

CONQUAIRE

Continuous quality control for research data - a first experiment

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

Open Science Conference

Tuesday, 21 March 2017, Berlin.
CC BY-NC-SA 4.0 International License.

About

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

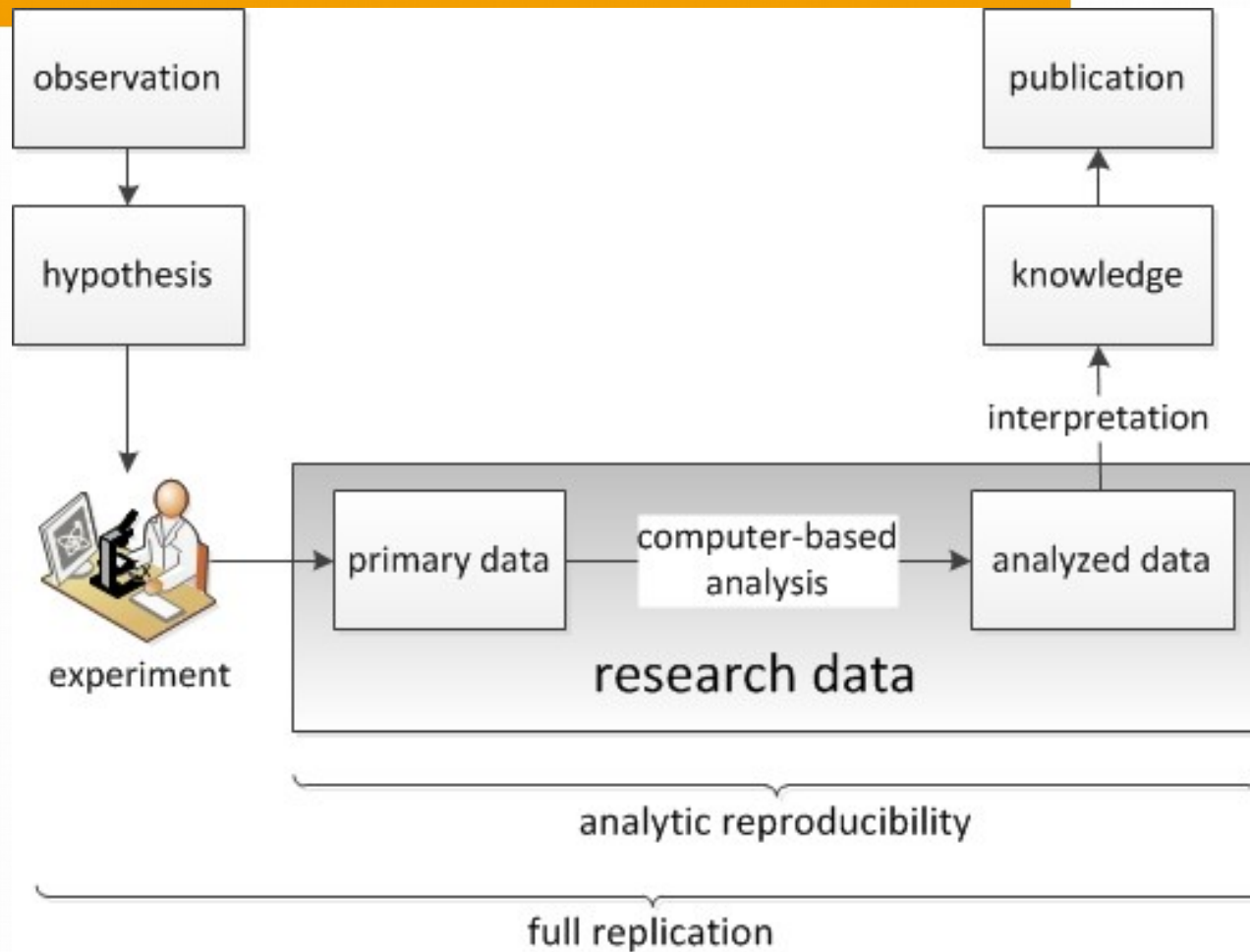
Disciplines

Computer
Chemistry
Applied Sports
Science
Linguistics
Neurobiology
Psychology
Economics
Computational Biology

(Ir)Reproducibility

- **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 ?

