



# Libraries as Curators of Open Citations

Perspectives of the Project LOC-DB in Germany

Anne Lauscher, Kai Eckert, Lukas Galke, Ansgar Scherp, Syed Tahseen Raza Rizvi,  
Sheraz Ahmed, Andreas Dengel, Philipp Zumstein, Annette Klein  
Workshop on Open Citations, 2018, Bologna



# The origins: What were libraries cataloging?



# The origins: What were libraries cataloging?



**Things you can put in a shelf.**

# Later: Resource Discovery Systems



**Now: Citations?**



# Agenda



1. Linked Open Citation Database
2. Reference Linking Workflow for Libraries
3. Infrastructure for Cataloging Citations
4. Conclusion



# Linked Open Citation Database

How much would it cost, with respect to resources, if libraries catalogued everything and curated the citation graph?

- Development of processes and tools based on linked data technologies to enable libraries to contribute to an open and interconnected citation graph
- Quantitative and qualitative evaluation, e.g., cost benefit analysis



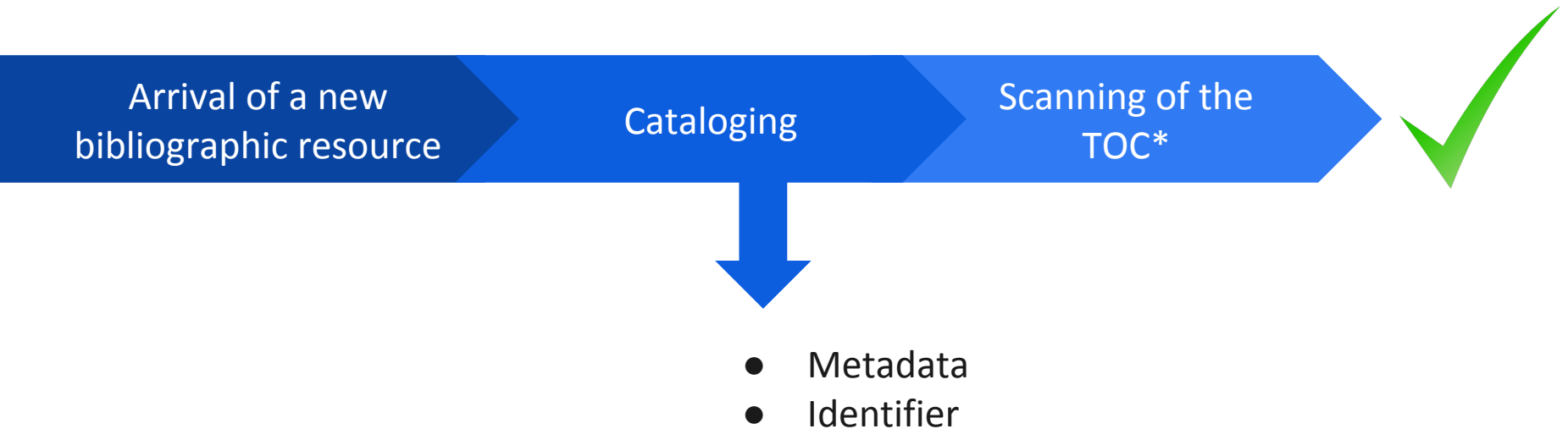
# Library Workflow

- Integrated into the standard library workflow
- Reuse of data from existing resources, e.g. from publishers, other projects, and standard library catalogs (high quality metadata)
- Automated as far as possible
  - Automatic reference extraction
  - Easy-to-use editorial system
- Distributed database and collaborative cataloging processes

