



Versionierungswerkzeuge für Forschungsdaten - Datenmanagement im Forschungsprozess

Sibylle Hermann

Universitätsbibliothek Stuttgart

17.06.2016



Inhaltsverzeichnis

① Problemstellung

② Ziel

③ Versionierung

④ Projekt

Nachbauen?!?



Nachbauen?!?



- Aus wieviel Teilen besteht das Schiff?
- Welche Teile brauche ich?
- Wie wurden diese zusammengebaut?
- Diese und weitere Informationen finden sich in der Aufbauanleitung
- ...und bei Veröffentlichungen?

Nachbauen?!?



- Aus wieviel Teilen besteht das Schiff?
- Welche Teile brauche ich?
- Wie wurden diese zusammengebaut?
- Diese und weitere Informationen finden sich in der Aufbauanleitung
- ...und bei Veröffentlichungen?

Nachbauen?!?



- Aus wieviel Teilen besteht das Schiff?
- Welche Teile brauche ich?
- Wie wurden diese zusammengebaut?
- Diese und weitere Informationen finden sich in der Aufbauanleitung
- ...und bei Veröffentlichungen?

Nachbauen?!?



- Aus wieviel Teilen besteht das Schiff?
- Welche Teile brauche ich?
- Wie wurden diese zusammengebaut?
- Diese und weitere Informationen finden sich in der Aufbauanleitung
- ...und bei Veröffentlichungen?

Nachbauen?!?



- Aus wieviel Teilen besteht das Schiff?
- Welche Teile brauche ich?
- Wie wurden diese zusammengebaut?
- Diese und weitere Informationen finden sich in der Aufbauanleitung
- ...und bei Veröffentlichungen?

Nachbauen?!?



- Aus wieviel Teilen besteht das Schiff?
- Welche Teile brauche ich?
- Wie wurden diese zusammengebaut?
- Diese und weitere Informationen finden sich in der Aufbauanleitung
- ...und bei Veröffentlichungen?

Veröffentlichungen (DH)

- Viele Veröffentlichungen basieren auf Daten, die mit einem Programm erstellt wurden. -> Metadaten für Programm sind bekannt -> Clarin-D -> nicht standardisiert
- Programm wurde an Problem angepasst -> nicht dokumentiert
- Veröffentlichungen basieren oft auf Zwischenstand -> nicht dokumentiert

Veröffentlichungen (DH)

- Viele Veröffentlichungen basieren auf Daten, die mit einem Programm erstellt wurden. -> Metadaten für Programm sind bekannt -> Clarin-D -> nicht standardisiert
- Programm wurde an Problem angepasst -> nicht dokumentiert
- Veröffentlichungen basieren oft auf Zwischenstand -> nicht dokumentiert

Veröffentlichungen (DH)

- Viele Veröffentlichungen basieren auf Daten, die mit einem Programm erstellt wurden. -> Metadaten für Programm sind bekannt -> Clarin-D -> nicht standardisiert
- Programm wurde an Problem angepasst -> nicht dokumentiert
- Veröffentlichungen basieren oft auf Zwischenstand -> nicht dokumentiert

Veröffentlichungen (DH)

- Viele Veröffentlichungen basieren auf Daten, die mit einem Programm erstellt wurden. -> Metadaten für Programm sind bekannt -> Clarin-D -> nicht standardisiert
- Programm wurde an Problem angepasst -> nicht dokumentiert
- Veröffentlichungen basieren oft auf Zwischenstand -> nicht dokumentiert

Veröffentlichungen (DH)

- Viele Veröffentlichungen basieren auf Daten, die mit einem Programm erstellt wurden. → Metadaten für Programm sind bekannt → Clarin-D → nicht standardisiert
- Programm wurde an Problem angepasst → nicht dokumentiert
- Veröffentlichungen basieren oft auf Zwischenstand → nicht dokumentiert



→ Nachnutzung schwierig

Arbeitsprozess



- zwei Stellschrauben: Tool und Inhalt
- Anpassung passiert parallel im Arbeitsprozess
- spätere Dokumentation schwierig
- Was wurde wann warum angepasst?

Arbeitsprozess



- zwei Stellschrauben: Tool und Inhalt
- Anpassung passiert parallel im Arbeitsprozess
- spätere Dokumentation schwierig
- Was wurde wann warum angepasst?

