

Aufbau eines Systems für die  
Erschließung verteilter  
Dokumente in der Internationalen  
Pädagogischen Forschung

Teilprojekt der SUB Hamburg in  
Kooperation mit dem Institute for  
Science Networking Oldenburg

# Umfang des Teilprojektes

- Index für den Nachweis dezentral gehaltener digitaler Dokumente mit Metadaten
- Index für den Nachweis dezentral gehaltener digitaler Dokumente ohne Metadaten
- Recherche

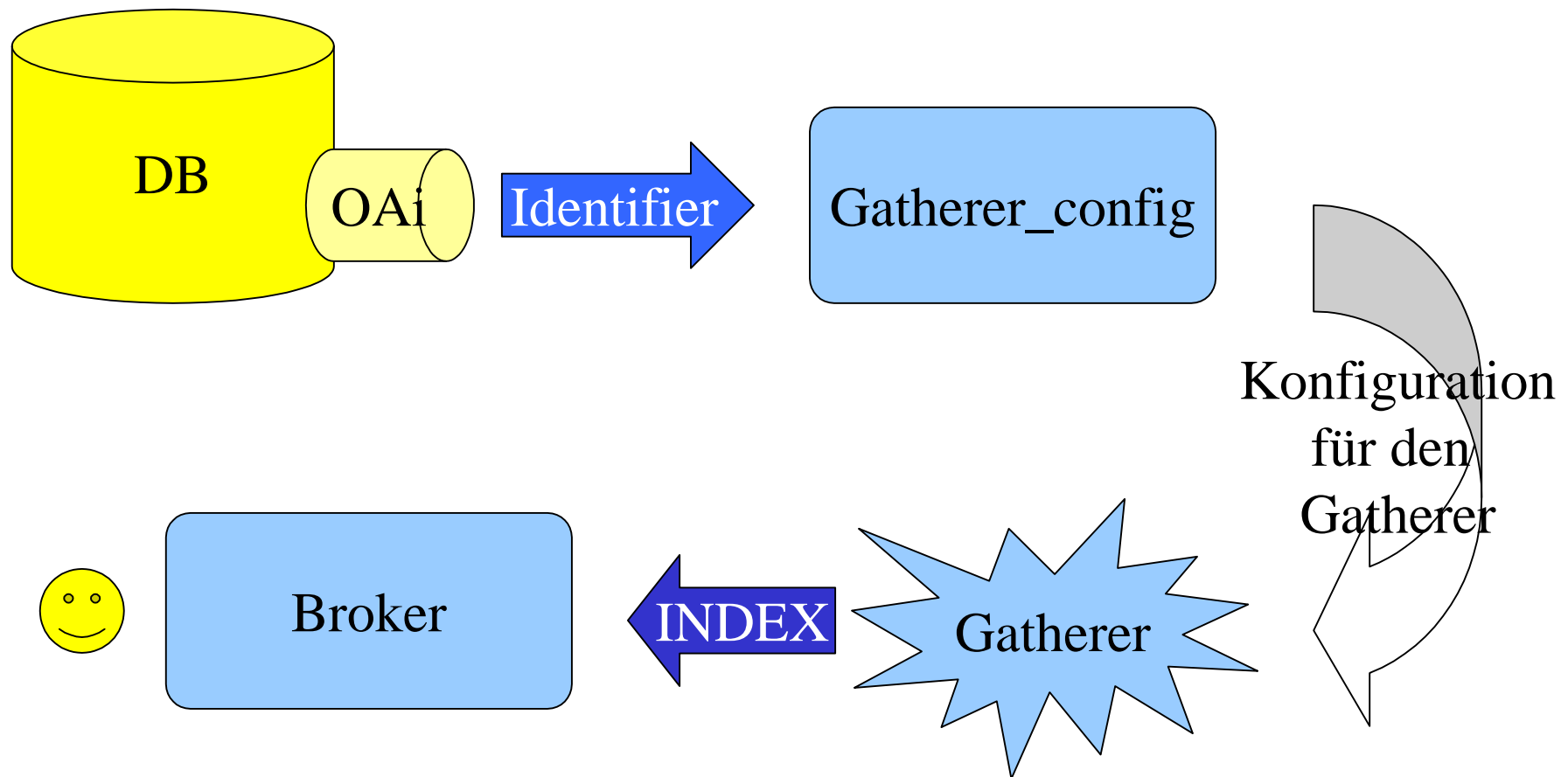
# Index dezentraler Dokumente mit Metadaten

