

Das BASE-Netzwerk

Services und Applikationen im Umfeld der Bielefeld Academic Search Engine

Mathias Lösch

Universitätsbibliothek Bielefeld – LibTec

Kolloquium Wissensinfrastruktur 17.5.2013

BASE

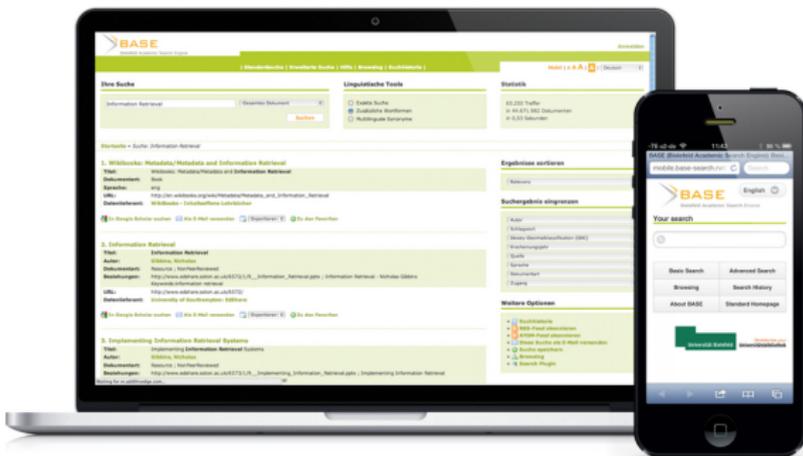
Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur



- ~45M Dokumente
- > 2.500 Quellen

Agenda

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

1. Voraussetzungen

2. Services

- Suchschnittstelle
- Rsync-Schnittstelle
- OAI-Schnittstelle
- Statistik-Schnittstelle
- Klassifikationsschnittstelle

3. Ausblick

4. Zusammenfassung

Agenda

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

1. Voraussetzungen

2. Services

- Suchschnittstelle
- Rsync-Schnittstelle
- OAI-Schnittstelle
- Statistik-Schnittstelle
- Klassifikationsschnittstelle

3. Ausblick

4. Zusammenfassung

Quellenakquise

- Repository-Registries
 - Open Archives Registry
 - OpenDOAR
 - ROAR
 - University of Illinois Registry
 - OJS Registry
- Bekanntheit in der Community
- Vorschlagsformular

Suchmaschine BASE - Base

www.base-search.net/about/et/suggest.php

BASE
Search Academic Search Engine

Anmelden

AAA [Deutsch]

Dienste

- BASE Suche
- Über BASE
- BASE Help
- FAQ
- BASE Lab
- Repository vorschlagen**
- Kontakt
- Impressum

Repository vorschlagen

Sie möchten uns auf Ihr Repository / O-Journal hinweisen?

Wir interessieren die Inhalte wissenschaftlicher Quellen, die Ihre Metadaten über OAI-PMH bereitstellen. Nur wenn Ihr Repository über eine OAI-PMH-Schnittstelle verfügt, ist es für uns möglich, die Metadaten aus Ihrem Repository / O-Journal zu indexieren.

Der **BASE OAI-PMH Validity Checker (OVAL)** testet schnell und einfach OAI-Schnittstellen auf Ihre Kompatibilität zu BASE.

Weitere Informationen zur Indexierung finden Sie in unserer **FAQ**.

Die mit * gekennzeichneten Felder bitte unbedingt ausfüllen!

