

Zur Integrität und Integration von elektronischen Publikationen – Modellüberlegungen im Umfeld der Digitalen Bibliothek

Philipp Mayr
GESIS / Informationszentrum Sozialwissenschaften, Bonn

Tagung der Gesellschaft für Wissenschaftsforschung 2007

23. März 2007



Agenda

1. Hintergrund

- Digitale Bibliotheken/Fachportale/vascoda
- Projekt Modellbildung/Heterogenitätsbehandlung
- Elektronische Publikationsformen

2. Elektronisches Publizieren (Kling/McKim, 1999)

- Dimensionen
 - Publizität
 - Zugänglichkeit
 - Vertrauenswürdigkeit
- Zusammenfassung

3. Integration von elektronischen Publikationen

4. Modelle

5. Fazit



Ausgangssituation

- Polyzentrische Informationsversorgung
- Trennung in kostenpflichtige und freizugängliche Materialien
- geändertes Informationsverhalten, siehe aktuelle Nutzerbefragungen/Studien*
 - Bündelung (Informationstypen)
 - Integrierte Suche
 - Direkter Zugriff (Volltext)
 - Interdisziplinärität (Fachcluster)
 - Relevante, qualitative Dokumente

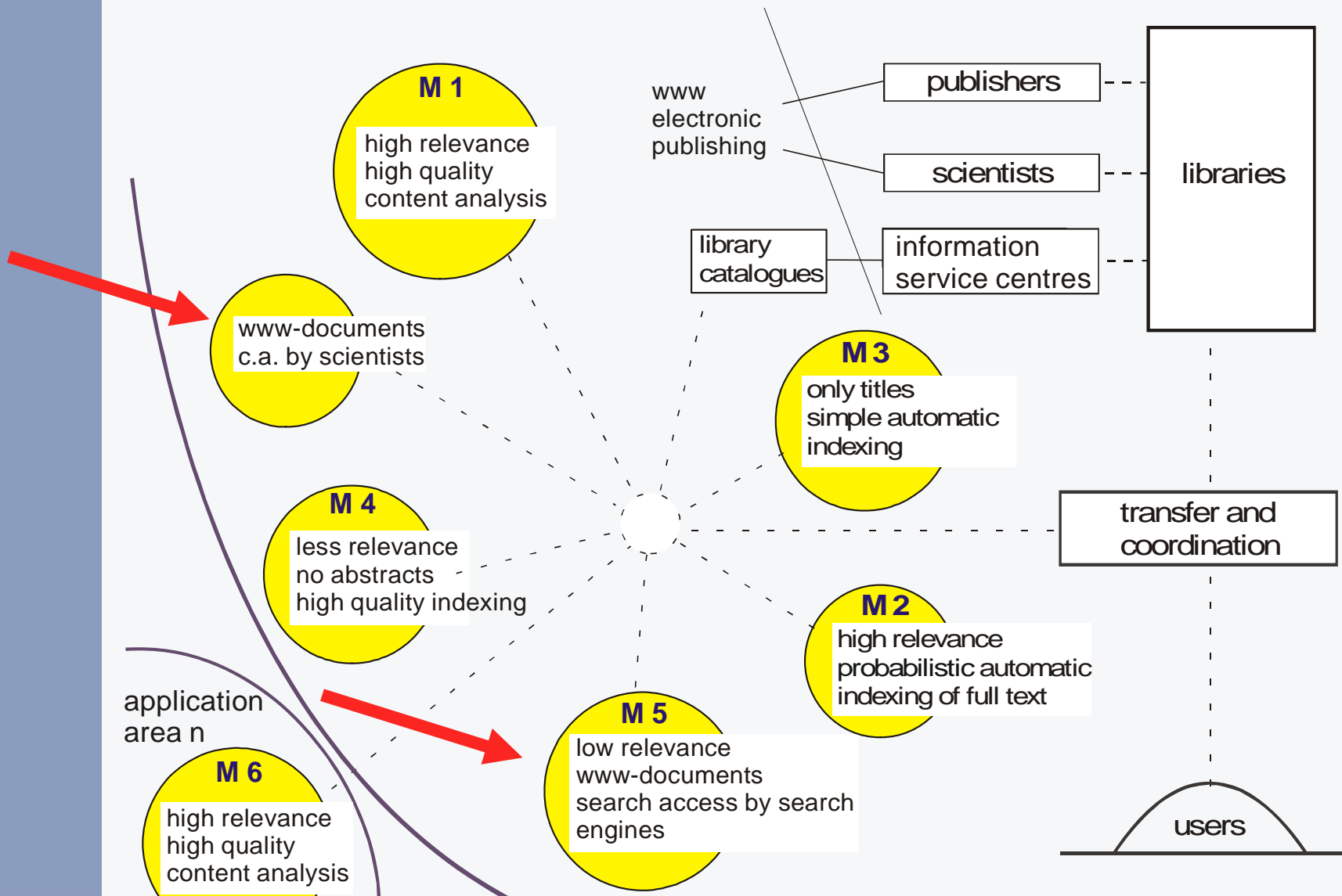
→ Agenda für Digital Library Research

* Quellen:

Boekhorst et al. 2003, Poll 2004, IMAC 2002, RSLG 2002, Binder et al. 2001, Stahl et al. 1998, zu WWW-Suchmaschinen: Machill & Welp 2003

Polyzentrischer Dokumentenraum

J. Krause





Aktuelle Entwicklungen

1. Open Access (OA) als alternative Publikationsform
 - steigende Zahl und Akzeptanz OA-Zeitschriften und Eprints-Archiven
 - Tendenz zu mehr verfügbaren Volltexten (u.a. durch self-archiving)
2. Anwendung von Suchmaschinen-Technologie auf wissenschaftlichen Dokumentenraum
 - Beispiele in Deutschland: BASE, Dreiländerkatalog, vascoda
 - Google Scholar
 - Scirus
 - ...



OA-Studien (Auswahl)

- Lawrence, 2001: “The mean number of citations to offline articles is 2.74, and the mean number of citations to online articles is 7.03, an increase of 157%.”
- Brody et al., 2004: “dramatic citation advantages for OA (...) even larger effects than those reported by Lawrence”
- Swan/Brown, 2005: “reach larger audiences, publish more rapidly and are more prestigious than the toll-access”
- Eysenbach, 2006: “has the potential to accelerate recognition and dissemination of research findings, but its actual effects are controversial (...) OA article on the journal site have higher impact than self-archived (...) OA articles are more immediately recognized and cited by peers than non-OA articles ”
