TECHNISCHE UNIVERSITÄT DRESDEN
Elektrotechnik & Informationstechnik, Institut für Automatisierungstechnik, Professur Prozessleittechnik

LDQ 2014

R43ples – Revisions for Triples

An Approach for Version Control in the Semantic Web

Markus Graube, Stephan Hensel, Leon Urbas

Leipzig, 02.09.2014

ComVantage

Professur für PROZESS LEITTECHNIK

DRESDEN concept Exzellenz aus Wissenschaft und Kultur

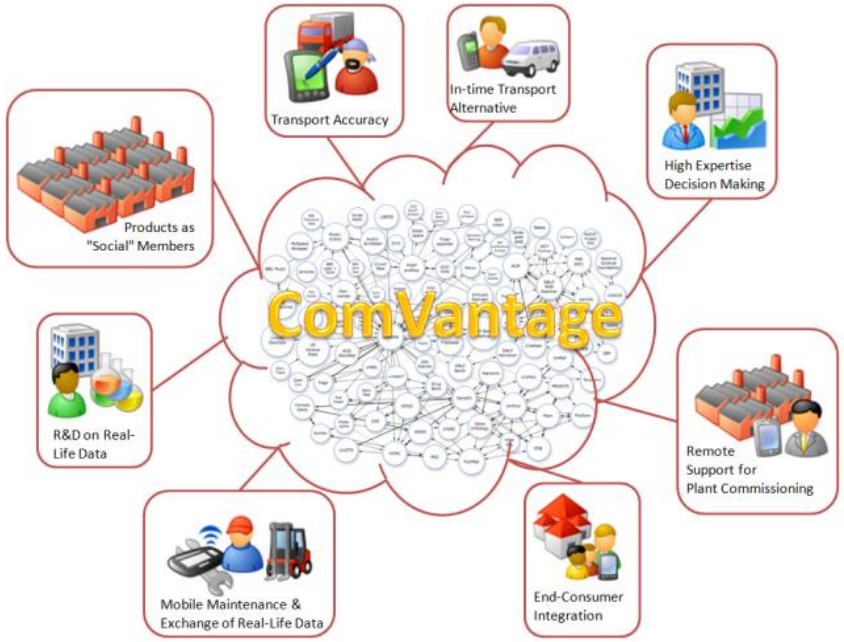
ComVantage - Motivation

Project-based collaborations…

- enable extended services
- require flexible, interlinked information models and
- trustful and convenient access to the information

ComVantage

- EU FP7 Project (2011 -2014)
- product-centric collaboration space for dynamic and flexible information exchange
- build on top of best practices from the Web for providing product-centric and workflow-based mobile apps



Revision Control System - Motivation

Professur für PROZESS LEITTECHNIK

Revision Control for Linked Data

- Record Changes and revise old version
- Track quality
- Branching / Merging
- Open nature of LD as challenge
- Complete checkout not possible

Approaches

- Temporal RDF
- File based (partly with Git, SVN, …)
- Reified Statements (Auer and Herre, 2007)
- Aggregated Deltas (Im et. Al., 2012)
- …

