

TRACKS

[Tasks, Code Reviews, Activities, Source Code, Knowledge about Software]

Data of Software Development

- Tracks of the development process
- Available from tools used in development
- Loosely coupled across these tools
- Repetitive data mining and collection
- Need to be linked together for reasoning



Reason across the data sources

Relations between projects, tasks, developers and their activity

Keep on tracks of software development

Single data source

Disambiguated entities

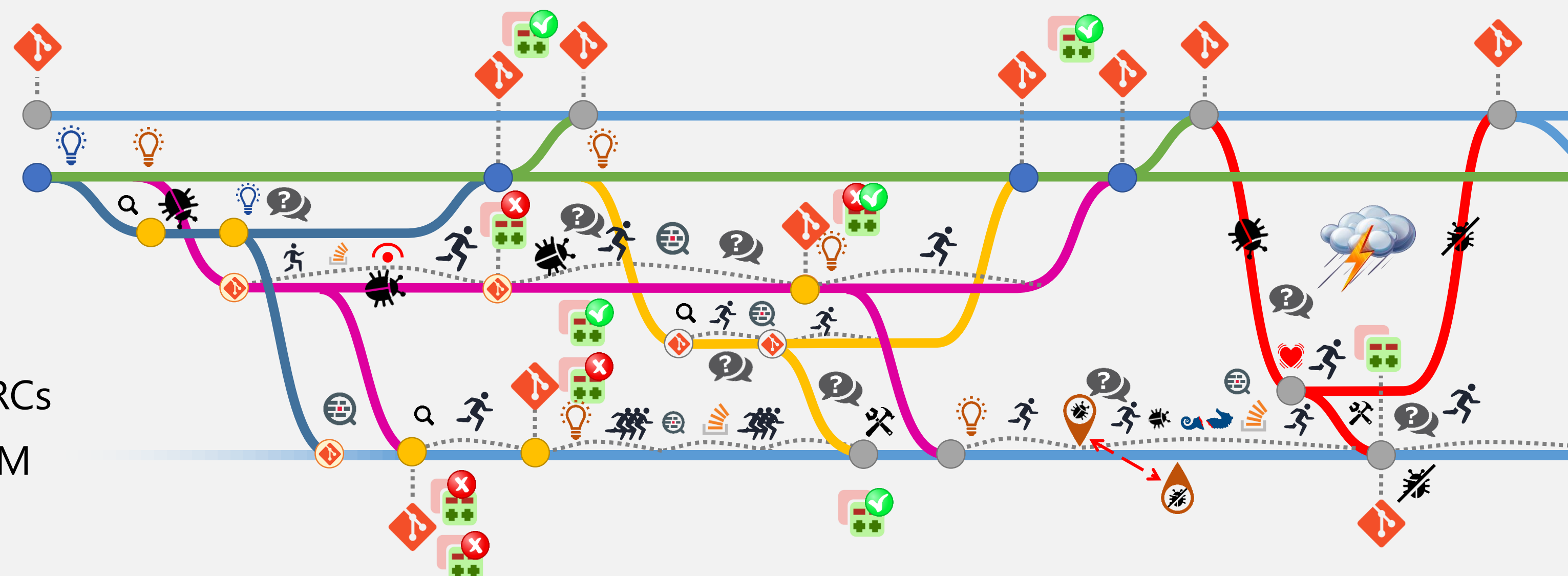
Open-sourced, contribute yourself

Be part of it

Join us now!

Data Sources

- Git, GitHub
- Bugzilla, Jira
- Gerrit, CodeFlow
- Mylyn, DevACTs
- StackOverflow, IRCs
- Eye Tracking, HRM



