

## **GBV FAG „Technische Infrastruktur“ Protokoll der 50. Sitzung vom 12.06.2013 in Göttingen**

**Teilnehmer:** Herr von der Ahe (ab 11:30 Uhr), Herr Dr. Blenkle, Herr Hahn (Moderation), Herr Schmitz (Protokoll), Herr Tuschick

**Gäste:** Herr Achtergarde (bis 12:30 Uhr), Herr Diedrichs, Herr Steilen (ab 13:30 Uhr)

**Tagungsort/-zeit:** VZG Göttingen - Seminarraum, 12.06.2013, 10:30 bis 15:15 Uhr

### **1. Protokoll / Tagesordnung / Folgetermin**

Das Protokoll der 49. Sitzung vom 10.04.2013 in Hannover ist bereits im GBV-Wiki veröffentlicht. Die Tagesordnung wird ohne inhaltliche Änderungen angenommen. TOP 4 und 5 werden in der zeitlichen Reihenfolge getauscht.

Als **Folgetermin** wird der **2. Oktober** vorgeschlagen. Eine Zustimmung durch die Teilnehmer wird eingeholt. Sitzungsort wäre die Verbundzentrale des GBV in Göttingen.

### **2. Archivierungsangebote für E-Medien der VZG [Herr Achtergarde]**

#### **[Hierarchisches Speichermanagement – HSM]**

Eingesetzt wird das Dateisystem SAM-QFS (Storage Archive Manager Quick File System<sup>1</sup>) gemeinsam mit einer hinzugekauften Index-Datenbank (SAM-WeB<sup>2</sup> vom HMK). SAM-QFS kann zum Archivieren von E-Medien per SAMBA oder NFS eingebunden werden. Die Speicherhierarchie besteht aus besonders schnellen SAS-HDDs (4 TByte, StorageTek D280), schnellen HDDs (2 Arrays mit je 30 TByte) und einer „Tape-Library“ aus HDDs. Ein Tape entspricht einer HDD. Ein ungenutztes Tape, d. h. eine ungenutzte HDD, bleibt ausgeschaltet (= hohe Standzeiten, geringer Energiebedarf, aber auch geringe Performance). Vorgaben zu Archivierungszielen werden über die „Head-Node“ (speichert alle „inode“s) definiert. Vorgaben sind u. a. Speichermedium in der Hierarchie oder die Speicherdauer von Daten (komplexes Regelwerk). Eine Konsistenzprüfung (per MD5) erfolgt, wenn Dateien eines Tapes abgefordert werden. Konsistenzprüfungen mittels vorgegebenen Zeitplan sind im SAM-QFS zurzeit nicht möglich. Eine Suche im QFS erfolgt über das Kommando „find“. Optimiert für die Suche im QFS ist der Befehl „sfind“<sup>3</sup>.

Nutzungsbeispiel für SAM-QFS ist die Landesbibliothek Oldenburg. Die dortigen digitalen Sammlungen auf Basis „Visual Library“ von semantics GmbH (in Kooperation mit Walter Nagel GmbH & Go. KG) werden im QFS abgelegt. Dafür sind zwei „Shares“ gemounted, in die Daten gespeichert werden (Server in der VZG). Die Einbindung eigener (lokaler) Speicherlösungen ist dabei grundsätzlich noch möglich.

Bei der Nutzung von SAM-QFS, so wie bei jedem Dateisystem, sind Grenzen – z. B. Anzahl der Dateien pro Verzeichnis – zu berücksichtigen. Eine Prüfung und Organisation der Archivierung durch die VZG ist zu empfehlen.

---

1 SAM-QFS (Oracle ← Sun Microsystems): [http://docs.oracle.com/cd/E22586\\_01/html/E22976/toc.html](http://docs.oracle.com/cd/E22586_01/html/E22976/toc.html)

2 SAM-WeB (von HMK ← <http://www.hmk.de/englisch/samtools.htm>) wird durch ein neues SAM-QFS abgelöst, da dort eine entsprechende Komponente integriert ist.