Repository Name*: _____

Repository URL*: _____

Open Access*: Alle Einige Keine

Open Access Set: _____

OAI Base URL: _____ [Mit OVAL prüfen]

Repository System: _____

Klassifikation (DDC, UDC etc.): _____

Straße: _____

PLZ: _____

Ort: _____

Land*: _____

Koordinaten (Suche): Breitengrad (Grad): _____ Längengrad (Grad): _____ (Nichtleere Werte)

Ihr Name*: _____

Ihre E-Mail-Adresse*: _____

Mithaltung: _____

Kopie gewünscht? Bitte schicken Sie mir eine Kopie meiner Nachricht

Datenqualität // Digital Curation

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

Kuratierung der Datenquellen

- Herkunftsland
- Kontinent
- Typisierung
- Repository-Software
- Open-Access-Anteil

OAI-PMH-Expertise

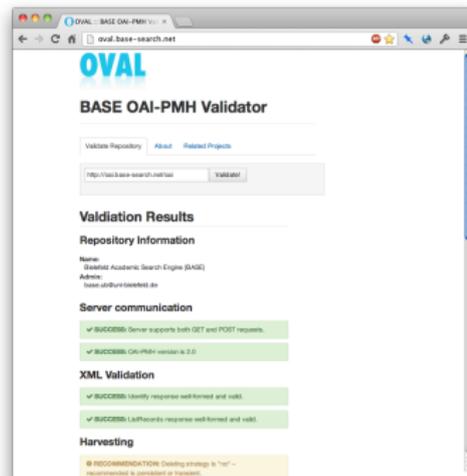
- Feedback, Beratung

- Validator

<http://oval.base-search.net/>

- Guidelines

- DINI
- DRIVER
- OpenAIRE



Datenqualität // Normalisierungen

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

Normalisierungen der Metadatenfelder

Sprache "eng", "en", "en_US", "English" → "eng"

Dokumenttyp "article", "doc-type:article" → "0001"

Publikationsjahr "2004", "2004-10-20", "April 2004" → "2004"

Automatische Klassifikation

- "The notion of a Hilbert space..." → DDC 510 Mathematics

Datenqualität // Normalisierungen

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

Normalisierungen der Metadatenfelder

Sprache "eng", "en", "en_US", "English" → "eng"

Dokumenttyp "article", "doc-type:article" → "0001"

Publikationsjahr "2004", "2004-10-20", "April 2004" → "2004"

Automatische Klassifikation

- "The notion of a Hilbert space..." → DDC 510 Mathematics

Schnittstellen // BASE-Architektur

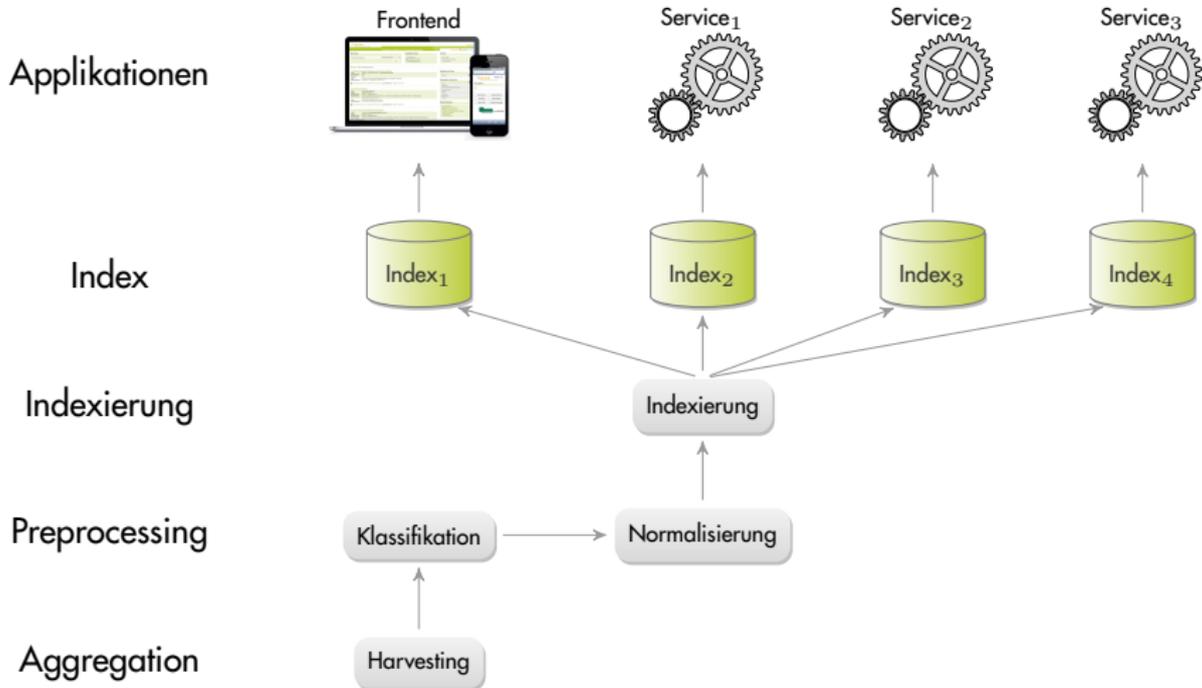
Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur



Agenda

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

1. Voraussetzungen

2. Services

- Suchschnittstelle
- Rsync-Schnittstelle
- OAI-Schnittstelle
- Statistik-Schnittstelle
- Klassifikationsschnittstelle

3. Ausblick

4. Zusammenfassung

Agenda

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

1. Voraussetzungen

2. Services

- Suchschnittstelle
- Rsync-Schnittstelle
- OAI-Schnittstelle
- Statistik-Schnittstelle
- Klassifikationsschnittstelle

3. Ausblick

4. Zusammenfassung

Suchschnittstelle

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

- Einbindung der BASE-Suche in externe Anwendungen
- z.B. Bibliothekskataloge:
 - Katalog.plus UB Bielefeld
 - Katalog d. UB Heidelberg
- Metasuchmaschinen
 - Metager
 - Karlsruher Virtueller Katalog (KVK)

The screenshot shows the 'Katalog.plus' search results page. The search criteria are 'Luhmann' and 'Web-Dokumente'. The results list includes:

- 1. **Luhmann and Marc: social theory and social freedom** / Thomàs, C. / Progress, 2012. [Link zum Dokument](#)
- 2. **The Author's Case of Hans-Joachim Friedrich Metzinger, Michel Foucault and Niklas Luhmann** / Reissbach, S. / European University (EUBL), 1993. [Link zum Dokument](#)
- 3. **Art as a social system: The sociological aesthetics of Niklas Luhmann** / Humpfer, M. (Hrsg.) / Tübingen, 2010. [Link zum Dokument](#)
- 4. **Niklas Luhmann's Theory of Politics and Law (Bulgarian translation 2008)** / King, M. / Progress, 2008, 2012. [Link zum Dokument](#)
- 5. **Niklas Luhmann y la sociología de la construcción** / RUI, Estrella. / Ediciones Universidad de Chile, 2012. [Link zum Dokument](#)
- 6. **Luhmann's political theory: politics after metaphysics?** / Thomàs, C.J. / HUP, 2008. [Link zum Dokument](#)
- 7. **Herzfeld's Ra-aptly. Niklas Luhmann's Social Systems Theory: Society and Law** / Franzen-Timmermans, L.M.A. / Utrecht University, 2009. [Link zum Dokument](#)
- 8. **Systemic Theory and Legal Theory: Luhmann, Heidegger and the False Side of Metaphysics** / Thomàs, C. / 2002. [Link zum Dokument](#)
- 9. **Niklas Luhmann: A Sociological Transformation of Political Legitimacy** / Thomàs, C. / 2008. [Link zum Dokument](#)

Agenda

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

1. Voraussetzungen

2. Services

- Suchschnittstelle
- Rsync-Schnittstelle
- OAI-Schnittstelle
- Statistik-Schnittstelle
- Klassifikationsschnittstelle

