

Protokoll des 6. Treffen der nestor AG Formaterkennung

20.09.2016, TIB Hannover, 11-16 Uhr

Teilnehmende: Andre, Heinz, Michelle, Martin, Yvonne, Arthur Spengler (Gast, TIB) und Franziska Schwab (Gast, TIB)

Entschuldigt: Claire, Marion, Mario, Rainer, Adnane, Stefan und Tim

1 Aktueller Stand aller Working Groups

AP 1 Tool Registries (Christine in Vertretung für Claire)

Für die Umfrage wäre es noch gut, folgende Frage unterzubringen:

Gibt es den Bedarf, statt des genauen Formats erst einmal die Formatfamilie (PDF, TIFF, JPEG, usw.) festzustellen und dann ggf. von dort aus das genaue Format zu bestimmen? Falls das genaue Format nicht bestimmbar wäre (fmt/19 oder fmt/20?), stünde wenigstens die Formatfamilie fest. Oder es ist überhaupt nur die Information für Familie wichtig, nicht zum detaillierten Format. (Das könnte man für die Umfrage natürlich noch besser und verständlicher formulieren.)

Andre und Heinz stellen fest, dass dieser Bedarf grundsätzlich besteht und auch in eigenen Workflows sinnvoll ist - z.B. wenn DROID keine Antwort liefert oder auch um Entscheidungen zu treffen ob andere Tools (z.B. Apache Tika) über Dateien laufen sollen.

Claire hat die Tabelle umgestellt und bittet darum, dass wir alle (nicht nur die an AP 1 beteiligten) unsere Erfahrungen und Verwendungsstand eintragen. Ebenso ist noch die Aufgabe aus der letzten Sitzung offen, in der eine neue Spalte "Benutzerschnittstelle" zur Kennzeichnung bzgl. GUI / Kommandozeilentool / Library vorgesehen ist.

Danach ist AP 1 abgeschlossen, abgesehen mal von der laufenden Aufgabe der Pflege und des Mit-Aufbaus von COPTR.

AP 2 Testsuites (André)

Keine aktuellen Neuigkeiten.

Es bietet sich an, Telkos für das AP durchzuführen (läuft in anderen APs erfolgreich).

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

Diskussion und Verabschiedung des [Fragebogens für die geplante Umfrage](#). Heinz stellt den Fragebogen im finalen Stand nochmal vor. Die Zielgruppe soll in Dienstreisentfernung liegen. Der Fragebogen versucht die Komplexität abzudecken, ohne die Zielgruppe zu sehr abzuschrecken.

Offen ist die Frage des Abfragetools (siehe oben). Die Vorstellung war, sich an die Umfrage aus AP1 zu hängen. Ebenso ist die Frage offen, wie viel "nestor Branding" die Umfrage haben muss / soll / kann (konkrete Anforderung an das Tool). Confluence fällt laut nestor Geschäftsstelle als Umfragetool aus, da geeignete Plugins kostenpflichtig / schlecht nutzbar sind. Unipark und Lime-Survey sind kostenpflichtig - hier müssten konkrete Kosten ermittelt

werden und die Finanzierbarkeit bei der nestor Geschäftsstelle erfragt werden. SoSci Survey wäre laut Website für unsere Umfrage kostenlos.

Nächster Schritt: Heinz setzt den Fragebogen in SoSci-Survey um und ruft zu internen Testlauf durch AG-Mitglieder auf.

Kontaktaufnahme zu [Mediopedia](#): Physical Format Carrier Resource (National Library of Australia): Michelle schreibt Douglas Elford (NLA) an und fragt, ob Links auf Holotypes / systematische Suchoptionen bestehen und ob wir die Holotype Ids / Thumbnails / Links für die Umfrage nutzen können.

Besteht von Seiten der NLA die Möglichkeit/Interesse, dass die AG neue Holotypes nachträgt?