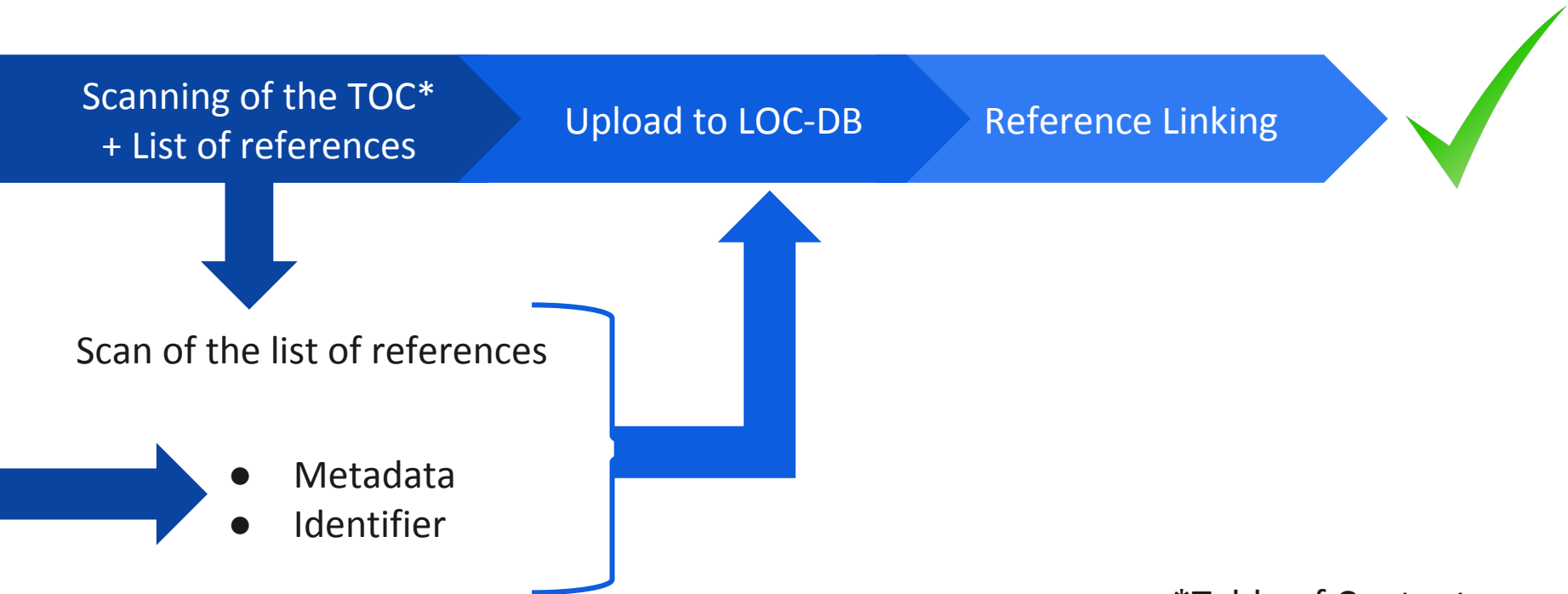




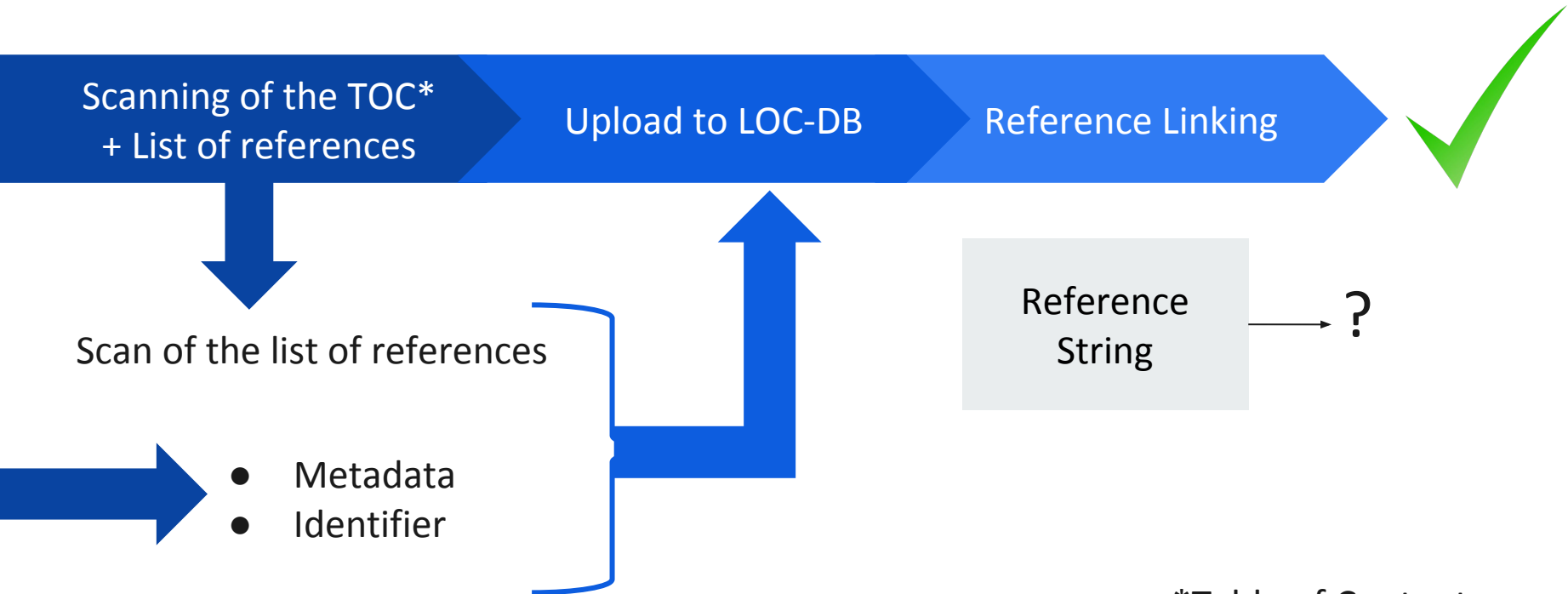
# Standard Library Workflow (UB Mannheim)



