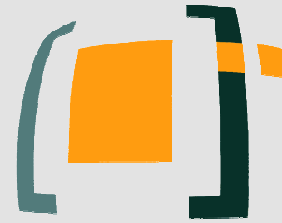




hbz

Wissen. Information. Innovation.



vascoda

Effizienzsteigerung durch Kooperation?

die vascoda- Suchmaschinenföderation

Ben Burkard (hbz)

98. Deutscher Bibliothekartag

02.-06.06.2009

Gliederung

- I. vascoda
- II. Warum Suchmaschinentechnologie?
- III. Zentrale vs. verteilte Indexierung
- IV. Technik der vascoda-Föderation
- V. Fazit & Ausblick

Gliederung

I. **vascoda**

II. Warum Suchmaschinentechnologie?

III. Zentrale vs. verteilte Indexierung

IV. Technik der vascoda-Föderation

V. Fazit & Ausblick

I. vascoda

- Interdisziplinäres Rechercheportal
- zunehmend auch Hintergrunddienste (z.B. Einbindung in lokale Angebote)
- Partner:
 - Virtuelle Fachbibliotheken / Fachportale
 - EZB, verschiedene Bibliotheksverbände
- Seit 2005 hbz als technischer Betreiber, bis Ende 2007 Förderung durch 2 BMBF-Projekte
 - Portalbetrieb und -weiterentwicklung
 - Einsatz von Suchmaschinentechnologie

Gliederung

I. vascoda

II. Warum Suchmaschinen- technologie?

III. Zentrale vs. verteilte Indexierung

IV. Technik der vascoda-Föderation

V. Fazit & Ausblick

II. Warum Suchmaschinentechnologie?

- Erster Ansatz im vascoda-Portal: klassische Metasuche
 - Lange Antwortzeiten
 - Unterteilung der Treffer in Blöcke
 - fehlende Aussagen zur Relevanz
 - ...
- Neuer Ansatz: Einsatz von Suchmaschinentechnologie
 - Kurze Antwortzeiten
 - Einheitliche Trefferliste
 - Drill-Downs
 - Linguistische Verfahren
 - Relevanz-Ranking
 - ...

Gliederung

- I. vascoda
- II. Warum Suchmaschinentechnologie?
- III. Zentrale vs. verteilte Indexierung**
- IV. Technik der vascoda-Föderation
- V. Fazit & Ausblick

III. Zentrale vs. verteilte Indexierung

- Frage in 2005: zentrale oder verteilte Indexierung?
- Vorteile zentraler Indexierung
 - Infrastrukturaufbau nur an einer Stelle notwendig
 - Hardware
 - Software
 - technisches Know-How (Betrieb Suchmaschine, Programmierung Konvertierungsroutinen etc.)
 - Bessere Kontrolle über das Ranking, linguistische Einstellungen etc.
 - Keine zusätzliche Komplexitätsebene für Verteilen von Anfragen und Zusammenführung von Treffermengen

III. Zentrale vs. verteilte Indexierung

- Nachteile zentraler Indexierung
 - Hoher Aufwand an einer Stelle konzentriert, ggf. Doppelarbeit, wenn ohnehin verteilte SM vorhanden
 - Lizenzrechtliche Bedingungen erlauben teilweise keine Weitergabe von Daten (so auch im vascoda-Kontext)



- Herausforderung in vascoda:
 - Verteilte Indexierung
 - Erhalt der Vorteile von Suchmaschinentechnologie
- Lösungsansatz: Föderierte Suche (= Metasuche über Suchmaschinen)

Gliederung

- I. vascoda
- II. Warum Suchmaschinentechnologie?
- III. Zentrale vs. verteilte Indexierung
- IV. Technik der vascoda-Föderation**
 - a. Allgemeines**
 - b. Die Anfragesprache
 - c. Das Antwortformat
 - d. Datenfluss (Anfrage/Antwort)
 - e. Beispiel-Anfrage im Portal
- V. Fazit & Ausblick

