMMEND](#) [Tweet](#) [EMAIL](#) [PRINT](#) [SHARE](#)

### FROM OUR BLOGS

Informed Archives: The  
Straphangers Campaign and the  
NYC Subway System

We are being held in the station due  
to train traffic ahead of us. We  
apologize for the inconvenience and  
should be moving [READ MORE](#)

# ÖNB Labs – Ziele

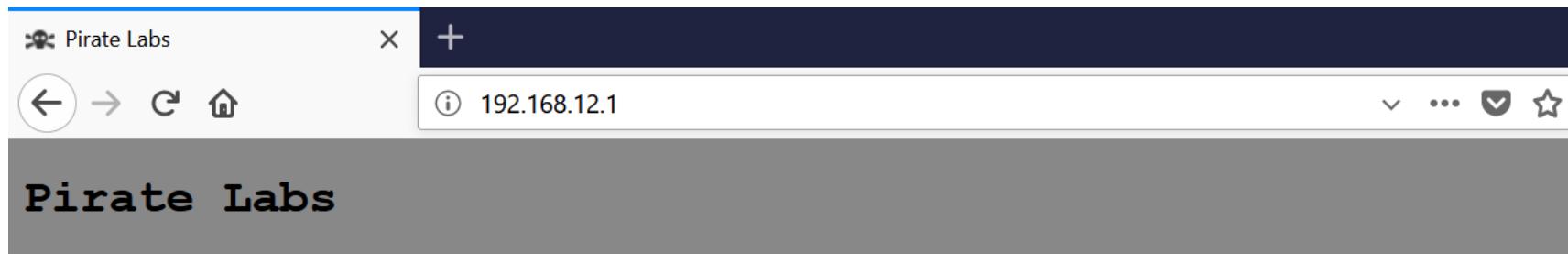
- Öffentlich sichtbare **innovative Dienste** entwickeln
  - Neue Nutzungsszenarien ermöglichen
  - Dienste im **Dialog mit Interessensgruppen** entwickeln
  - **Offen Lizenzierte** Digitalisate anbieten
  - Interaktiv Arbeit mit kuratierten Inhalten ermöglichen
- 
- **Start der Library Labs geplant für September 2018**

# ÖNB Labs

- Zugänglichkeit der Datensätze
  - (offene) Datenbestände auflisten und visualisieren
  - Raw Data Access gewährleisten
  - Schnittstellen anbieten
- Arbeit an den Daten



# 192.168.12.1





Wiener Bilder-ABC-Buch für brave und fleißige Kinder, ~1850, Wien

- [Download scans \(archive\)](#)
- [Download catalog entry - all formats \(archive\)](#)
  - [Download catalog entry - MARC json](#)
  - [Download catalog entry - MARC xml](#)
  - [Download catalog entry - MARC text](#)
- [Download OCR \(single pages, some pages missing\) - all formats \(archive\)](#)
  - [Download OCR as hOCR \(archive\)](#)
  - [Download OCR as ALTO \(archive\)](#)
  - [Download OCR as text \(archive\)](#)

Sign up for our newsletter!

Email:

 register

# ÖNB Labs

- Anreicherung der Daten
  - Annotationen
  - Kommentare
- Kollaboration
  - Kommunikation Projektteams
  - Kommunikation der Lab User
  - Erstellen und Teilen von Kollektionen
- Begleitende Events

# ÖNB Labs - Ausbaustufen

- Data Discovery
  - verborgene Schätze auffindbar machen
  - Durch User aufbereitete Daten teilen
- Datenanreicherung
  - Externe Daten einbeziehen
- Sharing is caring
  - Sharen von User Tools
- Feedbacksystem einbinden

# Herausforderungen

- Datenbeschaffenheit
  - Clean data – dirty data
  - Verfügbarkeit und Zitierbarkeit garantieren
- Rechtemanagement
  - Offene Quellen
  - Datenanreicherung
- Hausinterne Hürden
  - Migration aus Altsystemen

