

Collaborative Literature Work in the Scientific & Educational Publication Process: The Cogeneration of Citation Networks

Leon Burkard (burkard@kit.edu), Andreas Geyer-Schulz

Information Services and Electronic Markets



Outline



- A traditional publication process
- Challenges & Issues
- Use-Case & Requirements for a literature middleware
- Demonstration of the LitObject-Middleware
- Cogeneration of citation networks
- Future Work
- Summary

A traditional publication process - Retrieval / Indexing - Download in personal literature - Editorial sub-processes database. - Import of document's - Review sub-processes - Production editor subreference metadata (BibTeX, RIS, etc.) processes Scientific Quality Distribution Creation Consumption Management - Literature work - Final typesetting - Writing / Citing - Offline publication processes - Literature - Ordering & shipment processes Management - DOI Handling, reference & metadata reparation - Classification, keywords, indexing

Challenges

Creation

Scientific Quality

Management

Distribution

Consumption /

Analysis



How to rediscover past literature searches?

How to share and persist literature searches?

Time-consuming reference and content checks

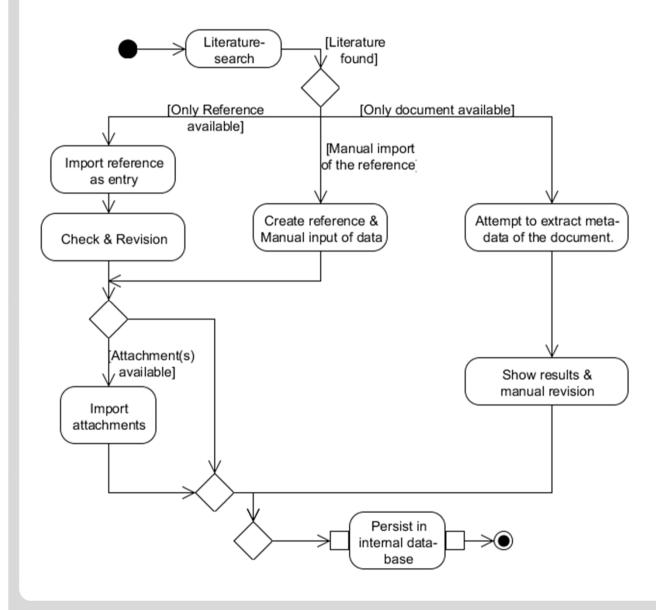
How to store cited literature permanently?

How to provide "good" reference metadata and links to attachments?

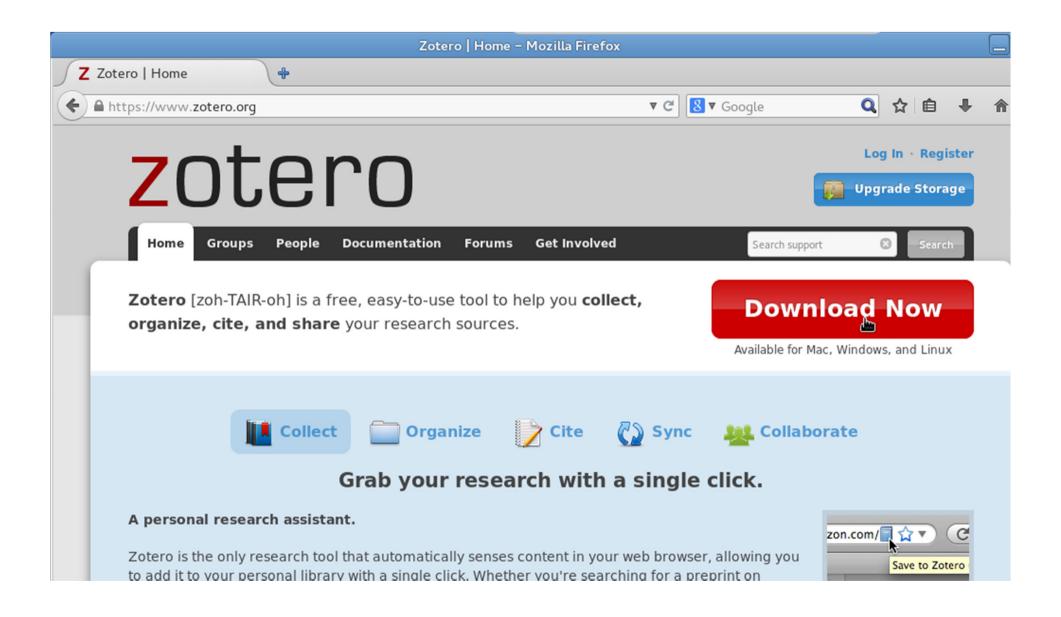
How to collect and use collected metadata for further usage e.g. citation networks?

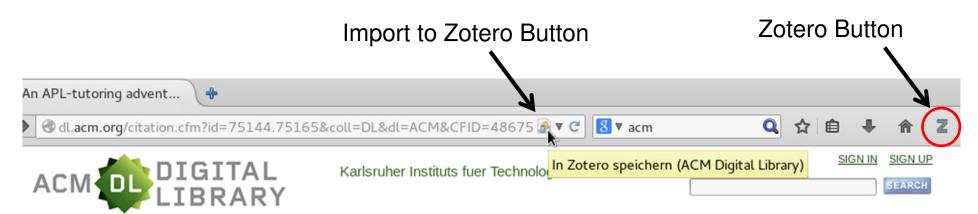
How to consume "good" reference metadata and links to attachments with various literature management approaches?