- SUB (später das DIPF) stellt eine DB mit den Identifiern (=URL) und anderen Metadaten der nachzuweisenden Dokumente.
- Gatherer indexiert die Volltexte der Dokumente und reichert die Indexe um die Metadaten an.
- Gatherer indexiert zusätzlich (z.Zt. 10 Dokumente) „im Umfeld“ dieses Dokumentes im Volltext

# Index dezentraler Dokumente ohne Metadaten

- SUB (später das DIPF) stellt eine DB mit den Identifiern (=URL) der nachzuweisenden Dokumente.
- Gatherer indexiert die Volltexte der Dokumente und legt einen Index an.
- Gatherer indexiert weitere Dokumente „im Umfeld“ (=derzeit 200 weitere URLs auf dem gleichen Server).

# Technischer Ablauf ohne Metadaten



```
Quelle für: http://paedox.sub.uni-hamburg.de/_test/oai/oai2.php?verb=ListRecords&metadataPrefix=oai_dc - Netscape
Datei Bearbeiten Anzeigen Hilfe

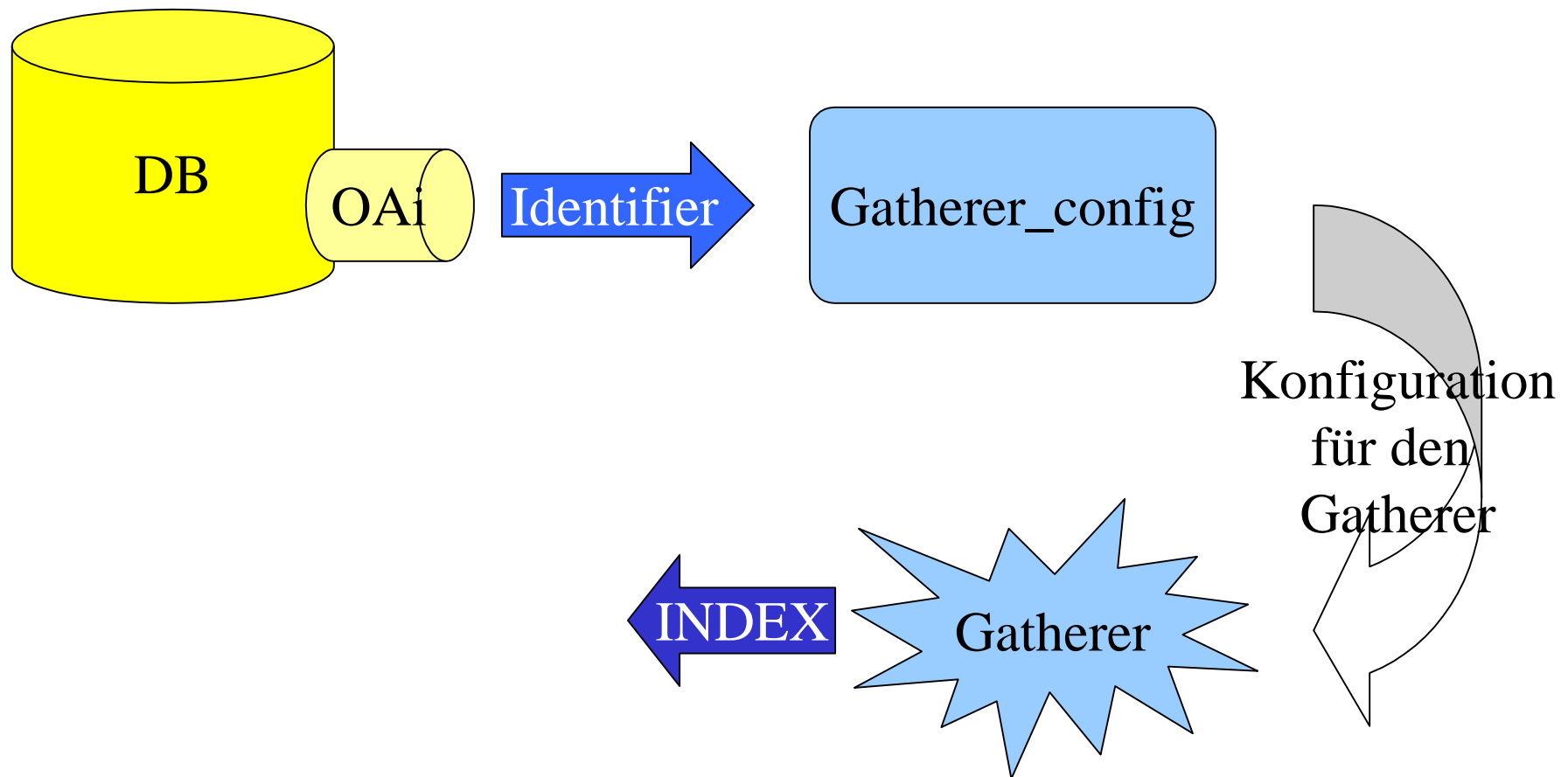
<record>
  <header>
    <identifier>oai:sub.uni-hamburg.de:http://www.sub.uni-hamburg.de/disse/745/titel.html</identifier>
    <timestamp>2002-01-01T00:00:00Z</timestamp>
    <setSpec>test1</setSpec>
  </header>
  <metadata>
    <oai_dc:dc
      xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
      xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
        http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
    </oai_dc:dc>
  </metadata>
</record>
<record>
  <header>
    <identifier>oai:sub.uni-hamburg.de:http://www.sub.uni-hamburg.de/disse/723/dissertation.pdf</identifier>
    <timestamp>2002-01-01T00:00:00Z</timestamp>
    <setSpec>test1</setSpec>
  </header>
  <metadata>