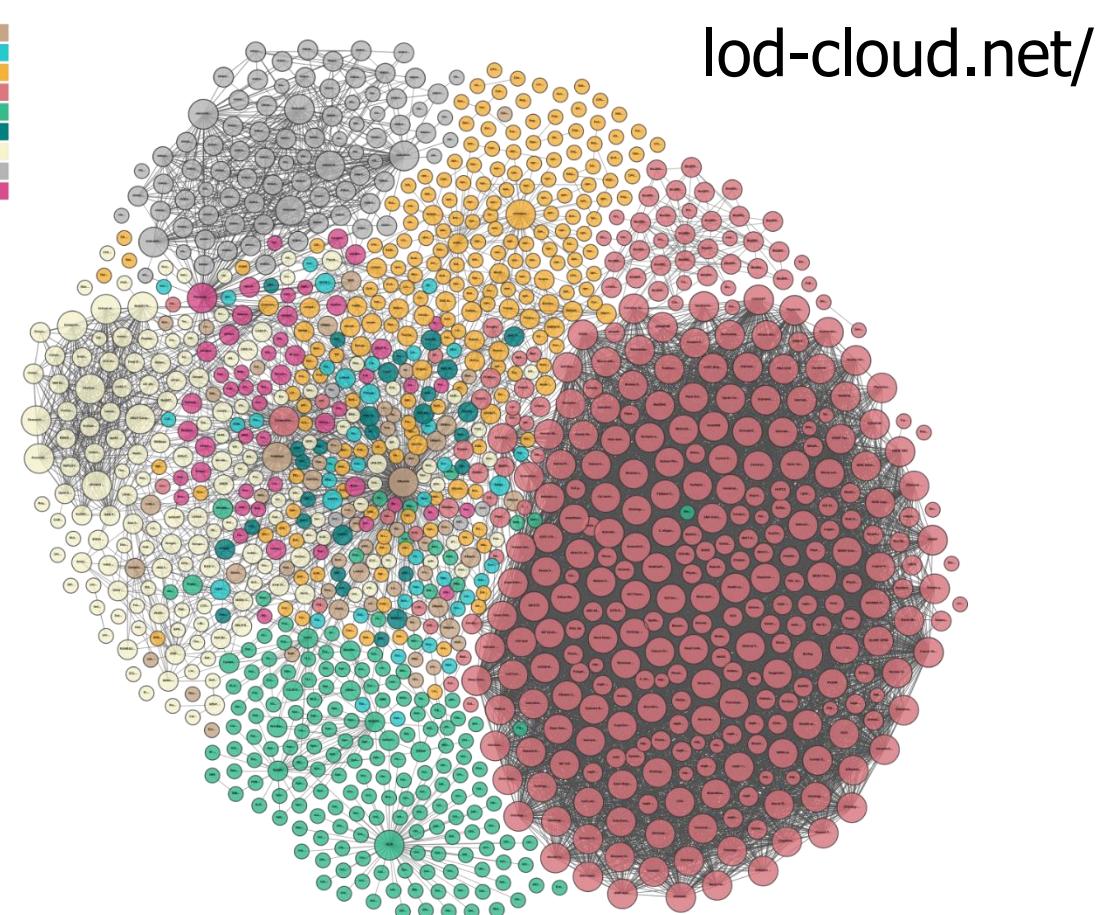
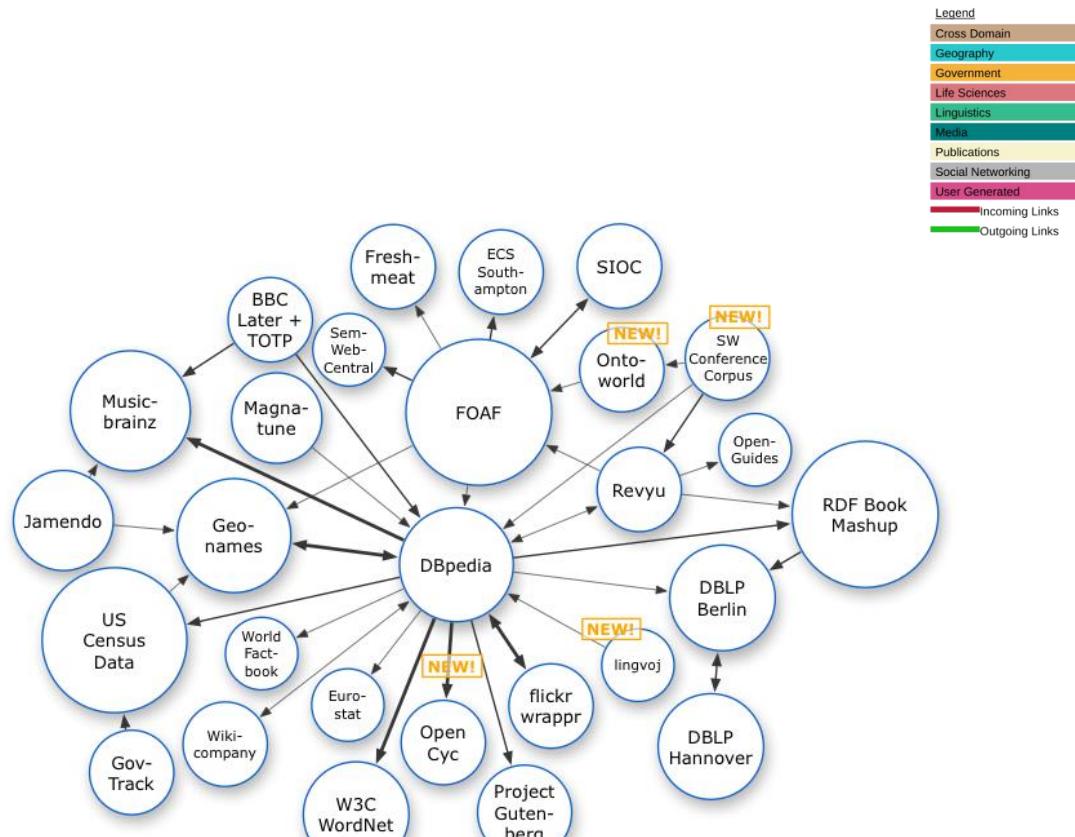
# Linked Open Data (LOD)