- Moed, 2006: “evidence of a strong quality bias and early view effect. no sign of a general ‘open access advantage’ (...) ArXiv accelerates citation, due to the fact that that ArXiv makes papers earlier available rather than that it makes papers freely available “



Fragen aus Perspektive der Digitalen Bibliothek

Wann kann ein elektronisches Angebot beispielsweise als wissenschaftlich integer bezeichnet werden?

Unter welchen Bedingungen können derartige Publikationen in eine Digitale Bibliothek integriert werden?

Kann die Integration erst nach vorheriger Integritätsprüfung erfolgen?

Kling, Rob; McKim, Geoffrey (1999): Scholarly Communication and the Continuum of Electronic Publishing. In: JASIS 50, Nr. 10, S. 890-906

- analytical approach for evaluating disciplinary conventions and for proposing policies about scholarly e-publishing ...
- to help conceptually clarify the relationships between various forms of paper and electronic publication ...

[Hervorhebungen durch den Autor]



Electronic publications

“We define an electronic publication as a document distributed primarily through electronic media. (...)”

According to this definition, an article posted on a Web page (under a variety of restrictions or conditions), an article distributed via e-mail, or via an e-mail-based distribution list are all electronic publications.” S. 891

“We refer to articles made publicly and electronically available in nonpeer-reviewed form, either as posted on an individual or organizational Web page, or on a server such as the Los Alamos Physics E-Print Archive (<http://www.arxiv.org/>), as electronic working articles.”



Electronic publications II

- Print & electronic journals (hybrid publications): e.g. JASIS
 - p-e: print is the main distribution form
 - e-p: limited distribution in paper form
- E-only journals: e.g. D-Lib Magazine
 - Peer reviewed e-journals: e.g. PLoS Biology
- Pre- und Postprint Server/Repository: e.g. ArXiv
 - Institutional
 - Subject-based

“We define an e-journal as an edited package of articles that is distributed to most of its subscribers in electronic form (Kling & Covi, 1995).”



E-Journals

“Articles from an e-journal may, and probably will, be printed for careful reading, and might be stored in libraries in a printed form, for archival purposes.”

“electronic publishing is dramatically less expensive than paper publishing; access to electronic publications is easier and wider; and electronic publishing can speed up scientific communication.”

