



Was ARNE alles gelernt hat!

Drei Jahre Webarchivierung
beim Deutschen Bundestag

Vortrag auf der Tagung des Arbeitskreises „Archivierung von
Unterlagen aus digitalen Systemen“

Frühjahrstagung der Fachgruppe 1 des VdA

Koblenz, 22. April 2008



Wer ist ARNE?

Archivierung von Netzressourcen des Deutschen Bundestages

- Eigenentwicklung des Deutschen Bundestages
- Vollständig basierend auf OpenSource
- Ausführliche Dokumentation offengelegt und verfügbar im Internet
- künftig Nutzung / Anpassung im Staatsarchiv Basel-Stadt aufgrund Überlassungsvereinbarung



Ausgangsüberlegungen zum System

- Archivierung kein Selbstzweck
- Übernahme und Sicherung nur ein Teil der Archivierung – Nutzbarmachung zweiter wichtiger Bestandteil
- Archivierte Snapshots ebenso zugänglich wie in vorarchivischem Lebensabschnitt



Schlussfolgerung

- Abbildung des gesamten Workflows von der Bewertung über Übernahme, archivtechnische Bearbeitung, Erschließung, Bereitstellung und Benutzung



Workflow der Archivierung

Archivische Bewertung



Netzressource X

Festlegung Archivierungsoptionen

Anlegen Metadaten

Start Archivierung

Download

Kopie

Konvertierung

Indexierung

Qualitätssicherung

Freigabe für Benutzung

ARNE

Referenzdatenbank

Sammlung /
Ergänzung
Metadaten



Ausgangsüberlegungen zur Systementwicklung

- 2002 kaum Erfahrungen mit Webarchivierung
- weltweite Projekte in „Kinderschuh“
- keine deutschsprachige Fachliteratur / keine deutschsprachigen archivtheoretischen Vorarbeiten
- keine Erfahrungen mit dieser Archivaliengattung
- keine Erfahrungen mit Nutzerverhalten und Nutzerbedürfnissen



Schlussfolgerungen

- zunächst wichtig: Sammeln von Erfahrungen, darauf aufbauend Entwicklung eines Gesamtkonzepts
- Einsatz von OpenSource-Werkzeugen als Bausteine des Systems mit geringen Anpassungen
- Eigenentwicklung eines Rahmens um diese Bausteine herum (ARNE)



Erkenntnisse, Erfahrungen, Thesen

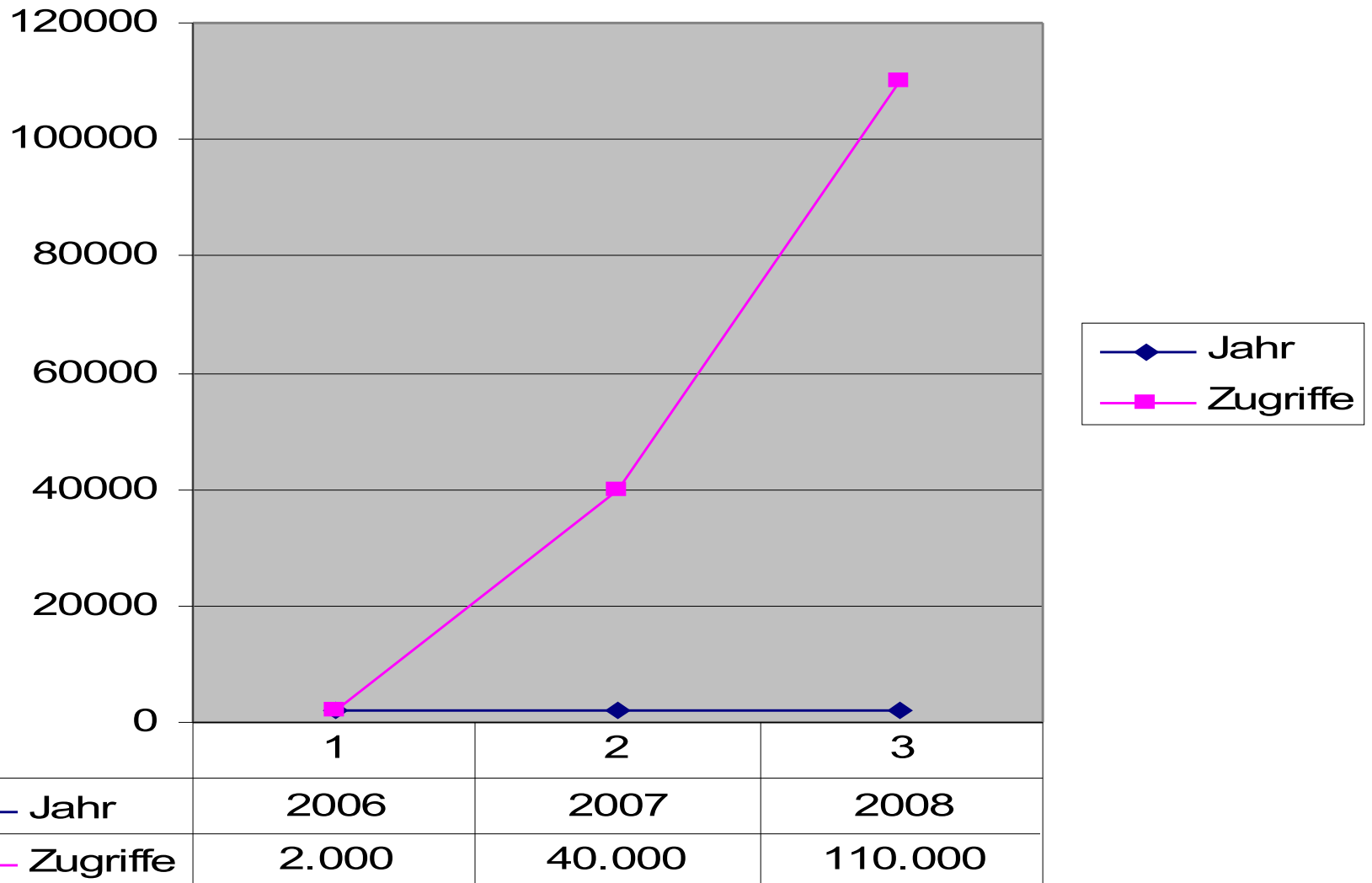
1. Netzressourcen sind mittlerweile als neue Archivaliengattung anerkannt. Webarchivierung ist Pflichtaufgabe (auch) der Archive.