3 QFS „sfind“: [http://docs.oracle.com/cd/E22586\\_01/html/E22976/glaia.html](http://docs.oracle.com/cd/E22586_01/html/E22976/glaia.html)

Zur Archivierung „digitale Kapseln“ (SIP-Archive (Submission Information Package) für E-Medien und Metadaten (z. B. im METS-Format)) wird von der VZG die Nutzung des „Digitalen Archiv NRW“ (s. u.) geprüft.

### **[Anbindung „Digitales Archiv NRW“ (DA-NRW<sup>4</sup>)]**

Verantwortlich für das Projekt DA-NRW und Ansprechpartner der VZG ist Prof. Dr. Manfred Thaller, Universität zu Köln, Historisch-Kulturwissenschaftliche Informationsverarbeitung. Grundlage des Projekts ist u. a. die „e-Pflicht NRW<sup>5</sup>“. Die Verstetigung des bisherigen Projektbetriebs erfolgt ab 2013. Dazu wird NRW drei Knoten betreiben: Universität zu Köln, hbz und Landschaftsverbund Rheinland (LVR). Die VZG ist am Betrieb eines vierten (oder auch weiterer) Knoten interessiert und will damit auch die Verstetigung des Projekts, insbesondere über Ländergrenzen hinweg, unterstützen. Die VZG baut aktuell eine Testinstallation (Knoten) zum DA-NRW auf. Insofern der Praxis-Test erfolgreich ist, kann sich die VZG weitere Angebote zum DA-NRW vorstellen, u. a. Anbindung / Zugang zu Goobi (<http://www.goobi.org>) und MyCoRe (<http://www.mycore.de>). Bezüglich Goobi ist aber die Weiterentwicklung abzuwarten.

SIP-Pakete können im BagIt („bag it and tag it<sup>6</sup>“) oder METS<sup>7</sup>-Format eingeliefert werden (per WebDAV). Archiviert wird AIP (Archival Information Package), abgerufen wird DIP (Dissemination Information Package; auch die Nutzung eines Präsentations-Server / von Präsentations-Kapseln ist möglich ← DFG-Viewer<sup>8</sup>). Die Suche kann über ein Web-Interface erfolgen. Maschinelle Nutzung ist z. B. über Harvesting möglich.

Generelles Interesse an einer Langzeitarchivierung besteht auf Länderebene für Niedersachsen, NRW und Thüringen, auf Verbandsebene für den GBV (VZG). Das Nds. MWK hat hierzu bereits den GBV zu Gesprächen eingeladen. Auf Dienstleisterseite besteht insbesondere für Semantics (<http://www.semantics.de>) und Walter Nagel (<http://www.walternagel.de>) ein Interesse.

An einem möglichem VZG-Angebot zur Langzeitarchivierung per DA-NRW haben bereits Herr Dr. Blenkle (), Herr Dr. Busse () und die Universität Jena (mit ihrem MyCoRe-Team) ein Interesse angemeldet.

Eine weitere LZA-Lösungen setzt z. B. die TIB Hannover mit ihrer „Rosetta“-Installation (Ex Libris: <http://www.exlibrisgroup.com/de/category/Rosetta>) ein.

### **[Eigene Entwicklungen]**

Die VZG strebt bezüglich der Langzeitarchivierung keine eigene Entwicklung an. Sie wird entsprechende kommerzielle Angebote und Open Source-Lösungen prüfen. Wichtiges Ziel dabei ist, zu archivierende Objekte möglichst einfach zu halten, damit eine Langzeitarchivierung überhaupt gewährleistet werden kann.