    <oai_dc:dc
      xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
      xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
        http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
      <dc:title>Einfluss von Gluconasturtiin und 2-Phenylethylisothiocyanat auf die Rhizosphärenmikroflora beim Raps (Brassic
      <dc:creator>Angelika Susanne Rumberger</dc:creator>
      <dc:description>Ziel der vorliegenden Arbeit war es, die Bedeutung des Senföles 2-Phenylethylisothiocyanat
      (PEITC) in dem Zusammenspiel von Boden, Mikroorganismen und Pflanze in der
      Rhizosphäre des Rapses zu untersuchen. Dazu wurden in Feld- und Gewächshausversuchen
      mit vier Rapsorten die zeitliche und räumliche Veränderung der PEITC-Konzentration in der
      Rhizosphäre untersucht. Parallel wurde die Artenzusammensetzung der pro- und
      eukaryotischen Rhizosphärenmikroflora mit der denaturierenden Gradienten
      Gelelektrophorese (DGGE) und Substratnutzungsmustern dargestellt. Die so gewonnenen
      Banden- und Substratnutzungsmuster wurden mit der Redundanz-Analyse (RDA) auf die
      PEITC-Konzentration der Rhizosphäre und andere Umweltparameter bezogen.
      Die PEITC-Konzentration in der Rhizosphäre ändert sich mit der Entwicklung der Pflanze.
      Bei Winterraps erreicht die PEITC-Konzentration jeweils vor Einbruch des Winters und zum
      Schossen ein Maximum. Bei Sommerraps wird die höchste PEITC-Konzentration während
      der Blüte erreicht. Bei Keimlingen und zur Reife ist die PEITC-Konzentration in der
```