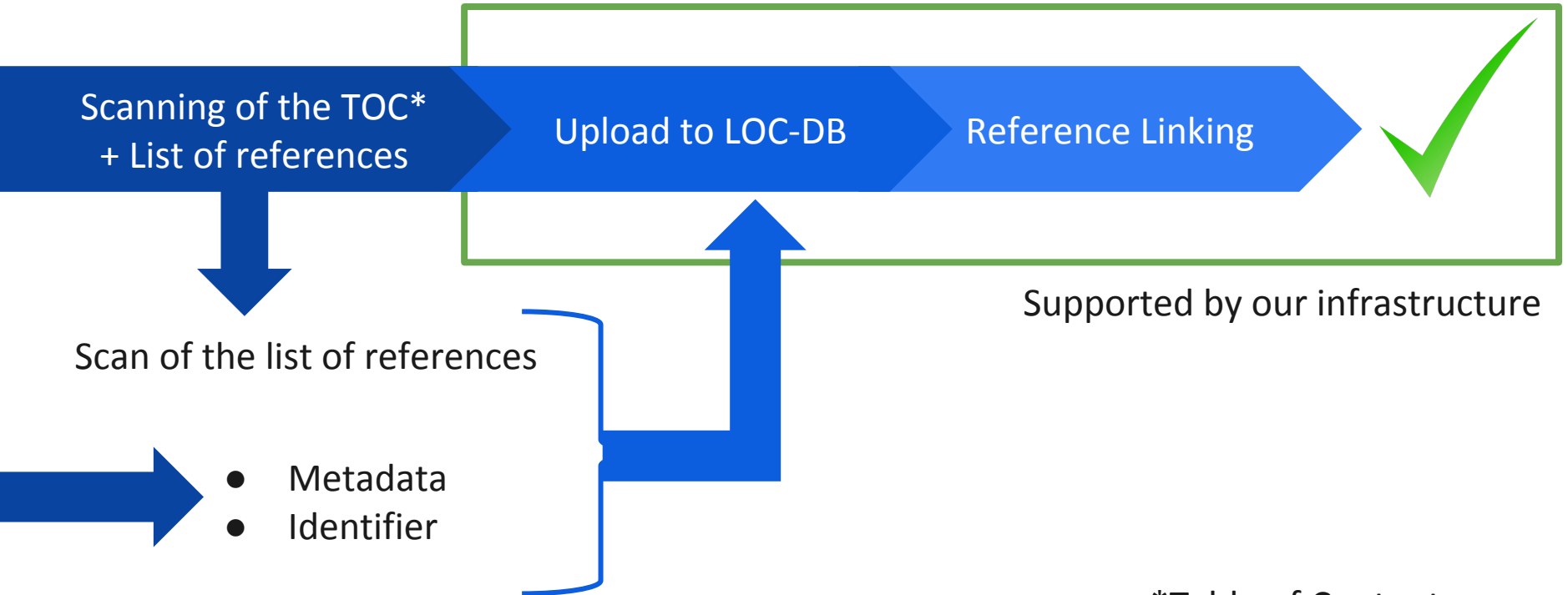
# Reference Linking

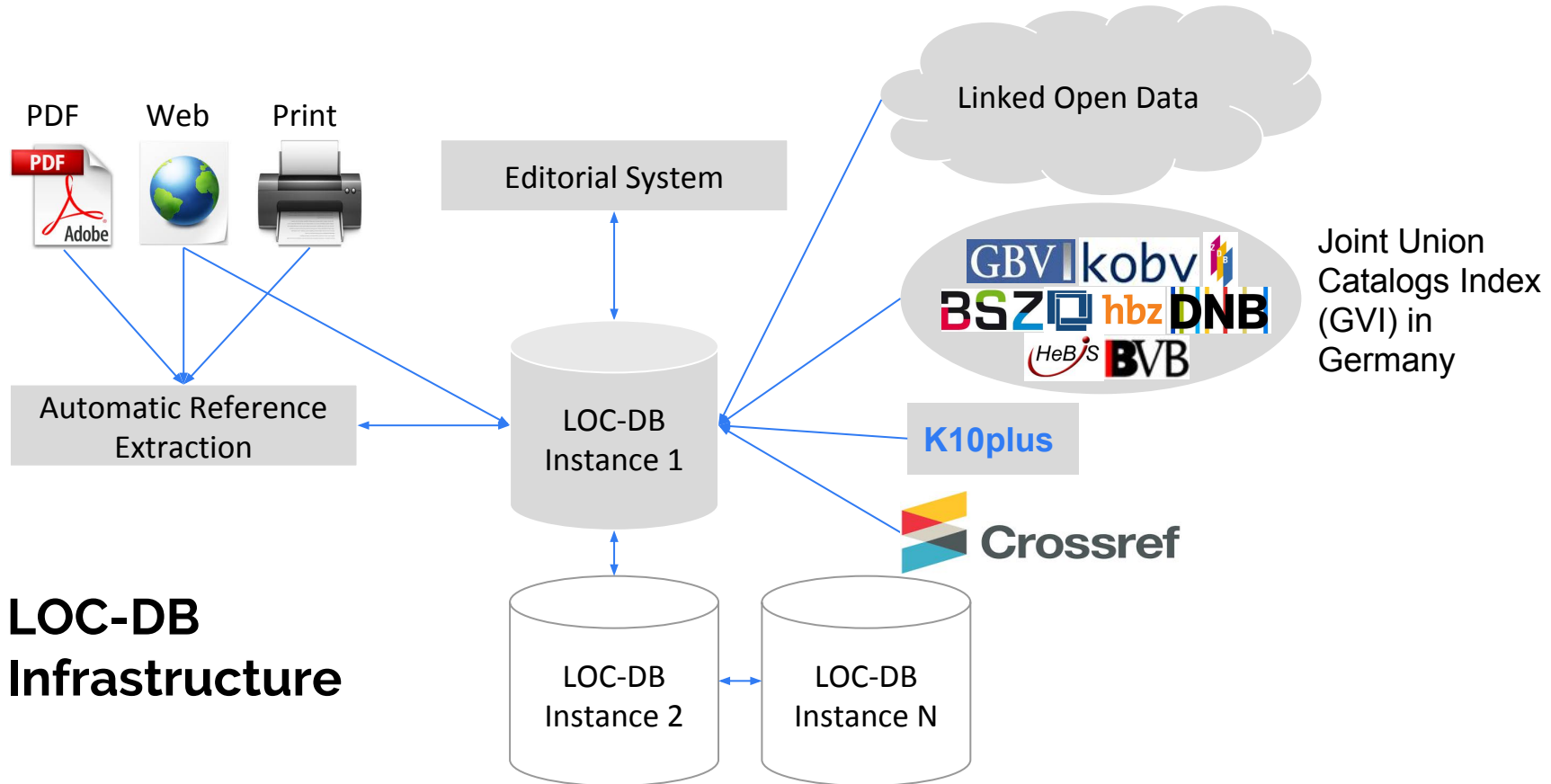


# Reference Linking

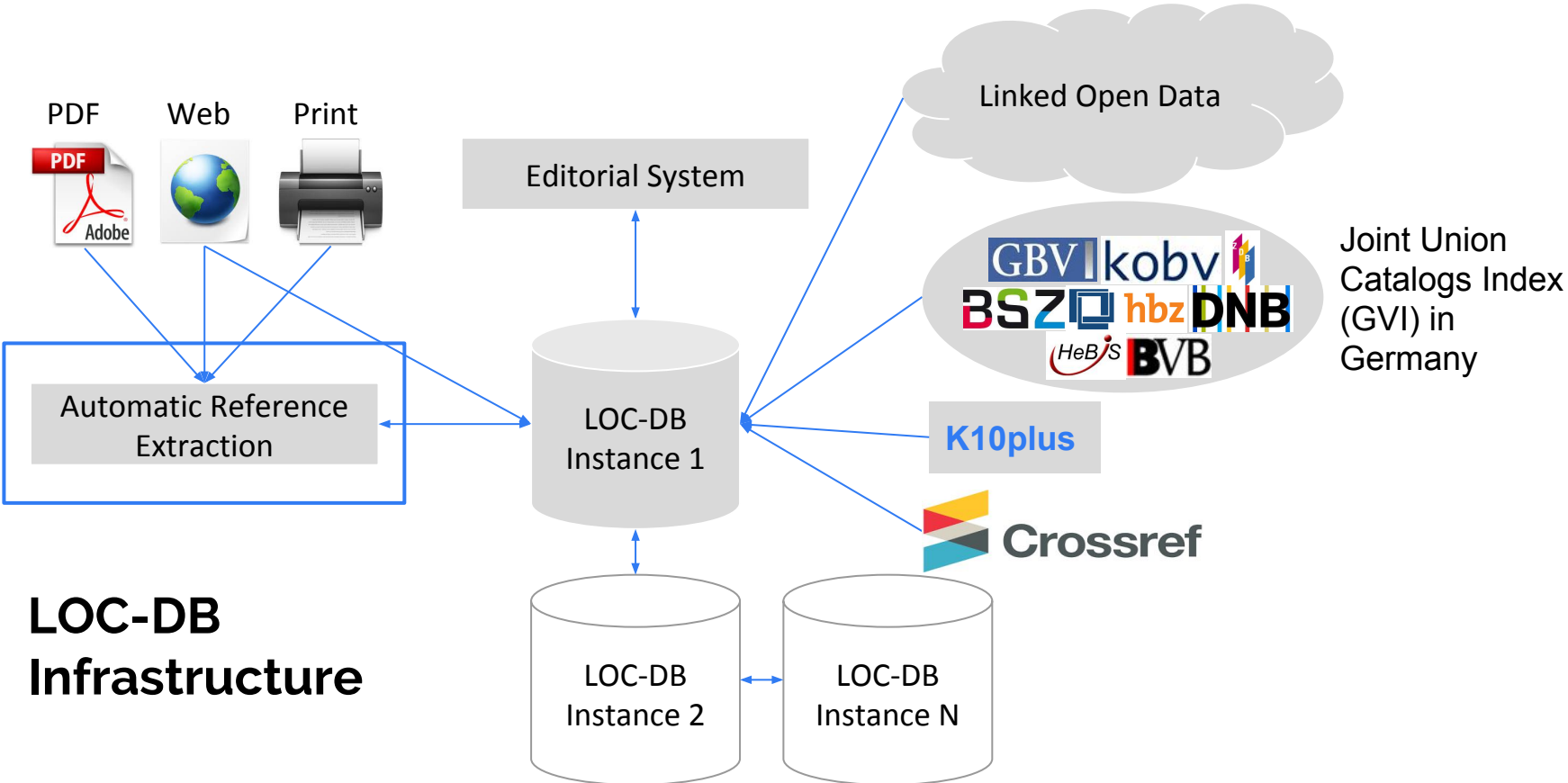


# Reference Linking





# LOC-DB Infrastructure



# LOC-DB Infrastructure

Ackermann, Josef, »Null Toleranz für Grauzonen. Josef Ackermann, Vorstandsvorsitzender der Deutschen Bank, über seine Moral, Luxusprobleme und Belcanto im Badezimmer«, in: *Zeitmagazin* 22/2007, S. 46–51.

Adorno, Theodor W., *Ästhetische Theorie*, Frankfurt/M. 1973.

Alemann, Ulrich von/Eckert, Florian, »Lobbyismus als Schattenpolitik«, in: *Aus Politik und Zeitgeschichte* (APuZ) 15–16 (2006), S. 3–10.

Anheier, Helmut K./Appel, Anja, »Stiftungen in der Bürgergesellschaft: Möglichkeiten und Grenzen«, in: *Aus Politik und Zeitgeschichte* (APuZ) 14 (2004), S. 8–15.

Arbeitsstab Forum Bildung in der Geschäftsstelle der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung, *Förderung von Chancengleichheit. Vorläufige Empfehlungen und Expertenbericht*, Bonn 2000.

Aristoteles, *Politik*, übers. u. hrsg. v. Olof Gigon, Zürich/Stuttgart 1971.