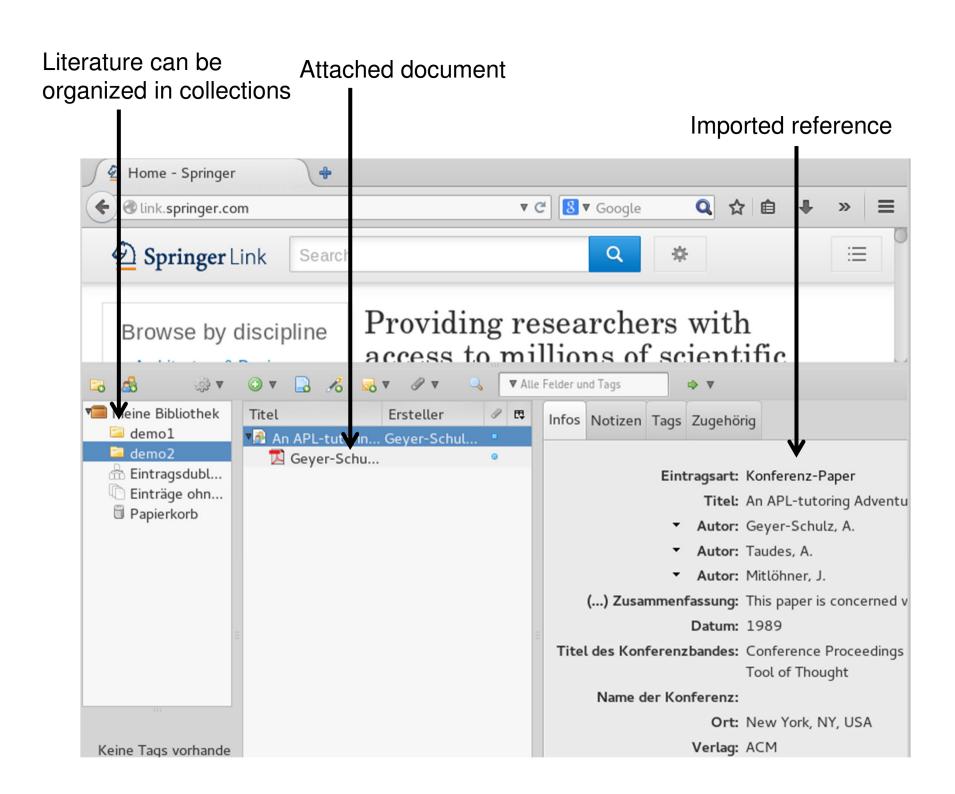


- a) Manual typing of reference, manual linking/importing attachments
- b) Attempt to extract information out of the document.
 - → Combination with a google search.
- c) Automatic import
 - a) Crawler / WebsiteTranslators
 - b) Embedded metadata in the website