- Voraussetzungen zur Verlinkung von Datenquellen & Schaffung des „Semantic Web“:
  - Metadaten in gut strukturierter Form
  - Verwendung von etablierten Vokabularen

# Bereitstellung von LOD

- Abruf von Datensätzen pro Titel/Person etc.
- Publikation des kompletten Datensatzes („Dump“)
- Zugriff via Datenbankabfrage mit der RDF-Abfragesprache SPARQL
- Zugriff via LOD API

# LOD cloud diagram 2007 – 2017



[lod-cloud.net/](http://lod-cloud.net/)



# CROWDSOURCING

# Ziele der Crowdsourcingplattform

- Intellektueller Input von „außen“, den die ÖNB selbst nicht erbringen kann
- Öffnung des Hauses
- Korrekte, standardkonforme Metadaten
- Bildung einer Crowdsourcing-Community

# Community Building

- Raum für Austausch schaffen
  - Laufender Austausch zwischen ÖNB-ModeratorInnen und Crowdworkern
  - Bereitstellung von Infrastruktur für den Austausches zwischen den Crowdworkern
- Persönliches Kennenlernen
  - „Appreciation Events“ für die aktivsten Crowdworker





# Ablauf Crowdaktion

- Digitalisierung
- Publish auf Plattform
  - Aufgaben & Tutorials für Community
- Crowdwork
  - Annotation, Transkription, Diskussion
- Präsentation der Ergebnisse
  - Feedback der Crowd
  - Überführen der Metadaten ins bibliothekarische Zielsystem

# UX Leitlinien

- Keep it simple
- Focus on the content
- Emotionale Aufwertung
- Motivation
- Don't make me think
- Don't bore

# User Roles

- Transcriber
- Contributor
- Aggregator
- Quality Assurer
- Influencer
- Viewer

# Derzeitiger Stand

- Aufbau Softwarearchitektur
  - Wahl geeigneter Technologien, u.a.: IIIF-Viewer
  - Konzeption Datenmodell
- Erste UI-Überlegungen
  - Kampagnenoverview
  - Aufgaben
  - Tagging der Bilder



