

Protokoll zum 3. Treffen

am 25. März 2015, 11.00-16.00 Uhr in Marbach

Teilnehmende

- Adnane Bousfiha
- Yvonne Friese (Leitung)
- Steffen Fritz (DLA, Gast)
- Marion Germies
- Tim Hasler
- Stefan Hein
- Martin Hoppenheit
- Heinz Werner Kramski (Protokoll)
- Stephanie Kuch (DLA, Gast)
- Andre Müller
- Mario Röhrle
- Simone Thierfelder
- Sina Westphal

Entschuldigt

- Georg Büchler
- Michelle Lindlar
- Joachim Rausch
- Claire Röthlisberger-Jourdan
- Christine Rost

Begrüßung, Protokoll der letzten Sitzung (Yvonne)

Das Protokoll des 2. Treffen kann nach einer kleinen stilistischen Änderung in den öffentlich sichtbaren nestor-Wiki-Bereich unserer AG verschoben werden.

TOP 1: Mission Statement, Abstimmung

Das Mission Statement und ein Drei-Jahres-Plan zur Arbeit der AG (*Aufgaben und Ziele 2015–2017*), den die nestor Geschäftsstelle erbeten hat, werden intensiv diskutiert und direkt im Wiki modifiziert bzw. angelegt.

Das [Mission-Statement](#) soll zum 30.4.2015 finalisiert werden, Änderungen und Diskussionen erfolgen direkt im Wiki.

Der Drei-Jahres-Plan muss zum 30.3.2015 (16:00) finalisiert werden. Sina Westphal übernimmt die Weitergabe an die Geschäftsstelle.

TOP 2: Stand der Arbeitspakete

In allen Arbeitspaketen werden nochmals die aktiven Mitglieder definiert. Federführend soll jeweils nur ein AG-Mitglied sein. Derartige Änderungen und inhaltliche Ergänzungen/Veränderungen werden direkt während der Sitzung in den Arbeitspaketen vorgenommen; sie werden hier idR. nicht wiederholt.

Arbeitspaket 1: Auflistung von Analysetools und Zuarbeiten zu COPTR

Dieses AP hieß bisher *Überblick über die vorhanden Tools*.

Hier wird u.a. die Frage diskutiert, ob Dateiformate eigene Einträge in COPTR bekommen sollen, so dass nicht nur bei einzelnen Tools erkennbar ist, welche Formate diese bearbeiten können, sondern umgekehrt auch bei einem Format, welche Tools dafür zur Verfügung stehen. Es gibt bereits einzelne Seiten zu Formaten in COPTR, z.B. http://coptr.digipres.org/Category:BMP_format. Für eine systematische Lösung wären ein hierarchische, eigene Kategorienstruktur für Formate nützlich (z.B. *Bildformate / TIFF / LZW-komprimiert*) sowie ein standardisiertes Eingabetemplate für Formate. Die Format-Einträge in COPTR auszubauen, könnte eine Aufgabe dieser AG sein. Dazu wäre ein kurzes englischsprachiges Konzept hilfreich. Eine deutschsprachige Anleitung hierfür befindet sich bereits im nestor wiki der AG.

Die Idee, die Fähigkeiten der in COPTR beschriebenen Tools automatisiert von deren Homepages etc. per Bot o.ä. einzusammeln, scheint zu groß gedacht. Vielmehr will sich die AG auf die tatsächlich verwendeten Tools und Formate konzentrieren. Insbesondere Tool-Bewertungen und -Empfehlungen durch die AG könnten hilfreich sein.

Als weitere Stichworte für AP 1 werden genannt (nicht verifiziert): PRONOM, GDFR, UDFR, Wikipedia, [format-reg-browser](#). Der auf der zweiten Sitzung in Köln entstandene Eindruck, Archivemata biete ebenfalls eine generelle File Format Registry, beruht wohl auf einem Missverständnis. Zwar gibt es dort einzelne Formatbeschreibungen [z.B. zu PDF](#), aber eher im Sinne von Use Cases, nicht mit umfassendem Anspruch.

Christine Rost hat zu AP1 umfangreiche schriftliche Bemerkungen übermittelt, die Yvonne Friese zusammenfassend vorträgt.

Arbeitspaket 2: Sammlung von Dateien

Dieses AP hieß bisher – missverständlich – *Archivemata*.