3. Ausblick

4. Zusammenfassung

■ Gesamte BASE-Datenbasis als Download

Anwendungen

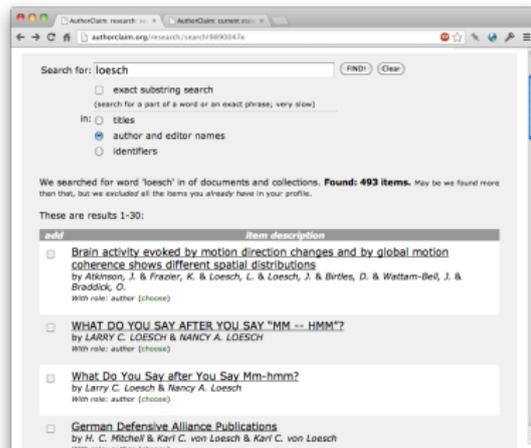
■ AuthorClaim

- Autoren „claimen“ ihre Publikationen (u.a.) aus dem BASE Pool

■ <http://authorclaim.org/>

■ High North Research Documents

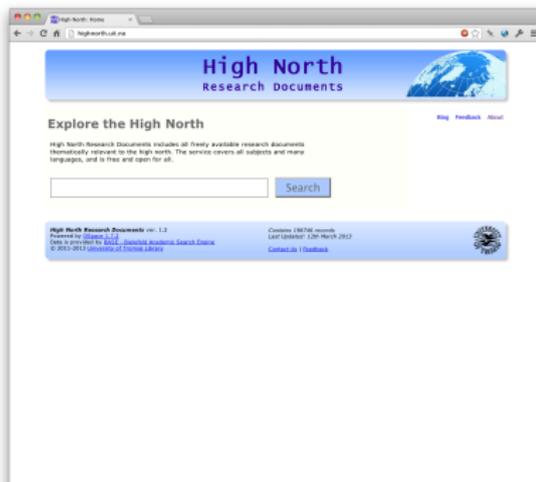
- BASE-Dokumente aus dem Gebiet der Polarforschung
- <http://highnorth.uit.no/>



■ Gesamte BASE-Datenbasis als Download

Anwendungen

- AuthorClaim
 - Autoren „claimen“ ihre Publikationen (u.a.) aus dem BASE Pool
 - <http://authorclaim.org/>
- High North Research Documents
 - BASE-Dokumente aus dem Gebiet der Polarforschung
 - <http://highnorth.uit.no/>



Agenda

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

1. Voraussetzungen

2. Services

- Suchschnittstelle
- Rsync-Schnittstelle
- **OAI-Schnittstelle**
- Statistik-Schnittstelle
- Klassifikationsschnittstelle

3. Ausblick

4. Zusammenfassung

Ziele

- Zugriff auf Daten-Ausschnitte aus BASE
 - Fachlich (DDC-basiert)
 - geographisch
 - nach Dokumentsprache
 - nach Quelle
 - Kombinationen
- Protokoll: OAI-PMH
 - Standard-Client-Software vorhanden
 - Breite Akzeptanz bei der Zielgruppe

OAI-Schnittstelle // Implementierung

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

Implementierung

- Erweiterung des OAI-PMH-Protokolls um „Dynamische OAI-Sets“ (inspiriert durch DataCite)
- Set-Spezifikation mit der Solr-Query-Syntax:
 - `set=ddc:330` vs.
 - `set=deweyfull:33* AND language:(eng OR ger)`

OAI-Schnittstelle // Spezifizierung von Ausschnitten

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

Beispiel: Spezifizierung fachlicher Ausschnitte

- Wirtschaftswissenschaften: deweyfull:33*
- Wirtschaftswissenschaften und BWL: deweyfull:33* AND deweyfull:65*
- Physik und Astronomie: deweyfull:53* AND deweyfull:52*

