Beitrags-ID: 131

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

UNIVERSITÄTSMEDIZIN GÖTTINGEN UMG

Anbindung eines elektronischen Laborbuchs an eine vorhandene Langzeitarchivierungsinfrastruktur

Bartlomiej Marzec[a], Philipp Weil[ab], Sara Y. Nussbeck[a]

[a]Universitätsmedizin Göttingen, Institut für Medizinische Informatik
[b]Universitätsmedizin Göttingen, Institut für Molekularbiologie

SFB 1002

1. Einleitung

Im Rahmen des Sonderforschungsbereiches wird eine Langzeitarchivierungsinfrastruktur (LZA) aufgebaut. Diese Infrastruktur wird aus einem Repository (GWDG CDSTAR [1]) bestehen, in welchem Forschungsdaten mit den dazugehörigen Metadaten abgelegt werden können. Die Eingabe der Daten soll über zwei Wege möglich sein: über ein Webportal des Repositories oder aber in einem elektronischen Laborbuch (ELN) (eCAT [2]), welches derzeit von drei ausgewählten Doktoranden aus unterschiedlichen Fachgebieten (Physik, Zellbiologie, Pharmakologie) pilothaft verwendet wird. Vom ELN aus müssen die Daten dann in das Repository exportiert werden, damit sie für andere Forscher des SFBs zur Verfügung gestellt und mit anderen Daten durchsuchbar gemacht werden (Export-Tool, siehe Abbildung 1). Dabei besteht die Schwierigkeit darin, dass die Datenerfassung im ELN sehr flexibel sein muss, um sich ändernden Bedingungen in Forschungsprojekten anpassen zu können, aber auf der anderen Seite die Daten in einer nachnutzbaren strukturierten Weise exportiert werden müssen. Diese Arbeit beschreibt wie ein solcher Datentransfer aus dem ELN in das Repository der LZA realisiert werden kann.

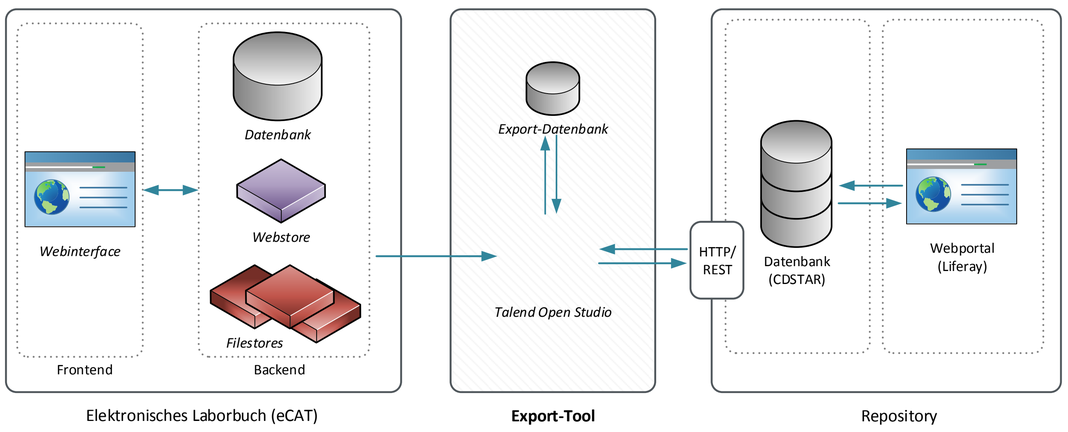

Abbildung 1: Übersicht des technischen Konzeptes

2. Methoden

Um zu analysieren, wie der Datentransfer aussehen kann, mussten mögliche Anforderungen identifiziert werden. Dazu wurden die ELN-Benutzer interviewt, das Quell- und Zielsystem analysiert und eine Literaturrecherche zur Ermittlung von Standards für die Speicherung von ELN-Daten durchgeführt. Aus der Verknüpfung dieser Anforderungen wurde dann das Konzept entwickelt, das anschließend prototypisch mithilfe eines sogenannten Extract-Transform-Load-Softwareproduktes (Talend Open Studio) umgesetzt wurde (Export-Tool in Abbildung 1).