R43ples – Revision for Triples

Professur für PROZESS LEITTECHNIK

Concept

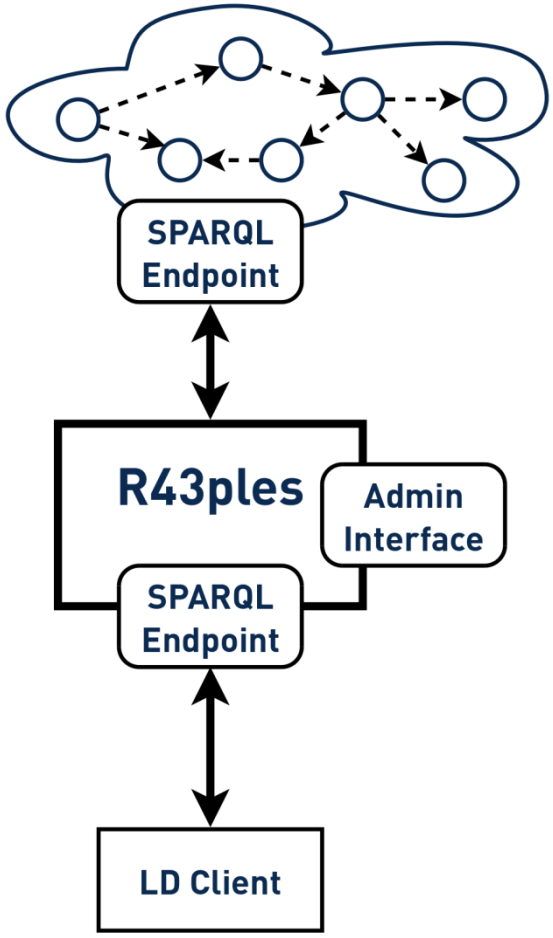
- Revision Control on Graph level
- Based on (Vander Sande et. al., 2013)
- R43ples as proxy SPARQL Endpoint
- Generation of old revisions on demand

Information space on external triplestore

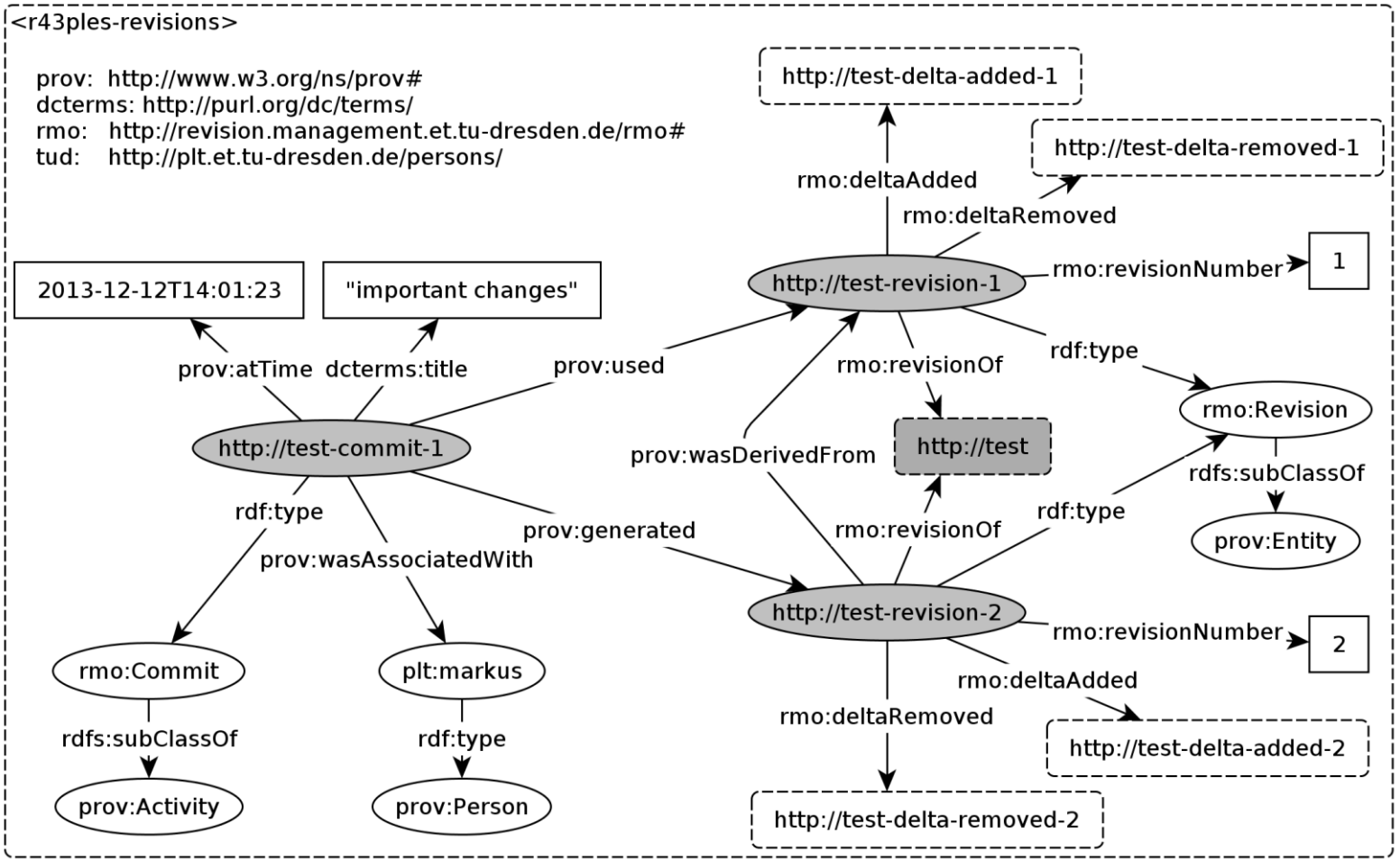
- Deltas in Named Graphs
  - Add-Sets
  - Delete-Sets
- Revision Information in additional graph
- Revision Management Ontology (RMO)