Anwender

- Virtuelle Fachbibliotheken
 - EconBiz: Virtuelle Fachbibliothek Wirtschaft (ZBW)
 - Fachportal Pädagogik: Erziehungswissenschaft (DIPF)
- OpenAIREplus
 - Trainingsdaten für Dokumentklassifikation und -clustering
- Europeana Cloud
 - Fachbereich „Social Sciences & Humanities“
 - `deweyfull:* AND -deweyfull:0* AND -deweyfull:5* AND -deweyfull:6* AND continent:ceu`

Anwender

- Virtuelle Fachbibliotheken
 - EconBiz: Virtuelle Fachbibliothek Wirtschaft (ZBW)
 - Fachportal Pädagogik: Erziehungswissenschaft (DIPF)
- OpenAIREplus
 - Trainingsdaten für Dokumentklassifikation und -clustering
- Europeana Cloud
 - Fachbereich „Social Sciences & Humanities“
 - `deweyfull:* AND -deweyfull:0* AND -deweyfull:5* AND -deweyfull:6* AND continent:ceu`

Anwender

- Virtuelle Fachbibliotheken
 - EconBiz: Virtuelle Fachbibliothek Wirtschaft (ZBW)
 - Fachportal Pädagogik: Erziehungswissenschaft (DIPF)
- OpenAIREplus
 - Trainingsdaten für Dokumentklassifikation und -clustering
- Europeana Cloud
 - Fachbereich „Social Sciences & Humanities“
 - `deweyfull:* AND -deweyfull:0* AND -deweyfull:5* AND -deweyfull:6* AND continent:ceu`

Agenda

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

1. Voraussetzungen

2. Services

- Suchschnittstelle
- Rsync-Schnittstelle
- OAI-Schnittstelle
- **Statistik-Schnittstelle**
- Klassifikationsschnittstelle

3. Ausblick

4. Zusammenfassung

Statistik-Schnittstelle

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

Beta

Motivation

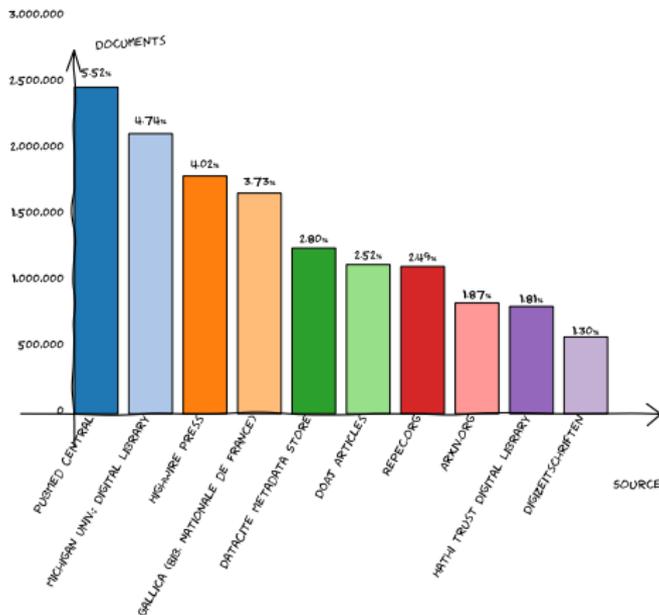
- BASE ist z.Zt. vermutlich das vollständigste Abbild der internationalen Repositorien-/Open Access-Landschaft
- Hohe Datenqualität durch Normalisierungen
- Forschungsinteresse:
 - „Dramatic Growth of Open Access“ (H. Morrison)
 - Census of Open Access Repositories (Vierkant et al., 2012)

Statistik-Schnittstelle // Funktionen

[Voraussetzungen](#)[Services](#)[Ausblick](#)[Zusammenfassung](#)[Literatur](#)

- Programmatische Abfrage von statistischen Daten aus dem aktuellen BASE-Index
- Tabellarische & graphische Ausgabeformate
- Verschiedene Dateiformate