“... they publish so few articles each year as to have minimal impact on scientific communication systems.”



Internet posting practices (APA policy 1996)

- 1) on a password protected Web page in order to share with coauthors;
- 2) to a private departmental e-mail discussion list for comments;
- 3) to a public e-mail discussion list for comments and feedback;
- 4) on a personal home page;
- 5) on an institutionally maintained working article or preprint server;
- 6) by a peer-reviewed electronic journal with no paper counterpart.

“But a document’s availability on the Web does not mean that it will be widely read by the relevant community.”

Internet posting practices

- 1) Weblog
- 2) Wiki
- 3) Web 2.0

The screenshot shows two browser windows. The left window displays a weblog entry from 16.3.2007 with the title "Chemical Abstracts, ein Jahrhundert alt" and another entry titled "Library 2.0, unter Beschuss". The right window shows the Wikipedia article "Digitale Bibliothek". The article text includes: "Dieser Artikel beschäftigt sich mit der digitalen Bibliothek als Sammlung digitaler Texte. Für die CD-ROM-Reihe Digitale Bibliothek sind Bedeutungen siehe Digitale Bibliothek (Begriffsklärung). Die Artikel **Digitale Bibliothek** und **Virtuelle Bibliothek** überschneiden sich thematisch. Hilf mit, die Artikel besser voneinander Diskussion über diese Überschneidungen findet hier statt. Bitte äußere dich dort, bevor du den Baustein entfernst." Below the article text, there are sections for "Möglichkeiten von digitalen Bibliotheken", "Volltextsuche", "Informationssammlung", and "Recherche und Bestellung".

The screenshot shows a del.icio.us search results page for the query "digitale bibliothek". The page header includes "del.icio.us search for 'digitale bibliothek' - Microsoft Internet Explorer". The search results are displayed as a list of links with associated tags and the number of people who saved the link. The results include:

- DigBib.Org - die freie digitale Bibliothek (saved by 64 people)
- KOBV Digitale Bibliothek Berlin-Brandenburg - Suche (saved by 13 people)
- Digitale Bibliothek (saved by 9 people)
- Digitale Bibliothek (saved by 6 people)
- Digitale Bibliothek - Münchener Digitalisierungszentrum (saved by 6 people)
- Fundgrube Internet - Digitale Bibliothek (saved by 5 people)
- DigiZeitschriften - Das deutsche digitale Zeitschriftenarchiv: DigiZeitschriften (saved by 27 people)
- DigBib.Org - die freie digitale Bibliothek (saved by 5 people)
- CESG - Codices Electronici Sangallenses (saved by 18 people)



When is a document effectively published?

When a scholarly document is effectively published within a scholarly community, it seems to satisfy three criteria:

- publicity,
- access,
- and trustworthiness.



Drei Dimensionen wissenschaftlichen Publizierens I

- Publizität (Publicity): The document is announced to scholars so that primary audiences and secondary audiences may learn of its existence. Publicity represents a continuum of activities from subscription, report lists, abstract databases, advertising and special issues, and citation.



Drei Dimensionen wissenschaftlichen Publizierens II

- Vertrauenswürdigkeit (Trustworthiness): The document has been vetted through some social processes that assure readers that they can place a high level of trust in the content of the document based on community-specific norms. Trustworthiness is typically marked by peer review, publishing house/journal quality, and sponsorship.



Drei Dimensionen wissenschaftlichen Publizierens III

- Zugänglichkeit (Accessibility): Readers must be able to access the document independent of the author, and, in a stable manner, over time. Accessibility is typically assured by institutional stewardship as practiced by libraries, publishing houses, clearinghouses, and is supported by stable identifiers, such as ISBN and ISSN.



Self-publishing on the Web

e.g. personal home page; no additional announcement

Publicity: little publicity, can change over time (e.g. search engines, other linkages)

Accessibility: can change over time (document formats, address change, people move)

Trustworthiness: depends almost entirely upon author's reputation within a community



Publishing in an E-journal

e.g. a document being published in a pure e-journal

Publicity: greater publicity (e.g. systematic subscriber list, or indexed by abstracting services)

Accessibility: depends upon stewardship of the journal, including their long-term funding, some redundancy and institutionalized stewardship

Trustworthiness: depends on the journal's quality, reputation (e.g. PLoS Biology impact factor is 14.7)