## XML-Output der OAi- Schnittstelle

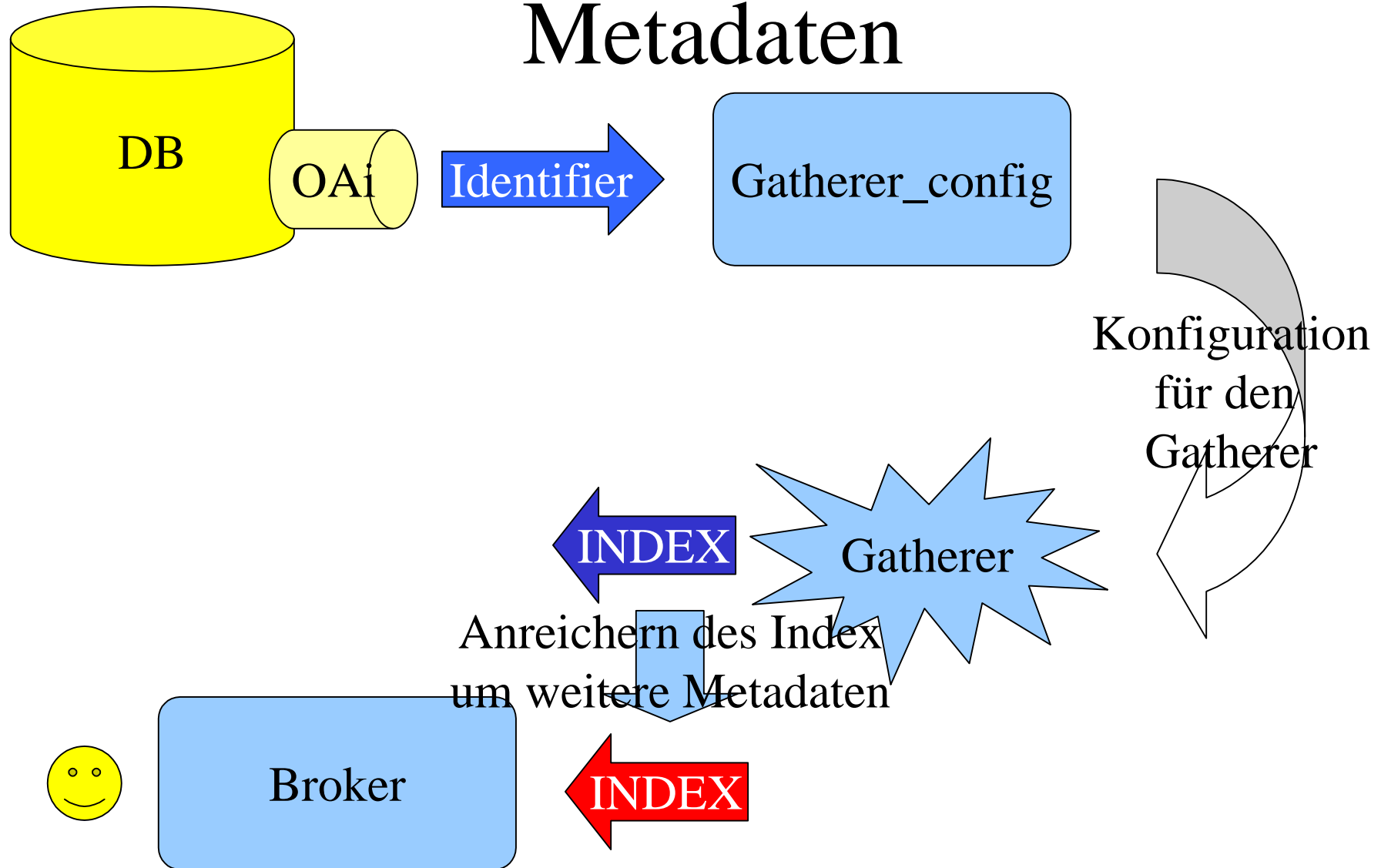
# Konfigurationsdatei des Gatherers

```
#  
# dipf-inst – configuration file for the collected information  
# about institutions listed in mysql database  
#  
Gatherer-Name: Collects information about institutions listed in mys  
Gatherer-Port: 8500  
Top-Directory: /home/harvest/Harvest/gatherers/dipf-inst  
Time-To-Live: 3888000  
Refresh-Rate: 86400  
<RootNodes>  
http://bildungplus.forumbildung.de/ URL=200 HOST=1  
http://bildungsforschung.bfz.de/ URL=200 HOST=1  
http://efh-bochum.de URL=200 HOST=1
```