10 größte Quellen



Statistik-Schnittstelle // Plots

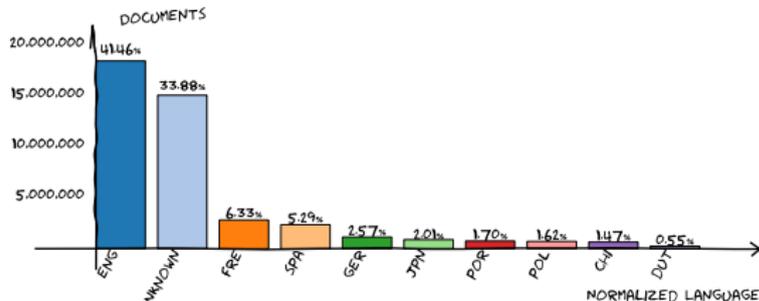
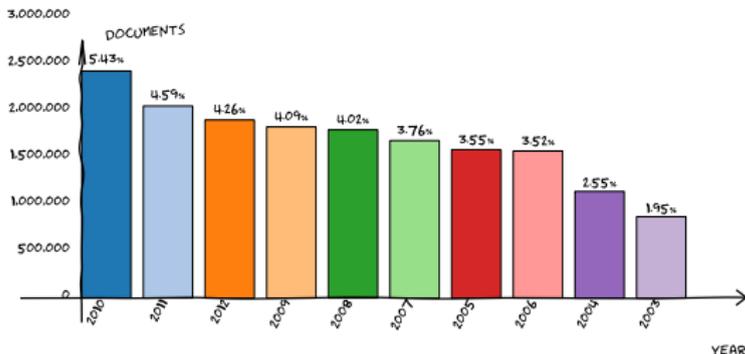
Voraussetzungen