Das Sample alter Office-Dateien wächst. Sina Westphal bemüht sich um weitere Beispiele aus Ihrem Bestand, alle anderen sind ebenfalls weiter aufgerufen. Heinz Kramski hat Applix Words (.aw) und KWord-Muster (.kw) eingestellt, die schwer oder gar nicht in heutige Formate konvertiert werden können und bittet um Hilfe dazu. Zugleich können diese beiden Formate (wie

andere Office-Formate auch) auf einer technisch niedrigeren Ebene korrekt als ZIP-Archive identifiziert werden, was nur begrenzt hilfreich ist. Es entspinnt sich eine Diskussion um schwierige, vielschichtige Formate, bei denen eine *Magic-Number*-Erkennung nicht ausreicht. Problematisch sind auch Dateien, die typischerweise mit anderen zusammen auftreten, um sinnvoll benutzt werden zu können, etwa Raw-Dateien mit XMP-Metadaten, die typische *normal.dot* oder Geodaten (###?### wieso Geodaten?).

Stefan Hein berichtet von dem Einsatz der Erkennungstools der [FITS-Familie](#). Adnane Bousfiha berichtet von [Ingestlist](#), das PRONOM und JHOVE benutzt.

Die Dateisammlung kann nicht nur qualitativ als Referenzquelle dienen, sondern auch, um die Performance von Konvertierungstools etc. zu beurteilen.

Die Musterdateien für dieses AP sollen einfach als Attachments dieser Seite hochgeladen und mit den Tabelleneinträgen verknüpft werden.

Arbeitspaket 3: Register und Kontaktbörse für obsoleszente Speichertechniken (»R.O.S.T.«)

Sitzungsvorlage, Grundsätze und vergleichbare Projekte siehe beim AP 3 selbst.

Arbeitspaket 4: Formatsammlung

Das bisherige AP 4.1: *Zuarbeit zu Format Registries* wird als neues, eigenes AP 6 weitergeführt.

Damit wird AP 4 zu einer rein internen Informationssammlung, welche AG-Mitglieder mit welchen Dateiformaten Erfahrung haben. Die Teilnahme könnte noch intensiver sein.

Arbeitspaket 5: Forschungsfragen zum Thema Formatvalidierung

Dieses AP hieß bisher *Validierungsfragen*.

Die Open Preservation Foundation hat eine [Documents interest group](#) mit dem Schwerpunkt Office-Formate etabliert, die von Yvonne Friese geleitet wird. Synergien sind also zu erwarten.

DNB und BA berichten, dass beim Ingest von PDF/A häufig schwer interpretierbare JHOVE-Fehlermeldungen auftreten. Beide verfahren pragmatisch und brechen nur ab, wenn der Import tatsächlich scheitert.

[Das PDF/A Competence Center](#) gibt in solchen Fällen gerne Auskunft zu den Hintergründen und Implementierungseinheiten.

TOP 3: Konferenzen, Workshops und sonstige Veranstaltungen

Auf folgende Termine wird hingewiesen:

- Workshop "Bilder und Digitalisate – Herausforderungen bei der Langzeitverfügbarkeit von Bildformaten" am 30.9. und 1.10.2015 an der [ZB MED](#) in Köln (kostenpflichtig, Anmeldungen bei Yvonne Friese)
- [PDF Days Europe](#) 2015 am 8. und 9.6.2015 in Köln
- [Nestor Praktikertag 2015](#) am 24. und 25.6.2015 im Landesarchiv NRW Duisburg.

TOP 4: Sonstiges

Wiki dieser AG

Die Arbeitspakete sollten keine Unterseiten von Wissensdatenbank sein, sondern einen eigenen Zweig bilden (inzwischen erledigt.)

Organisatorisches

Um aussagekräftige Beschreibungen bei Wiki-Änderungen wird gebeten. *Notify watchers* bitte zugleich nur auf die größeren Änderungen beschränken.

Andere relevante Entwicklungen und Projekte

Stefan Hein berichtet, dass das Projekt TREASURES, eine Initiative der National Library of New Zealand, als EU-Projekt abgelehnt wurde.

Er berichtet ferner von dem Projekt [PREFORMA](#) (Ziel: Entwicklung und Bereitstellung einer Open-Source-Software lizenzierter Referenzimplementierungen für verschiedene Standards von Dateiformaten zur Durchführung von Konformitätstests bezgl. spezieller Standards. [Präsentation \(deutsch\)](#). [FactSheet \(deutsch\)](#)).

TOP 5: Nächstes Treffen

Das nächste Treffen findet statt am 24.08.2015 im Konrad Zuse Institut Berlin.