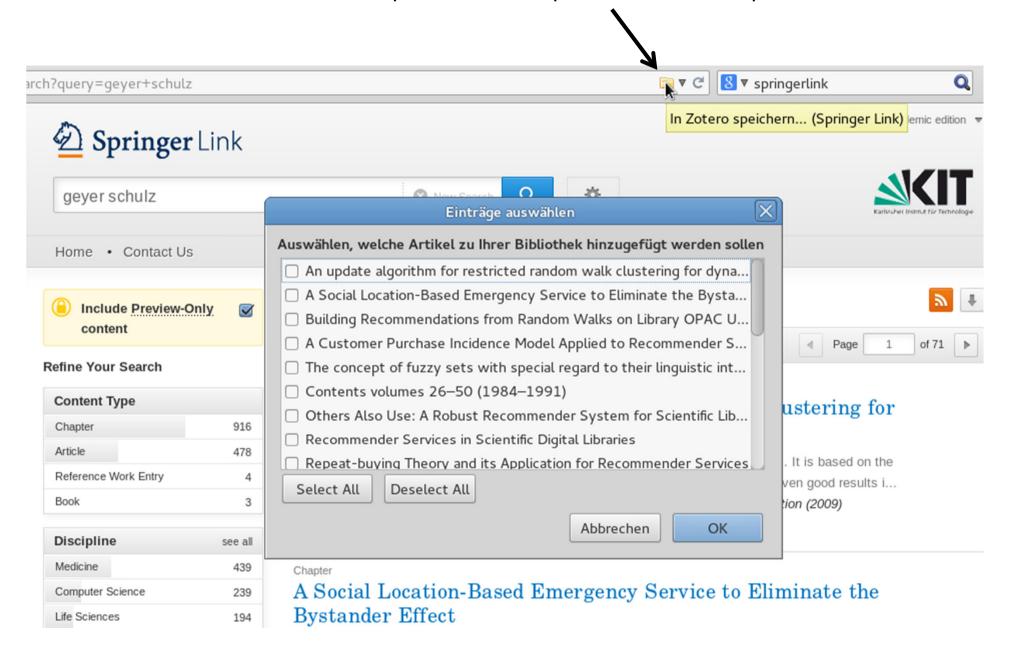




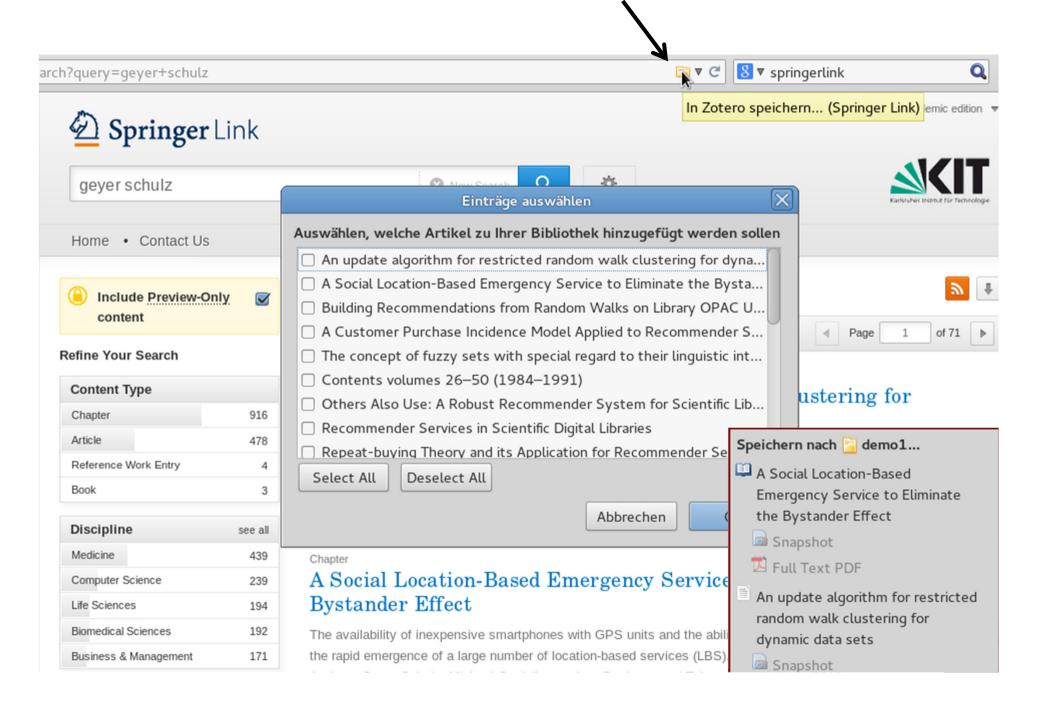


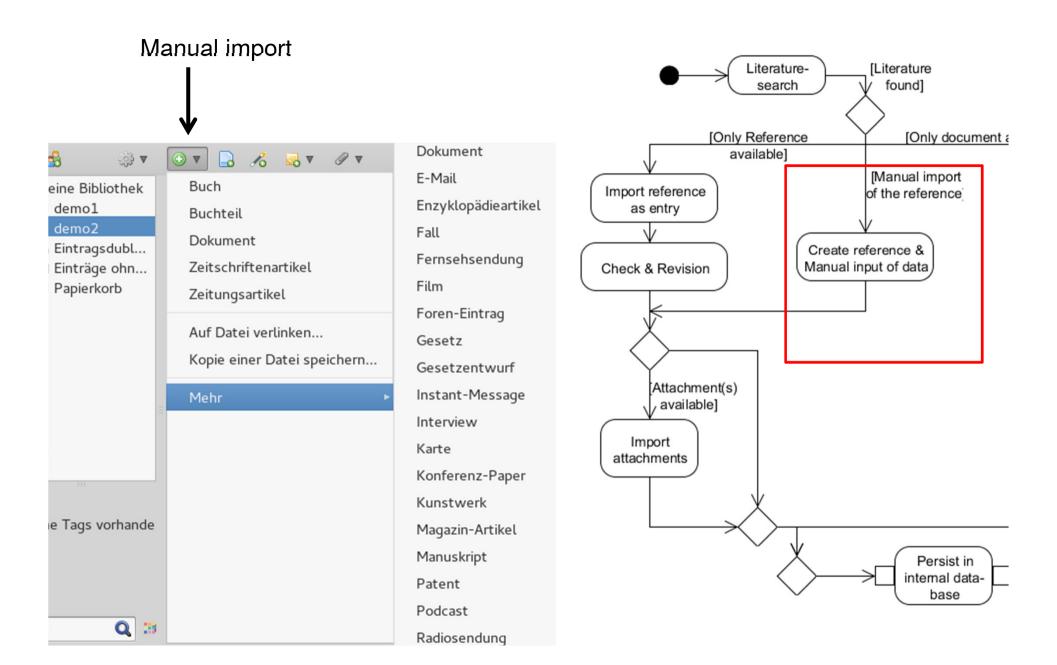


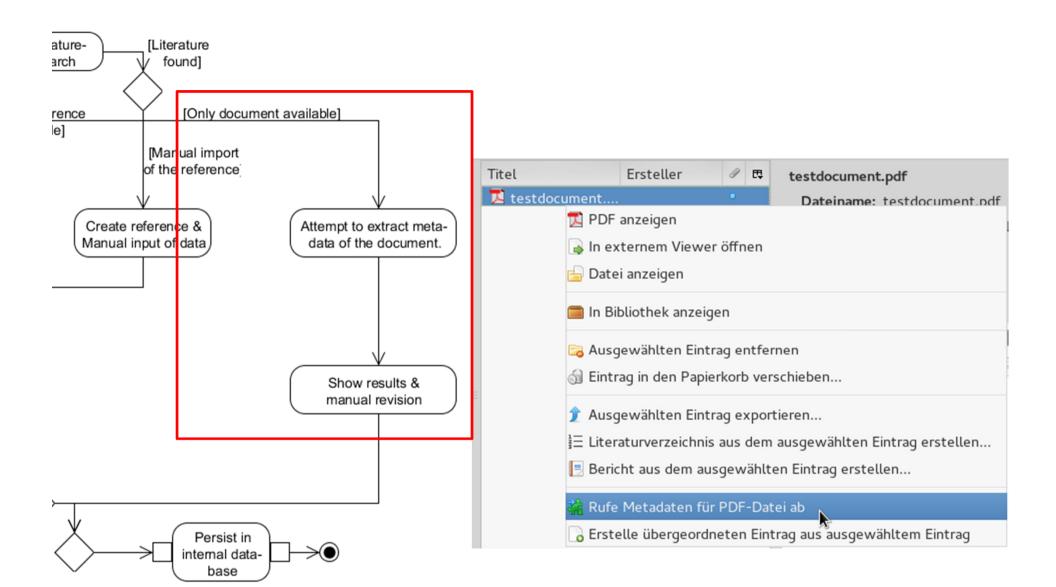
Multiple downloads possible in one step



Multiple downloads possible in one step









Quick-Copy

Quick Copy erlaubt Ihnen, ausgewählte Literaturangaben durch Drücken einer Tastenkombination (Strg+Umschalt+C) in die Zwischenablage zu kopieren oder Einträge in ein Textfeld auf einer Webseite zu ziehen.

Für Bibliografie-Stile können Sie Zitationen oder Fußnoten kopieren, indem Sie Strg+Umschalt+A drücken oder die Umschalttaste gedrückt halten, bevor Sie Einträge mit Drag-and-Drop einfügen.

Standardausgabeformat

BibTeX

Chicago Manual of Style 16th edition (note)

Vancouver

Export-Formate

TEI

Bibliontology RDF

MODS

Simple Evernote Export

BibLaTeX

Bookmarks

CSL JSON

Endnote XML

Refer/BibIX

RefWorks Tagged Format

RIS

Wikipedia Citation Templates

Various export options

Quick Export:

- 1. Select one or many literature entries in Zotero.
- 2. Press **Ctrl+Shift+C** for copying the reference to the clipboard.
- 3. At the desired destination document **press Ctrl+V** (paste) and the corresponding literature entries' references are exported in the selected style.

