



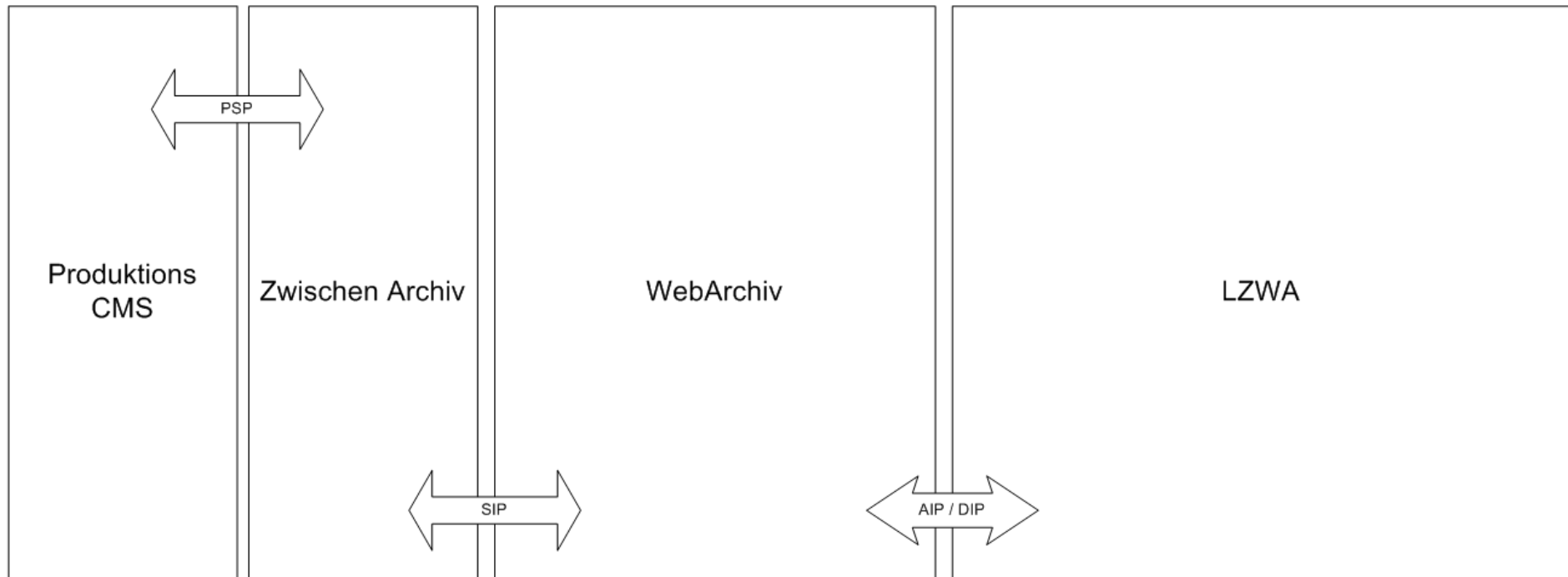
Webarchivierung
NESTOR März 2012



Ansätze
„Produktionsarchiv“

Webarchivierung OAIS

Systemübersicht

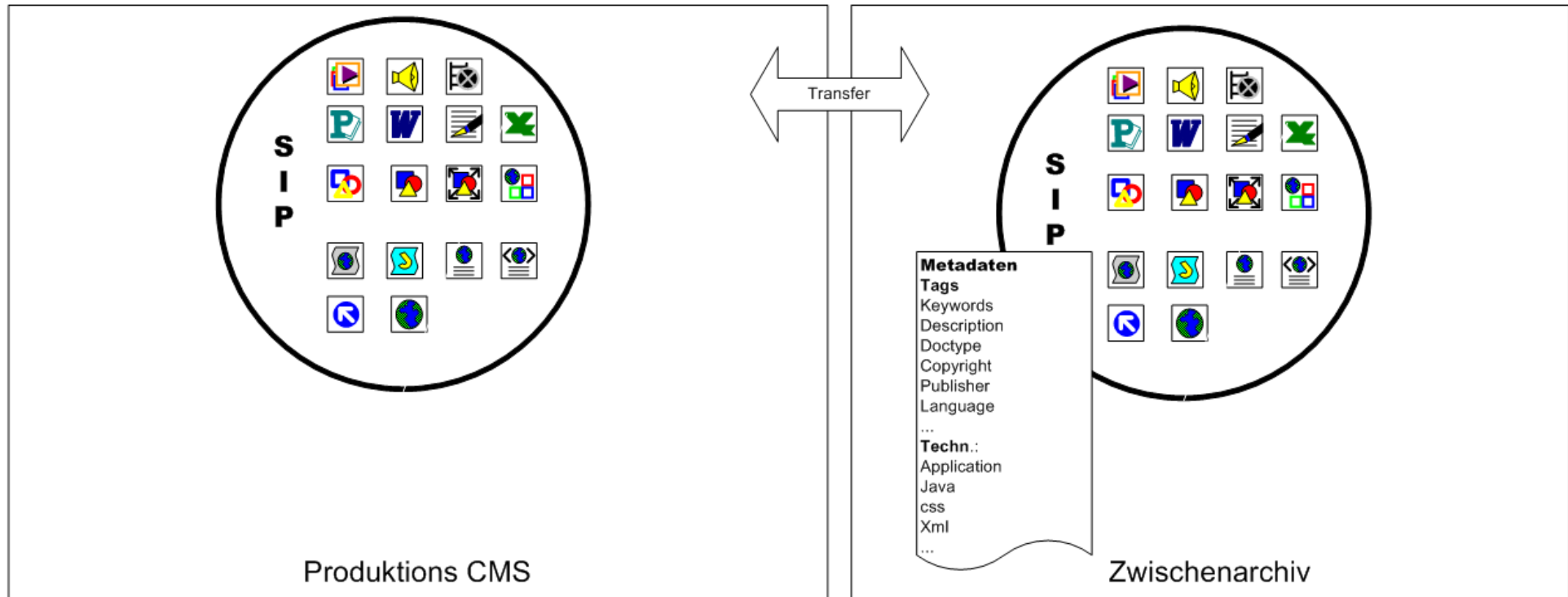


Glossar:

SIP	Submission Information Package	Aus CMS exportiertes Paket aller Daten/Dateien zu einem Web-Objekt
PSP	Producer Submission Package	Sammlungen von mehreren SIP's = mehrerer Web-Objekte
AIP	Archival Information Package	Archivisch bearbeitetes SIP
DIP	Dissemination Information Package	AIP mit allen für die Anzeige im Archivbrowser notwendigen Browse-Formaten