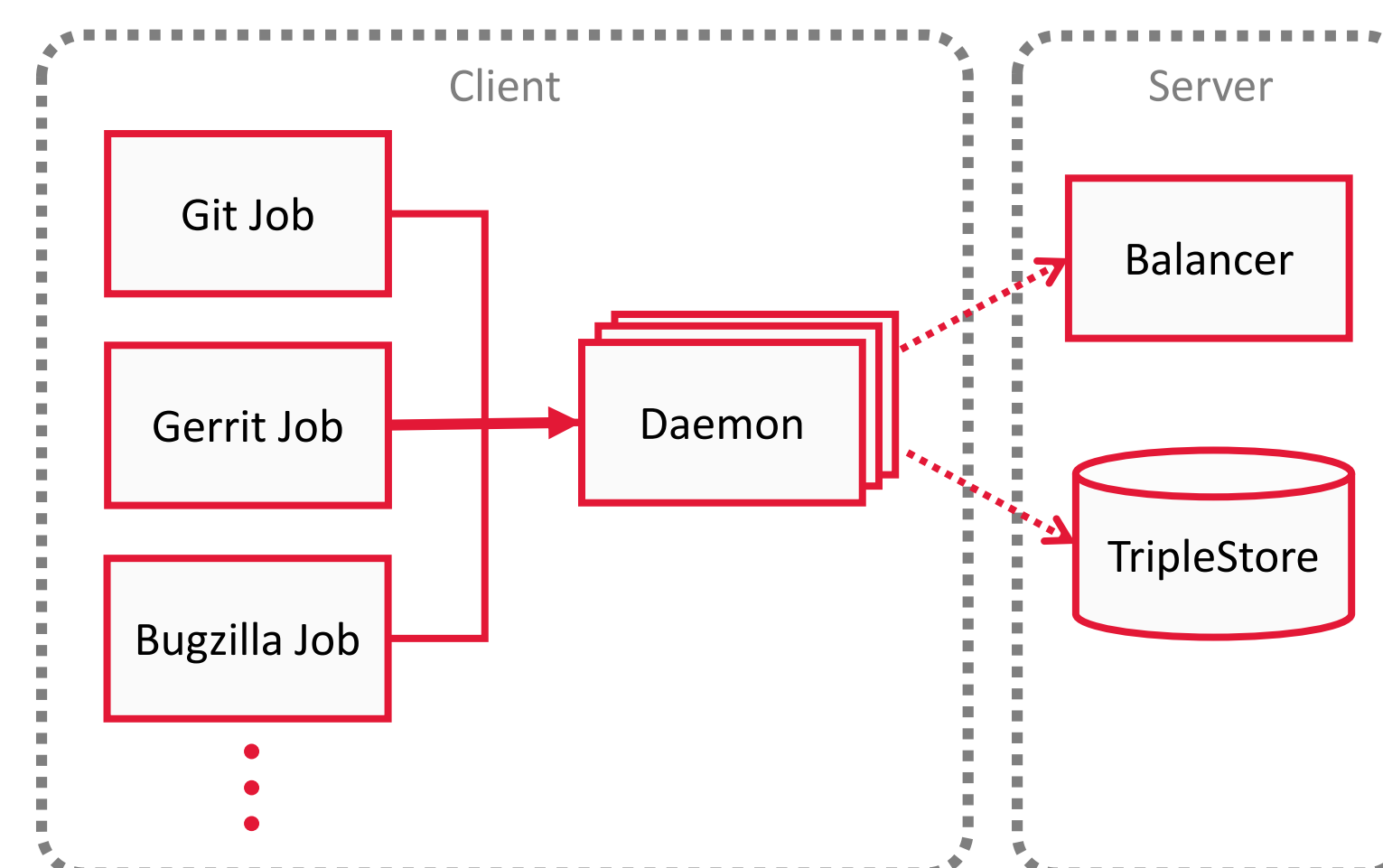
Solution Linked Data of Software Development

- Collecting data from all of the data sources
- Ontologies for each of the data sources
- Collected data transformed into RDF triples
- Linking ontologies between each other
- Based on same e-mails, IDs, projects, artifacts, etc.