Educational Publication Process: The

Example BibTeX export

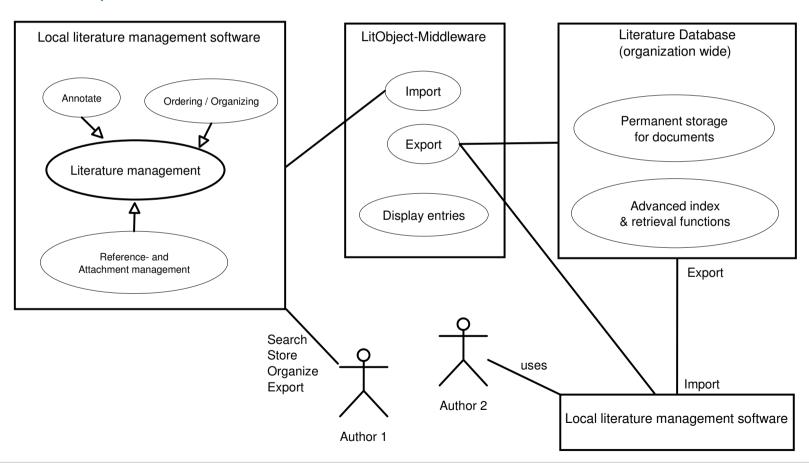
But....what about collaborative working with literature?

```
@inproceedings{geyer-schulz apl-tutoring 1989,
         address = {New York, {NY}, {USA}},
         series = {{APL} '89},
         title = {An {APL}-tutoring Adventure Game},
         isbn = \{0-89791-327-2\},\
         url = {http://doi.acm.org/10.1145/75144.75165},
         doi = {10.1145/75144.75165},
         abstract = {This paper is concerned with the development of a
         like to embed the learning component in an object oriented ad
         communicating automata, can be freely defined, using a "game
         actor who is presented with a number of exercises that are se
         entered into the system are analysed by the teachers, who use
         urldate = {2014-06-25},
         booktitle = {Conference Proceedings on {APL} As a Tool of Tho
12
         publisher = {{ACM}}},
         author = {Geyer-Schulz, A. and Taudes, A. and Mitlöhner, J.},
         year = \{1989\},\
         pages = \{148-157\},
         file = {Geyer-Schulz et al. - 1989 - An APL-tutoring Adventur
         Adventure Game.pdf:application/pdf}
     @article{franke update 2009,
         title = {An update algorithm for restricted random walk clust
         volume = \{3\},
         issn = {1862-5347, 1862-5355},
         url = {http://link.springer.com/article/10.1007/s11634-009-00
         doi = \{10.1007/s11634-009-0039-6\},
         abstract = {In this article, we present a randomized dynamic
         given good results in past studies. We discuss different appr
         efficient for a small percentage of changes in the data set a
         thus no need for any reclustering ever. In addition, the meth
         language = {en},
         number = \{1\},
         urldate = {2014-06-25},
         journal = {Advances in Data Analysis and Classification},
         author = {Franke, Markus and Geyer-Schulz, Andreas},
         month = jun,
         year = {2009},
         keywords = {60320, 62H30, Clustering, Dynamic clustering, Ran
         Physics, Computer Science, Chemistry and Earth Sciences, Stat
         pages = \{63--92\},
         file = {Full Text PDF:/home/leon/.mozilla/firefox/atcxfiq7.De
         pdf;Snapshot:/home/leon/.mozilla/firefox/atcxfiq7.Demo_LIS Wo
     @incollection{geyer-schulz social 2012,
         series = {Communications in Computer and Information Science}
```

Use-Case Scenario

Karlstule Institute of Technology

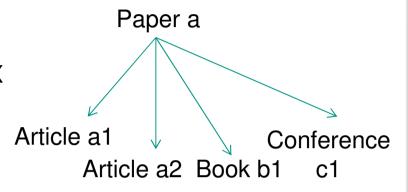
- Rediscover
- Share
- Persist
- Less duplication of work
- Continuous transfer of literature objects
- Basis for growing, co-generated linked literature objects

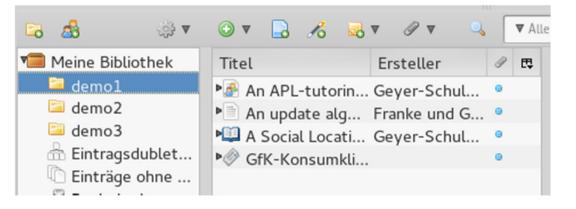


Requirements for the LitObject Middleware



- Save literature objects
 - Reference in a data format, e.g. BibTeX
 - One or many attachments
 - Notes & Annotations to the literature object
- Save links between literature objects
 - Literature object a references b,c,d
- Save collections of literature objects
 - All literature to a sub-topic

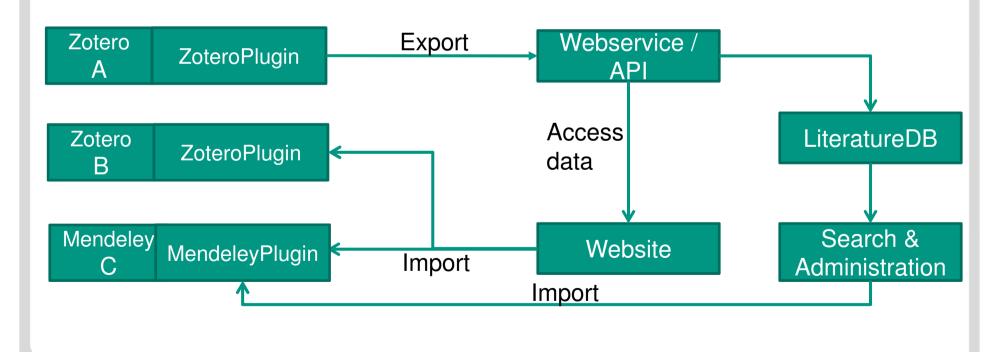




Karlsruhe Institute of Technology

Requirements for the LitObject Middleware

- Save literature objects
- Save links between literature objects
- Save collections of literature objects



Why not use one of the integrated solutions?

- Commercial services = Limited research possibilities
 - Closed source, difficult to extend for new applications
 - Often too expensive for intensive collaborative literature work
- Vendor Lock-In
 - No literature object sharing, interoperability nearly non-existent
- Extensibility very limited
 - No further services like citation graph generation or recommendation services
 - No seamless integration in publication processes
- We need a solid basis for further development projects in the field of open library and information services Create plan



Personal library space

For education & non-profit use

(including VAT)

