

# WikiCite als kollaborativer Ansatz für offene Zitationsdaten

Philipp Zumstein (UB Mannheim)    Jakob Voß (VZG)



{ } wikicite



13. Juni 2018, Bibliothek\*tag Berlin

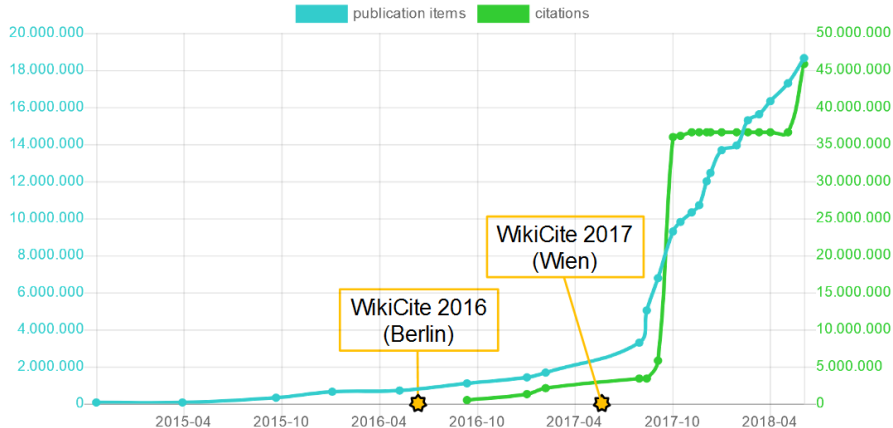
# Einführung & Motivation

# Was ist WikiCite?

- ▶ Freie bibliographische Datenbank in Wikidata
  - ▶ Bibliographische Metadaten
  - ▶ Zitationsdaten
- ▶ Katalogisierung selbstständiger + unselbstständiger Werke + Beziehungen
  - ▶ kollaborativ (Social Cataloging)
  - ▶ offene Daten (CC0)

# Umfang

- ▶ **Publikationen:** 19 Millionen Wikidata-Einträgen
- ▶ **Zitationen:** 46 Millionen Zitationen in Wikidata



# Motivation für Wikipedia

- Aussagen mit Belegen versehen, z.B.

Berlin gilt als [Weltstadt](#) der Kultur, Politik, Medien und Wissenschaften.<sup>[15][16][17][18]</sup> Die [Universitäten](#), [Forschungseinrichtungen](#), Sportereignisse und Museen Berlins genießen internationalen Ruf.<sup>[19]</sup> Die [Metropole](#) trägt den UNESCO-Titel *Stadt des Designs* und ist eines der meistbesuchten Zentren des Kontinents.<sup>[20]</sup> Berlins Architektur, Festivals, Nachtleben und vielfältige Lebensbedingungen sind weltweit bekannt.<sup>[21]</sup>

15. ↑ [Global Power City Index 2015](#) (PDF; 1,2 MB) Abgerufen am 10. April 2016.
16. ↑ [Revealed: Cities that rule the world – and those on the rise.](#) (CNN (englisch), abgerufen am 14. April 2010).
17. ↑ [Leading cities in cultural globalisations/Media](#) (englisch), GaWC, abgerufen am 14. April 2010.
18. ↑ [Berlin – Stadt der Nobelpreisträger.](#) In: [luisse-berlin.de](#). Abgerufen am 3. September 2015.
19. ↑ [London crowned as SportBusiness' 2012 Ultimate Sports City](#) (Memento vom 18. August 2012 im [Internet Archive](#)). In: [sportsbusiness.com](#), abgerufen am 26. November 2012.
20. ↑ [Tourismusreport 2014.](#) IHK Berlin.
21. ↑ [Berlin, the Big Canvas.](#) In: [nytimes.com](#). [The New York Times](#), abgerufen am 16. April 2010 (englisch).