IV. Technik der vascoda-Föderation

(a) Allgemeines

- Zentrale Herausforderung: Kommunikation zwischen verteilten Suchmaschinen
- Klass. Metasuchprotokolle reichen nicht aus (z.B. für Transport von Drill-Down-Informationen)
- Entwicklung eines herstellerunabhängigen Protokolls
- realisiert als Web-Service
- Föderationsmitglieder
 - Gesis Bonn, hbz -> FAST ESP
 - TIB Hannover, ZPID Trier -> Lucene

Gliederung

- I. vascoda
- II. Warum Suchmaschinentechnologie?
- III. Zentrale vs. verteilte Indexierung
- IV. Technik der vascoda-Föderation**
 - a. Allgemeines
 - b. Die Anfragesprache**
 - c. Das Antwortformat
 - d. Datenfluss (Anfrage/Antwort)
 - e. Beispiel-Anfrage im Portal
- V. Fazit & Ausblick

IV. Technik der vascoda-Föderation

(b) Die Anfragesprache (ECQL)

- Entwicklung in Anlehnung an Contextual Query Language (CQL) der LoC
- Erweiterungen für Suchmaschinen
 - Extended CQL (ECQL)
 - Filter-Kontext: Einschränkung der Suche auf bestimmte Kollektionen (z.B. Fach-Cluster oder Fachportale)
 - Facet-Kontext: Eingrenzung der Treffermenge
 - Option-Kontext: für Steuerung der Suchmaschinen, z.B. linguistische Einstellungen

Gliederung

- I. vascoda
- II. Warum Suchmaschinentechnologie?
- III. Zentrale vs. verteilte Indexierung
- IV. Technik der vascoda-Föderation**
 - a. Allgemeines
 - b. Die Anfragesprache
 - c. Das Antwortformat**
 - d. Datenfluss (Anfrage/Antwort)
 - e. Beispiel-Anfrage im Portal
- V. Fazit & Ausblick

IV. Technik der vascoda-Föderation

(c) Das Antwortformat (ATOM)

- Basis = ATOM (RFC 4287)
- XML-Format für plattformunabhängigen Austausch von Informationen
- Mechanismus für universell eindeutige Treffer-IDs (Universally Unique Identifier)
- Integration von beliebigen Namensräumen möglich (in vascoda insbesondere Dublin Core)