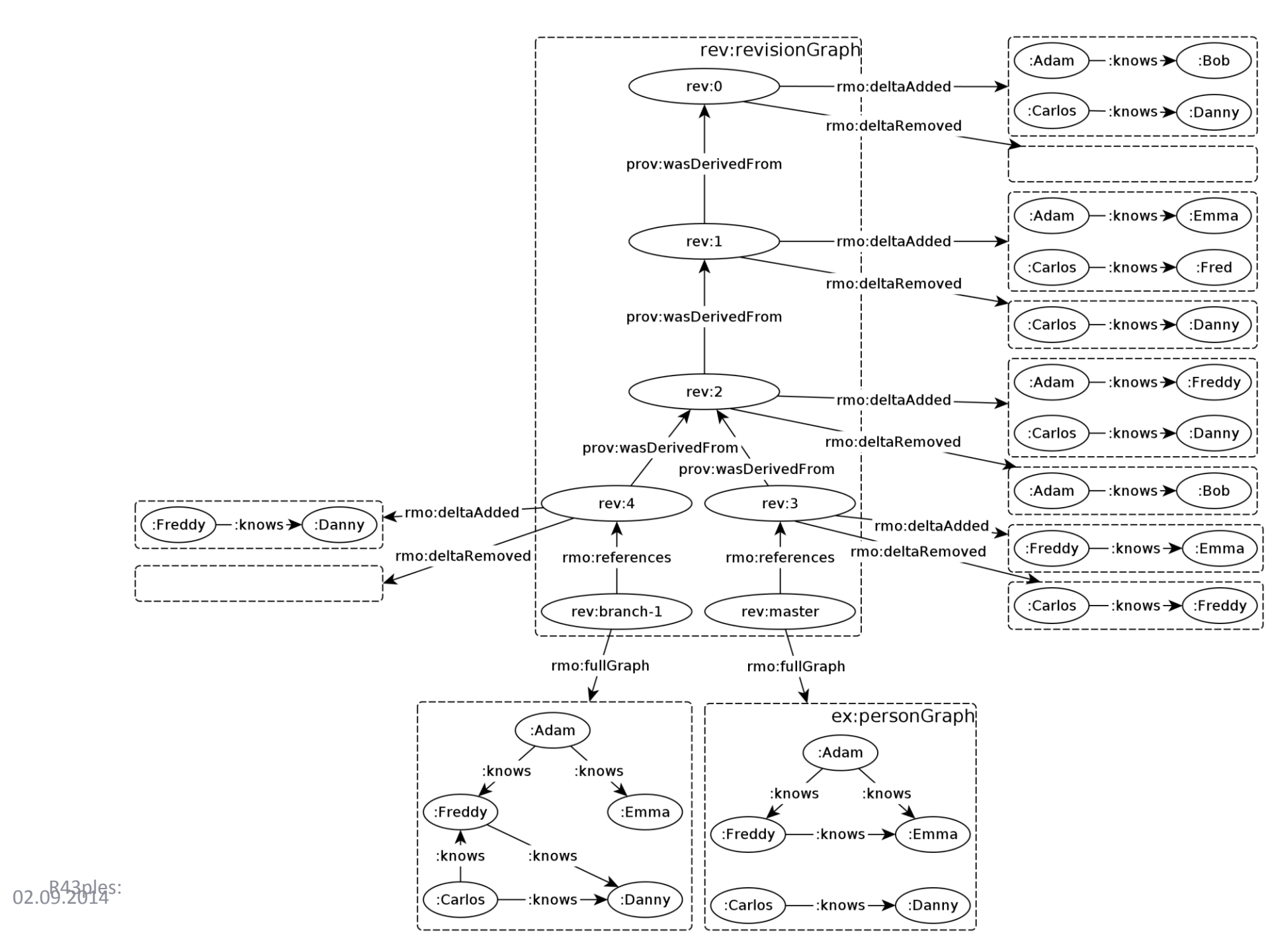
Some revisions as full graphs

- Tag, Branch, Master



R43ples – Revision Model




rev:revisionGraph
rev:0
rmo:deltaAdded
rmo:deltaRemoved
:Adam
:knows
:Bob
:Carlos
:knows
:Danny
prov:wasDerivedFrom
rev:1
rmo:deltaAdded
rmo:deltaRemoved
:Adam
:knows
:Emma
:Carlos
:knows
:Fred
prov:wasDerivedFrom
:Carlos
:knows
:Danny
rev:2
rmo:deltaAdded
rmo:deltaRemoved
:Adam
:knows
:Freddy
:Carlos
:knows
:Danny
prov:wasDerivedFrom
prov:wasDerivedFrom
rev:4
rev:3
rmo:deltaAdded
rmo:deltaRemoved
:Adam
:knows
:Bob
:Freddy
:knows
:Danny
rmo:deltaAdded
rmo:deltaRemoved
:Freddy
:knows
:Emma
rmo:references
rmo:references
:Carlos
:knows
:Freddy
rev:branch-1
rev:master
rmo:fullGraph
rmo:fullGraph
:Adam
:knows
:knows
:Freddy
:Emma
:knows
:knows
:Carlos
:knows
:Danny
ex:personGraph
:Adam
:knows
:knows
:Freddy
:knows
:Emma
:Carlos
:knows
:Danny

R43ples – SPARQL Interface

Professur für PROZESS LEITTECHNIK

Select queries

- SELECT * FROM <graph> **REVISION „23“** WHERE {?s ?p ?o.}
- SELECT * WHERE {
  GRAPH <xyz> **REVISION „master“** { ?s ?p ?o.}
  }

Update

- **MESSAGE „commit xyz“**
  INSERT INTO <graph> **REVISION „tag-v0.8“**
  { <a> <b> <c>. }

Branching/Tagging

- BRANCH GRAPH <xyz> REVISION „23“ TO <new-branch>
- TAG GRAPH <xyz> AS „new-tag“

R43ples – Interface

Professur für PROZESS LEITTECHNIK

Semantic description allows other client to recreate old revisions

$$TG_{g,r} = g_{\text{nearestBranch}} + \sum_{\text{revision } i=r}^{\text{nearestBranch}} (deleteSet_{g,i} - addSet_{g,i})$$

R43ples Endpoint transparent for SPARQL clients

- Default revision = MASTER

SPARQL 1.1. Service Description

- Extension of *sd:feature* with *sd:r43ples*

R43ples – Implementation

Professur für PROZESS LEITTECHNIK

Proxy-Server

- Current Master revision in HTTP header response
- SPARQL as interface to triplestore

Java

- Jersey (REST)
- Grizzly as webserver
- (Jena)