Aristoteles, *Nikomachische Ethik*, übers. u. hrsg. v. Ursula Wolf, Reinbek b. Hamburg 2006.

Aristoteles, *Rhetorik*, übers. u. hrsg. von Gernot Krapinger, bibliogr. erg. Ausg., Stuttgart 2007.

Aron, Raymond, *Hauptströmungen des modernen soziologischen Denkens*, Reinbek b. Hamburg 1979.

Aron, Raymond, »Zur politischen Theorie« (1962), in: ders.: *Zwischen Macht und Ideologie. Politische Kräfte der Gegenwart*, Wien 1974.

Aron, Raymond, »Regierende und Regierte«, in: ders.: *Fortschritt ohne Ende? Gütersloh*, 1970, S. 50–70.

Bachrach, Peter, *Die Theorie demokratischer Elitenherrschaft. Eine kritische Analyse*, Frankfurt/M. 1970.

Backhaus-Maul, Holger/Bidermann, Christine/Nährlich, Stefan/Polterauer, Judith (Hg.), *Corporate Citizenship in Deutschland. Gesellschaftliches Engagement von Unternehmen. Bilanz und Perspektiven*, 2. akt. u. erw. Aufl. Wiesbaden 2010.

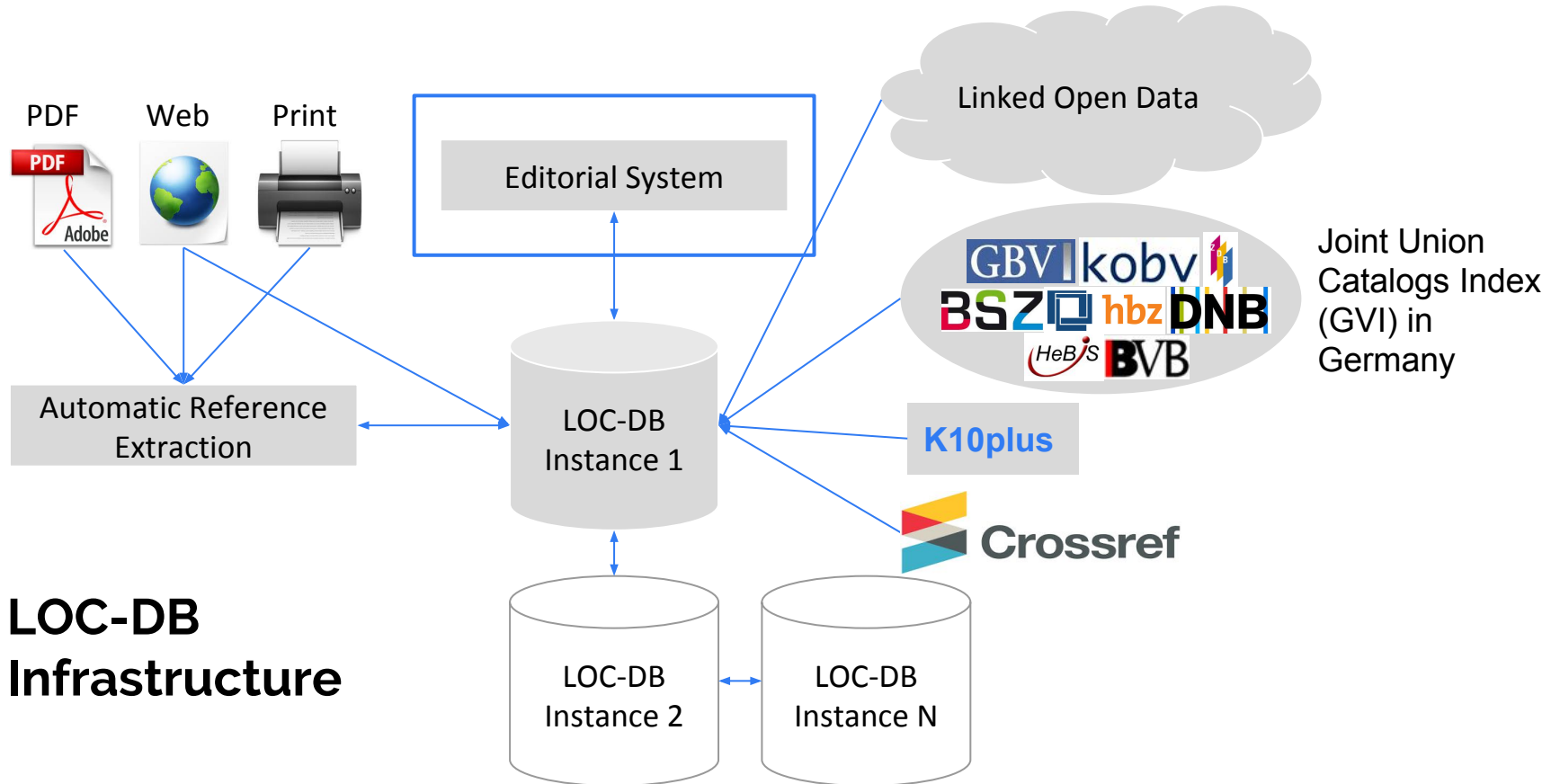
Baum, Manfred: »Probleme der Begründung Kantischer Tugendpflichten«, in: *Jahrbuch für Recht und Ethik* VI (1998), S. 41–56.

Benda, Julien, *Der Verrat der Intellektuellen* (1927), Frankfurt/M. 1988.

Berchem, Theodor, *Elitebildung in der Massenuniversität? Bilanz und Perspektiven der Hochschulreform*, Köln 1985.

# Automatic Reference Extraction

Combination of text-driven and layout-driven extraction using deep learning techniques (Bhardwaj et al., 2017)



# LOC-DB Infrastructure



## Ingest

Add Batch of Files

Add Electronic Resource

DOI: 10.1145/3197026.3197050 **Electronic** x

Resource Type

JOURNAL\_ARTICLE

DOI ▾

10.1145/3197026.3197050 |

Clear

Upload

## Agenda

Waiting    In Queue    Ready    Electronic Journals

**PROCEEDINGS** **Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries - JCDL '18**, ACM Press

[Edit](#)

**PROCEEDINGS ARTICLE** **2018** Lauscher, Anne; Eckert, Kai; Galke, Lukas; Scherp, Ansgar; Rizvi, Syed Tahseen Raza; Ahmed, Sheraz; Dengel, Andreas; Zumstein, Philipp; Klein, Annette: **Linked Open Citation Database Enabling Libraries to Contribute to an Open and Interconnected Citation Graph**, ACM Press. [CrossRef](#) [DOI](#)

[References](#)

[Edit](#)

## Reference Inspector

**PROCEEDINGS\_ARTICLE** **2018** Lauscher, Anne; Eckert, Kai; Galke, Lukas; Scherp, Ansgar; Rizvi, Syed Tahseen Raza; Ahmed, Sheraz; Dengel, Andreas; Zumstein, Philipp; Klein, Annette: **Linked Open Citation Database Enabling Libraries to Contribute to an Open and Interconnected Citation Graph** in: **PROCEEDINGS** Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries - JCDL '18, ACM Press. [CrossRef](#) [DOI](#)

Akansha Bhardwaj, Dominik Mercier, Andreas Dengel, and Sheraz Ahmed. 2017. DeepBIBX: Deep Learning for Image Based Bibliographic Data Extraction. Springer International Publishing, Cham, 286--293. [https://doi.org/10.1007/978--3--319--70096-0\\_30](https://doi.org/10.1007/978--3--319--70096-0_30)