- Bibl. Metadaten zentral für alle Artikel und Sprachversionen



# Motivation für bibliometrische Auswertungen

## ► Auswertung und Visualisierung von Publikationsdaten

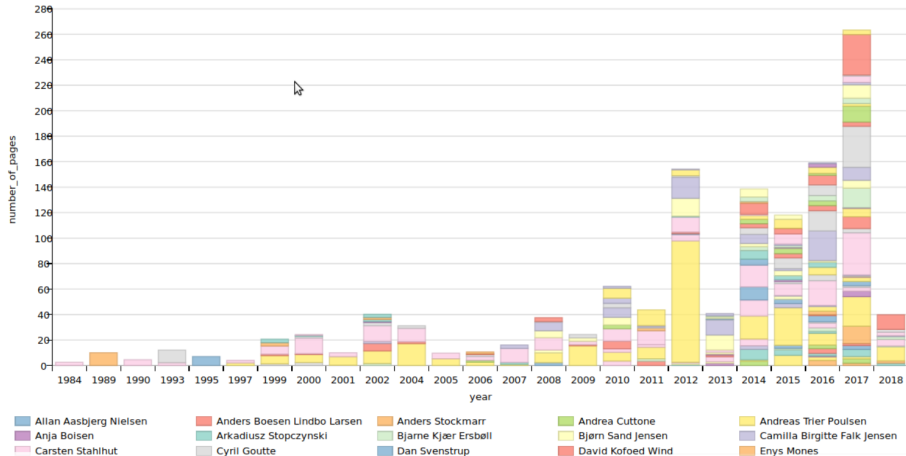
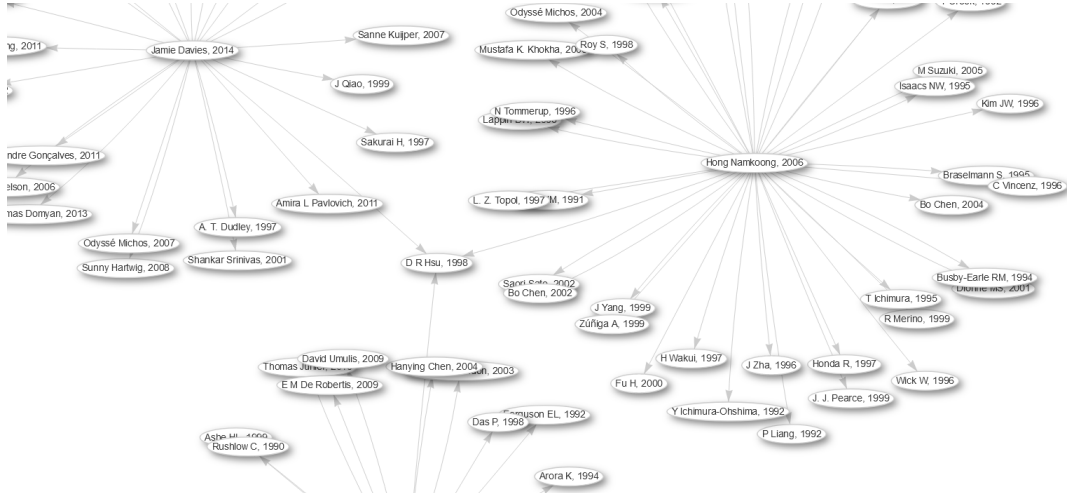


Abb. Screenshot von Scholia zur Anzahl produzierter Seiten von Publikationen mit Autoren der DTU

# Motivation für bibliometrische Auswertungen

## ► Auswertung und Visualisierung von Zitationsdaten





# Umsetzung von WikiCite

**Bezeichnung:** The Synergistic Activity of Thyroid Transcription Factor 1 and Pax 8 Relies

**ID:** (Q30000000)...

**Beschreibung:** wissenschaftlicher Artikel aus dem Jahr 2002

» In weiteren Sprachen

**mehrsprachig**

**Bearbeitungsmöglichkeiten**

**1. Aussage:**

**Aussagen**

ist ein(e)

wissenschaftlicher Artikel ...