Publishing in ND LTD (edoc-Server equivalent)

e.g. publishing a dissertation as part of the Networked Digital Library of Theses and Dissertations (ND LTD)

Publicity: fairly weak, rarely publicized in research communities

Accessibility: potentially stronger, institutionalized and stable over time

Trustworthiness: is certainly met, because determined by the evaluation procedures of the university



Print publication practices

a) High-Impact journal; b) minor or regional journal

Publicity:

a) high;

b) decreased, depending on subscribership to journals

Accessibility:

a) high;

b) often very low (not available at most libraries)

Trustworthiness:

a) high, but still depends on reputation of author;

b) high to more peripheral participants



Zusammenfassung Kling/McKim, 1999

This analysis shows how the common claim that e-publishing “substantially expands access” is oversimplified.

Web posting and effective scholarly communication are loosely coupled rather than strongly related.

... that scholarly communication can be conceptualized as a communicative practice anchored in three dimensions: publicity, access, and trustworthiness.

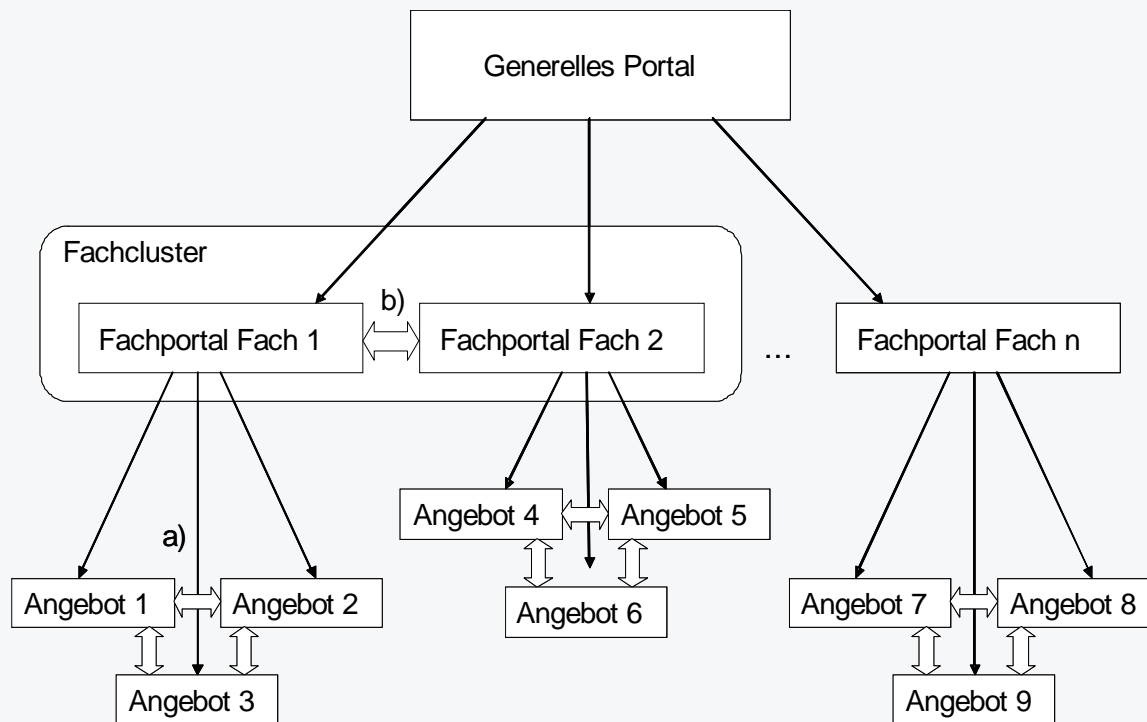
Teilprojekt innerhalb des Kompetenznetzwerks „Neue Dienste, Standardisierung, Metadaten“ (BMBF)

- Modellbildung: übergreifende Modellbildung für komplexe Informationsinfrastrukturen, u. a. am Beispiel des Wissenschaftsportals vascoda mit allen nachgeschalteten Ebenen
 - Heterogenitätsbehandlung: als notwendige Ergänzung zur Standardisierung durch einheitliche Metadaten
- ➔ Vorschläge (Modelle) zur Strukturierung wissenschaftlicher Informationsangebote
- ➔ Schaffung von Rahmenbedingungen und Infrastrukturen zur Förderung der semantischen Integration heterogener Informationssammlungen

Projektlaufzeit September 2004 bis August 2007

Leitlinien vascoda

- Integration: alle wichtigen Informationsanbieter und Datenkollektionen (Ausgangspunkt “deep web”)
- Einsatz kontrollierter Vokabulare
- Behandlung semantischer Heterogenität
- Fachcluster/Fachportale als Basis von vascoda
- Antwort auf Google/Google Scholar



Kaskadierendes
Modell einer
Portalinfrastruktur
(Mayr et al., 2005)

Was sind Crosskonkordanzen?