Arbeitsprozess



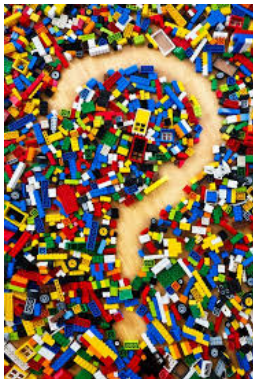
- zwei Stellschrauben: Tool und Inhalt
- Anpassung passiert parallel im Arbeitsprozess
- spätere Dokumentation schwierig
- Was wurde wann warum angepasst?

Arbeitsprozess



- zwei Stellschrauben: Tool und Inhalt
- Anpassung passiert parallel im Arbeitsprozess
- spätere Dokumentation schwierig
- Was wurde wann warum angepasst?

Arbeitsprozess



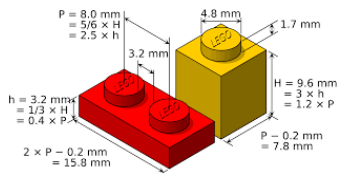
- zwei Stellschrauben: Tool und Inhalt
- Anpassung passiert parallel im Arbeitsprozess
- spätere Dokumentation schwierig
- Was wurde wann warum angepasst?



Inhaltsverzeichnis

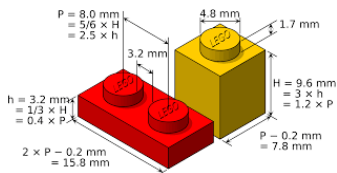
- ① Problemstellung
- ② Ziel
- ③ Versionierung
- ④ Projekt

Ziel



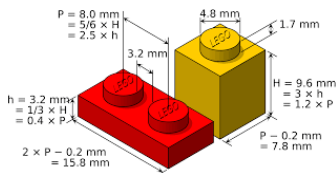
- Metadaten bereits bei der Genese und nicht erst bei der Archivierung erfassen
- Metadaten standardisieren
- Nachnutzen der Daten und Nachvollziehen der Änderungen ohne erheblichen Mehraufwand
- DOIs für Zwischenstände (zitierbar)

Ziel



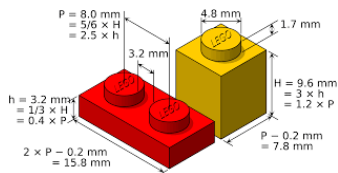
- Metadaten bereits bei der Genese und nicht erst bei der Archivierung erfassen
- Metadaten standardisieren
- Nachnutzen der Daten und Nachvollziehen der Änderungen ohne erheblichen Mehraufwand
- DOIs für Zwischenstände (zitierbar)

Ziel



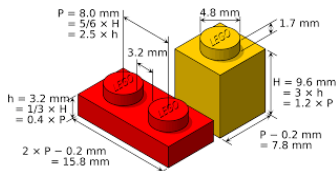
- Metadaten bereits bei der Genese und nicht erst bei der Archivierung erfassen
- Metadaten standardisieren
 - Nachnutzen der Daten und Nachvollziehen der Änderungen ohne erheblichen Mehraufwand
 - DOIs für Zwischenstände (zitierbar)

Ziel



- Metadaten bereits bei der Genese und nicht erst bei der Archivierung erfassen
- Metadaten standardisieren
- Nachnutzen der Daten und Nachvollziehen der Änderungen ohne erheblichen Mehraufwand
- DOIs für Zwischenstände (zitierbar)

Ziel



- Metadaten bereits bei der Genese und nicht erst bei der Archivierung erfassen
- Metadaten standardisieren
- Nachnutzen der Daten und Nachvollziehen der Änderungen ohne erheblichen Mehraufwand
- DOIs für Zwischenstände (zitierbar)

Lösung

Versionierungsprogramm

- bekannt aus der Softwareentwicklung
- Änderungen können schnell erkannt und nachvollzogen werden
- Warum wurde etwas geändert statt wann

Problem

- muss erlernt werden
- Programme, wie Git erzwingen zwar eine Beschreibung → Freitext

Lösung

Versionierungsprogramm

- bekannt aus der Softwareentwicklung
- Änderungen können schnell erkannt und nachvollzogen werden
- Warum wurde etwas geändert statt wann