[Edit](#)

DOI: 10.3789/isqv27no1.2015.04

[Edit](#)

Thomas M. Breuel. 2008. The OCRopus open source OCR system. In Document Recognition and Retrieval XV, part of the IS&#38;T-SPIE Electronic Imaging Symposium, San Jose, CA, USA, January 29--31, 2008. Proceedings (SPIE Proceedings), Berrin A. Yanikoglu and Kathrin Berkner (Eds.), Vol. 6815. SPIE, 68150F. <https://doi.org/10.1117/12.783598>

[Edit](#)

DOI: 10.1109/ICDAR.2013.140

[Edit](#)

Isaac Councilil, C. Lee Giles, and Min-Yen Kan. 2008. ParsCit: an Open-source CRF Reference String Parsing Package. In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC-08). European Language Resources Association (ELRA),

[Edit](#)


 Source:  Type:  Contained:

### New resource from reference metadata

[Create Resource](#)[Commit Link](#)

## Reference Inspector

**PROCEEDINGS\_ARTICLE** **2018** Lauscher, Anne; Eckert, Kai; Galke, Lukas; Scherp, Ansgar; Rizvi, Syed Tahseen Raza; Ahmed, Sheraz; Dengel, Andreas; Zumstein, Philipp; Klein, Annette: **Linked Open Citation Database Enabling Libraries to Contribute to an Open and Interconnected Citation Graph** in: **PROCEEDINGS** Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries - JCDL '18, ACM Press. [CrossRef](#) [DOI](#)

Akansha Bhardwaj, Dominik Mercier, Andreas Dengel, and Sheraz Ahmed. 2017. DeepBIBX: Deep Learning for Image Based Bibliographic Data Extraction. Springer International Publishing, Cham, 286--293. [https://doi.org/10.1007/978--3--319--70096-0\\_30](https://doi.org/10.1007/978--3--319--70096-0_30)

[Edit](#)

DOI: 10.3789/isqv27no1.2015.04

[Edit](#)

Thomas M. Breuel. 2008. The OCRopus open source OCR system. In Document Recognition and Retrieval XV, part of the IS&#38;T-SPIE Electronic Imaging Symposium, San Jose, CA, USA, January 29--31, 2008. Proceedings (SPIE Proceedings), Berrin A. Yanikoglu and Kathrin Berkner (Eds.), Vol. 6815. SPIE, 68150F. <https://doi.org/10.1117/12.783598>

[Edit](#)

DOI: 10.1109/ICDAR.2013.140

[Edit](#)

Isaac Councilil, C. Lee Giles, and Min-Yen Kan. 2008. ParsCit: an Open-source CRF Reference String Parsing Package. In Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC-08). European Language Resources Association (ELRA),

[Edit](#)Source:  Type:  Contained:

## Suggestions

### Suggestions from external sources

**JOURNAL\_ARTICLE** **2015** Breeding, Marshall: **Future of Library Discovery Systems** in: **JOURNAL\_ISSUE** Information Standards Quarterly, volume 27, issue 1, National Information Standards Organization (NISO). [CrossRef](#)

[Edit](#)[DOI](#) [CROSSREF](#)

### New resource from reference metadata

[Create Resource](#)[Commit Link](#)

## Scan Inspector

**BOOK CHAPTER** in: **EDITED BOOK** **2017** Görtler, Michael; Lotz, Mathias; Partetzke, Marc; Poma Poma, Sara; Winckler, Marie, Kritische politische Bildung: Standpunkte und Perspektiven /, Wochenschau, pp. 81-83.

[Scan view](#)[Reference list](#)[Previous](#)[1](#)[2](#)[3](#)[Next](#)

Weißeno, Georg (1989): Lernertypen und Lernerdidaktiken im Politikunterricht. Ergebnisse einer fachdidaktisch motivierten Unterrichtsforschung. Frankfurt/M., Haag und Herchen.

Weißeno, Georg/Deitjen, Joachim/Juchler, Ingo/Massing, Peter/Richter, Dagmar (Hrsg.) (2010): Konzepte der Politik. Ein Kompetenzmodell. Schwalbach/Ts.: Wochenschau.

Widmaier, Benedikt/Overwien, Bernd (Hrsg.): Was heißt heute Kritische Politische Bildung? Schwalbach/Ts.: Wochenschau.

Winckler, Marie (2014): Von Geschlechterdifferenzen zu Differenzierungsprozessen. In: Zeitschrift für Didaktik der Gesellschaftswissenschaften 1/2014, S. 116-127.

Zeuner, Christine (2012): Intentionen in der politischen Bildung – Notwendigkeit einer Diskursverschiebung. [http://www.boecler.de/pdf/v\\_2012\\_03\\_01\\_zeuner.pdf](http://www.boecler.de/pdf/v_2012_03_01_zeuner.pdf) [28.4.2016].

Lernertypen und Lernerdidaktiken im Politikunterricht. Ergebnisse einer fachdidaktisch motivierten Unterrichtsforschung. Frankfurt/M 1989

Source:  Type:  Contained:

## Suggestions

**EDITED BOOK** **2009** Borgetto, Bernhard: **Stress- und Schmerzursachen verstehen : Gesundheitspsychologie und -soziologie in Prävention und Rehabilitation** / , Thieme,.

[Edit](#)

## Suggestions from external sources

**MONOGRAPH** **1989** Weißeno, Georg: **Lernertypen und Lernerdidaktiken im Politikunterricht : Ergebnisse einer fachdidaktisch motivierten Unterrichtsforschung** / , Haag u. Herchen, **SWB**