Relationen

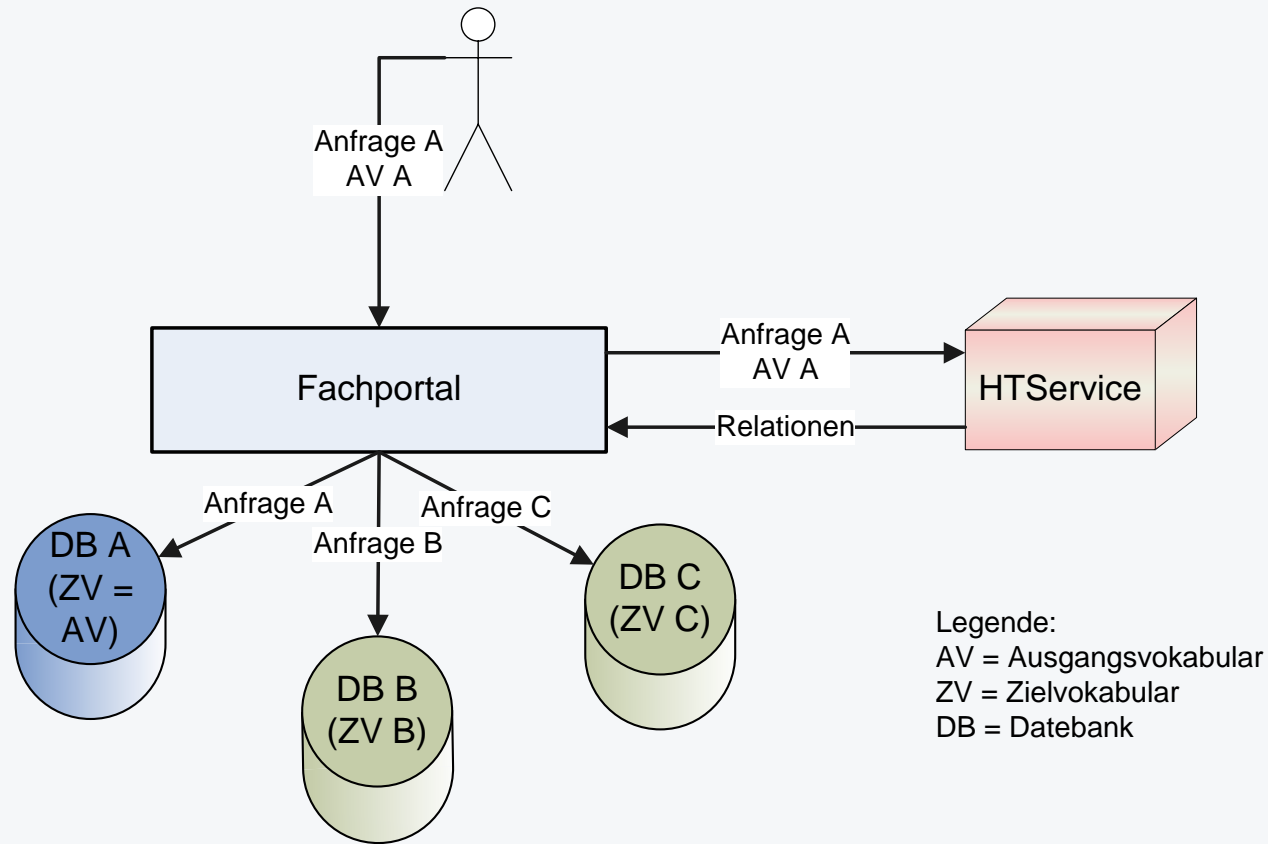
Ausgangsterm	Relation	Relevanz	Zielterm
Schifffahrtsrecht	=+o (Äquivalenz)	m	Schiffahrt + Verkehrsrecht
Schifffahrtsrecht	<o (Oberbegriff)	g	Verkehrsrecht
Schifffahrtsrecht	^o (verwandt)	m	Seerecht
Regionalstruktur	> (Unterbegriff)	m	Regionale Wirtschaftsstruktur