Gliederung

I. vascoda

II. Warum Suchmaschinentechnologie?

III. Zentrale vs. verteilte Indexierung

IV. Technik der vascoda-Föderation

a. Allgemeines

b. Die Anfragesprache

c. Das Antwortformat

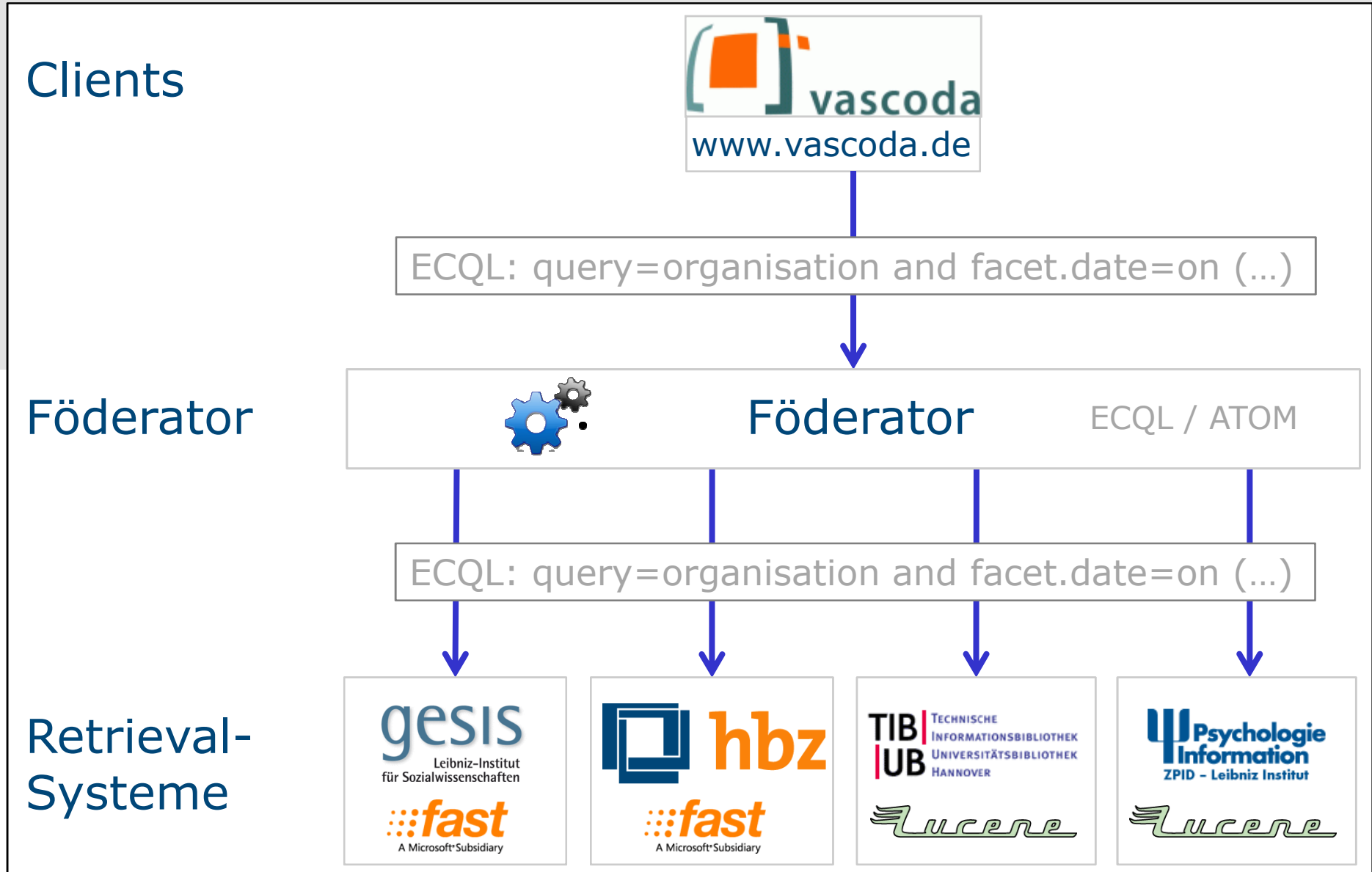
d. Datenfluss (Anfrage/Antwort)