Webarchivierung ohne CMS Schnittstelle

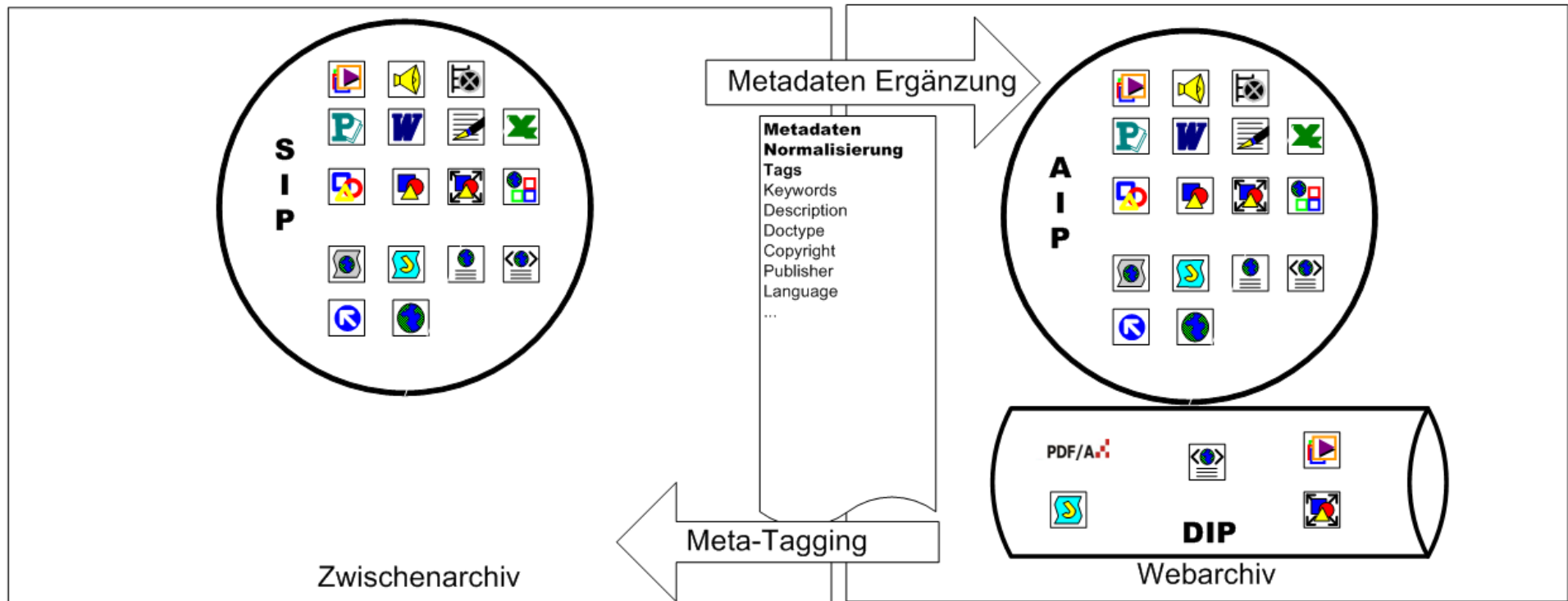
Submission Information Package SIP



Beim Transferschritt vom Produktions-CMS zum Zwischenarchiv werden die im SIP mit geführten Objektdaten automatisch ausgelesen und generieren einen Metadatenatz, der inhaltliche und technische Metadaten beinhaltet. Der internationale Standard für das SIP ist der WARC-Container, für den Metadatenatz ist das Format offen (WARC, Dublin Core, ...). Die Deutsche Nationalbibliothek hat hierfür den Standard LMER entwickelt.

Webarchivierung mit CMS Schnittstelle

Submission IP zu Archival Information Package AIP



Bei der Übernahme des SIP in das Webarchiv werden die Metadaten überarbeitet und normalisiert. Auf diese Weise entsteht ein Archival Information Package AIP. Bei der Rückgabe an ein produktives CMS (eigene RFA oder Austausch) werden die verbesserten Meta-Tags generiert.

Außerdem werden Browsingformate erzeugt und als Dissemination Information Package DIP abgelegt. Archivbrowserzugriffe werden vom DIP bedient.