Problem

- muss erlernt werden
- Programme, wie Git erzwingen zwar eine Beschreibung → Freitext

Lösung

Versionierungsprogramm

- bekannt aus der Softwareentwicklung
- Änderungen können schnell erkannt und nachvollzogen werden
- Warum wurde etwas geändert statt wann

Problem

- muss erlernt werden
- Programme, wie Git erzwingen zwar eine Beschreibung → Freitext

Lösung

Versionierungsprogramm

- bekannt aus der Softwareentwicklung
- Änderungen können schnell erkannt und nachvollzogen werden
- Warum wurde etwas geändert statt wann

Problem

- muss erlernt werden
- Programme, wie Git erzwingen zwar eine Beschreibung → Freitext

Lösung

Versionierungsprogramm

- bekannt aus der Softwareentwicklung
- Änderungen können schnell erkannt und nachvollzogen werden
- Warum wurde etwas geändert statt wann

Problem

- muss erlernt werden
- Programme, wie Git erzwingen zwar eine Beschreibung → Freitext

Lösung

Versionierungsprogramm

- bekannt aus der Softwareentwicklung
- Änderungen können schnell erkannt und nachvollzogen werden
- Warum wurde etwas geändert statt wann

Problem

- muss erlernt werden
- Programme, wie Git erzwingen zwar eine Beschreibung → Freitext



Inhaltsverzeichnis

- ① Problemstellung
- ② Ziel
- ③ Versionierung**
- ④ Projekt

Versionierung

Konventionell

20151216 document.txt
 20160415_document_bearb.txt
 20160420_document_dr_ke.txt
 20160423_document_neu.txt
 20160430_document_final.txt



Versioniert

commit 067ed126...
 Author: hermann
 Date: Wed Dec 16 08:11:29 2015
 final Version
 commit 8b53f4d5...
 Author: droessler
 Date: Tue Dec 15 09:02:53 2015
 Änderungen von Stefan
 commit d8f478e4...
 Author: hermann
 Date: Thu Nov 26 07:41:33 2015
 erste Version



Inhaltsverzeichnis

- ① Problemstellung
- ② Ziel
- ③ Versionierung
- ④ Projekt

Lösungsansätze

Ein Werkzeug, mit dem gearbeitet werden soll

- reduziertes Git für Nichtinformatiker
- GUI mit Auswahlfeldern und standardisierten Metadaten
- Vergabe von DOIs für wichtige Zwischenstände
- Einbindung in existierende Repositorien ?!?

Lösungsansätze

Ein Werkzeug, mit dem gearbeitet werden soll

- **reduziertes Git für Nichtinformatiker**
 - GUI mit Auswahlfeldern und standardisierten Metadaten
 - Vergabe von DOIs für wichtige Zwischenstände
 - Einbindung in existierende Repositorien ?!?

Lösungsansätze

Ein Werkzeug, mit dem gearbeitet werden soll

- reduziertes Git für Nichtinformatiker
- GUI mit Auswahlfeldern und standardisierten Metadaten
- Vergabe von DOIs für wichtige Zwischenstände
- Einbindung in existierende Repositorien ?!?

Lösungsansätze

Ein Werkzeug, mit dem gearbeitet werden soll

- reduziertes Git für Nichtinformatiker
- GUI mit Auswahlfeldern und standardisierten Metadaten
- Vergabe von DOIs für wichtige Zwischenstände
- Einbindung in existierende Repositorien ?!?

Lösungsansätze

Ein Werkzeug, mit dem gearbeitet werden soll

- reduziertes Git für Nichtinformatiker
- GUI mit Auswahlfeldern und standardisierten Metadaten
- Vergabe von DOIs für wichtige Zwischenstände
- Einbindung in existierende Repositorien ?!?

Replay-DH

Realisierung einer **Plattform** und begleitender Dienste zum Forschungsdatenmanagement für die Fachcommunity **Digital Humanities**

- Eins von sieben BW-Projekten zum Thema Forschungsdatenmanagement

Projektpartner

- Universität Ulm
 - Institut für Organisation und Management von Informationssystemen
 - Kommunikations- und Informationszentrum
- Universität Stuttgart
 - Universitätsbibliothek Stuttgart
 - Institut für maschinelle Sprachverarbeitung (Clarin-D-Zentrum)