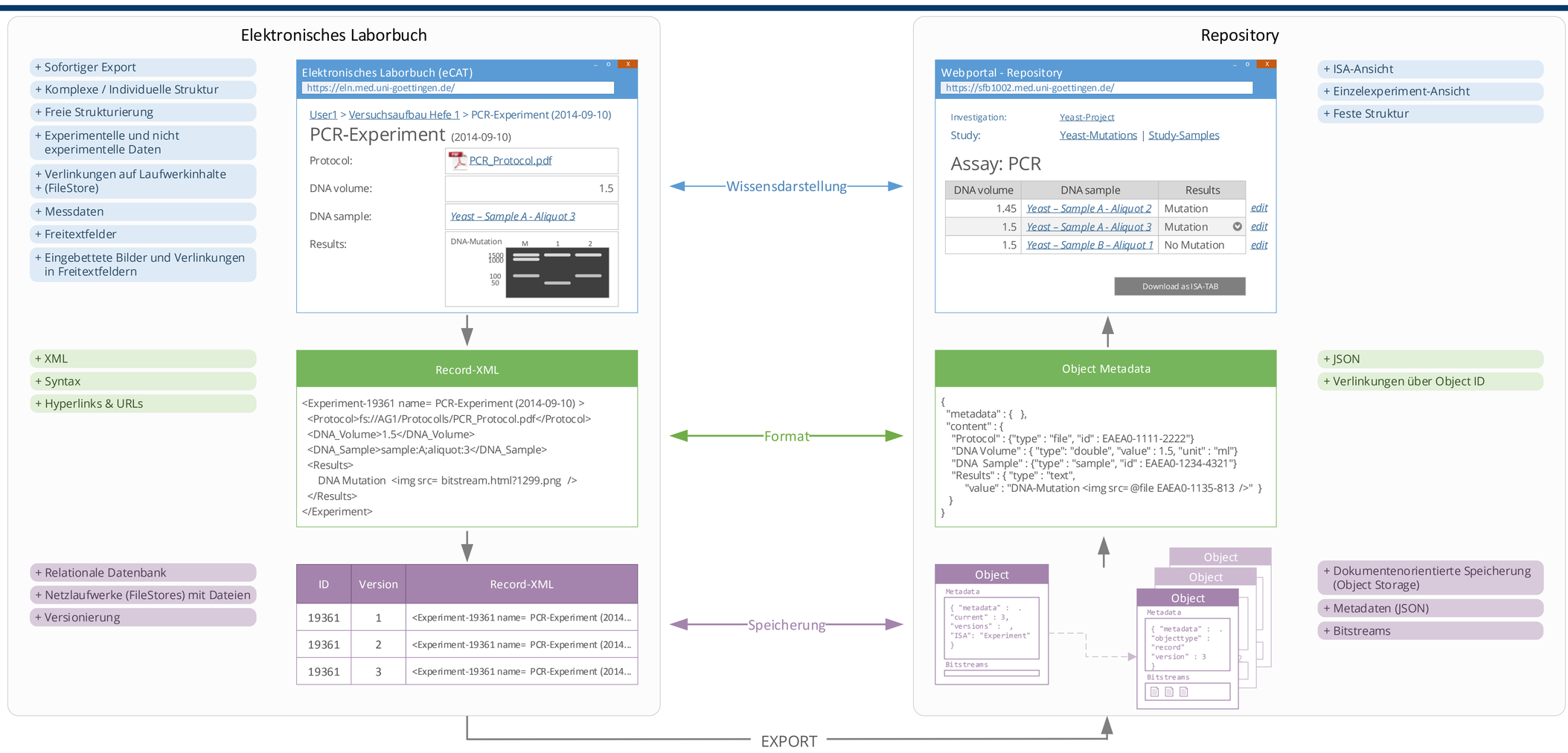

Abbildung 2: Drei Ebenen des Konzeptes jeweils mit Anforderungen aus den Interviews und der Quell- und Zielsystemanalyse.