Relevanzen bezogen auf die Treffermenge und Dokumentrelevanz

1:1- oder 1:n-Verknüpfungen

Biologieunterricht	<	Unterricht	DZI
Biologieunterricht	<	Unterricht	Standard Thesaurus Wirtschaft
Biologieunterricht	=	Biologieunterricht	Schlagwortnormdatei
Biologieunterricht	<+	Biology + Teaching	CSA
Biologieunterricht	=+	Naturwissenschaftlicher Unterricht + Biologie	Psyndex Terms
Biologieunterricht	=+	Fachunterricht/Unterrichtsfach + Biologie	IBLK
Biologieunterricht	=+o	Biologie + Schulfach	BISp-Liste
Biologieunterricht	<+o	Biologie + Unterrichtsstunde	BISp-Liste
Biologieunterricht	<+	Biologie + Schule	DZA
Biologieunterricht	^+	Biologie + Unterricht	FES

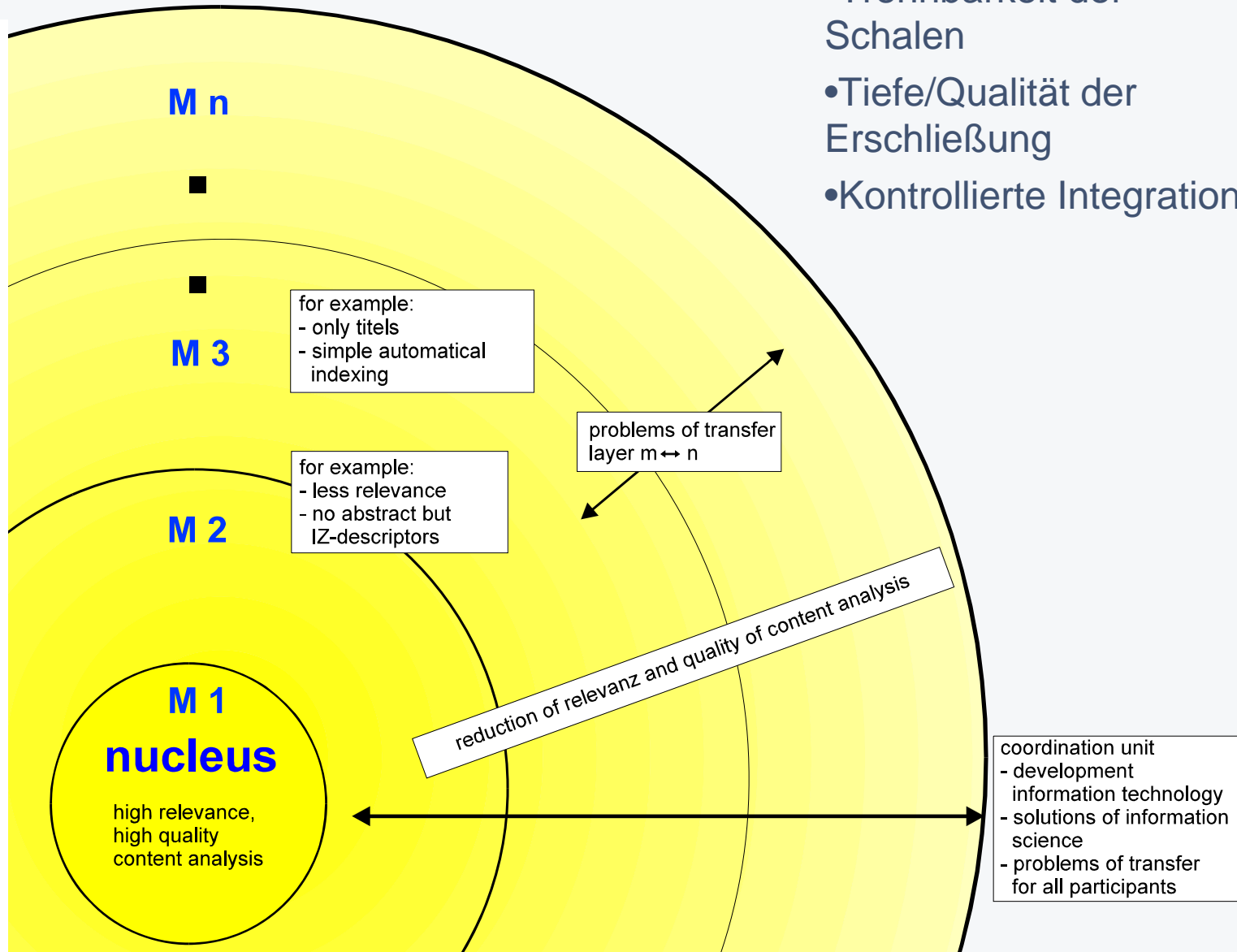
Einsatzszenario: automatische Transformation der Anfrage