Group collaborators

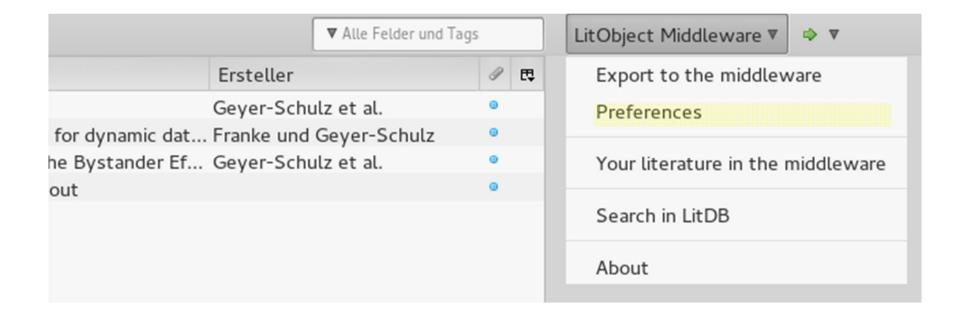


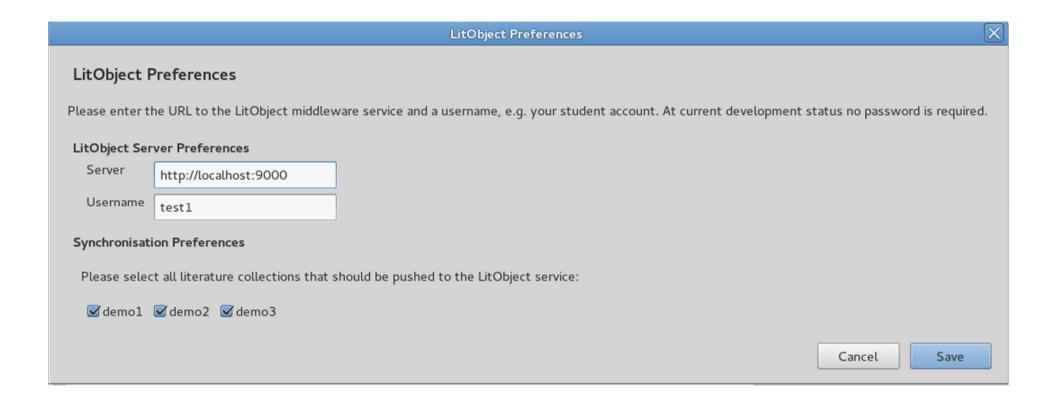
Demo Part 2

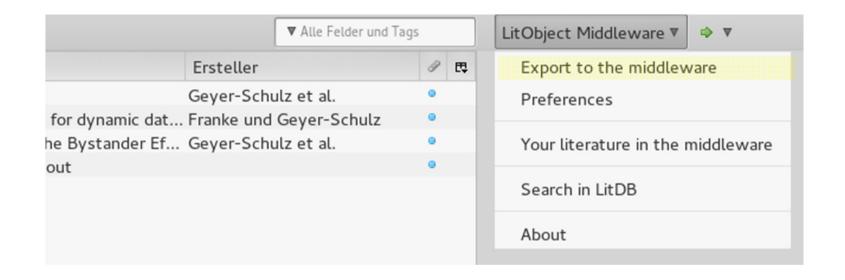
- Save literature objects
- Save links between literature objects
- Save collections of literature objects

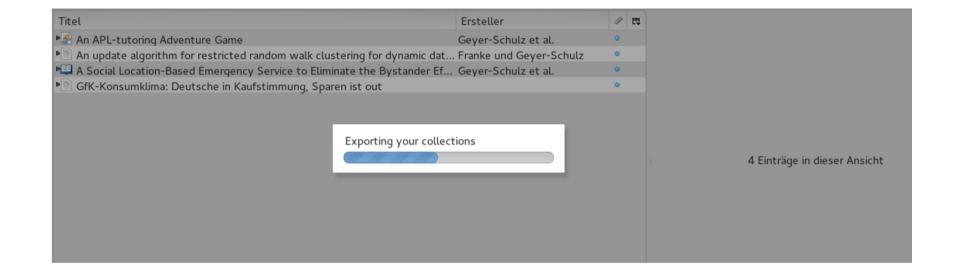


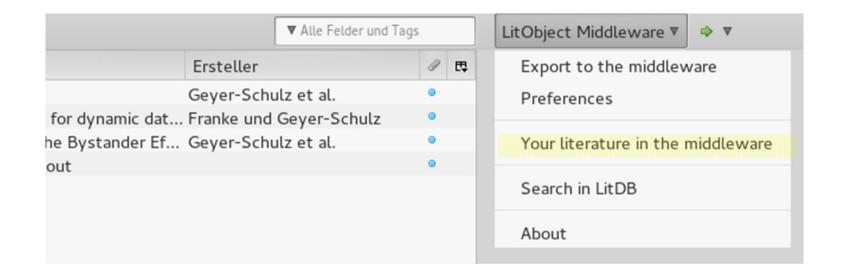
- LitObject Middleware
 - Developed as a Mozilla Firefox extension
 - Extends the literature management software Zotero
 - Provides the export process step to the LitObject Middleware Service













Workshop/zotero/storage/SC5P9W2N/content.pdf:application/pdf

o year: 2009

o urldate: 2014-06-25

o title: An update algorithm for restricted random walk clustering for dynamic data sets

o pages: 63-92

incollection: A Social Location-Based Emergency Service to Eliminate the Bystander Effect

publisher: Springer Berlin Heidelberg
 isbn: 978-3-642-25205-1, 978-3-642-25206-8

o language: en

o copyright: @2012 Springer-Verlag {GmbH} Berlin Heidelberg

author: Geyer-Schulz, Andreas and Ovelgönne, Michael and Sonnenbichler, Andreas C.

url: http://link.springer.com/chapter/10.1007/978-3-642-25206-8_7

o series: Communications in Computer and Information Science

• abstract: The availability of inexpensive smartphones with {GPS} units and the ability to run 3rd party software facilitated the rapid emergence and navigation related application scenarios, in this contribution we want to motivate the use of mobile devices for personal safety services. The Obstacles in the social help process have been summarized in the concept of the bystander effect. This contribution aims to show how a {LBS} caservice that notifies nearby contacts of a victim of the incident and guides them to the victim so that they can provide help. Furthermore, we will closeness of service participants from different sources of data. Also, we will discuss and compare approaches for the system design we consider benefits in practice.

title: A Social Location-Based Emergency Service to Eliminate the Bystander Effect

number: 222pages: 112-130

keywords: Algorithm Analysis and Problem Complexity, Bystander effect, Coding and Information Theory, Computer Communication Networks, Computer Systems, Privacy, Social network analysis

o editor: Obaidat, Mohammad S. and Tsihrintzis, George A. and Filipe, Joaquim

file: A Social Location-Based Emergency Service to Eliminate the Bystander Effect - Springer:/home/leon/.mozilla/firefox/atcxfiq7.Demo_LIS Worl Workshop/zotero/storage/BR22HKVW/content.pdf:application/pdf

o vear: 2012

o urldate: 2014-06-25

o month: #jan#

booktitle: e-Business and Telecommunications

inproceedings: An {APL}-tutoring Adventure Game

publisher: {ACM}
 doi: 10.1145/75144.75165
 isbn: 0-89791-327-2

o author: Geyer-Schulz, A. and Taudes, A. and Mitlöhner, J.