Heinz nimmt Kontakt zu EMiL Projekt / Universität Freiburg auf zwecks bestehender Arbeiten an Hardwareregistries.

Finetuning, produktive Umfrage/Laufzeit, Auswertung und Präsentation diskutieren und festlegen.

AP 4 Format Registries (Martin)

Keine aktuellen Neuigkeiten. Auch hier bietet sich eine Telko an 😊

AP 5 Formatvalidierung (Stefan, Yvonne & Michelle unterstützen, da Stefan bei der letzten Telko nicht dabei war)

Erste Aufgabe der Unter-AG: PDF/A-Validatoren testen. Zurzeit werden der Korpus zusammengestellt und erste Toolzuständigkeiten verteilt. Martin ergänzt noch um ein paar problematische PDF/A.

Zwischendiskussion:

Obsolet ist 0 oder 1, Obsoleszenz ist ein Prozess und kann z.B. einen Wert zwischen 0 und 100 % einnehmen. Beispiel für den Prozess der Obsoleszenz: Ein Laufwerk ist am Markt nicht mehr verfügbar, aber die Medien können noch gekauft werden. Die nächste Stufe wäre: weder Medien noch Laufwerk sind verfügbar. Die finale Stufe wäre: Niemand kennt es mehr.

AP 6 Best Practices in der Identifizierung (Yvonne in Vertretung für Claire)

AP 6 soll ja praktische Beispiele und Tools beschreiben und könnte daher sogar das Grundgerüst für den nestor Workshop in 06/2017 bilden.

Ziel: Mehrere Szenarien, welche in der AG angewendet werden, beschreiben und erläutern. Pro Szenario ein Bericht oder eine Subsite, welche später veröffentlicht werden könnte.

- Übernahme von Unstrukturierten Dateisammlung (NSRL) [Heinz?]
- Integrierte Formaterkennung (technische Umsetzung der Formaterkennung in KOST-Val)
- Wie müssen DROID-Erkennungen interpretiert werden (mehrere Treffer, verschachtelte Dateiformate) [???

Zurzeitige To-Dos

- Eine Subseite müsste noch erstellt werden rund um „droidsfin“ respektive diese sollte das Szenario droidsfin beschreiben inkl. den Geschwindigkeitstests zusammengefasst [Martin?]
 - Dort möchte ich noch ein einfaches GUI machen, sobald ich wieder Luft habe (Claire)
- Formaterkennung mit Jhove? [???] (bzw. die Grenzen der Formaterkennung mit JHOVE) (Michelle und Yvonne, da eh IDCC Paper?)
- Jeder soll sein Szenario beschreiben, wenn er ein anderes Vorgehen hat.
- nestor Kurzartikel (max 500 Wörter) zu DROID vs. Siegfried Ergebnisse (Michelle schreibt Entwurf und schickt an Martin & Yvonne) (Heinz fragt, ob Artikel generell noch willkommen sind. Sein Vorschlag ist ein Kurzartikel zu archiveready. Yvonne meint, dass Artikel grundsätzlich noch willkommen sind.) Kurzartikel zu PREMIS (Einstieg) - Martin oder Michelle ?

2 Beitrag zum Leitfaden

Welchen Input wollen wir liefern? Georg spricht dann für uns mit Herrn Keitel. Herr Keitel weiß auch schon Bescheid. Eigentlich ruht die AG sogar gerade, sie warten quasi nur noch auf uns. Die AG beschließt, hier auf Rückmeldung von Herrn Keitel bzgl. gewünschtem Umfang und Inhalt eines Leitfadenbeitrags zu warten. Für eine Kapitelaktualisierung im nestor Handbuch müsste Stefan Strathmann angesprochen werden.