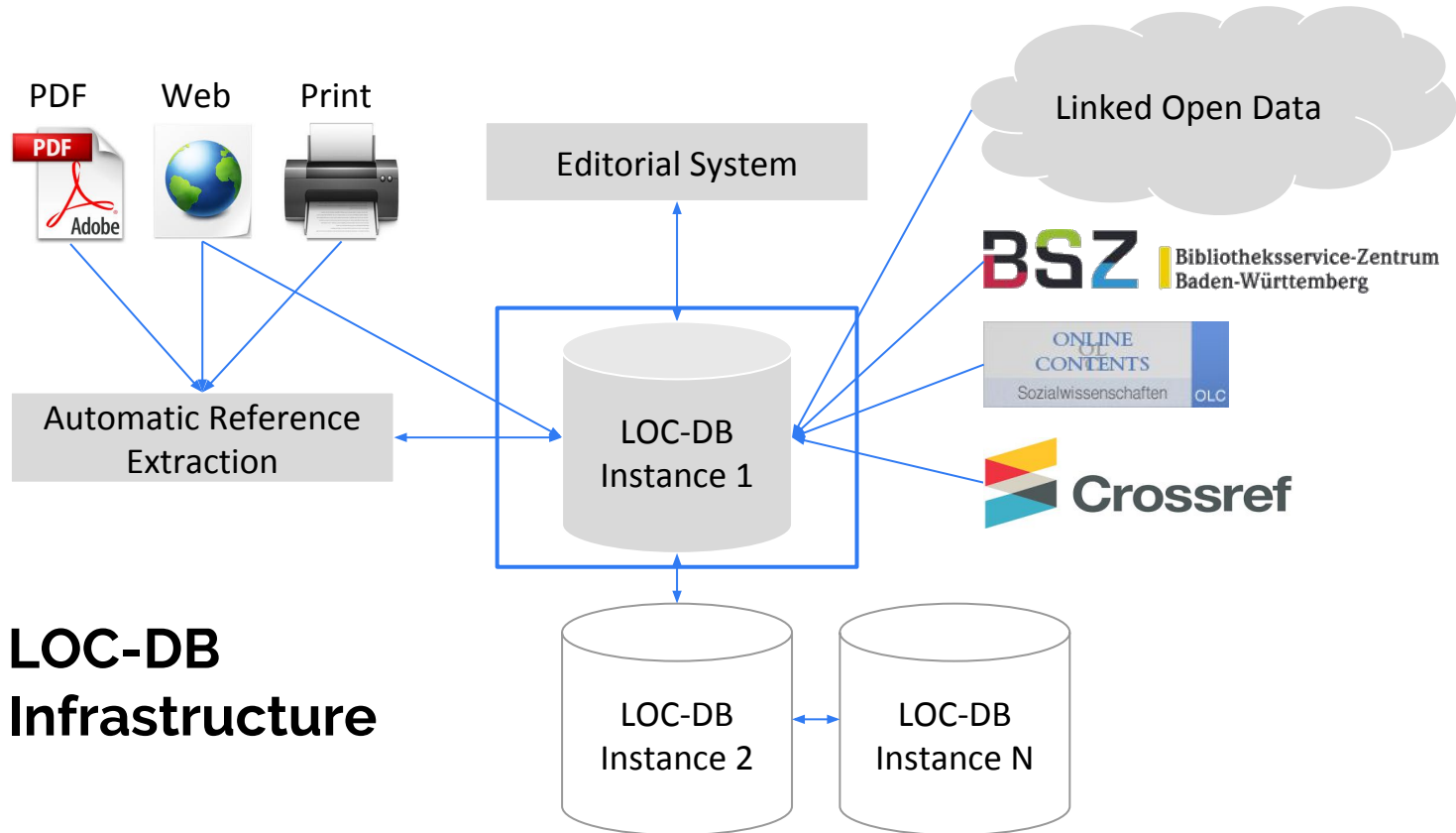
[Edit](#)

**MONOGRAPH** **1989** Weißeno, Georg: **Lernertypen und Lernerdidaktiken im Politikunterricht Ergebnisse einer fachdidaktisch motivierten Unterrichtsforschung** , Haag u. Herchen, **GVI**

[Edit](#)

**MONOGRAPH** **1989** Weißeno, Georg: **Lernertypen und Lernerdidaktiken im Politikunterricht Ergebnisse einer fachdidaktisch motivierten**

[Edit](#)



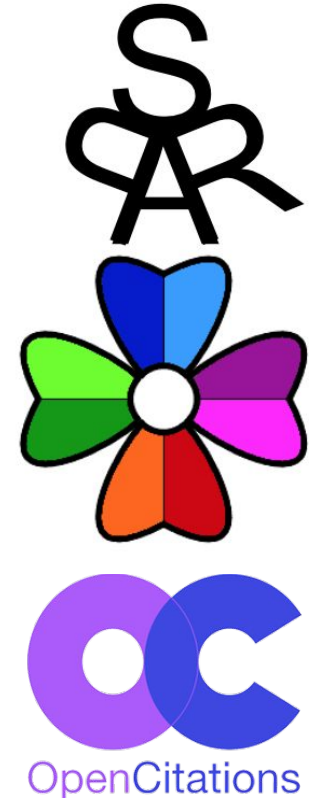
# LOC-DB Infrastructure



# Data Model and Publishing

Ensuring optimal reusability and interoperability of the produced data

- Adaption of the OpenCitations metadata model (Peroni and Shotton, 2016)
- Publishing of the Data in RDF format by using the Semantic Publishing and Referencing (SPAR) Ontologies (Peroni, 2014)



**How much would it cost if libraries catalogued everything and curated the citation graph?**

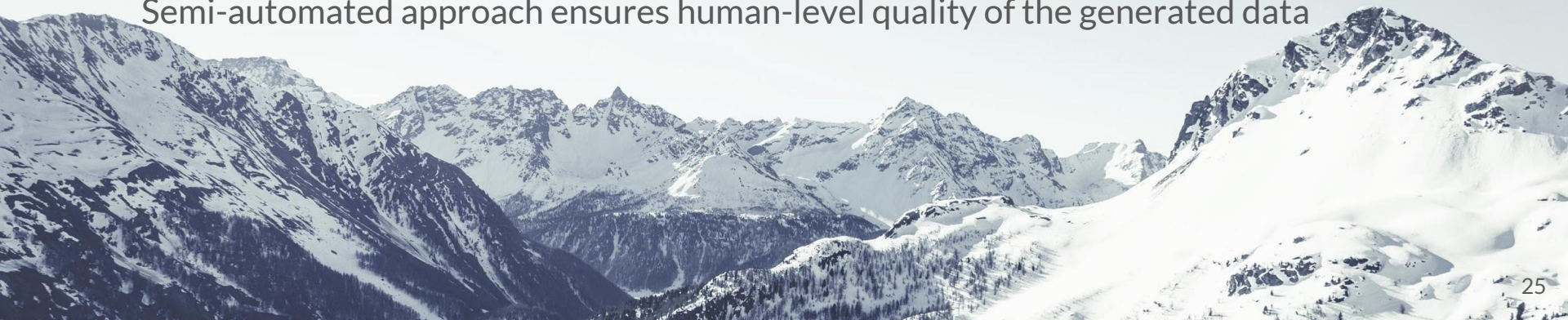




# How much would it cost if libraries catalogued everything and curated the citation graph?

Preliminary results suggest general feasibility of the approach  
Continuous improvement on the infrastructure and processes ongoing

Semi-automated approach ensures human-level quality of the generated data



**Citations are ahead!**



# Citations are ahead!

For more information please visit

<https://locdb.bib.uni-mannheim.de>

<https://github.com/locdb/>



6th November, 2018



# Bibliography

- Marshall Breeding. 2015. Future of Library Discovery Systems. *Information Standards Quarterly* 27, 1 (2015), 24. <https://doi.org/10.3789/isqv27no1.2015.04>
- Christian Wilke and Regina Retter. 2017. Zitationsdaten extrahieren: halbautomatisch, offen, vernetzt. Ein Workshopbericht. *Informationspraxis* 3, 2 (Dec. 2017). <https://doi.org/10.11588/ip.2017.2.43235>
- Silvio Peroni and David Shotton. 2016. Metadata for the OpenCitations Corpus. Technical Report. <https://dx.doi.org/10.6084/m9.figshare.3443876>
- Akansha Bhardwaj, Dominik Mercier, Andreas Dengel, and Sheraz Ahmed. 2017. DeepBIBX: Deep Learning for Image Based Bibliographic Data Extraction. Springer International Publishing, Cham, 286–293. [https://doi.org/10.1007/978-3-319-70096-0\\_30](https://doi.org/10.1007/978-3-319-70096-0_30)



# Bibliography

- Silvio Peroni. 2014. The Semantic Publishing and Referencing Ontologies. In Semantic Web Technologies and Legal Scholarly Publishing. Springer, Cham, 121–193. [https://doi.org/10.1007/978-3-319-04777-5\\_5](https://doi.org/10.1007/978-3-319-04777-5_5)



# Appendix



# Data

- Social sciences collection of Mannheim University Library
- 522 print books and collections acquired by in 2011:  
~271,000 references
- Articles published in 2011 in 101 (mostly electronic) journals:  
~298,000 references
- New print acquisitions of the social sciences branch library  
from July 2017 on

## Agenda

NOT\_OCR\_PROCESSED  OCR\_PROCESSING  OCR\_PROCESSED  EXTERNAL

[Refresh](#)

**EDITED\_BOOK** **2017** **Kritische politische Bildung: Standpunkte und Perspektiven /**

, Wochenschau, *In*: Görtler, Michael; Lotz, Mathias; Partetzke, Marc; Poma Poma, Sara; Winckler, Marie (ed.), Kritische politische Bildung: Standpunkte und Perspektiven /, Wochenschau, .

