Technical and less technical success factors for open access

perspectives from a public service provider for academic libraries

Institutional Background (1/2)

hbz - centre for academic libraries in NRW

- ➔ NRW: state with the largest population in Germany
  over 20% of Germany, nearly 4% of Europe
- ➔ operational range of the Ministry of Science NRW
- ➔ dozens of academic institutions served
  - universities, universities of applied sciences and others in NRW and beyond (DigiBib, Vascoda)

Institutional Background (2/2)

hbz - services

- digital libraries ("DigiBib")
- distance document delivery
- administration of consortial acquisition of subscriptions (e.g. GASCO licenses for BMC)
- implementations of OAI compliant DMS at academic institutions
- operator of the open access initiative for eJournals "Digital Peer Publishing" DiPP
- implementation of OAI-protocols in general MeInD

hbz

Disclaimer

- hbz activities broader than OA
  - focus here
    "the challenge of OA research papers"
- practical perspective
  - personal observations (scholars | librarians)
- definitions
  - scholar ~ scholars, scientists of each discipline
  - library ~ institutional information services
    (incl. computing centres, university publishers, university administration etc.)

hbz

Focus

- research papers
  - most important document type in many scientific areas
  - domain with the most severe access restrictions

- But! Monographs, thesis or other research results (datasets, programs, simulations etc.) are still important on the long run

- *Application scenario:* scholar@internet
  - searching publications for retrieval and referencing
  - publishing his or her scholarly work on the internet

hbz

OA use cases

- Publishing
  - Pre-print → Quality Assurance → Formatting → Post-Print (→ Publisher's version)
- Indexing and Distribution
- Search and Retrieval
- Application
- Impact

☝ institutional services (e.g. libraries) can provide information systems infrastructure and support

hbz

Status quo

- → Conceptual, legal & technical framework is given
- → But! Insufficient quantity of high quality content
  - → in institutional OA archives as well as OA eJournals
- → Explanation #1: "extra work and risk"
- → Explanation #2: "prestige paradox" [Crow and Goldstein]
  - → OA eJournals: second choice
  - → institutional archives: added value not understood
- → Explanation #3: "blackboard problem" (?)
  - → criteria for scholars: ratio between workload and added value
    - → *reputation leverage*
  - → criteria for libraries: ratio between workload and prices
    - → *economic leverage*

hbz

Can we quantify and specify the success of OA?

- → Working hypotheses
  - → success ~ quantity of high quality content
  - → quantity ~ number of publications
  - → quality ~ number and range of citations, downloads, traffic
  - → overall ~ % of OA-Versions of all publications

hbz

How do metrics vary with the different facets of Open Access **?**
Examples and experiences (1/5)

Institutional repositories

- Success as …
  - Coverage = repositories / region almost 100% in NRW, 10 run by hbz
  - Quantity of high quality content

- Experiences OPUS
  - ⊕ accompanying measures for institutions and scholars
  - ⊖ content quantity, resources at libraries for 'low threshold' support

- Perspectives
  - relaunch for dms with focus on research papers

hbz

How do metrics vary with the different facets of Open Access **?**
Examples and experiences (2/5)

OA eJournals

- Success as …
  - Portion of OA eJournals in start-ups
  - Number of existing eJournals converting to OA
  - Quantity of high quality content
    - number of publications * rejection rate, citations

- Experiences DiPP
  - ⊕ infrastructure setup, ideological uptake
  - ⊕ convince scholars with high reputation
  - ⊕ value added services (e.g. NIH author requirements)
  - ⊖ practical uptake, duration of setup

- Perspectives
  - incubation and transfer centre for OA eJournals

hbz

How do metrics vary with the different facets of Open Access ?
Examples and experiences (3/5)

Indexing

- Success as …
  - Indexer's perspective
    Coverage of indexed OA content, indexing quality
  - Publisher's perspective
    Distribution of indexed OA content in the web
- Experiences MeInD
  - ⊕ aggregation of different servers, high indexing quality
  - ⊖ links to fulltext, international reach
- Perspectives
  - larger aggregation, central indexes (e.g. hbz-search, vascoda)

hbz

How do metrics vary with the different facets of Open Access ?
Examples and experiences (4/5)

Search & Retrieval

- Success as …
  - Portion of OA content found by scholars
  - Portion of OA in full content access
- Perspectives
  - multiple versions of documents in one-stop-shop (vascoda)

hbz

How do metrics vary with the different facets of Open Access ?
Examples and experiences (5/5)

Application *uptake by scholars*

- ➔ Success as …
  - ➔ Portion of OA documents used by scholars (e.g. as compared to consortial subscription based content)
  - ➔ Portion of OA documents cited
- ➔ Perspectives
  - ➔ advanced web-based metrics and citation analysis needed (who?)

hbz

Measuring the visibility of OA documents in the web!