## **3. FAG Workshop im Rahmen der Verbundkonferenz [Herr Hahn]**

Rund um das Thema Schnittstellen werden folgende Vorträge angeboten:

- Jürgen Hofmann: Die PAIA Schnittstelle, Stand und Ausblick.
- Jakob Voß: Stand und Planungen im Bereich der Schnittstellen in der VZG.
- Dr. Matthias Finck: "Mobiler Zugriff auf Katalog und Konto – Umsetzung der PAIA

---

4 DA-NRW: <http://www.danrw.de>

5 e-Pflicht NRW: [https://recht.nrw.de/lmi/owa/br\\_vbl\\_detail\\_text?anw\\_nr=7&vd\\_id=10912&menu=1&sg=0&keyword=Abgabe%20amtlicher%20Ver%F6ffentlichungen%20an%20Bibliotheken](https://recht.nrw.de/lmi/owa/br_vbl_detail_text?anw_nr=7&vd_id=10912&menu=1&sg=0&keyword=Abgabe%20amtlicher%20Ver%F6ffentlichungen%20an%20Bibliotheken)

6 BagIt: <http://en.wikipedia.org/wiki/BagIt>

7 METS: [http://www.loc.gov/standards/mets/METSOverview.v2\\_de.html](http://www.loc.gov/standards/mets/METSOverview.v2_de.html)

8 DFG-Viewer: <http://dfg-viewer.de/ueber-das-projekt/>

Schnittstelle für die BibApp und deren Nachnutzungsszenarien".

### **[BibApp]**

Die BibApp kann von jeder Bibliothek nachgenutzt werden. Die Anpassung<sup>9</sup> an die eigene Umgebung kann selbst erfolgen oder als Dienstleistung<sup>10</sup> (Fa. Effective WEBWORK, Dr. Matthias Finck) beauftragt werden.

### **[PAIA – Herr Diedrichs]**

Die PAIA-Entwicklung ist an Fa. Shiphrah (Herrn Hofmann) beauftragt, wird von Herrn Voß begleitet und ist quasi abgeschlossen (SW Version 0.9). Die Schnittstelle basiert auf dem PICA3-Protokoll, ist als REST-basierter Web-Service implementiert und läuft am LBS3, kann – bei parallelem Betrieb – auch mit LBS4 genutzt werden, solange der LBS4-Betrieb die parallele Nutzung von LBS3 noch ermöglicht (nicht mehr ab LBS4 mit Unicode-Unterstützung!). PAIA für LBS4 kann durch Shiphrah nur dann realisiert werden, wenn OCLC definierte Java-Klassen stabil hält. PAIA für LBS4 wird es erst ab der LBS4-Unicode-Variante geben. Weitere Entscheidungen stehen noch bis Ende 2013 aus.

Informationen zur Nachnutzung von PAIA finden sich unter:

- GitHub: <https://github.com/gbv/paia>
- GBV PAIA Server: <https://paia.gbv.de>
- GBV-Wiki: <http://www.gbv.de/wikis/cls/PAIA>

### **[LBS4 Unicode – Herr Diedrichs]**

Nach der Unicode-Version wird die funktionale Weiterentwicklung weitgehend eingestellt. Es wird Erweiterungen der Funktionalität über Schnittstellen geben (z. B. PAIA, NCIP).

## **4. Stand des Hosting-Angebots „Discovery“ [Herr Steilen]**

### **[VuFind<sup>11</sup> an der UB Magdeburg]**

Die UB Magdeburg hat eine umfassende Anforderungsspezifikation (ca. 75 Seiten) für einen neuen Online-Katalog erstellt, die die bekannten OPAC-Funktionen wie Boole'schen Such-Termen und speziellen Suchschlüsseln mit den Funktionen eines Discovery-Systems, wie Volltextsuche und Facettierung, vereint. Die Entwicklung eines Prototypen wurde bei der VZG beauftragt.

Basierend auf Erfahrungswissen vorheriger Installationen wurde durch Hr. Steilen ein Gegenvorschlag mit einer sehr reduzierten Form eines Discovery-Systemes eingebracht. Der Leistungsschwerpunkt liegt auf dem Volltext-Index (Solr<sup>12</sup>/ Lucene) und dem Ranking-Algorithmus bei quasi Verzicht auf Facetten. Für dieses veränderte Konzept war entsprechende Überzeugungsarbeit notwendig, jedoch wurde schnell deutlich, dass OPAC und Discovery komplett unterschiedliche Recherche-Strategien verfolgen und der OPAC neben dem Discovery-System verfügbar bleiben wird. Entsprechend wurde dem Gegenvorschlag zu gestimmt.

