Scientific Software Engineering: Mining Repositories to gain insights into BACARDI

Lynn von Kurnatowski¹, Martin Stoffers¹, Martin Weigel², Michael Meinel¹, Yi Wasser², Kathrin Rack¹, Hauke Fiedler²

German Aerospace Center (DLR)
¹Institute for Simulation and Software Technology
²Space Operations and Astronaut Training
BACARDI – Backbone Catalogue of Relational Debris Information

BACARDI is a software platform to register and track orbital objects like space debris and satellites.

Background information:

- software was initially developed as a demonstrator
- during this phase the software was developed without a formal development process in place
- after the successful development of a prototype
  => the project has been continued as a long-term project
- to balance the character of a research project and operational requirements
  a software engineering process has been designed and applied
Motivation

Highly distributed project and working environment
- 2 Institutes
- 4 Departments
- 3 physical locations
- 4-6 team members

Different personal focus on the project based upon:
- research goals
- personal background

Communication difficulties due to:
- technologies available to team members
- distributed workplaces of team members
BACARDI – Software Engineering Process

Describes how and where issues are handled

Planning using Issue Tracker

Assign Issue → Design → Coding/Unit testing → Integration Testing → (Review) → Resolve Issue

Prepare work:
- Update from VCS
- Checkout feature branch

Daily Work:
- Run local tests
- Commit to VCS (Mention issue ID)

Resolved when:
- DoD checklist OK
- Code Review OK
- Successful integration build

Reviewer is assigned, who checks the DoD

Definition of Done:
- Review the new code against code guidelines
- Ensure that new tests are adequate
- All documentation is available
- New code is correct

Reviewer considers all discussion as resolved, the merge request is approved

Team members pick the issues they want to implement themselves freely
Can we measure a change?
A first Pattern

Number of commits from July 2015 to September 2019
Solution:
Mining Repositories
Mining Software Repositories

- **Repositories**
  - Version Control
  - Issue Tracker
  - Merge Request

- **Common Data Storage**

- **Data Integration and Provisioning**

- **Analysis and Usage**
  - Graph Analytics
```
MATCH (n:MergeRequest)
WITH* WHERE '2018-01-01' <= n.created_at <= '2018-03-31'
WITH* WHERE
  ((n)-[:IS_ASSIGNED]->(:User)) and
  ((n)-[:APPROVED_BY]->(:User)) and
  ((n)-[:WAS_ASSIGNED]->(:User)) and
  ((n)-[:MERGED_BY]->(:User)) and
  ((n)-[:CREATED_BY]->(:User))
WITH* WHERE n.approved = true and n.state = 'merged'
WITH*
MATCH (u1:User)<-[::CREATED_BY]-:(n)-[::APPROVED_BY]-(u2:User)
WHERE u1.name <> u2.name
WITH*
MATCH (n)-[:ISRELATED]->(:Issue)
WITH*
MATCH (n)-[:HAS_MILESTONE]->(m:Milestone)
WHERE n.merged_at <= m.due_date
RETURN count(*)
```
Results Issue Tracker

MATCH (n:Issue)
WHERE
  '2018-01-01' <= n.created_at <= '2018-03-31'
WITH*
WHERE
  n.weight IS NOT NULL
RETURN count(*)
Results Communication

MATCH (n:MergeRequest)-[:HAS_NOTE]->(o:Note)
WHERE '2018-01-01' <= n.created_at <= '2018-03-31'
RETURN count(o)

MATCH (n:Issue)-[:HAS_NOTE]->(o:Note)
WHERE '2018-01-01' <= n.created_at <= '2018-03-31'
RETURN count(o)
Summary

• More regular and direct communication due to process
• Better introspection into different parts of the software for each team member
• General acceptance of the process within the team
And now?
Future Work

• Utilize metrics within CI build to provide developer feedback
• Find methods to raise the acceptance of the software development process
• Extended the new concept of GitLab2PROV
GitLab2Prov

GitLab2PROV generates retrospective provenance of GitLab projects

• Utilizes the W3C PROV standard
• Includes provenance models for
  • Git Commits
  • GitLab Issues
  • GitLab Merge Requests
GitLab2Prov

PROV Activity Count

- **ssas/BACARDI**
  - Jan 2016 to Jul 2020
  - 10k-100k activity count

- **ssas/RK4-Propagation**
  - Jul 8 2018 to Aug 5 2020
  - 50-150 activity count

- **ssas/ld-code-optimisation**
  - Mar 2019 to Mar 2020
  - 50-150 activity count

- **ssas/tracklet-correlation**
  - May 2019 to Jul 2020
  - 1000-3000 activity count

Legend:
- Green: commit-related
- Cyan: issue-related
- Black: merge-request-related
Thanks!

@DLR_software

https://www.dlr.de/sc/
https://www.dlr.de/rb/