3. Ergebnisse

Die Abbildung 2 stellt die Ergebnisse der Anforderungsanalyse und das Konzept dar. Es wurde herausgefunden, dass bei der Entwicklung des Konzeptes drei Ebenen betrachtet werden müssen: die Wissensdarstellungs-, die Format- und die Speicherungsebene. Die dazugehörigen Anforderungen aus den Interviews und den Systemanalysen wurden in der Abb. 2 jeweils links und rechts zu jeder Ebene aufgelistet. Aus den ca. 540 gefundenen Standards wurde ISA-TAB [3] für die Entwicklung des Konzeptes ausgewählt, da er die drei oben genannten Ebenen abdeckt.

Das sich daraus ergebende Konzept ist eine hybride Lösung aus einer „Eins-zu-Eins"-Abbildung der Daten aus dem ELN und dem gleichzeitigen Aufbau der ISA-Struktur im Repository. Bei der 1:1-Abbildung werden alle Einträge und die geforderte Ordner- und Dateistruktur versioniert abgebildet. Dies erlaubt später die Daten im Repository-Webportal in der Form anzusehen, wie diese auch im ELN gespeichert waren. Damit im Repository zumindest der Aufbau der ISA-Beziehungen stattfinden kann, müssen im ELN spezielle Formulare (Investigation- und Study-Templates) verwendet werden, die dem Export-Tool signalisieren, dass es sich bei der ELN-Struktur um eine ISA-Struktur handelt. Der Aufbau der ISA-Struktur ermöglicht im Repository-Webportal eine ISA-TAB-konforme Ansicht der Daten und bei eventueller Erweiterung der Funktionalität des Webportals, auch einen Export der Daten nach ISA-TAB-Standard.

4. Diskussion

Zu den positiven Eigenschaften des Konzeptes gehören:

- Die Funktionalität des ELNs wird durch die Verwendung der ISA-Struktur nicht beeinträchtigt
- Die ELN-Benutzer werden nicht mit einer Reihe von Regeln beim Anlegen der Daten konfrontiert
- Die Verwendung der ISA-Struktur im ELN ist optional
- Das Verhalten des Export-Tools kann für jeden ELN-Eintragstyp angepasst werden
- Die Experimente eines Versuchsaufbaus können tabellarisch angesehen und heruntergeladen werden

Es konnten jedoch folgende Probleme festgestellt werden:

- Eine ISA-TAB-konformer Export war nicht möglich, da bestimmte Datenstrukturen aus dem ISA-TAB im ELN nicht erstellt werden können
- ISA-TAB-Standard sieht keine Speicherung von Freitextfeldern, insbesondere mit eingebetteten Bildern und Verlinkungen vor.

Im weiteren Projektverlauf muss das Konzept mit den Benutzern ausdiskutiert und entsprechend angepasst werden. Eine neue Version des ELNs soll über neue Funktionalitäten verfügen, die möglicherweise den ISA-TAB-konformen Datenexport schließlich ermöglichen wird.

Danksagung

Wir danken allen Wissenschaftlern, die an den Interviews teilgenommen haben. Insbesondere Markus Kilisch, Sumon Sur und Daniel Hornung.

Literatur

[1] Schmitt O et al. Integrierte Portalumgebung und verteilte Echtzeitsuche für medizinische Langzeitarchivierung. 2014, GMDS 2014, Beitrags-ID: 184
[2] Goddard NH et al. eCAT: Online electronic lab notebook for scientific research. Autom Exp. 2009;1:4.
[3] Sansone S et al. The First RSBI (ISA-TAB) Workshop: "Can a Simple Format Work for Complex Studies?". OMICS: A Journal of Integrative Biology.2008

Funded by DFG Deutsche Forschungsgemeinschaft

E-Mail des Korrenspondenzautors
bartlomiej.marzec@med.uni-goettingen.de

Zusammenfassung online unter:
http://www.egms.de/static/de/meetings/gmds2014/14gmds121.shtml