url: http://doi.acm.org/10.1145/75144.75165

series: {APL} '89

• abstract: This paper is concerned with the development of a program system that assists the teaching of {APL}2. Rather than presenting the teaching adventure game environment written itself in {APL}2. The game's actors, objects and places, which are implemented as hierarchical claby a game simulator. The teaching aspect is integrated as follows: The adventurer/student is a specific actor who is presented with a number of exobtained by the student so far. The solutions entered into the system are analysed by the teachers, who use an {LR}(1)-parser generator and "cri-

o title: An {APL}-tutoring Adventure Game

year: 1989
 pages: 148-157

file: content:/home/leon/.mozilla/firefox/atcxfiq7.Demo LIS Workshop/zotero/storage/NB5DZ666/content.pdf:application/pdf

o address: New York, {NY}, {USA}

o urldate: 2014-06-25

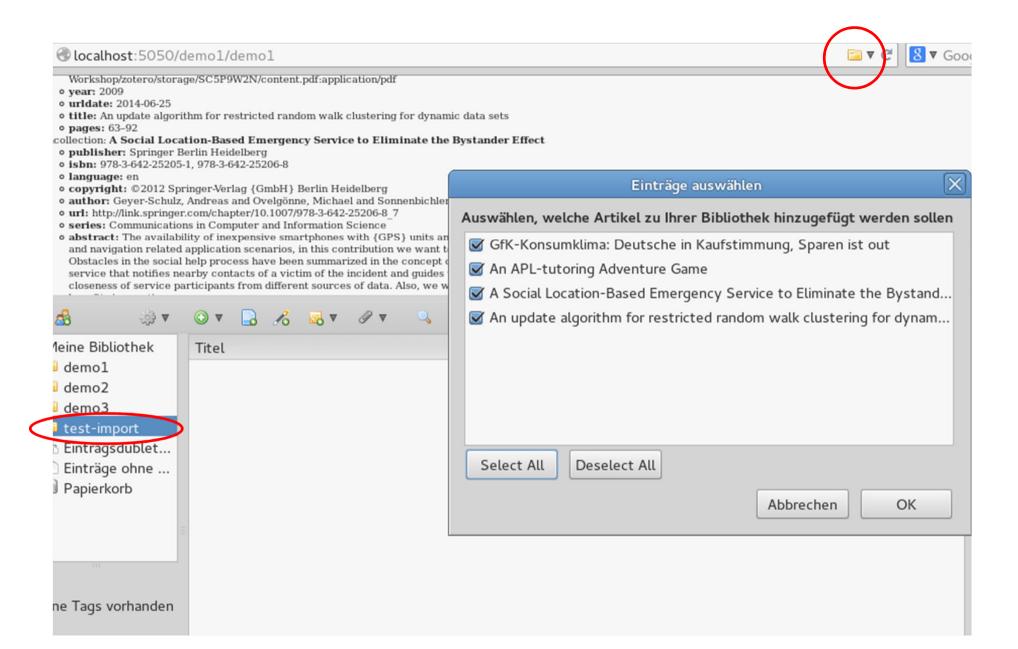
o booktitle: Conference Proceedings on {APL} As a Tool of Thought

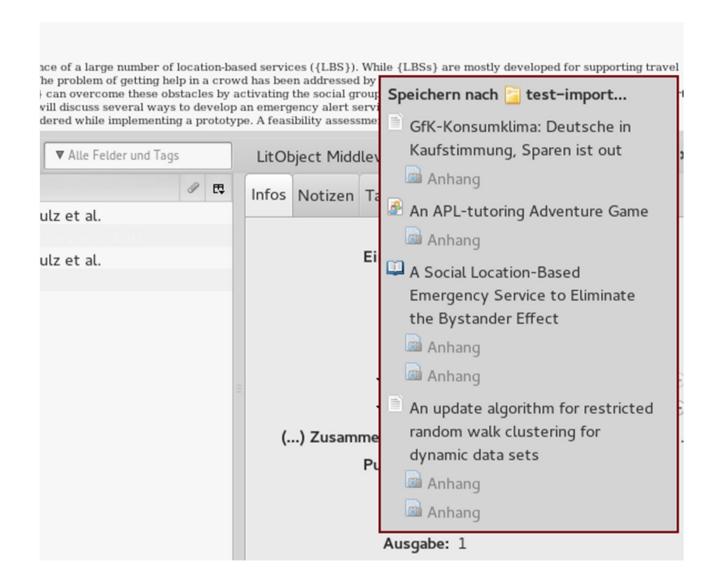
article: {GfK}-Konsumklima: Deutsche in Kaufstimmung, Sparen ist out

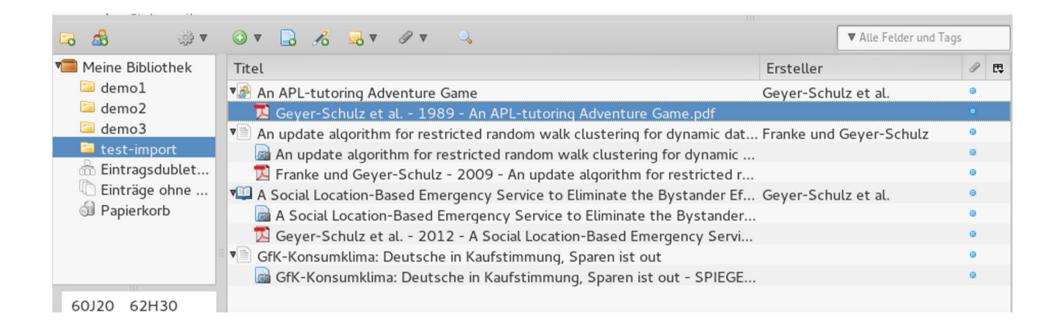
o shorttitle: {GfK}-Konsumklima

o title: {GfK}-Konsumklima: Deutsche in Kaufstimmung, Sparen ist out

url: http://www.spiegel.de/wirtschaft/unternehmen/gfk-konsumklima-deutsche-in-kaufstimmung-sparen-ist-out-a-977318.html