- Fachportal fragt HTService nach Termtransformationen
- Danach erfolgt Abfrage der Datenbanken

Schalenmodell

J. Krause



- Trennbarkeit der Schalen
- Tiefe/Qualität der Erschließung
- Kontrollierte Integration



Empfehlung Modellbildung

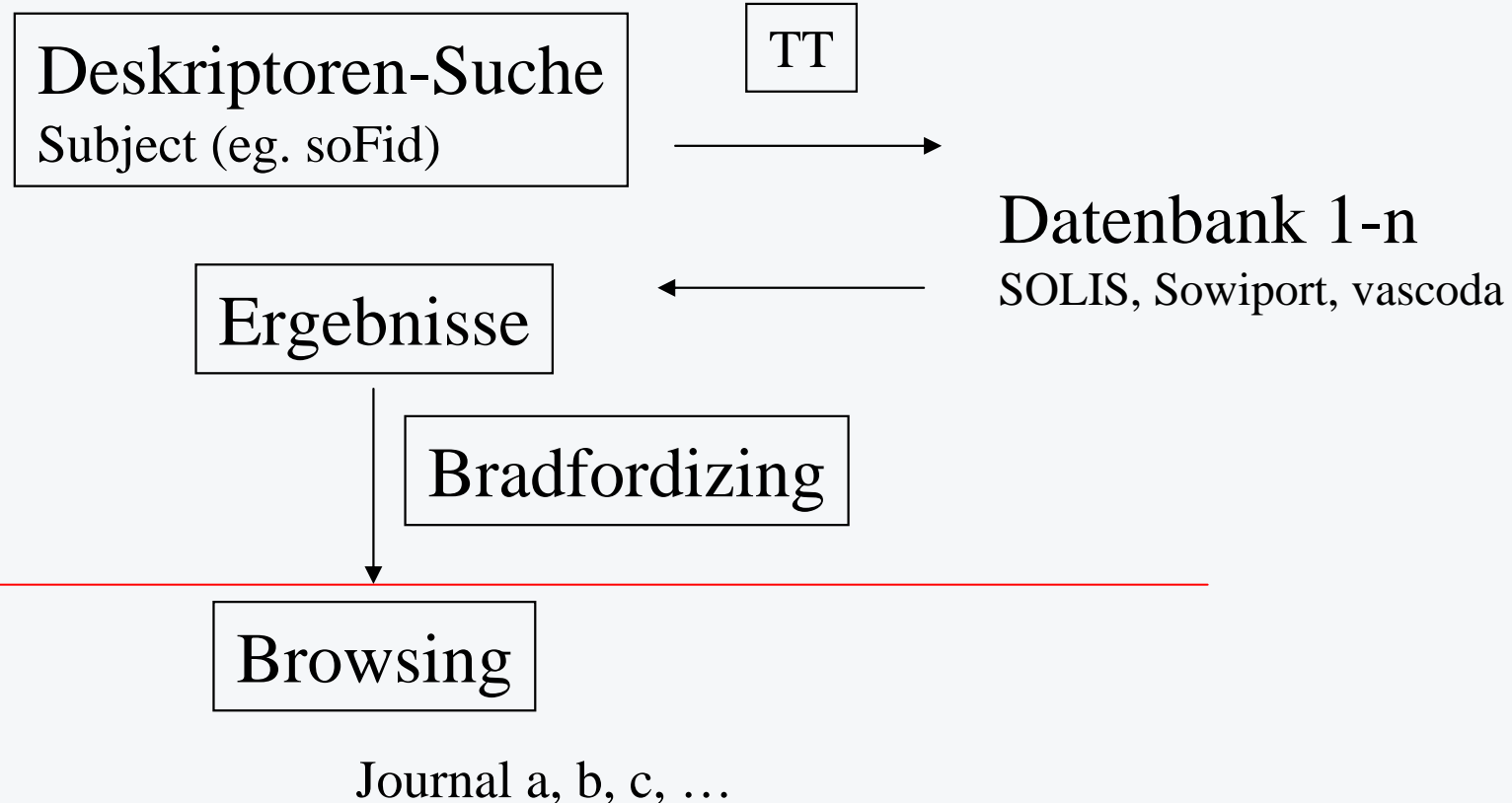
Schalenmodell + Heterogenitätsbehandlung