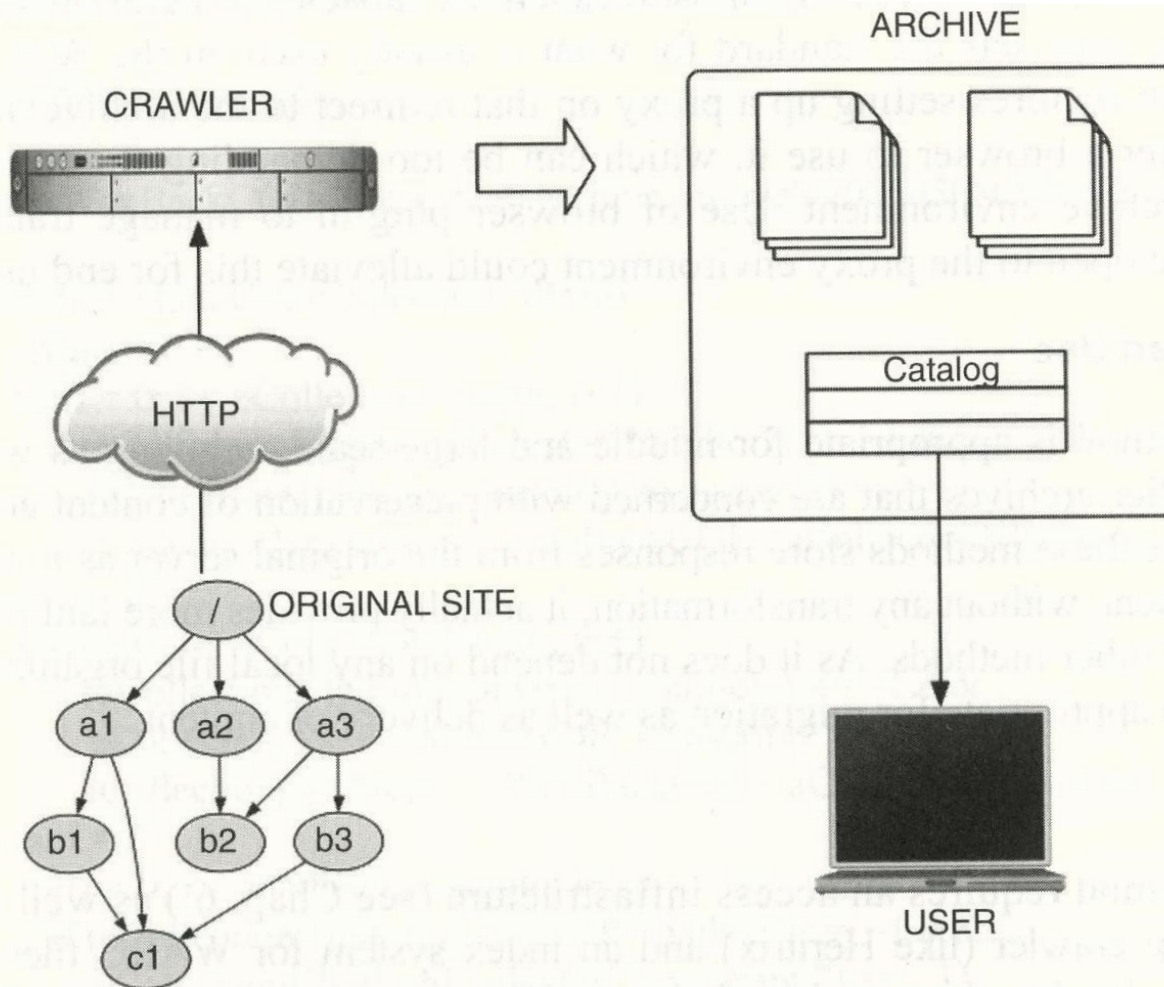
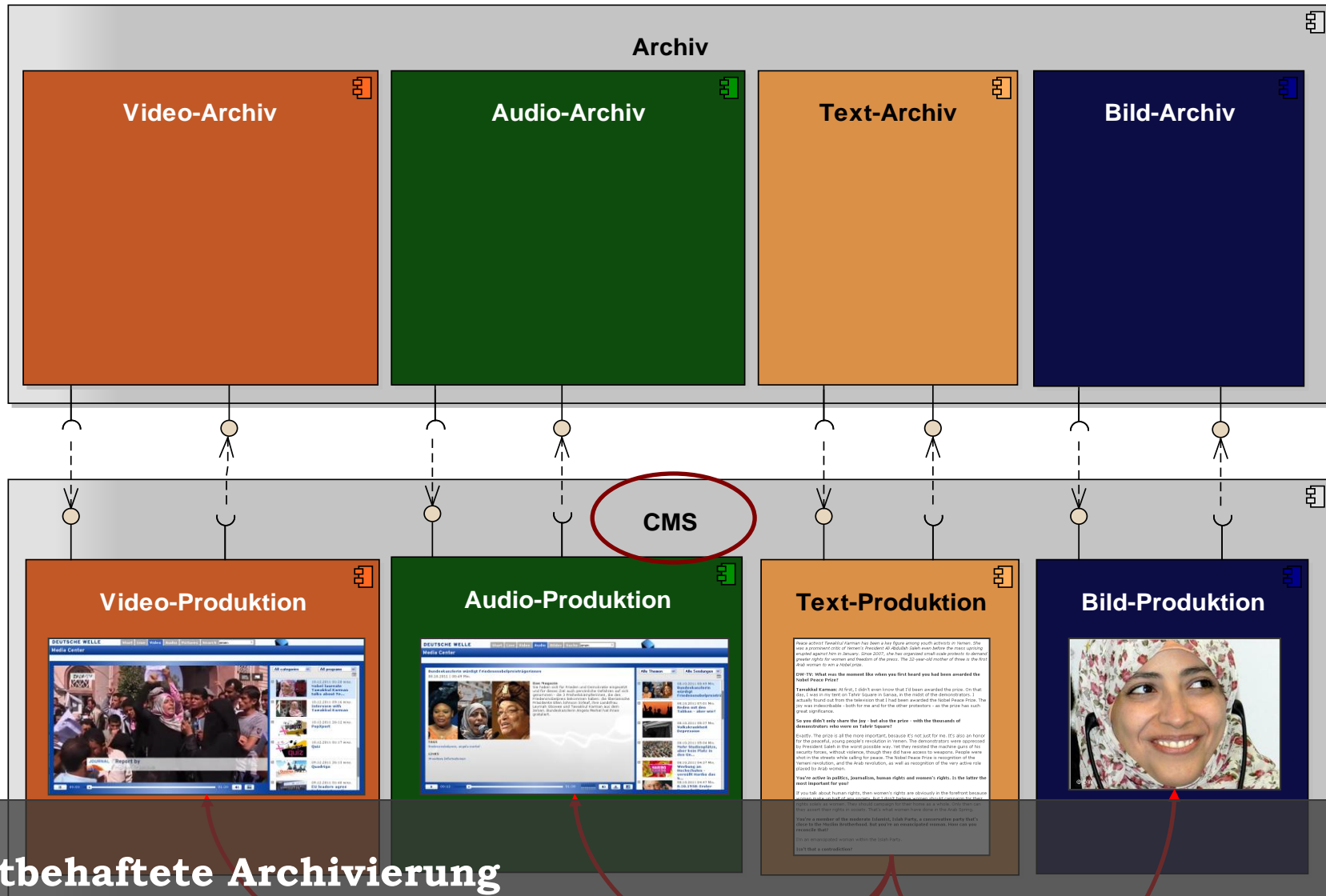


Fig. 1.7. Documents from the original site are re-organized in the archive, following a non-Web structure, using for instance a catalogue providing access to individual documents



Verlustbehaftete Archivierung

Die redaktionell erstellten Verknüpfungen bleiben nicht erhalten. Das produzierte und distribuierte Objekt ist nicht mehr rekonstruierbar.