CONQUAIRE

Enabling Git based research data quality control for institutional repositories!

Vidya Ayer, Christian Pietsch, Johanna Vomprnas, Jochen Schirrwagen, Cord Wiljes, Vitali Peil, Philipp Cimiano. CITEC, Bielefeld University, Germany

IG Repository Platforms for Research Data @RDA Plenary9
Friday, 07 April 2017, Barcelona, Spain.
CC BY-NC-SA 4.0 International License.
About

- CITEC + Bielefeld University Library
- Pilot research: 9 partners – interdisciplinary labs
Irreproducibility

- **Reproducibility** – basic principle of Science!
- **Irreproducibility::**
  - Nature survey – scientists (50-70%)
  - Psychology Reproducibility Project (61% fail)
  - Pharma clinical drug trials (82% fail)
- Reproducibility == Full Replication?
Analytical Reproducibility

- Observation
- Hypothesis
- Experiment
- Primary Data
- Computer-based Analysis
- Analyzed Data
- Research Data
- Analytic Reproducibility
- Full Replication

Publication
- Knowledge
- Interpretation
Goals

- Research Data Management (RDM) - generic infrastructure
- Support analytical reproducibility of research results
- Storage + DVCS Versioning → data sharing & reuse
- Data Quality = Open formats + Validation (raw, results) + scripts
- Continuous integration → quality data
Architecture

Version Control System (VCS):
stores versioned research data

Data Repository:
makes data publicly available for other researchers

Conquaire Server:
co-ordinates data storage + data quality evaluation

Quality Control Middleware:
evaluate data quality

Database:
stores reports + Logfiles

Researcher

upload research data

data quality reports
set up preferences

invoke quality evaluation
evaluation results

Quality Control Services

Q1 ... Qn
PUB!

Publications at Bielefeld University

- 5193 Publication References
- 3427 Individual Author Pages
- 179 Data Publications
- 9174 Open Access Publications
- 1853 Theses

Open Access at Bielefeld University
Research Data Management
PUB Theses
Date Seal of Approval

Fri-07Apr2017 | RDA-P9@BCN | Vid Ayer, Conquaire, CC-BY-NC-SA-4.0
PUB Overview

- **Management of Institutional research output:**
  - Scientific literature & research data at #UniBi
- **Built with LibreCat:**
  - Joint effort of Lund, Gent, Bielefeld libraries.
- **Provides:**
  - Author publication lists, supports ORCID
  - Mints DOI / URN for permanent, reliable citation
  - Interfaces (OAI, SRU, CQL)
  - Formats (DC, MODS, DataCite, XMetaDissPlus)
  - Linking of research data with literature
  - Partner in the SAFE Private LOCKSS Network - distributed preservation network.
  - Data Seal of Approval certification
Connecting Gitlab, Conquaire & PUB

e.g. research data, software

User

Gitlab
as JSON data with
* commits
* user info

Conquaire Webhook Listener

PUB Repository

send URL of HTML page

send web hook

generate HTML
* sum of JSON data
* CI test results
* current state

case a) quality control

search for tagged commits

case b) publishing

release for publishing

* mint DOI
* create landing page
* and metadata

feedback to user
Thank You!

Questions? Contact:

• Email: ayer@uni-bielefeld.de
• Twitter: @svaksha
• Email (project): conquaire-contact@lists.uni-bielefeld.de
• Website: http://conquaire.uni-bielefeld.de