Proximately Analysis of website logfiles (e.g. AWstats)
- ➔ general: (different) visitors, pages, hits
- ➔ timecourse: correlation of traffic to new publications
- ➔ specific: downloads/hits of publication URL (HTML, PDF)

But! No control of spread of OA content

Ultimately web citations / web based metrics

hbz

Arguments for what is impact and how to catch this (1/2)

Note! Disciplinary differentiation necessary

Dominant example: ISI

Journal Impact Factor, Web of Science (Citation indexes)
but also Immediacy Index, Cited Half-Life

- ⊕ standardized, thorough and professional
- ⊕ wide acceptance
- ⊕ concerned with OA [1,2]
- ⊖ restricted access to metrics
- ⊖ restricted disciplinary scope ("nomothetic" only)
- ⊖ restricted lingual scope
- ⊖ restricted journal scope
- ⊖ restricted content-types (e.g. no data, programs)
- ⊖ rather slow

hbz

Arguments for what is impact and how to catch this (2/2)

- ➔ ISI causes uptake gap for OA content
  - ➔ self-archiving: remote uptake through publishers versions
  - ➔ existing Journals in ISI: unlikely to change to OA
  - ➔ new OA Journals: 3 years latency minimum (rather 10)
- ➔ web-based metrics necessary for OA
- ➔ examples
  - ➔ transparent and open citation analysis
    - ➔ NOTE! indexing of OA in scholar.google is premature
  - ➔ public OA/non-OA retrieval services
    - ➔ e.g. vascoda (work in progress)
  - ➔ OA-specific retrieval services
    - ➔OAIster, "BASE" but still metadata problems

hbz

How do you assess relevance of material?

- Scholars do quality assurance (review)
- Librarians check for technical standards

hbz

What are the specificities of OA in your scientific community?

NOTE! hbz works across disciplines

Disciplinary differences (coarsely)

- publication culture
  - between STM, HSS …: e.g. acceptance of electronic media
  - even within STM: e.g. 'green' in Physics, 'gold' in Biology
- publication format
  - peer reviewed papers (science), working papers (economics), conference proceedings (engineering), monographs (HSS)
- 'regimes'
  - e.g. traditional publishers in natural sciences, IEEE in Computer Science

Common ground technology and human-computer interaction

hbz

Which actions to sooth prejudices proved successful? (1/2)

General prejudice attribution of lower quality than traditional

- ➔ objection: "content quality"
  - ➔ "Quality selection that is provided by traditional publication system in terms of peer-review and journal formats as quality filters are missing in OA."
- ➔ answer: "scholarly self-control"
  - ➔ eJournals: peer review and use editorials as aggregators of content
  - ➔ archives: peer-review for post-prints, threat of loss in reputation for pre-prints (what about data etc.)

hbz

Which actions to sooth prejudices proved successful? (2/2)

- ➔ objection: "technical quality"
  - ➔ "Electronic documents are not secure in terms of authenticity and integrity. Platforms are not sustainable.
- ➔ answer: "professional systems"
  - ➔ authenticity: national libraries (checksums, signatures, LTP)
  - ➔ sustainability: public operators, open systems, certificates

hbz

Summary: Success factors for OA

- technical factors
  - smooth and thorough systems (e.g. assuring citability)
  - easy GUIs (e.g. restriction to DC)
  - value added services (e.g. web based metrics, new media)
  - technical aggregation (e.g. integrated search and retrieval)
- less technical factors
  - social
    - *direct communication between scholars and librarians*
  - organizational
    - *clear workload distribution between scholars and librarians*
    - *funding for extra work*
  - psychological
    - *respect disciplinary characteristics*
    - *sooth quality prejudices (best practice examples)*
    - *individual leaders (see e.g. "cream of science")*

hbz