3 Öffentlichkeitsarbeit der nestor AG

nestor Praktikertag 2017

Dieser findet am 28. Juni in Kiel statt. Am Vortag ist das Treffen der Operativrunde/Koordinationsrunde. Die Idee unseres Workshops wurde von der Operativrunde positiv aufgenommen. Sie möchten dies aber nicht parallel, sondern lieber am Folgetag machen, z.B. am 29. Juni vormittags. Ggf. könnten wir dann am Nachmittag unser reguläres Treffen abhalten, damit sich die Reise so richtig lohnt.

Das Oberthema des Praktikertags wird vermutlich sogar die Formaterkennung sein. Herr Keitel und Herr Wolf haben zugesagt, den Tag inhaltlich zu gestalten (und hoffen aber auf unsere Unterstützung) und Referentinnen und Referenten zu finden. Die nestor Geschäftsstelle liegt demnächst leider wieder einige Zeit lang brach, da sowohl Sabine als auch Sabrina ca. ab Herbst erst einmal im Mutterschutz sind. Zum Praktikertag 2017 sind aber beide wieder in Teilzeit zurück. Wenn möglich, im Oktober mit Geschäftsstelle klären, wie viele Vorträge von uns besetzt werden sollen. Letztes Jahr gab es 5 Vorträge + 1 Vortrag Geschäftsstelle + Kurzvorträge Posteraussteller + 2 Austauschrunden - somit könnten wir zwischen 5-7 Vorträge besetzen. Gibt es eine Postervorstellung? Sonstige Rahmenparameter (9:30 - 16:30 Uhr)?

Wer von uns mag einen Vortrag halten? Claire hat Interesse signalisiert - gut wäre ein Einstiegsvortrag zum Thema Identifizierung (warum? wie? was? womit? Hinweis auf viele Dateien lassen sich identifizieren, für manche können technische Metadaten extrahiert werden, wenige können validiert werden ... und ganz wenige repariert). Heinz kann einen Vortrag zu NSLR halten. Michelle kann einen Vortrag zu PUIDs halten. Yvonne kann einen Vortrag zu JHOVE halten. Ggf. können wir auch einige der Vorträge in unseren Workshop verschoben werden. Es bietet sich an, das Programm vom Praktikertag und Workshop aufeinander aufbauen zu lassen. Neben kurzen Vorträgen sollte der Workshop auch "hands-on" sessions haben - wir können z.B. TeilnehmerInnen dazu aufrufen, problematische Files mitzubringen. Weitere Themenvorschläge: Wie kann ich DROID / JHOVE Reports auf meine Bedürfnisse hin anpassen (z.B. Martins XSLTs für DROID)? PRONOM Signaturen verstehen

und schreiben.

Workshopgröße: max. 25 Teilnehmer (Schätzung Praktikertagteilnahme ca. 80). Workshop soll kostenlos angeboten werden. Bei Dauer > 4 Stunden müssen wir Mittagspause mit einplanen (Catering? Yvonne klärt mit Geschäftsstelle).

Archivtag

Mario hält dort in diesem Monat einen Workshop zu Dateiformaten

4 Sonstiges

nestor wiki:

- Ab einer gewissen Befüllung mal Rundmail an die nestor-Mailingliste?
Uns fehlt eine Einstiegsseite / Einstiegsinformation. Martins PRONOM Anleitung kann als Einstieg für den öffentlich einsehbaren PUID Artikel verwendet werden.
- das wiki sollte im Rahmen des Praktikertags als Ressource beworben / genannt werden.
- Die nestor Koordinierungsgruppe kann ggf. gezielt um Feedback gebeten werden.
- nestor Stammtische um Feedback bitten?
- Jonas Recker fragen, ob Wikiseiten im Rahmen von CESSDA Training hilfreich sein könnten?
- Bei dem Treffen wurde "Brunhilde" erwähnt zum erstellen von Reports bei DROID. Kann jemand den Link zur Verfügung stellen? --> <https://github.com/timothyryanwalsh/brunnhilde>

nächstes Treffen: Duisburg im März - Daten für foodle am 14.3. von 11 bis 16 Uhr