



nestor Handbuch:  
**Eine kleine Enzyklopädie  
der digitalen Langzeitarchivierung**

9.5 File Format Registries

## Herausgeber

Heike Neuroth  
Hans Liegmann †  
Achim Oßwald  
Regine Scheffel  
Mathias Jehn  
Stefan Strathmann

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

## Im Auftrag von

nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit  
digitaler Ressourcen für Deutschland  
nestor – Network of Expertise in Long-Term Storage of Digital Resources  
<http://www.langzeitarchivierung.de>

## Kontakt

[editors@langzeitarchivierung.de](mailto:editors@langzeitarchivierung.de)

c/o

Niedersächsische Staats- und Universitätsbibliothek Göttingen

Dr. Heike Neuroth

Forschung und Entwicklung

Papendiek 14

37073 Göttingen

Tel. +49 (0) 55 1 39 38 66

Der Inhalt steht unter folgender Creative Commons Lizenz:  
<http://creativecommons.org/licenses/by-nc-sa/2.0/de/>



## 9.5 File Format Registries

*Andreas Aschenbrenner, Thomas Wollschläger*

### 1. Zielsetzung und Stand der Dinge

Langzeitarchive für digitale Objekte benötigen aufgrund des ständigen Neuerscheinens und Veraltens von Dateiformaten aktuelle und inhaltlich präzise Informationen zu diesen Formaten. File Format Registries dienen dazu, den Nachweis und die Auffindung dieser Informationen in einer für Langzeitarchivierungsaktivitäten hinreichenden Präzision und Qualität zu gewährleisten. Da Aufbau und Pflege einer global gültigen File Format Registry für eine einzelne Institution so gut wie gar nicht zu leisten ist, müssen sinnvollerweise kooperativ erstellte und international abgestimmte Format Registries erstellt werden. Dies gewährleistet eine große Bandbreite, hohe Aktualität und kontrollierte Qualität solcher Unternehmungen.

File Format Registries können verschiedenen Zwecken dienen und dementsprechend unterschiedlich angelegt und folglich auch verschieden gut nachnutzbar sein. Hinter dem Aufbau solcher Registries stehen im Allgemeinen folgende Ziele:

- Formatidentifizierung
- Formatvalidierung
- Formatdeskription/-charakterisierung
- Formatlieferung/-ausgabe (zusammen mit einem Dokument)
- Formatumformung (z.B. Migration)
- Format-Risikomanagement (bei Wegfall von Formaten)

Für Langzeitarchivierungsvorhaben ist es zentral, nicht nur die Bewahrung, sondern auch den Zugriff auf Daten für künftige Generationen sicherzustellen. Es ist nötig, eine Registry anzulegen, die in seiner Zielsetzung alle sechs genannten Zwecke kombiniert. Viele bereits existierende oder anvisierte Registries genügen nur einigen dieser Ziele, meistens den ersten drei.

Beispielhaft für derzeit existierende File Format Registries können angeführt werden:

- (I) die File Format Encyclopedia,  
<http://pipin.tmd.ns.ac.yu/extra/fileformat/>
- (II) FILExt,

<http://filext.com/>

(III) Library of Congress Digital Formats,

[http://www.digitalpreservation.gov/formats/fdd/browse\\_list.shtml](http://www.digitalpreservation.gov/formats/fdd/browse_list.shtml)

(IV) C.E. Codere's File Format site,

<http://magicdb.org/stdfiles.html>

(V) PRONOM,

<http://www.nationalarchives.gov.uk/pronom/>

(VI) das Global Digital Format Registry,

<http://hul.harvard.edu/gdfr/>

(VIIa) Representation Information Registry Repository,

<http://registry.dcc.ac.uk/omar>

(VIIb) DCC RI RegRep,

<http://dev.dcc.rl.ac.uk/twiki/bin/view/Main/DCCRegRepV04>

(VIII) FCLA Data Formats,

<http://www.fcla.edu/digitalArchive/pdfs/recFormats.pdf>

## 2. Bewertung von File Format Registries

Um zu beurteilen bzw. zu bewerten, ob sich spezielle File Format Registries für eine Referenzierung bzw. Einbindung in das eigene Archivsystem eignen, sollten sie sorgfältig analysiert werden. Sinnvoll können z.B. folgende Kriterien als Ausgangspunkt gewählt werden:

- Was ist der Inhalt der jeweiligen Registry? Wie umfassend ist sie aufgebaut?
- Ist der Inhalt vollständig im Hinblick auf die gewählte Archivierungsstrategie?
- Gibt es erkennbare Schwerpunkte?
- Wie werden Beschreibungen in die Registry aufgenommen? (Governance und Editorial Process)
- Ist die Registry langlebig? Welche Organisation und Finanzierung steckt dahinter?
- Wie kann auf die Registry zugegriffen werden?, Wie können ihre Inhalte in eine lokale Archivierungsumgebung eingebunden werden?

Künftig werden File Format Registries eine Reihe von Anforderungen adressieren müssen, die von den im Aufbau bzw. Betrieb befindlichen Langzeit-Archivsystemen gestellt werden. Dazu gehören u.a. folgende Komplexe:

## **I) Vertrauenswürdigkeit von Formaten**

Welche Rolle spielt die qualitative Bewertung eines Formats für die technische Prozessierung? Braucht man beispielsweise unterschiedliche Migrationsroutinen für Formate unterschiedlicher Vertrauenswürdigkeit? Wie kann dann ein Kriterienkatalog für die Skalierung der confidence (Vertrauenswürdigkeit) eines Formats aussehen und entwickelt werden? Unter Umständen müssen hier noch weitere Erfahrungen mit Migrationen und Emulationen gemacht werden, um im Einzelfall zu einem Urteil zu kommen. Es sollte jedoch eine Art von standardisiertem Vokabular und Kriteriengebrauch erreicht werden und transparent sein.

## **II) Persistent Identifier**

Wie können Persistent Identifier (dauerhafte und eindeutige Adressierungen) von File Formats sinnvoll generiert werden? So kann es bestimmte Vorteile haben, Verwandtschafts- und Abstammungsverhältnisse von File Formats bereits am Identifier ablesen zu können. Die Identifizierung durch „Magic Numbers“ scheint zu diesem Zweck ebenso wenig praktikabel wie die anhand eventueller ISO-Nummern. Die vermutlich bessere Art der Identifizierung ist die anhand von Persistent Identifiers wie URN oder DOI.

## **III) ID-Mapping**

Wie kann ein Mapping verschiedener Identifikationssysteme (Persistent Identifier, interne Identifier der Archivsysteme, ISO-Nummer, PRONOM ID, etc.) durch Web Services erreicht werden, um in Zukunft die Möglichkeit des Datenaustausches mit anderen File Format Registries zu ermöglichen?

## **IV) Integration spezieller Lösungen**

Wie kann in die bisherigen nachnutzbaren Überlegungen anderer Institutionen die Möglichkeit integriert werden, spezifische Lösungen für den Datenaustausch bereit zu halten? Dies betrifft beispielsweise die Möglichkeit, lokale Sichten zu erzeugen, lokale Preservation Policies zuzulassen oder aber mit bestimmten Kontrollstatus von eingespielten Records (z.B. „imported“, „approved“, „deleted“) zu arbeiten.

### **3. Bibliografie**

- Abrams, Seaman: Towards a global digital format registry. 69th IFLA 2003. [http://www.ifla.org/IV/ifla69/papers/128e-Abrams\\_Seaman.pdf](http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf)
- Representation and Rendering Project: File Format Report. 2003. <http://www.leeds.ac.uk/reprend/>
- Lars Clausen: Handling file formats. May 2004. <http://netarchive.dk/publikationer/FileFormats-2004.pdf>