UNTERSTÜTZEN ▾

NOCH BIS 12. FEBRUAR 2018

## Bibliothekarische Archivierung

– erstellt vor 90 Tagen von  
Österreichische Nationalbibliothek

### AUFGABE 1 – TRANSKRIBIEREN

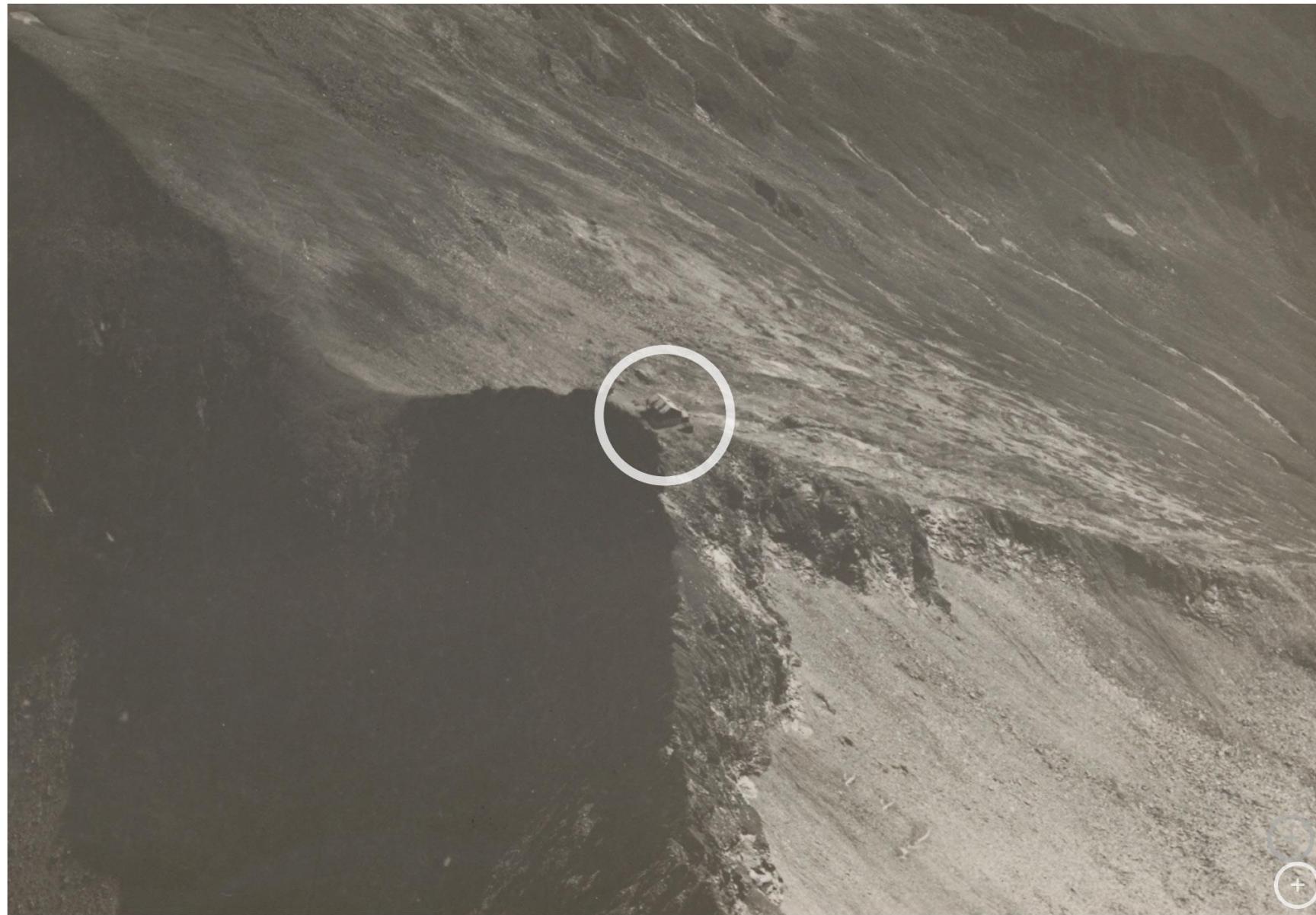
#### Transkribieren der Bildbeschreibung

Helfen Sie der österreichischen Nationalbibliothek, die Beschreibungstexte für Luftbilder zu transkribieren.

12. OKT –  
30. NOV



GEWÄHLTE AUFGABE  
5 Schlagworte vergeben ▾





GEWÄHLTE AUFGABE  
5 Schlagworte vergeben ▾



# Ausblicke

- **Ab 2018:**
  - DE/LL/CS: Interner Test
- **Herbst 2018: Launch**
  - Okopenko-Hybridedition und Blotius-Teiledition publiziert
  - Launch der ÖNB Labs Plattform mit ersten Beständen
  - 1. Kampagne „Luftbilder“ CS
- **2019: Laufender Betrieb**
  - DE: Modularisierung der Infrastruktur
  - LL/CS: Veröffentlichung weiterer Bestände, Events

# Anregungen? Fragen?

## Kontakt

*Library Labs:* [sophie.wagner@onb.ac.at](mailto:sophie.wagner@onb.ac.at)

*Linked Open Data:* [georg.petz@onb.ac.at](mailto:georg.petz@onb.ac.at)

*Digitale Editionen:* [christiane.fritze@onb.ac.at](mailto:christiane.fritze@onb.ac.at)

*Crowdsourcing:* [paul.sommersguter@onb.ac.at](mailto:paul.sommersguter@onb.ac.at)

# Vielen Dank

Christiane Fritze | Stefan Frühwirth | Stefan Karner |  
Georg Petz | Christoph Steindl | Paul Sommersguter |  
Sophie-Carolin Wagner | Can Yilmaz