Services

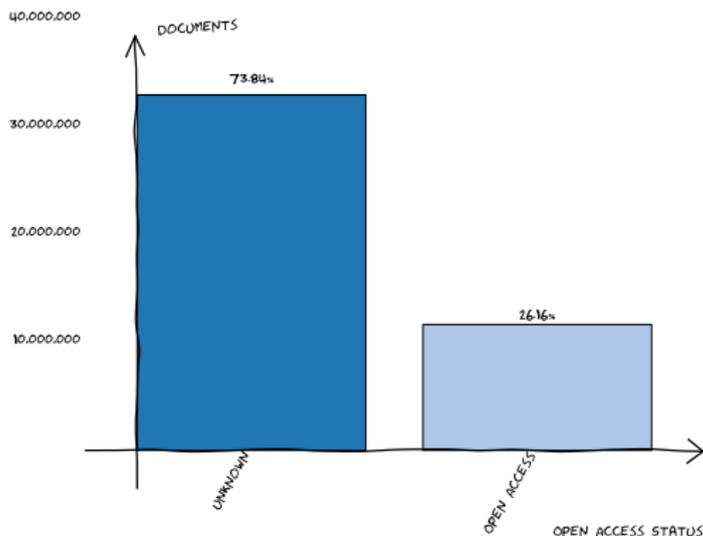
Ausblick

Zusammenfassung

Literatur

Top 10 Sprachen
(normalisiert)Top 10
Publikationsjahre

Open Access-Anteil



Statistik-Schnittstelle // Tabellen

Voraussetzungen

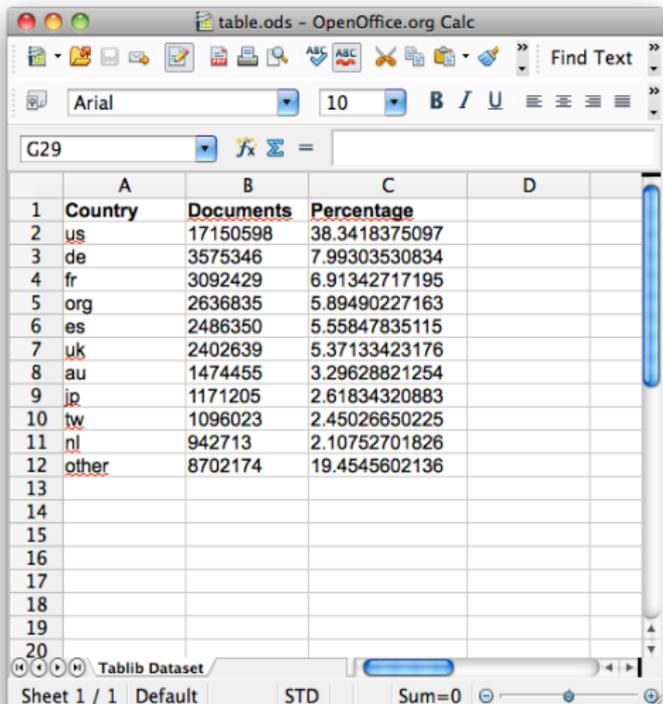
Services

Ausblick

Zusammenfassung

Literatur

Top Herkunftsländer
(OpenOffice Spreadsheet)



The screenshot shows an OpenOffice Calc window titled 'table.ods - OpenOffice.org Calc'. The spreadsheet displays a table with the following data:

	A	B	C	D
1	Country	Documents	Percentage	
2	us	17150598	38.3418375097	
3	de	3575346	7.99303530834	
4	fr	3092429	6.91342717195	
5	org	2636835	5.89490227163	
6	es	2486350	5.55847835115	
7	uk	2402639	5.37133423176	
8	au	1474455	3.29628821254	
9	jp	1171205	2.61834320883	
10	tw	1096023	2.45026650225	
11	nl	942713	2.10752701826	
12	other	8702174	19.4545602136	
13				
14				
15				
16				
17				
18				
19				
20				

The spreadsheet interface includes a toolbar with various icons, a 'Find Text' search box, and a status bar at the bottom showing 'Sheet 1 / 1', 'Default', 'STD', and 'Sum=0'.

Statistik-Schnittstelle // Datenformate

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

Grafikformate

- SVG
- PNG
- JPG
- TIF
- PostScript/EPS
- PDF

Tabellarische Formate

- CSV
- ODS
- Excel
- JSON
- YAML
- HTML

Agenda

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

1. Voraussetzungen

2. Services

- Suchschnittstelle
- Rsync-Schnittstelle
- OAI-Schnittstelle
- Statistik-Schnittstelle
- **Klassifikationsschnittstelle**

3. Ausblick

4. Zusammenfassung

Klassifikationsschnittstelle // <http://clfapi.base-search.net/doc/>

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

- Bereitstellung der Automatischen DDC-Klassifikation für externe Applikationen
- Nutzer:
 -  OA-Netzwerk (Kindling und Schirmbacher, 2013)
 - PUB (DDC-Vorschlagsfunktion)

Agenda

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

1. Voraussetzungen

2. Services

- Suchschnittstelle
- Rsync-Schnittstelle
- OAI-Schnittstelle
- Statistik-Schnittstelle
- Klassifikationsschnittstelle

3. Ausblick

4. Zusammenfassung

Mögliche zukünftige Services

- Linked Open Data
 - Vergabe von URIs für alle indexierten Dokumente
 - „BASE-Identifizier“?
- Identifier-Resolver
 - Ausgabe aller Identifier zu einer Publikation
 - Z.B.: arXiv:1105.0557 → doi:10.2140/pjm.2012.259.421
- Repository Registry
 - Exponierung der Metadaten über Datenquellen

Mögliche zukünftige Services

- Linked Open Data
 - Vergabe von URIs für alle indexierten Dokumente
 - „BASE-Identifizier“?
- Identifier-Resolver
 - Ausgabe aller Identifier zu einer Publikation
 - Z.B.: arXiv:1105.0557 → doi:10.2140/pjm.2012.259.421
- Repository Registry
 - Exponierung der Metadaten über Datenquellen