Open Source

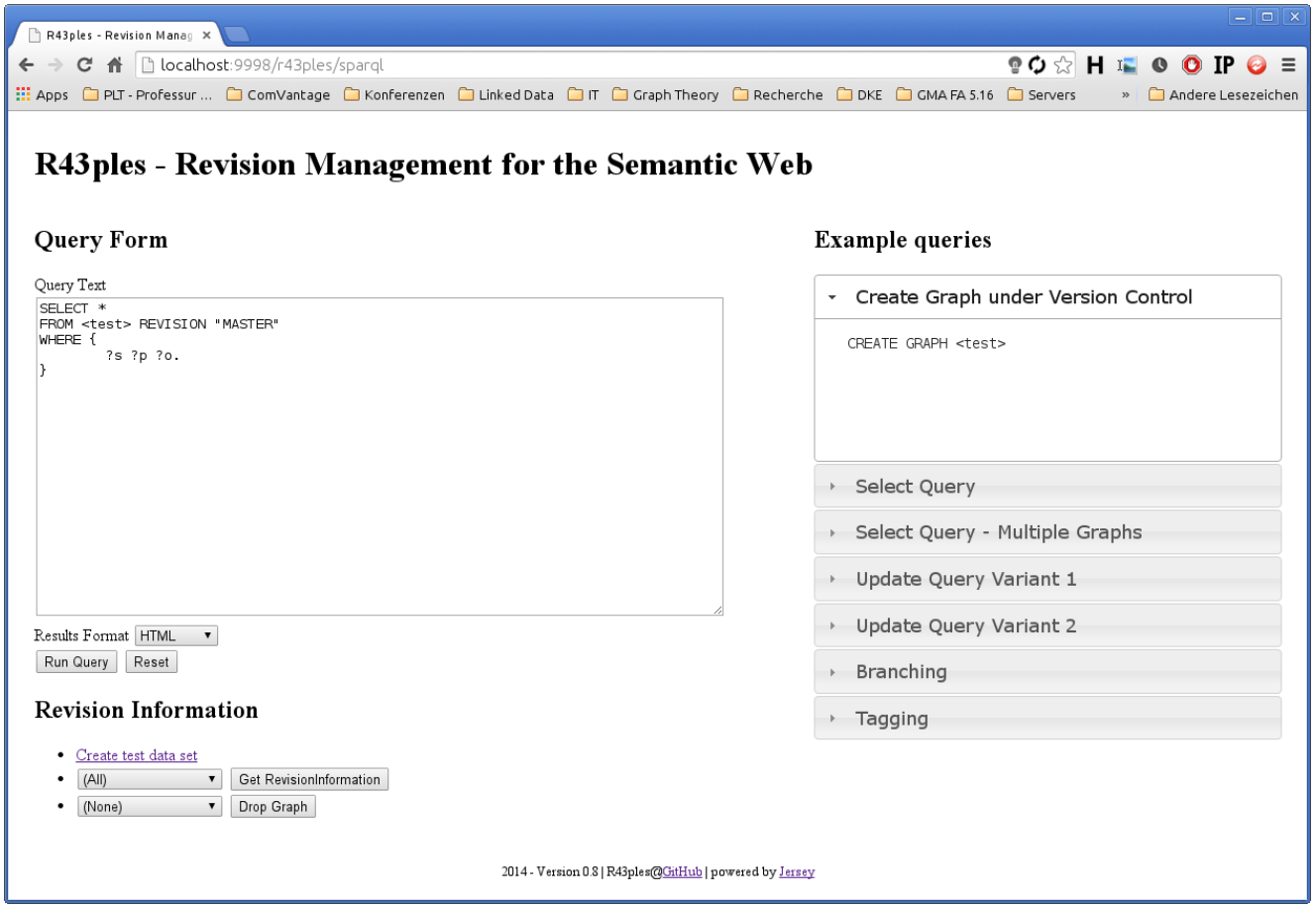
- EUPL licence
- GitHub: https://github.com/plt-tud/r43ples