e. Beispiel-Anfrage im Portal

V. Fazit & Ausblick

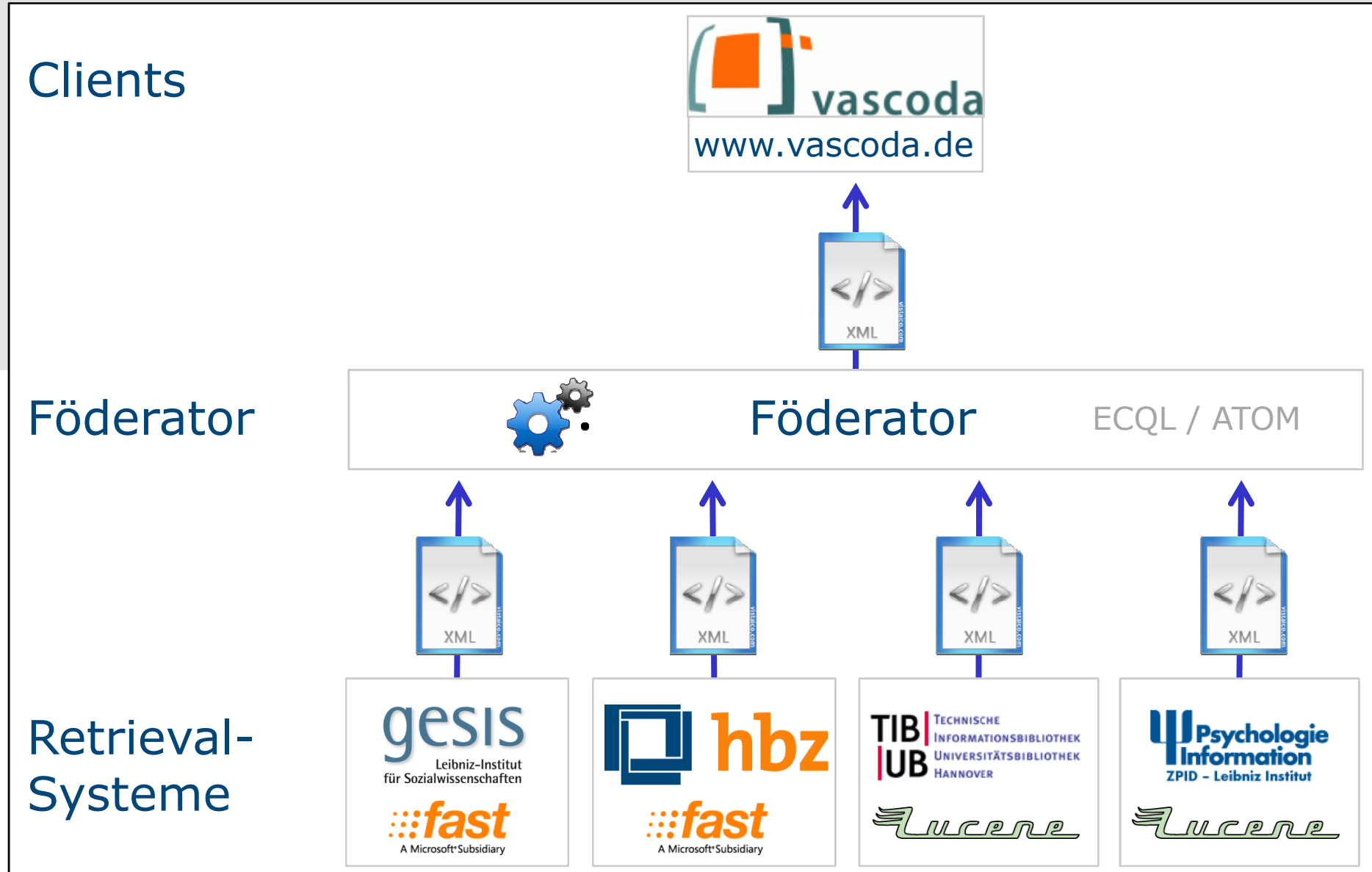
IV. Technik der vascoda-Föderation

(d) Datenfluss (Anfrage)



IV. Technik der vascoda-Föderation

(d) Datenfluss (Antwort)



Gliederung

- I. vascoda
- II. Warum Suchmaschinentechnologie?
- III. Zentrale vs. verteilte Indexierung
- IV. Technik der vascoda-Föderation**
 - a. Allgemeines
 - b. Die Anfragesprache
 - c. Das Antwortformat
 - d. Datenfluss (Anfrage/Antwort)
 - e. Beispiel-Anfrage im Portal**
- V. Fazit & Ausblick

IV. Technik der vascoda-Föderation

(e) Beispiel-Anfrage im Portal

The screenshot displays the vascoda portal search interface. At the top, the logo and tagline "Das Internetportal für wissenschaftliche Inform" are visible. A navigation bar includes "Startseite", "Suchen", "Fachzugänge", "Mein vascoda", and "Hilfe / FAQ". The search bar contains the text "organisation" and a "SUCHEN" button. Below the search bar, there is a checkbox for "nach verwandten Begriffen suchen". A blue arrow points from the search bar to the search results area.