- Das Thema ist angekommen...
- Webarchivierung, Netzpublikationen und das Gesetz über die Deutsche Nationalbibliothek



Exkurs zu 1: Nutzungsstatistik Webarchiv

Monate / Jahr	01-12 2006	01-12 2007	01-03 2008
Snapshots im Archiv	36	47	50
Aufrufe von Snapshots aus dem Internet	2.000	40.000	27.000





Erkenntnisse, Erfahrungen, Thesen

2. Webarchivierung erfüllt klassische Archivfunktionen.

- Entlastung des vorarchivischen Bereiches
- Sicherung, Erhaltung („Preservation Planning“) und Nutzbarmachung historischer Quellen
- Beweis- und Gedächtnisfunktion



Erkenntnisse, Erfahrungen, Thesen

3. Webarchivierung bietet Mehrwert für Archivträger.

- Logdateien und Dokumentationen zum Archivierungsvorgang offenbaren Fehler in der Live-Version



Erkenntnisse, Erfahrungen, Thesen

4. Webarchivierung setzt eine noch engere Zusammenarbeit zwischen dem „Produzenten“ und dem Archiv voraus.

- Keine Quelle ist in Hinblick auf Inhalt, Gestaltung und technische Realisierung so vielen und schnellen Veränderungen unterworfen wie Netzressourcen.



Erkenntnisse, Erfahrungen, Thesen

5. Für den Einsatz eines CMS und die Archivierung von Netzressourcen sind ähnliche Qualitätsvorgaben wie bei einem DMS wünschenswert.

- Vgl. DOMEA
- bislang nur rudimentäre Metadaten in CMS



Exkurs zu 5: Suchabfrage in ARNE

Bearbeitung einer Suchanfrage im Webarchiv

Suchbegriff: **Nichtraucherschutz**

Folgende Snapshots wurden durchsucht:

Snapshot vom **19.09.2005**

Snapshot vom **25.06.2007**

Im Snapshot vom **19.09.2005** wurden die folgenden **11** Suchergebnisse ermittelt:

Name oder Titel der Datei	Dateigröße
Deutscher Bundestag - Blickpunkt 05/2000 Nichtraucherschutz verbessern	10.66KB
aufsaeetze_06_2003.pdf	291.37KB
26-Suchtbeauf_d_Berufsverb_d_Kinder-u_Jugendaerzte.pdf	76.67KB
20-Prof__Hurrelmann.pdf	92.28KB
ThemenGesundheitSozialpolitik.pdf	189.99KB
buecher_200502.pdf	386.64KB



Exkurs zu 5: Metadaten in einem CMS

Allgemein

Pfad:	/htdocs/wissen/archiv
Name: *	archiv
Vorlage: *	Standard-Übersicht mit Untermenü
Dateityp:	Ordner
Wird exportiert werden:	Ja
Fehler in der Version:	keine

Suche

Kurzbeschreibung:	< ... >
Schlagwörter:	Deutscher Bundestag, Verwaltung, Archiv, Parlam
Ausschluss aus der Suche:	< ... >
Empfehlung / Suche / Adwords:	< ... >



Erkenntnisse, Erfahrungen, Thesen

6. Die Bewertung von Netzressourcen ist ständig zu überprüfen und ggf. zu modifizieren.

- „*Permanente Bewertung*“
- Generelle Entscheidung
- Anlassbezogene oder Turnusarchivierung
- Downloadzeitpunkt



Erkenntnisse, Erfahrungen, Thesen

7. Die Umwidmung aktueller Informationen zu Archivgut erfolgt in der Informationsgesellschaft in immer kürzeren Zeitspannen.

- *„Heute im Netz, morgen im Archiv“*
- Archivierung noch aktiver Dokumente und Dateien



Erkenntnisse, Erfahrungen, Thesen

8. Für die Archivierung und Präsentation aller digitalen Quellengattungen ist **ein** System nicht geeignet.

- Bilder, Datenbanken, Webseiten, Akten, Drucksachen ...
- Funktionalitäten zur archivtechnischen Bearbeitung und zur Präsentation
- Integration verschiedener Workflows
- Systemlandschaft zur Archivierung?



(Fast) Alles nachzulesen und vieles anzuschauen ...

www.bundestag.de/archiv
Veröffentlichungen

www.webarchiv.bundestag.de