[Edit](#)[References](#)

**BOOK\_CHAPTER** *In*: Görtler, Michael; Lotz, Mathias; Partetzke, Marc; Poma Poma, Sara;

Winckler, Marie (ed.), Kritische politische Bildung: Standpunkte und Perspektiven /, Wochenschau, .

[Edit](#)[References](#)

14 - 16 PRINT

- [5b13ffbd48b24e6c788da1f3.pdf](#)



## Edit Resource

**JOURNAL\_ARTICLE** Platt, Jennifer **1974** **The Phenomenon of Sociology: A Reader in the Sociology of Sociology** , SAGE Publications .

Type

JOURNAL\_ARTICLE ▼

Title

The Phenomenon of Sociology: A Reader in the Sociology of Sociology

Subtitle

Contributors

×

AUTHOR ▼

Platt, Jennifer

×

PUBLISHER ▼

SAGE Publications

Add Contributor

Die Zukunft Europas :

[Back](#)

[Reference list](#)

[Scan view](#)

## Literaturverzeichnis

Albers, Detlev / Haseler, Stephen / Meyer, Henning: Social Europe - A Continent's Answer to Market Fundamentalism; European Research Forum at London Metropolitan University, London 2006.

Baum-Ceising, Alexandra / Faber, Anne (Hrsg.): Soziales Europa? Perspektiven des Wohlfahrtsstaates im Kontext von Europäisierung und Globalisierung: Festschrift für Klaus Busch; Vs Verlag, Wiesbaden 2005.

Beck, Ulrich / Grande, Edgar: Das kosmopolitische Europa; Suhrkamp Verlag, Frankfurt am Main 2004.

Bundeszentrale für politische Bildung: Vertrag von Lissabon, Bonn 2008.

Collignon, Stefan: Bundesrepublik Europa? Die demokratische Herausforderung und Europas Krise: Vorwärts Buch, Berlin 2007.

Delors, Jacques: Erinnerungen eines Europäers; Parthas Verlag, Berlin 2004.

Dullien, Sebastian / Herr, Hansjörg / Kellermann, Christian: Der gute Kapitalismus; Transcript Verlag, Bielefeld 2009.

Eichenhofer, Eberhard: Geschichte des Sozialstaats in Europa: Von der „sozialen Frage“ bis zur Globalisierung; C.H. Beck Verlag, München 2007.

Europäische Kommission: Der Vertrag von Lissabon: Einführung, Vertrag über die Europäische Union, Vertrag über die Arbeitsweise der Europäischen Union, Charta der Grundrechte der Europäischen Union, Protokolle, Erklärungen, Deutsches Recht; Autoren der Einführung: Brok, Elmar/Leinen, Jo; Bundeszentrale für politische Bildung, Bonn 2008.

Europäische Union: Amtsblatt der Europäischen Union C 306, 17.12.2007: Vertrag von Lissabon; Anne Lauscher, Librarian at Curators of Open Citations

Bundesrepublik Europa? Die demokratische Herausforderung und Europas Krise; Vorwärts Buch, Stefaan CCollignon

Confidence threshold for internal ... and for external Suggestions

[Search](#)

[Ask Others](#)

## Suggestions

### External suggestions

Stefan Collignon: Bundesrepublik Europa? Die demokratische Herausforderung und Europas Krise. [CrossRef](#)

### New resource from reference metadata

Bundesrepublik Europa? Die demokratische Herausforderung und Europas Krise; Vorwärts Buch, [Local](#)

[Edit](#)

[Scan view](#) [Reference list](#)

[BOOK CHAPTER](#)

Theodor W Adorno. 1967. Resume über Kulturindustrie. In: Adorno,Theodor W. Ohne Leitbild. Parva Aesthetica. FrankfurtIM.: Suhrkamp,. In: S. [Edit](#)

Theodor W Adorno. 1971. Nas bedeutet: Aufarbeitung der Vergangenheit. In: Adorno,Theodor W.: Erziehung zur Mündigkeit. Vorträge und Gespräche mit Hellmut Becker 1959-1969. FrankfurtIM.: Suhrkamp,. In: S. [Edit](#)

Ien Ang. 1999. Culture and Communications,'Towards an EthnographicCritique of MediaConsumption in theTransnational Media System. In: [Edit](#)

Ien Ang. 2005. Who needs cultural earcher? In: Leistyna, Pepi (Hrsg.): Cultural Studies, From 'Theory to Action. Osford: [Edit](#)

Carsten Büniger. 2010. Poliische Bildung nach dem ‚Tod desSubjekts‘.. In: In: Losch, Bettina,Thimmel, Andreas(Hrsg.) Kritischepolitische Bildung. Ein Handbuch. Schwalbachl'Te.: Wochenschau, S. [Edit](#)

Halil Can. 2011. Demokratiarbeit und Empowerment gegen Diskriminierung und Rassismus in selbstbestimmten People of Color-Räumen. In: Castro Varela, Maria do MarlDhawan, Nikita (Hrsg.): Soziale (Un)Gerechtigkeit. Kritische Perspektiven auf Diversity, Intersektionalität und Antidiskriminierung. [Edit](#)

Resume über Kulturindustrie. In: Adorno,Theodor W. Ohne Leitbild. Parva Aesthetica. FrankfurtIM.: Suhrkamp, 1967

Confidence threshold for internal  ... and for external Suggestions

[Ask Others](#)