The search results area shows a list of categories with checkboxes and expandable icons:

- Ingenieur- und Naturwissenschaften
- Medizin und Biowissenschaften
- Rechts-, Wirtschafts- und Sozialwissenschaften
- Geistes- und Kulturwissenschaften
- Regionen und Kulturräume
- Fächerübergreifende Datenbestände

On the right side, there is a sidebar with the following sections:

- Das ist vascoda**: A brief description of the portal's purpose as an entry point for scientific literature research.
- Aktuelles**: A section for news and updates, including links to the RSS feed, newsletter, and blog.

At the bottom of the page, there are links for "Impressum", "Kontakt", and "Über vascoda", along with a "Permanenter Link" and "vascoda empfehlen" button. The footer also mentions "Gefördert durch: Deutsche Forschungsgemeinschaft DFG".

IV. Technik der vascoda-Föderation

(e) Beispiel-Anfrage im Portal

The screenshot shows the vascoda portal search results for the query 'organisation'. The search results are displayed in a list format, with each result including a title, year, and the source Fachportal. Blue boxes and arrows highlight specific results and their corresponding Fachportal sources.

1. Building New Labour: The Politics of Party Organisation
2009
gefunden in: Fachportal: SOWI/PORT.de

2. Bericht Nr. 8 der Weltgesundheits-Organisation der Vereinten Nationen (WHO) zur AIDS-Verbreitung in Europa (Statistik)
31.12.1985
1987
gefunden in: Fachportal: Psychologie

3. „Enth. außerdem“ Organisation de la recherche dans le domaine du froid, commission 9 – Organisation of research in the domain of refrigeration)
unknown
1962
gefunden in: Fachportal: GetInfo

4. The organisation of capabilities
Loasby, Brian J.
gefunden in: Fachportal: EconBiz

5. ORGANISATION IN WIRTSCHAFT UND VERWALTUNG
Dworatschek, Sebastian ; Arbeitsgruppe Organisation ; Görg, Elisabeth
1977
gefunden in: Fachportal: Fachübergreifende Datenbestände

Gesis-Treffer
gefunden in: Fachportal: SOWI/PORT.de

ZPID-Treffer
gefunden in: Fachportal: Psychologie

TIB-Treffer
gefunden in: Fachportal: GetInfo

hbz-Treffer
gefunden in: Fachportal: EconBiz

Gliederung

- I. vascoda
- II. Warum Suchmaschinentechnologie?
- III. Zentrale vs. verteilte Indexierung
- IV. Technik der vascoda-Föderation
- V. Fazit & Ausblick**

V. Fazit & Ausblick

Effizienzsteigerung durch Kooperation?

→ Nutzerperspektive

→ + Vorteile gegenüber klass. Metasuche (z.B. einheitliche Trefferliste, Drill-Downs)

→ + Vorteile gegenüber zentraler Indexierung (lizenzrechtliche Bedingungen)

→ - Bisher unbefriedigendes Relevanz-Ranking („round-robin“)

→ Betreiberperspektive

→ + Teilen von Aufwand für Indexpflege, bes. wenn ohnehin verteilte Suchmaschinen vorhanden