Die Umsetzung für Magdeburg erfolgt bis August 2013 durch die VZG (Abrechnung nach entsprechenden Tagessätzen). Das Ergebnis der dortigen VuFind-Installation wird eine mögliche Basis für weitere Installationen im Verbund sein. Es wird ein Werkstattbericht<sup>13</sup> der UB Magdeburg auf der Verbundkonferenz Ende August (organisiert durch die FAG Fernleihe) angeboten.

---

9 BibApp Anpassung: <http://www.gbv.de/wikis/cls/BibApp> ← BibAppDoku.pdf

10 Effective WEBWORK / BibApp / Folien: <http://www.db-thueringen.de/servlets/DocumentServlet?id=22249>

11 VuFind: <http://vufind.org>

12 Apache Projekt Solr / Lucene: <http://lucene.apache.org/solr/>

13 Beta-Version VuFind UB Magdeburg: <http://ubfind.ovgu.de>

Werkstattbericht (mit Folien und Videos) unter der Veranstaltung „Workshop FAG Fernleihe und Endbenutzer“ auf der Website der 17. Verbundkonferenz: [http://verbundkonferenz.gbv.de/?page\\_id=2973](http://verbundkonferenz.gbv.de/?page_id=2973)

### **[Präsentation des Screen-Layout der VuFind-Installation für die UB Magdeburg]**

Die Funktionalitäten, insbesondere Facetten, wurden stark reduziert. Neben der obligatorischen Übersicht der Erscheinungsjahre eines Suchergebnisses wird nur eine „Format“-Facette, auf die ggf. auch verzichtet werden kann, angeboten. Das Suchergebnis wird als „endlose Liste“ präsentiert. Die Titeldaten werden direkt angezeigt (Expansion des Seiteninhalts über einen „+“-Button).

### **[Geplantes GBV / VZG-Angebot / Infrastruktur]**

VuFind kann als Hosting-Lösung sowie als lokale Installationen (nach Aufwand / Tagessätze) durch die VZG angeboten werden. Der Abschluss eines Wartungsvertrags wird möglich sein. Eine eigene Weiterentwicklung ist ebenfalls möglich.

*[Hr. Diedrichs]*

Die gestiegene Nutzung des „GBV Zentral“-Index macht eine Erweiterung der Hardware und ein neues Verteilungskonzept des Index dringlich. Der Index soll dupliziert und auf bis zu fünf Server verteilt werden (Solr-Cloud<sup>14</sup> + Zoo-Keeper<sup>15</sup>). Neue Server mit SSD-Caches und schnellen lokalen Platten sind bereits beschafft worden.

### **[Weitere Punkte / Diskussion]**

*[Hr. Steilen]*

Bei proprietären Lösungen von Summon, Ebsco und Exlibris ist der Trend zu einem „Super-OPAC“ zu beobachten, einer Lösung, die alle bekannten Funktionen eines Online-Katalogs vereint, ähnlich wie ihn die UB Magdeburg anfangs spezifiziert hat.

*[Hr. Diedrichs]*

Abgrenzbare Bestände der Bibliotheken werden nach unterschiedlichen Kriterien (Filtern) in der Suche und Präsentation mit einem Discovery-Service kombiniert werden können.

Eigene Daten, die noch nicht im zentralen Index enthalten sind, können selber erworben und für die Nutzung eingespielt werden. Import-Format<sup>16</sup> für den zentralen Index ist das Format MARC21<sup>17</sup> in der Ausprägung der VZG (Erweiterungen in der Kategorie 9XX). Nähere Informationen dazu können bei Hr. Diedrichs erfragt werden.

## **5. Themen des Fachbeirats [Herr Diedrichs]**

Die Fachbeiratssitzung ist, bedingt durch die Nachwirkungen der Flutkatastrophe, auf einen Termin nach dem 12. Juni verlegt worden. Der Bericht auf dieser Sitzung der FAG-TI entfällt damit.