Test System

- http://eatld.et.tu-dresden.de:9998/r43ples/sparql

R43ples – Implementation

Professur für PROZESS LEITTECHNIK



R43ples - Evaluation

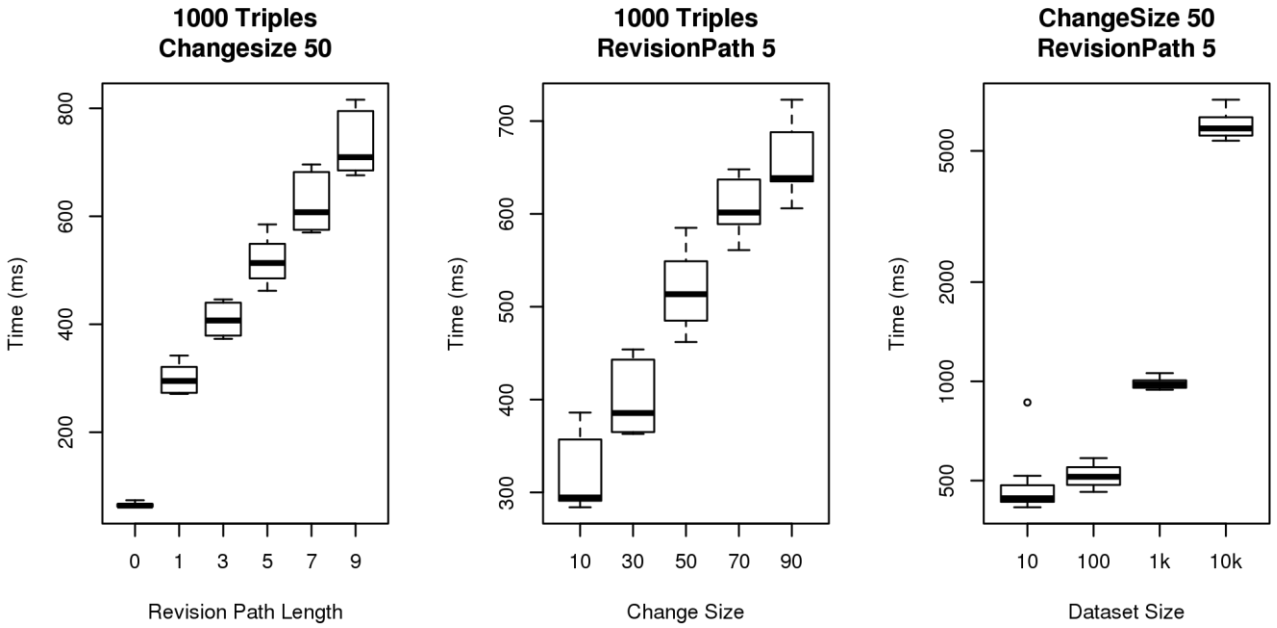
Professur für PROZESS LEITTECHNIK

Response time dependant on

- dataset size
- revision path length
- size of changes

Additional Storage overhead

- 12 triples + changes for update
- Full graph for tag and branches



Current Work in Progress

Professur für PROZESS LEITTECHNIK

Join of revision and change sets in SPARQL query

- No generation of temporary graph
- Pre-process original SPARQL query
  - Remove Revision Information
  - Embed full graph and changesets
- Perform query direct on triplestore

Merging Functionality

- Interface (Conflict Model)
- User Support

Graphical Representation of Revision Tree

- GraphViz

Conclusion

Professur für PROZESS LEITTECHNIK

Revision Control on Graph level

- Change sets in named graphs
- Blank Nodes problem

Semantic Description

- No implicit knowledge necessary

SPARQL interface

- Extendend keywords

R43ples Endpoint transparent for SPARQL clients

- Default-revision = MASTER

Performance depandant on dataset size

- Alternative approach under evaluation

Thank you for your attention!

markus.graube@tu-dresden.de

Professur
für **PROZESS**
LEITTECHNIK

Prof. Dr. habil. Leon Urbas

Technische Universität Dresden
Fakultät Elektrotechnik und
Informationstechnik
Institut für Automatisierungstechnik

Tel.: +49 351 463-34604
Fax: +49 351 463-39681

Besucheradresse:
Barkhausen-Bau
Georg-Schumann-Str. 11
01187 Dresden

Postanschrift (Briefe):
Technische Universität Dresden
Fakultät Elektrotechnik und
Informationstechnik
Institut für Automatisierungstechnik
01062 Dresden