→ - Overhead für Kommunikation zwischen Suchmaschinen

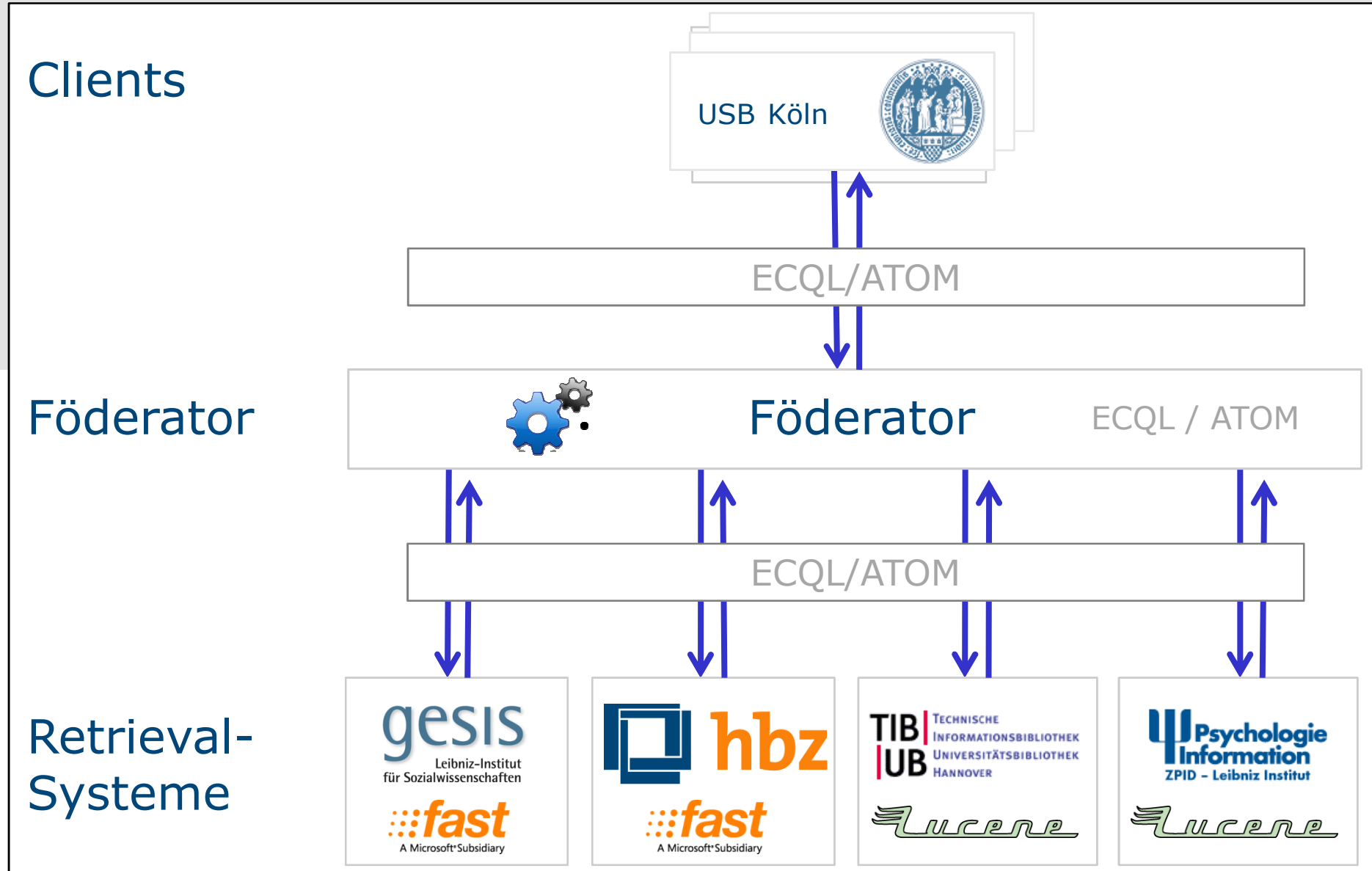
V. Fazit & Ausblick

Herausforderungen und weitere Schritte

- Optimierung des föderationsübergreifenden Relevanz-Rankings
- Nächster Schritt: Einbindung des vascoda-Suchraums (oder Ausschnitten davon) in lokale Angebote → ...

V. Fazit & Ausblick

Einbindung in lokale Angebote



→ Vielen Dank!

→ Kontakt:

- Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen (hbz)
- Gruppe Portale / vascoda
- Ben Burkard
- burkard <at> hbz-nrw.de