- kontrollierte Integration
- Trennbarkeit der Schalen
- semantische Heterogenitätsbehandlung (kontrollierte Vokabulare)
- Mehrwertdienste
 - Autorennetzwerke
 - Search Term Recommender
 - Bradford Zonen (Kernzonen-Generierung)
- Anschluss an das Semantic Web

Thesen

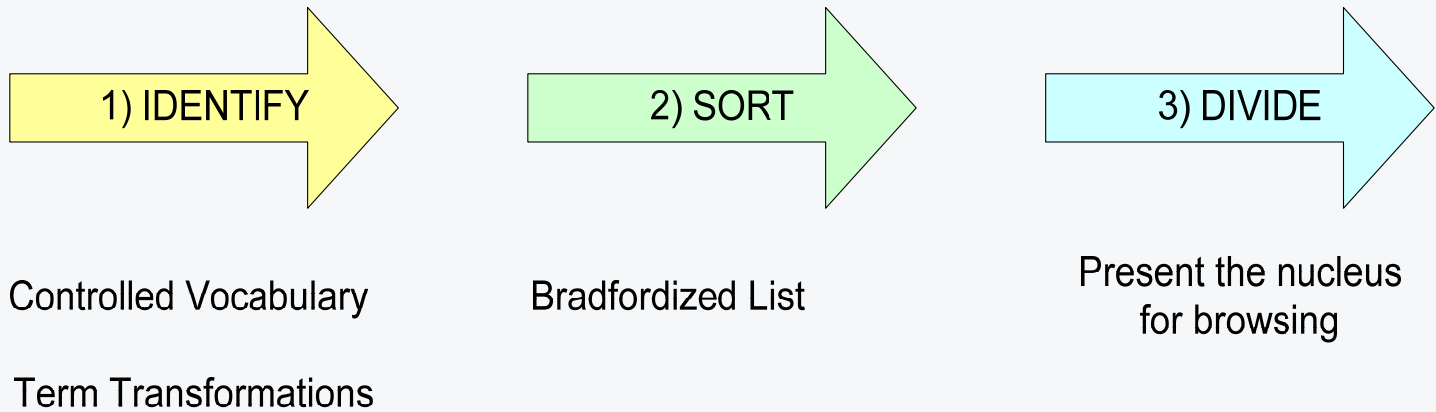
- Kernzone/Kernjournals (hohe Artikelhäufigkeit) eines Subject haben für den Nutzer positive Auswirkungen beim IR (vgl. Berry Picking Model von Bates 1989, 2002)
- Publikationen der Kernzone sind i.d.R. relevanter als Publikationen aus den weiteren Zonen
- TT erhöhen die Interdisziplinarität der Treffermenge
- Überstieg vom Suchtyp Deskriptorensuche zu Browsing kann anhand des BLS automatisch erfolgen

Vorgehensweise Ansatz Browsing



Überstieg von herkömmlicher Deskriptoren-Suche in den Browsing-Modus

Vorgehensweise Suchüberstieg (Bradford)





Fazit

Keine Integritätsprüfung im Sinne wissenschaftlicher Korrektheit aus der Perspektive der Digitalen Bibliothek, aber Eingliederung vertrauenswürdiger Angebote

- Digitale Bibliothek als Publikationsort (z.B. sowiport/Zeitschrift FQS)
- Integration von peer-reviewed OA-Journals über pull-Mechanismus auf der Fachportal-Ebene

Elektronischen Publikationen fehlt es häufig noch an Publizität und Vertrauenswürdigkeit, beides kann durch kontrollierte Integration (Modellbildung) in digitalen Bibliotheken erreicht werden

Bradford Law als ein mögliches Modell

Vielen Dank für die Aufmerksamkeit!

Philipp Mayr

Informationszentrum Sozialwissenschaften (IZ)

Abt. Forschung und Entwicklung

Lennéstr. 30

53113 Bonn

Tel. 0228 / 22 81 - 0

email mayr@bonn.iz-soz.de

<http://www.gesis.org/IZ>