

Elisabeth Niggemann

Archiving-on-demand - a cure against link rot in scholarly web-citations?

The legal framework for the German National Library (DNB)

- Since 1912, DNB collects everything published in Germany and in German (text and music)
- Since 2006: Legal deposit for digital publications
 - Born-digital e-books, e-journals, newspapers, dissertations, music files...
 - Digitized works
 - Webharvesting: selective harvesting
 - Webharvesting: .de-domain-crawls

Current digital collection at DNB

- More than 2 Mio. digital publications:
 - 482.973 e-books
 - 168.983 online-dissertationen
 - 2.531 e-journal titles (44.720 issues)
 - 1.202 e-paper titles (910.984 issues)
 - 4.170 digital music scores
 - 4.811 audio-books as audiofiles

Current collection from web-harvesting (selective and .de-domain-crawl)

- DNB is harvesting about 900 websites on a regular basis
- DNB harvested about 370 websites representing specific events
- Altogether 7.641 individual snapshots
- In 2014 first .de domain-crawl: 6 Mio sites, 120 TB

Partnerships and networking

- DNB is a member of the International Internet Preservation Consortium (IIPC)
- Event harvesting: European Archive
- Workflow development: oia (German company)
- .de domain crawl: Internet Memory Foundation
- Research projects: national and international partners
- nestor: German network of libraries, archives, museums, universities, ...

EVENTS

Workshops
Presentations



QUALIFIZIERUNG

QUALIFICATION

nestor-School
nestor Handbook

PUBLICATIONS

Criteria Catalogues
nestor Handbook
Info Sheets
nestor edition
Newsletter

INFORMATION

News
Calendar of Events
Experts Database
Projects Database
Information Database



STANDARDISIERUNG

STANDARDISATION

Guidelines
Recommendations
DIN Standards

INTERNATIONAL COOPERATION

Alliance Permanent Access
DPC
NCDD
NDIIPP
IFLA
EU Projects

NETWORKING & COLLABORATION

WG Media
WG Cooperation & Networking
WG Digital Preservation
WG Rights
WG Emulation
WG Certification
WG Cost Models
WG Policy



VERNETZUNG

nestor

Web-harvesting: workflow

- Librarians in DNB use special tool to select URLs, parameters and metadata of web sites
- oia use their own crawler to harvest web pages, check the quality and store the data on their own servers
- Metadata will be automatically integrated into DNB's catalogue for exclusive access in the DNB reading rooms via metadata in the catalogue and full text search
- Interface for long-term preservation in DNB's archival system

Curation of digital objects at DNB

- Automatic workflows for collecting, storing and indexing
- Dealing with a huge variety of file formats and fast changes in software and hardware. Provision for long-time archiving and long-time accessibility:
 - Bit stream preservation: Regular carrier migration and refreshing
 - Maintaining access to content: File format migration and emulation
- User access for works under copyright only on-site in the reading rooms. Access for copyright-free objects worldwide via internet

Archiving websites on-demand

- Organisations and individuals distributing digital publications can ask DNB to collect and archive these works -> archiving-on-demand is possible for publishers, publishing bodies and publishing individuals
- DNB has no service for researchers who want a particular digital object archived
- Good example in the US: Perma.cc (Harvard Law Library project)

Archiving websites on-demand as a service for researchers?

- Websites and their content become relevant objects when cited by researchers
- Citations need permanency
- Link rot makes citations unreliable
- Persistent Identifiers (PIs) linking to trusted archives would be the solution
- A need for a (distributed) service?

On-demand-archiving workflow

- Researcher wants to make sure that the citation of a digital object is permanent
- Researcher goes to the website of an on-demand-archiving service, makes the request by citing the URL, gets a PI and puts this PI in her or his (digital) article
- On-demand-archiving service harvests the document behind the URL, creates a PI for the copy and stores it in its long-time-archive

Legal challenge

- DNB is not allowed to offer free internet access to its web-archives – not even to those consisting of harvested collections from free websites
- Clear licensing policies needed for open access repositories: machine readable and standardized
- Need for two campaigns: towards politics and research and towards open access repositories communities

Archiving-on-demand as part of a research information infrastructure

- Global? European? German?
- Collecting digital publications in a network of reliable partners with a shared infrastructure
- Giving a mandate to researchers to prioritize the collection policy of heritage organisations
- Cooperation has to start by creating awareness for the volatility of the web and the need for permanency in digital research – citation and reuse of results

Thank You for Your Attention! Your Questions Please ...

- Now
- Or to:
- e.niggemann@dnb.de