Research Workflow



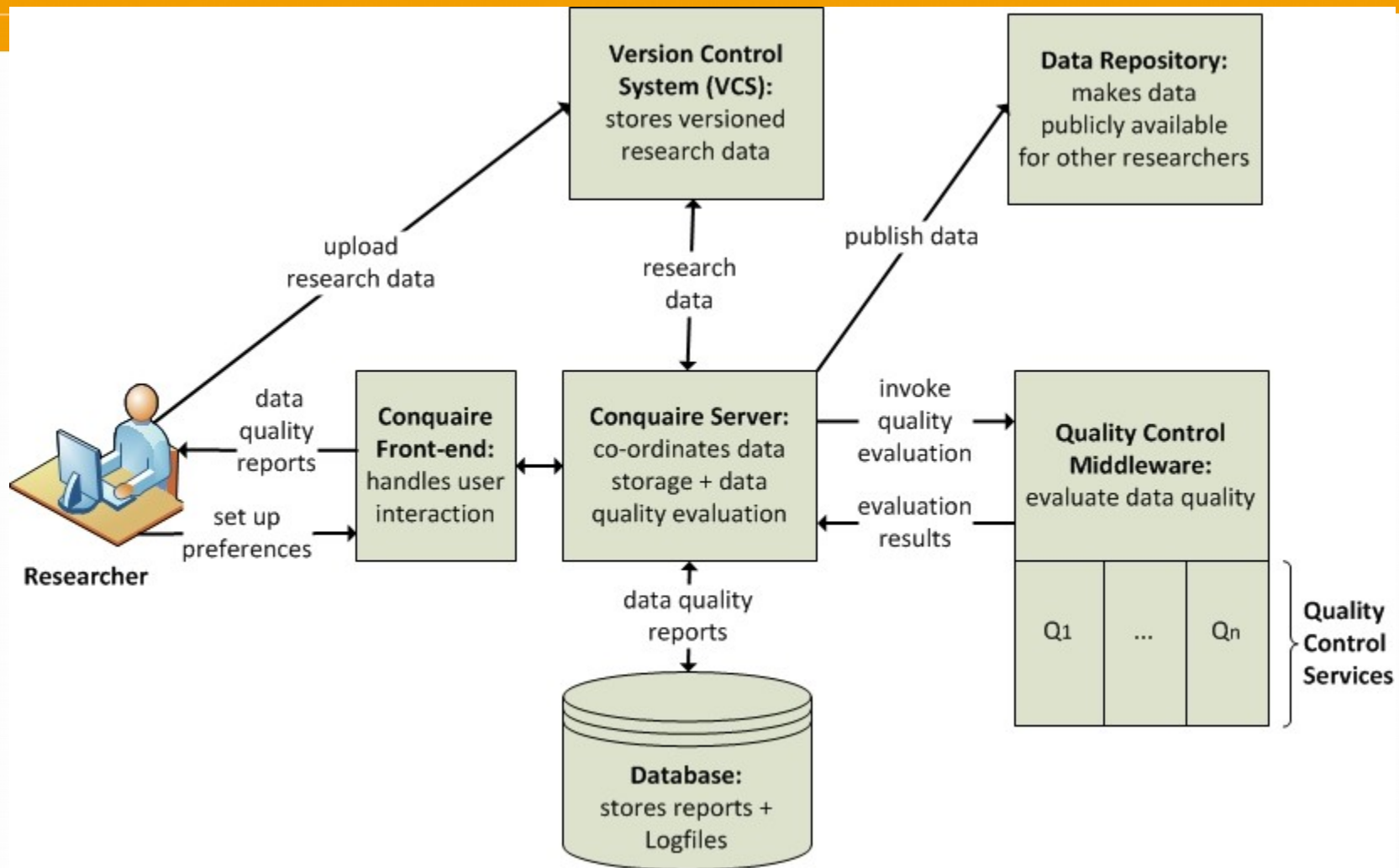
Computational Reproducibility

- Analytical Reproducibility == Mathematical !
- Continuous Quality Analysis - research data
- **Challenges** :: Research **diversity** (data, discipline, tools, formats, workflows, etc..)

Goals

- Research Data Management System (RDMS) - generic infrastructure architecture
- Support computational/ analytical reproducibility of research data
- Storage + DVCS Versioning – data sharing/ reuse
- Data Quality = Open formats + Validation (raw, results) + scripts
- Continuous integration - quality data

Architecture



First Implementation

- **Quality Analysis (.csv sample file):**
 - data type matching
 - checks out-of-range numeric values & NaN/ Null values.
 - records (one per line)
 - column headers
 - field delimiters (reserved character: comma, semicolon, or tab)
 - character set - ASCII, Unicode character sets (e.g. UTF-8)
- **CI Runners:** Gitlab's CI runners monitor commits made into Gitlab.
- **Flask:** Front-end displays the CSV quality results, emails commit results to users.

Thank You!



Contact

- Email: ayer@uni-bielefeld.de
- Email (project list):
conquaire-contact@lists.uni-bielefeld.de
- Website: <http://conquaire.uni-bielefeld.de>