» 0 Fundstellen

bearbeiten

+ Fundstelle hinzufügen

+ Wert hinzufügen

**2. Aussage:**

**Titel**

The Synergistic Activity of Thyroid Transcription Factor 1 and Pax 8 Relies on the Promoter/Enhancer Interplay (Englisch)

» eine Fundstelle

bearbeiten

+ Wert hinzufügen

**Fundstellen für Aussagen**

**3. Aussage:**

**Schlagwort**

Transkription ...

» eine Fundstelle

bearbeiten

+ Wert hinzufügen

...

# Wikidata Datenmodell

- ▶ IDs
  - ▶ [www.wikidata.org/entity/Q30000000](http://www.wikidata.org/entity/Q30000000)
- ▶ Eigenschaften
  - ▶ P31 *ist ein(e) / instance of / est ...*
  - ▶ P50 *Autor / author / scriptor ...*
- ▶ Datenmodell ist dynamisch
  - ▶ Neue Eigenschaften vorschlagen, diskutieren und erstellen

# Tools zur Bearbeitung

- ▶ **sourceMD:** Import Metadaten anhand DOI
- ▶ **zotkat:** Import aus Zotero (Literaturverwaltung)
- ▶ **Mix'n'Match:** Abgleich von Personendaten mit externen Identifikatoren
- ▶ **OpenRefine 3.0** mit erweiterter Wikidata-Anbindung
- ▶ ...

QuickStatements, wikidata-cli ...

# Qualitätssicherung

- ▶ Eigene Beobachtungslist für Wikidata-Einträge mit Benachrichtigung bei Änderungen
- ▶ Einschränkungen von Werten von Eigenschaften
  - ▶ ISBN (P212/P957) folgen bestimmten Muster
  - ▶ Autoren (P50) müssen Personen oder Organisation sein
- ▶ Regelverletzungen werden regelmässig protokolliert (constraint reports) und können dann verbessert werden

# Normdaten-Verlinkung

- ▶ Wikidata ist selbst eine universelle Normdatei
- ▶ Verknüpfungen mit externen Einträgen
  - ▶ DOI, ISBN, PMID, arXiv-ID, ... (Publikationen)
  - ▶ GND, VIAF, ORCID, dblp-ID, ... (Autoren, Organisationen...)
  - ▶ Geonames, ... (Orte...)

X von Y Wikidata-Eigenschaften dienen der Verlinkung mit anderen Datenbanken

# Abfragemöglichkeiten

- ▶ <https://query.wikidata.org/> (SPARQL-Editor)
- ▶ Beispielabfragen
  - ▶ Bücher zusammen von Elter(n) und Kinder(n) geschrieben:  
<http://tinyurl.com/y7oe8uo8>
  - ▶ Autoren, deren Werke 2018 gemeinfrei wurden:  
<http://tinyurl.com/yc8v6l9c>
  - ▶ Meist zitierte Wissenschaftlerinnen:  
<http://tinyurl.com/y8ell8nw>>

# Perspektiven für Bibliotheken




# Nachnutzung von Daten aus Wikidata

- ▶ Kataloganreicherungen bei Normdaten
- ▶ mehrsprachige Suche nach Schlagwörtern
- ▶ Artikelindex für Discovery aus Wikidata?
- ▶ Kataloganreicherungen mit offenen Zitationsdaten?

# Katalogisierungstools

- ▶ Kann man Ansätze zum Datenerfassen in Wikidata auch bei der Katalogisierung in Bibliotheken adaptieren?
- ▶ Sind nicht auch stärker Tools für Bearbeitungen im Batch gefragt?
- ▶ Was ist mit Kontroll- und Prüftools?

# Umfang der Katalogisierung

- ▶ Aufsatzkatalogisierung
  - ▶ Thema für Bibliotheken? FIDs?
  - ▶ WikiCite zeigt auf, dass es möglich ist
- ▶ Katalogisierung von Zitationen
  - ▶ Können Bibliotheken zu offenen Zitationsdaten beitragen?
  - ▶ Zitationsdaten in Verbundkatalog speichern?
  - ▶ Linked Open Citation Database (LOC-DB): Clip 

**Fragen / Diskussion /  
Antworten ?**