## Suggestions

## External Suggestions

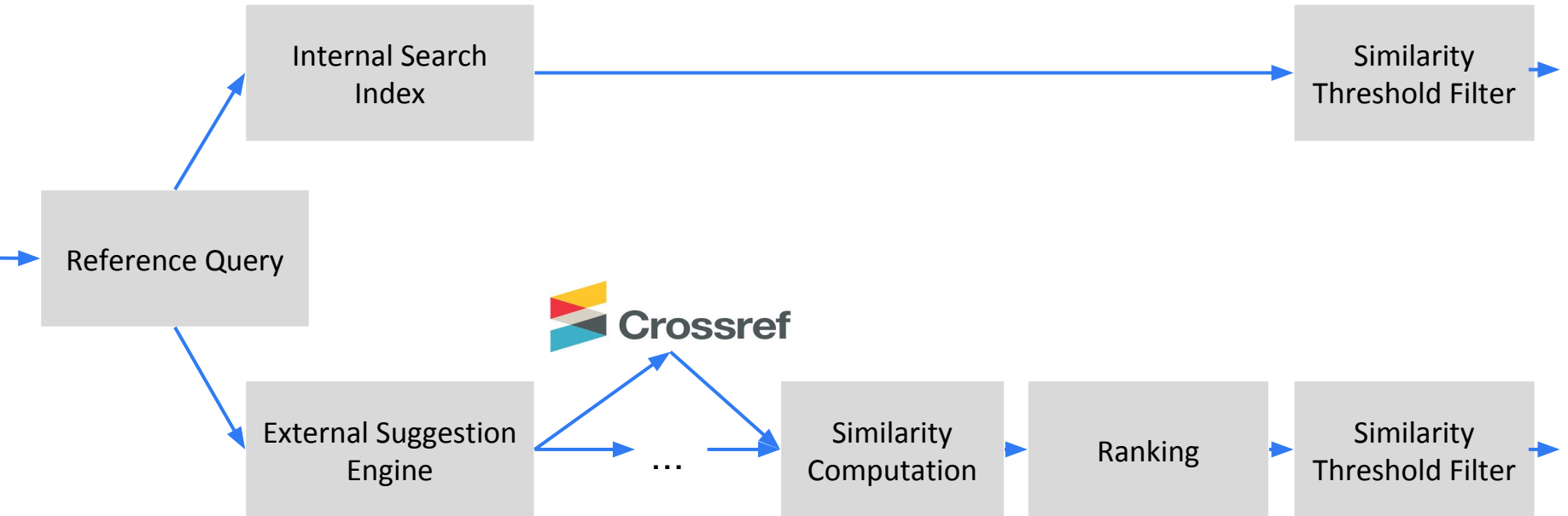
[BOOK](#) Adorno, Theodor W. [1970](#) **OHNE LEITBILD. PARVA AESTHETICA. VON THEODOR W. ADORNO. 4. AUFL. FRANKFURT A.M.: SUHRKAMP 1970. 193 S.** , [GVI](#) [Edit](#)

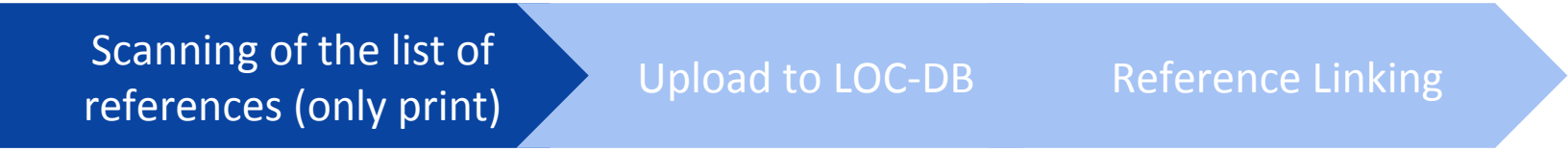

## New resource from reference metadata

[Create Resource](#)

# Reference Target Suggestions

## Speeding up the linking process



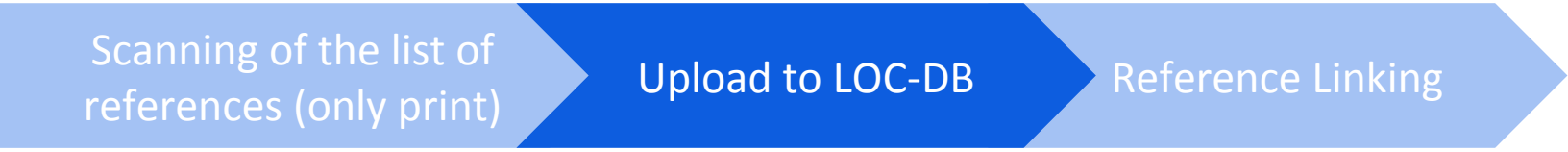



Scanning of the list of references (only print)

Upload to LOC-DB

Reference Linking

- > 100 pages per person per hour
- Upper bound ~ 15 minutes for scanning for an average book (26 pages of references)
- Additional scanning time does not significantly affect other processes in the library
- Prolongs the processing of a book on average by only 3 minutes




```
graph LR; A[Scanning of the list of references (only print)] --> B[Upload to LOC-DB]; B --> C[Reference Linking];
```

Scanning of the list of references (only print)

Upload to LOC-DB

Reference Linking

- Batch upload
- Background processing for meta data retrieval and reference extraction  
→ Does not affect the process in the library



Scanning of the list of references (only print)

Upload to LOC-DB

Reference Linking

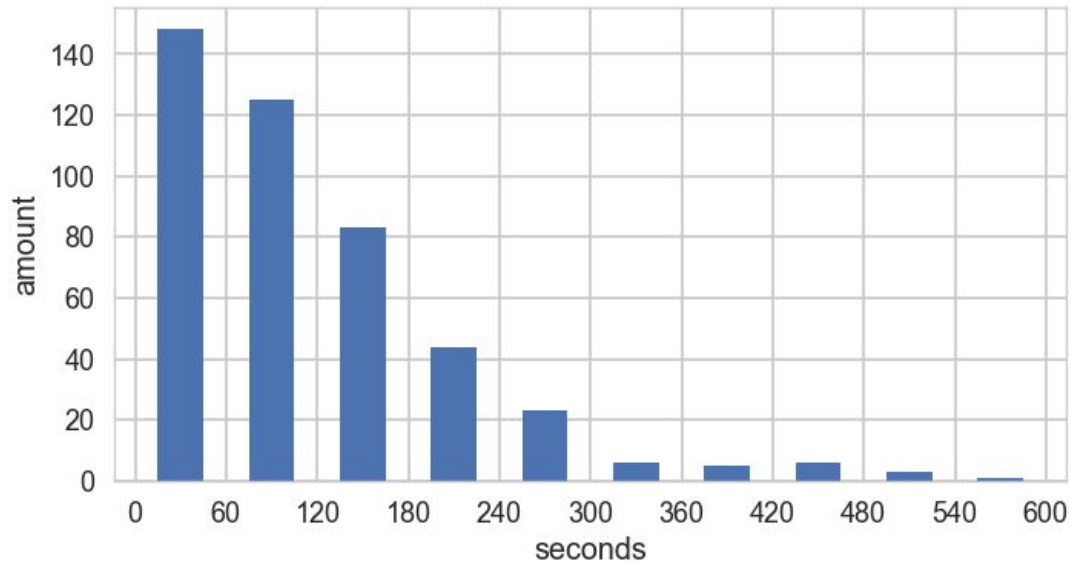
Criterion	Minimum	Maximum	Median
Citation Linking (s)	9.93	557.20	89.45
Internal Suggestion Retrieval (s)	0.02	0.5	0.06
External Suggestion Retrieval (s)	0.50	95.65	0.89
# Searches per Reference	1	36	2

Minimum, maximum, and median time in seconds for the reference linking step

Scanning of the list of references (only print)

Upload to LOC-DB

Reference Linking



Histogram of reference linking times





# How much would it cost if libraries catalogued everything and curated the citation graph?

Estimation about the number of full-time employees needed to process all literature of social sciences bought in 2011 by Mannheim University Library, depending on the time  $t$  in seconds to resolve a reference.

$t$	1	5	10	20	30	60	120
# employees	0.1	0.5	1	2	3	5.9	11.9