---

14 Solr-Cloud: <http://wiki.apache.org/solr/SolrCloud>

15 Zoo-Keeper: <http://zookeeper.apache.org> und: <https://cwiki.apache.org/confluence/display/ZOOKEEPER/Index>

16 Solr-Konfiguration: <https://github.com/gbv/findex-config> und

Index-Kategorien: [https://github.com/gbv/findex-config/tree/master/solr3/solrmarc\\_config](https://github.com/gbv/findex-config/tree/master/solr3/solrmarc_config)

17 MARC21: <http://www.loc.gov/marc/bibliographic/>

## **[Weitere Berichtspunkte]**

*[Herr Dr. Blenkle]*

Die DFG will die Sondersammelgebiete neu strukturieren. Mögliche Auswirkungen auf den Verbund sollen überlegt werden. Stichworte: GBV-Zentral (Index), Nationallizenz-Infrastruktur, Virtuelle Fachbibliotheken, zukünftiger Umfang elektronischer Ressourcen bei den Bibliotheken.

*[Herr Diedrichs]*

Die VZG wird mit einer Projektstelle in der DFG-Ausschreibung<sup>18</sup> „Neuausrichtung überregionaler Informationsservices“, Themenfeld 3 „Electronic Resource Management“ und zwei Arbeitspaketen (1. Marktübersicht bestehender Angebote, 2. Vorschlag für eine Nationallösung [Lizenz-Manager Nds.-Konsortium]) beteiligt.

Die OA-Aktivitäten der VZG werden verstetigt. Die Kopie<sup>19</sup> des „Directory of Open Access Journals“ (DOAJ<sup>20</sup>) wird jetzt wöchentlich automatisch aktualisiert (die Suche nach Personen-Namen ist ggf. problematisch, da eine gewollte Trennung in Namensbestandteile nicht möglich war ← Indexierung).

Der Verlag „Walter de Gruyter“ zeigte bezüglich der Zweitverwertung von Allianz-Lizenzen (Lizenzen mit OA-Komponenten) kein besonderes Interesse.

Die VZG kann mit der OCLC XML-Library mittels C beliebige XML-Formate gegen PICA direkt prüfen und verarbeiten. Dies wird u. a. schon für die DOAJ-Aktualisierung (GVK, nicht GVK-PLUS!) genutzt.

Bezüglich des Themas „Dark Archive“ strengt die VZG aktuell Überlegungen an, wie der Eigentumsübergang elektronischer Dokumente zu regeln ist, wenn z. B. ein Dienstleister abgewickelt wird.

Aufgrund der gestiegenen Nutzung des „GBV Zentral“-Index wird momentan die Leistung der Hardware in der VZG ausgebaut, um bereits bestehende Performance-Probleme zu lösen. Der Wunsch nach Normdaten im Index, u. a. Herr Hahn, muss erst zurückgestellt werden. Von Seiten der VZG ist Feedback bezüglich der Index-Kategorien im „GBV Zentral“ gewünscht. Die Suche nach originalsprachlichen Dokumenten im GBV Zentral-Index ist häufig problematisch, da bei der Suche entweder die Zeichen / Symbole der Sprachen, z. B. Chinesisch, zu verwenden sind oder die (teilweise) als Transliterationen vorliegenden Daten (deren Erstellungs-Regeln) dem Suchenden bei seiner Anfrage für die eigene Transliteration bekannt sein müssen.

Ein GVK-Zugang per SSL-Verschlüsselung ist in Vorbereitung

## **6. Verschiedenes**

Als möglicher Folgetermin wird der 2. Oktober 2013 vorgeschlagen (siehe auch Pkt. 1).

Schmitz, 01.10.2013 – Korrektur: 28.01.2014

---

18 Ergebnisse der DFG-Ausschreibung „Neuausrichtung überregionaler Informationsservices“:  
[http://www.dfg.de/foerderung/info\\_wissenschaft/archiv/2013/info\\_wissenschaft\\_13\\_11/](http://www.dfg.de/foerderung/info_wissenschaft/archiv/2013/info_wissenschaft_13_11/)

19 Kopie DOAJ: <http://gso.gbv.de/DB=1.91/LNG=DU/>

20 DOAJ: <http://www.doaj.org>