Mögliche zukünftige Services

- Linked Open Data
 - Vergabe von URIs für alle indexierten Dokumente
 - „BASE-Identifizier“?
- Identifier-Resolver
 - Ausgabe aller Identifier zu einer Publikation
 - Z.B.: arXiv:1105.0557 → doi:10.2140/pjm.2012.259.421
- Repository Registry
 - Exponierung der Metadaten über Datenquellen

Agenda

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

1. Voraussetzungen

2. Services

- Suchschnittstelle
- Rsync-Schnittstelle
- OAI-Schnittstelle
- Statistik-Schnittstelle
- Klassifikationsschnittstelle

3. Ausblick

4. Zusammenfassung

BASE-Netzwerk

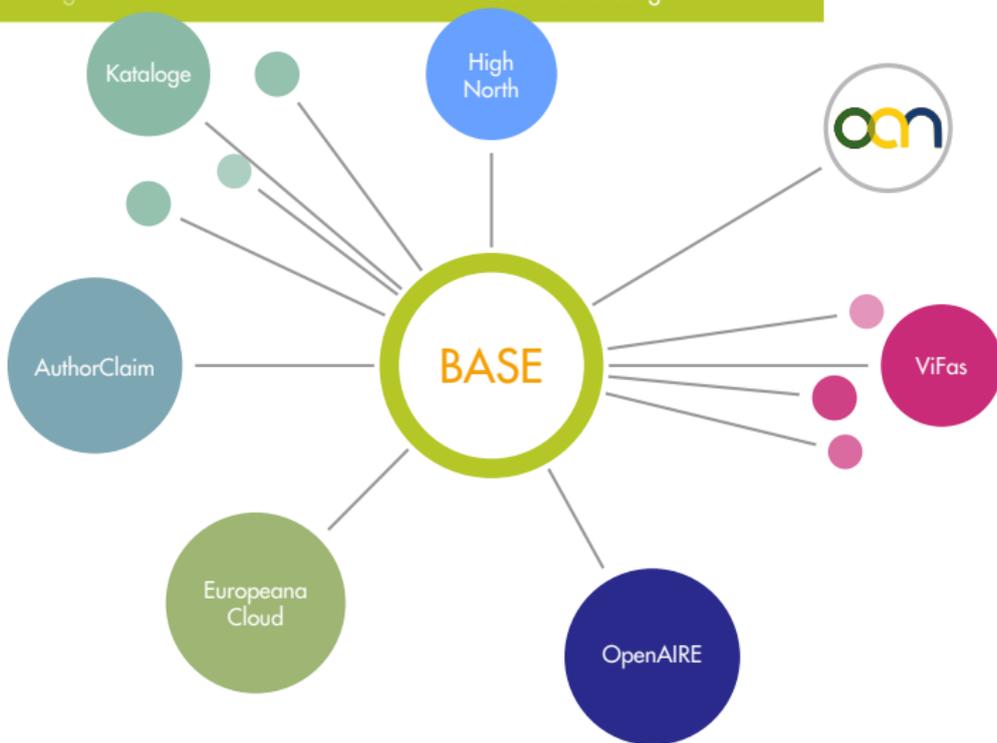
Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur



Vielen Dank für die Aufmerksamkeit!

Literatur

Voraussetzungen

Services

Ausblick

Zusammenfassung

Literatur

Kindling, M. und P. Schirnbacher (2013). Die digitale Forschungswelt als Gegenstand der Forschung. *Information – Wissenschaft & Praxis* 64(2-3), 127–136.

Vierkant, P., M. Voigt, J. Dupski, S. David und M. Lösch (2012). 2012 Census of Open Access Repositories. Poster presented at Open Access Tage, Vienna (AT), 26.–27.9.2012.