Challenges

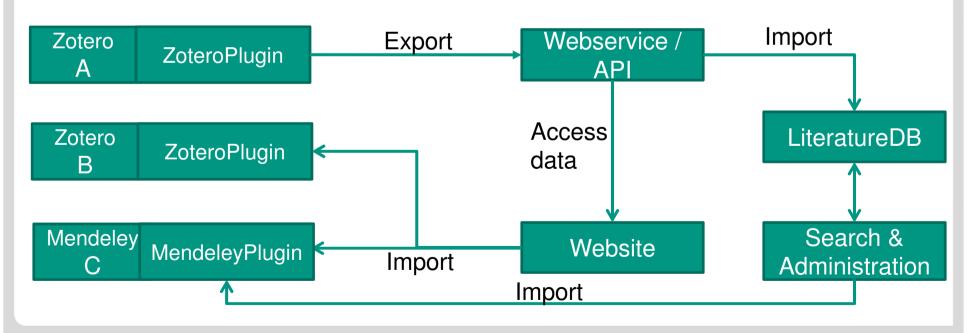
Scientific Quality

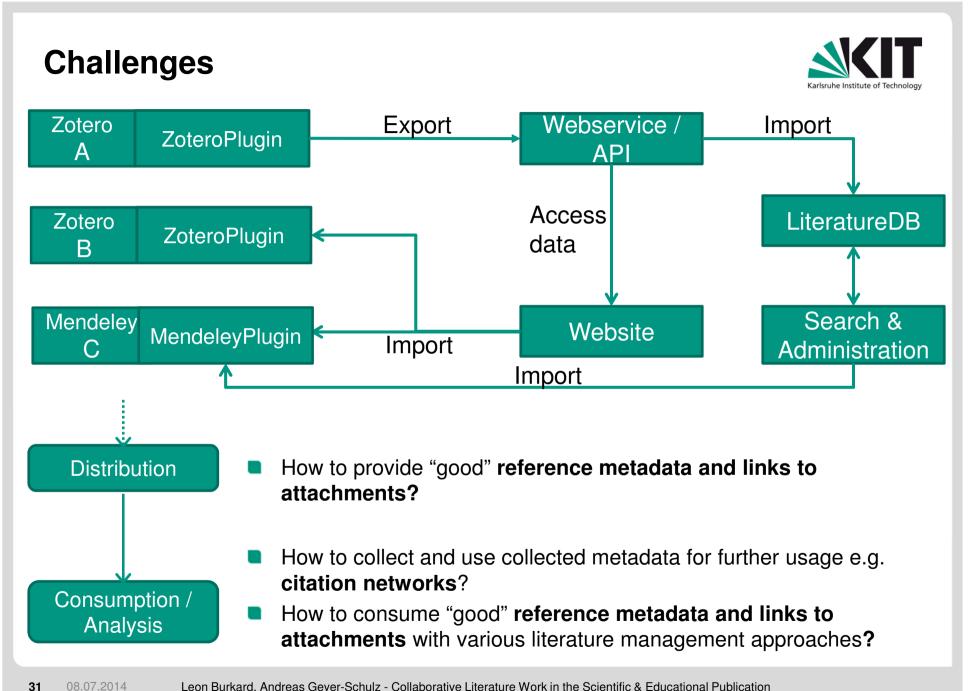
Management





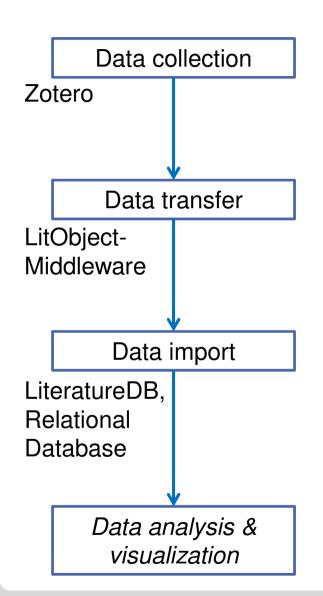
- How to **share** and **persist** literature searches?
- How to work together with regard to literature?
- Do the work of the author(s) again for reviews?
- Time-consuming reference and content checks
- What about **interoperability** between various software tools?
- How to **store cited literature** permanently?





Cogeneration of citation networks – Generic process





Authors/Users: Search, retrieve, check and organize literature

→ Quality management shifted to the individual

Client-side software: Export references and attachments to a web-service for further use

→ Software supported transfer of structured data (links)

Server-side software: Import to link tables as data basis for citation graph analysis and algorithms

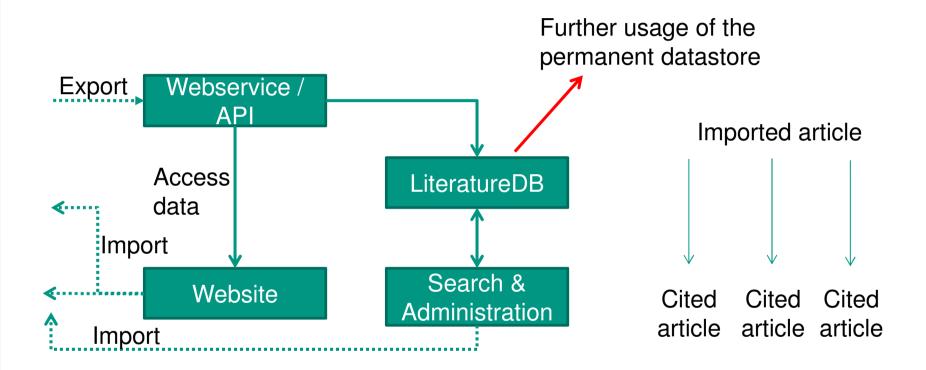
→ Using the LitObject Middleware for a growing data basis with individuals' support in checking and organizing

Using a co-generated growing data basis for analysis, visualization

→ Not covered today

Cogeneration of citation networks





Step-by-Step co-generation



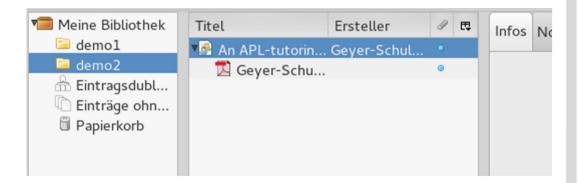


Organize literature



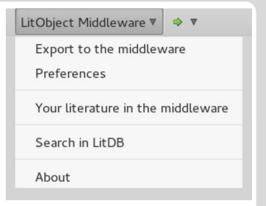


- Save all cited literature in one collection
- Save all collected literature for the topic in another collection