Platform Main Features

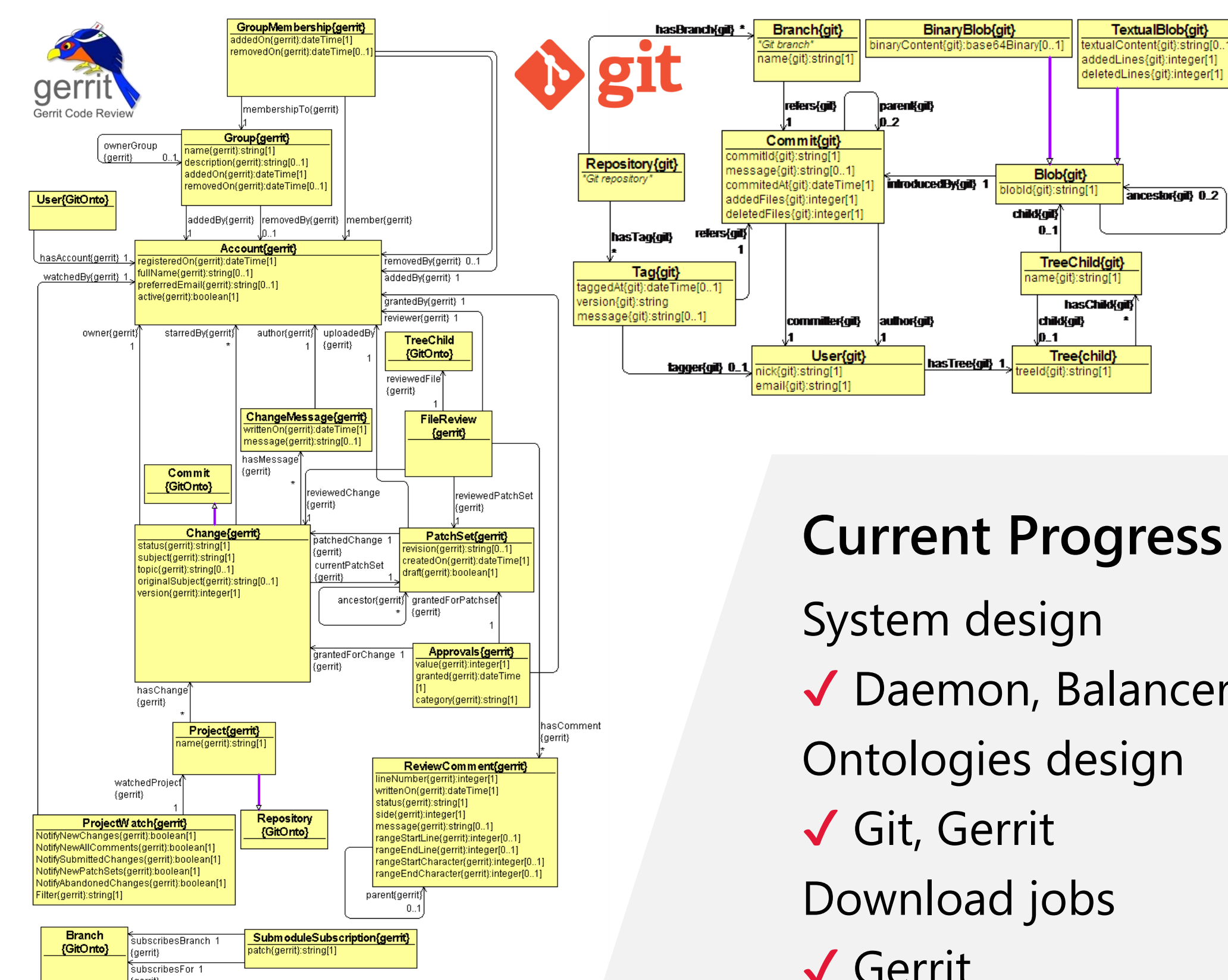
- Distributed and gentle data collection
- Incremental collection – up-to-date data
- Linked Data representation – ontologies
- Support for static and stream reasoning
- Extensibility – add your own tool

Architecture



- Botnet-like network of Daemons
- Jobs assigned by the Balancer
- Storing RDF triples in the TripleStore

Ontologies



Current Progress

System design

✓ Daemon, Balancer

Ontologies design

✓ Git, Gerrit

Download jobs

✓ Gerrit

"There's so much data of software development, though sparse, scattered, not linked together." – "Let's put them all together!"

On the Tracks of Software Development

Martin Konôpka, Karol Rástočný | Supervisors: Pavol Návrát, Mária Bieliková