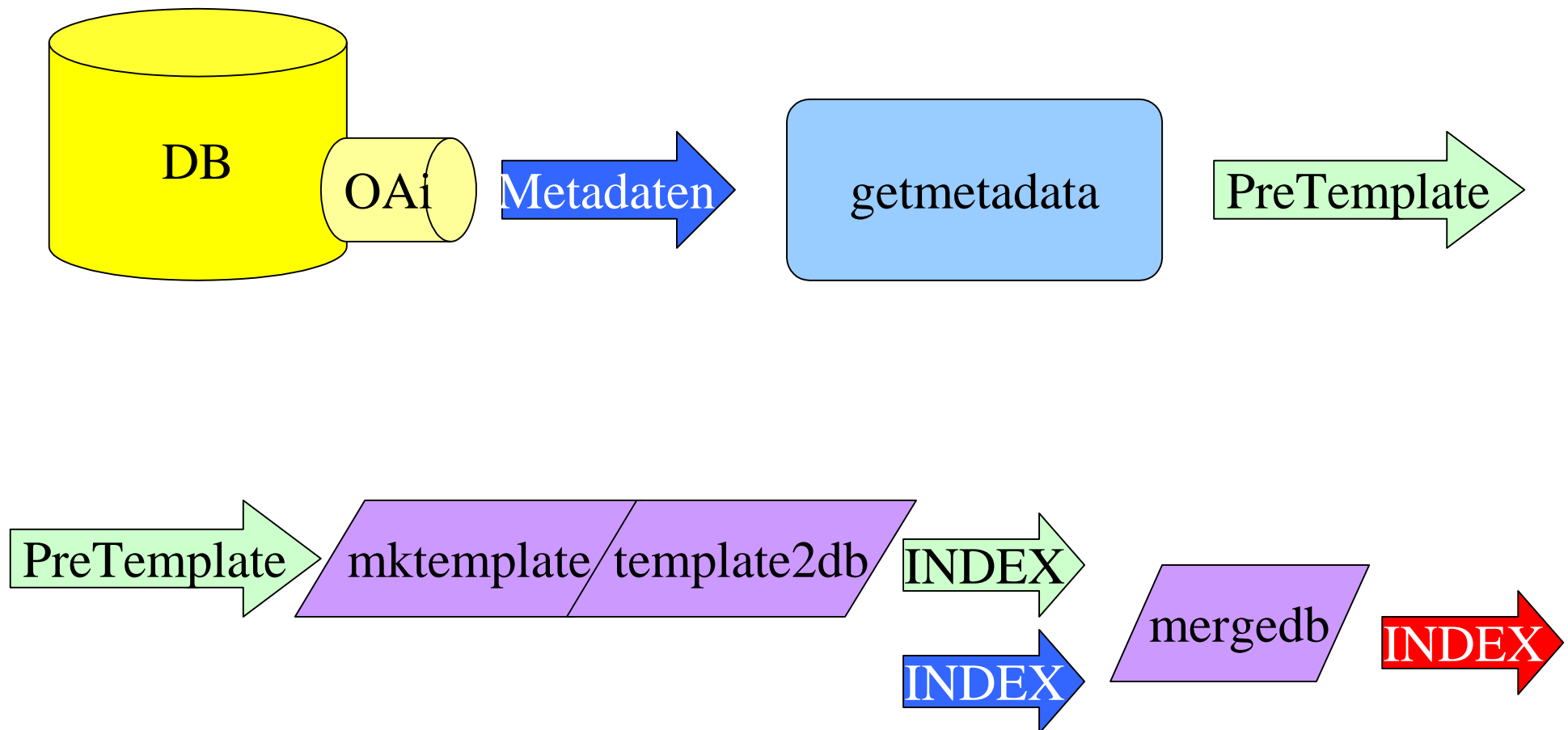
# Technischer Ablauf mit Metadaten



# Technischer Ablauf mit Metadaten



# Anreichern des Index



# Recherche

- Recherche über die Indexe
- Kurztitelanzeige
  - Ranking
  - Markierung der angereicherten Objekte
- Als Module zum Einbinden in andere Web-Seiten

Fulltext-Search

submit new query

Searchresults for:

**Fulltext: name**

5105 Results

- Title: **Pro46\_HEGISS.PDF**  
Author: Ihr **Name**  
URL: [http://www.inbas.com/projekte\\_pdf/pro46\\_hegiss.pdf](http://www.inbas.com/projekte_pdf/pro46_hegiss.pdf)  
Host: [www.inbas.com](http://www.inbas.com)  
Keywords: hegiss pdf pro  
Description: Projektname: HEGISS-Planungsdaten Foerderung durch: Stadt Offenbach
- Title: **MPIfG: Willkommen bei der EDV**  
Author: **Name** des Autors  
URL: <http://www.mpi-fg-koeln.mpg.de/edv/>  
Host: [www.mpi-fg-koeln.mpg.de](http://www.mpi-fg-koeln.mpg.de)  
Keywords: admin bruno burtscheidt dfn edv egger email esser fuer gesellschaftsforschung http institut koeln manuel markus max mpg mpi planck schueren schwarz services stichworte susanne sysadm win www  
Description: Willkommen bei der EDV, EDV, Max-Planck-Institut fuer
- Title: **Kommentiertes Vorlesungsverzeichnis SS 2002**  
Author: **Name**  
URL: <http://www.uni-koeln.de/hp-fak/psychiatrie/informationen/kommwvss03q.pdf>  
Host: [www.uni-koeln.de](http://www.uni-koeln.de)  
Keywords: kommentiertes vorlesungsverzeichnis  
Description: KOMMENTIERTES VORLESUNGSVERZEICHNIS SS 2003 VERANSTALTUNGEN VERTRE...
- Title: **Namenlos**  
URL: <http://www.fh-fulda.de/fb/sw/beratung/>  
Host: [www.fh-fulda.de](http://www.fh-fulda.de)  
Keywords: diesem informationen thema weitere  
Description: **Namen** ein Anliegen darzustellen und in den
- Title: **MPIfG: Newsletters & Mailings**  
Author: **Name** des Autors

## Recherche über die Institutionen (ohne Metadaten)

DIPF Searchresults - Netscape 6

name Fulltext-Search  
Author  
Title  
Keyword

submit new query Searchresults for:  
**Fulltext: name**

26 Results

- Title: **Dissertation Meike L. Schipper - Ergebnisse**  
URL: <http://www.sub.uni-hamburg.de/disse/745/ergebnisse.html>  
Host: [www.sub.uni-hamburg.de](http://www.sub.uni-hamburg.de)  
Keywords: inhalt weiter zurueck  
Description: Der 5' flankierende Bereich von Exon 1 (proximaler Promoter) wurde in sieben seriellen Deletionen amplifiziert, deren 5'-Ende immer naeher an die Transkriptionsstartstelle rueckte. Das 3'-Ende schloss in allen PCR-Produkten die Transkriptionsstartstelle ein, deren Position durch Primer Extensionsanalyse und 5' RACE bestimmt worden war (18). Die PCR-Produkte wurden in den pGL3-Basic Vektor kloniert, in dem sie die Expression des Luziferase-Reportergens steuerten. Die Vektorkonstrukte aus dem 5' flankierenden Bereich von Exon 1 erhielten die **Namen** LP1 bis LP7. Ausgehend von der Transkriptionsstartstelle ATG endeten LPI und LP2 an Position +92, LP3 bis LP7 an Position +24. Damit ergaben sich die folgenden seriellen Deletionsbereiche.
- Title: **Dissertation Meike L. Schipper - Material und Methoden**  
URL: <http://www.sub.uni-hamburg.de/disse/745/material.html>  
Host: [www.sub.uni-hamburg.de](http://www.sub.uni-hamburg.de)  
Keywords: inhalt weiter zurueck  
Description: zurueck
- Title: **Dissertation Meike L. Schipper - Diskussion**  
URL: <http://www.sub.uni-hamburg.de/disse/745/diskussion.html>  
Host: [www.sub.uni-hamburg.de](http://www.sub.uni-hamburg.de)  
Keywords: inhalt weiter zurueck  
Description: zurueck | Inhalt | weiter In
- \* Title: **Microsoft Word - DrArbeitEnd.doc**  
URL: <http://www.sub.uni-hamburg.de/disse/751/dissertation.pdf>  
Host: [www.sub.uni-hamburg.de](http://www.sub.uni-hamburg.de)

\* Markiert die Anreicherung

Dokument fertig (3.35 Sek.)

# Nächste Teilprojekte (Vorschläge)

- Begleitung der Software beim Überführen der DB ans DIPF:
  - Notwendige Anpassungen der Konfiguration
  - Tests
  - Bugfixes und Anpassung an evtl. geänderte Metadaten

# Nächste Teilprojekte (Vorschläge)

- Anpassung der Software an die Ergebnisse einer Nutzerevaluation
  - Strukturierte Recherche
  - Anfrage-Syntax
  - Ranking-Algorithmen
  - ...

Herzlichen Dank für Ihre  
Aufmerksamkeit

Eberhard R. Hilf,  
Thomas Severiens

Institute for Science Networking  
Oldenburg GmbH