Step-by-Step co-generation

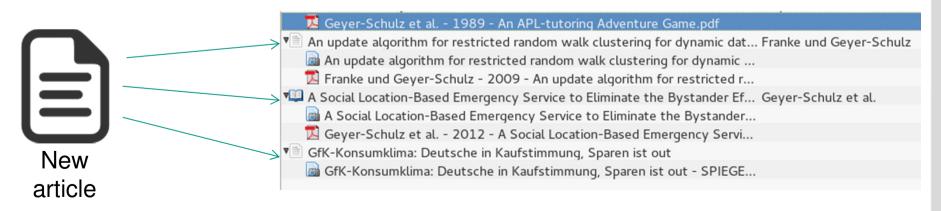
Export both collections to the webservice



Export LiteratureDB

Search & Administration

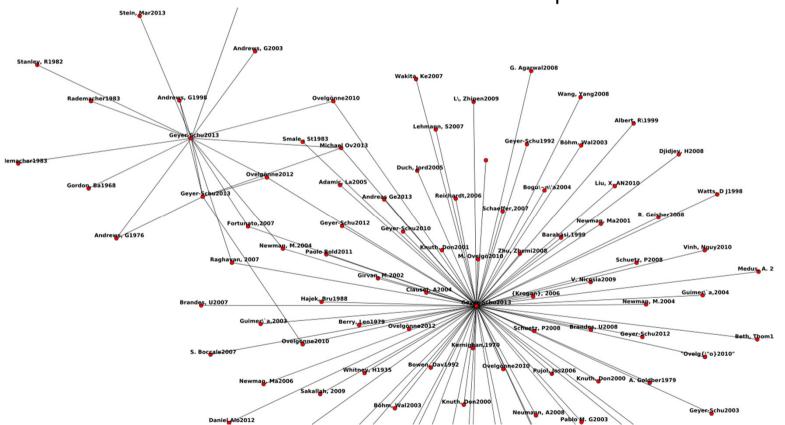
- Save links between documents
- During the import in the LiteratureDB save additionally a new document - that is the article newly written.
 - Save additionally a link from the newly written document (the additional document) to each document in both collections



Step-by-Step co-generation



- Use the growing, cogenerated link database for further applications and visualizations
- Work in progress:
 - Provide an API for automatic graph generation
 - Provide an API for automatic citation data export in various formats



Future Work



- Imported references are not validated
 - E.g. how to detect/solve misspelled author names?
 - Sometimes reference and attachment do not match
- Duplicates detection is future work
 - Currently limited to a single user duplicates detection on client-side
- User Management exists only on an elementary level
- Process improvements using the LitObject middleware
 - Use the LitObject middleware for advanced publication and review processes.
 - Use the LitObject middleware for private institutional repositories
 - Provide an automated citation network graph generation and structured linked data export.

Future Work



- Support more literature management tools
 - Mendeley is on top of the agenda.
 - An HTML5 online interface for organizing literature is planned.
- Developing new services on top of the LitObject middleware
 - Implement classification algorithms that use the LitObject middleware objects (reference and fulltext document)
 - Implement and test recommendation algorithms.
- Discuss/implement how to provide references and links to fulltext documents for common literature management software
- Publish a small open-source literature database as organization internal persistence unit.

Summary



- The LitObject middleware helps to share literature objects consisting of references and attachments
 - Basis tool for developing advanced processes that include literature
 - Basis tool for developing new applications that use the structured literature data in the webservice
- Possibility to shift workload to authors/literature organizers and reduce duplicate work
 - No manual reference transfer throughout the publication process chain.
 - Chance to build up a database of checked, correct references that can be used from authors again.
- Import step for literature management software through websites needs further development instead of new web translators every time.
 - Maybe unAPI v2





Backup

How to import LitObjects



- There exist various metadata formats for exposing metadata to literature management tools through websites:
 - Colns (ContextObjects in Spans)
 - DublinCore Metadata tags: dc.*
 - Eprints tags: eprints.*, PRISM tags: prism.*, etc.
 - MODS XML, Marc21, MAB, etc.
 - BibTex, RIS
- Highwire Press Keys: citation *
 - Field for linking to the fulltext attachment: citation pdf url
- **Lesson learnt:** Either good metadata and no link to literature attachments or bad metadata and links to literature attachments
 - Good = Export a BibTeX reference to another format, import and convert it back to BibTeX without too much information loss

A small extension for unAPI



- unAPI by Daniel Chudnow, 2006
 - Link to common reference formats instead of translating to other microformats before

1. Embed a resolver service

2. Link by using identifiers to the service

New: Add links to documents/fulltext attachments





Embedded Metadata 400

DOI 300

Colns 250

unAPI / Highwire 200

Specific Web Translator (individual simple crawler for each specific site)



Embedded Metadata (<meta> keys in <head> section)



Resolve DOI



Colns



Example for ContextObjects in spans



```
<span class='Z3988' title='url ver=Z39.88-2004&amp;</pre>
  ctx ver=Z39.88-2004&
  rfr id=info%3Asid%2Fzotero.org%3A2&
  rft id=urn%3Aisbn%3A978-3-642-25205-1%2C%20978-3-642-25206-8&
  rft val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Abook&
  rft.genre=bookitem&
  rft.atitle=A%20Social%20Location-
Based%20Emergency%20Service%20to%20Eliminate%20the%20By[...]ct&
  rft.btitle=e-Business%20and%20Telecommunications&
  rft.publisher=Springer%20Berlin%20Heidelberg&
  rft.series=Communications%20in%20Computer%20and%20I[...]&
  rft.aufirst=Andreas&
  rft.aulast=Gever-Schulz&
  rft.au=Andreas%20Geyer-Schulz&
  rft.au=Michael%200velg%C3%B6nne&
  rft.au=Andreas%20C.%20Sonnenbichler&
  rft.au=Mohammad%20S.%200baidat&
  rft.au=George%20A.%20Tsihrintzis&
  rft.au=Joaquim%20Filipe&
  rft.date=2012-01&
  rft.pages=112%E2%80%93130&
  rft.isbn=978-3-642-25205-1%2C%20978-3-642-25206-8&
  rft.language=en'>
</span>
```



unAPI Details

```
<link rel="unapi-server" type="application/xml" title="unAPI"
href="http://localhost:5050/unapi/demo1" />
```

Get the available formats

- Get the metadata in one of the available formats
- → http://localhost:5050/unapi/demo1?id=7QZM2MXJ&format=bibtex

A full graph of our private link database



