Alambic: An Open-Source Platform for Software Engineering Data Management
The Case of Embedded Software Development

Boris Baldassari
boris.baldassari@castalia.solutions
http://castalia.solutions

Check the full article!
Plan

- Software quality, open data, and Alambic
- Alambic
  - Introduction: why Alambic
  - Architecture
  - Plugins: MetricsGrimoire, Hudson CI, PMD, Stack Overflow
- Use Cases
  - The PolarSys dashboard
  - The Eclipse forge
- Alambic 3.0 – one step forward
- Conclusion & Perspectives
Software quality, open data and Alambic

**Software quality** has different meanings for different contexts:

- Open-source projects (e.g. includes community attributes),
- Embedded software (e.g. stresses reliability and testability),
- Desktop software (e.g. stresses usability and maintainability).

► Defining a **quality model** helps people realise what quality is in their context, and how to measure and improve it.

**Transparency** (open data) is a key feature of software development.

- For open-source projects,
- For proprietary projects – Agile methodology explicitly promotes it.
Why Alambic?

- A fairly big amount of data is generated along the development process, including requirements, code, issues, builds, delivery, support, and more...

- Alambic automatically retrieves, curates, and organises information from these data sources and presents it to the user in a unified vision:
  - Dedicated plugins can run checks, actions, and reports.
  - Visualisations show various aspects of the project behaviour.
  - Recommendations are proposed to improve practices and quality.

- All data is made available in a neat format (JSON, CSV) so data enthusiasts can easily play with it, and use external tools for reporting, visualisation, data mining, post-processing.
Alambic – Architecture

Git / SVN

scm_commits_1M
scm_committers_1M

Hudson / Jenkins

ci_jobs_failed_1M
ci_builds_vol_1M

Bugzilla / Jira / GitHub

its_opened_1M
its_closed_1M

Forums / MLs

mls_threads_1M
mls_authors_1M

Compute model
Publish results

scm_commits_1M
scm_committers_1M
ci_jobs_failed_1M
ci_builds_vol_1M
its_opened_1M

Compute model
Publish results

scm_commits_1M
scm_committers_1M
ci_jobs_failed_1M
ci_builds_vol_1M
its_opened_1M
Alambic outputs

Alambic provides:

- A dashboard to browse all information online.
- Downloads for metrics, attributes, recommendations (JSON, CSV).
- Exportable figures: PNG/SVG, HTML snippets, interactive plots.
- Visualisations: quality model, data, history, specific data sources analysis.

Alambic 3.0

- PDF reports.
- Plotly graphics for online editing.
Alambic plugins – Hudson CI engine

This plugin connects to a Hudson CI instance and provides

- Metrics
  number of jobs, jobs failed/successful/unstable, builds failure rate..
- Figures
  builds history, jobs results
- Data
  builds, jobs (JSON, CSV)
- Recommendations.

![Graph showing Hudson CI Builds for modeling.mdt.papyrus](image1)

![Graph showing CI builds for modeling.mdt.papyrus](image2)
A set of plugins to retrieve information from the Eclipse dashboard [EDA].

Community metrics are computed using Bitergia's MetricsGrimoire [BMG], a generic tool to analyse software development repositories and tools metadata.

- **Issue Tracking System** (ITS)
  - Metrics: its_closed_7, its_closed_30, its_closed_365, its_bugs_closed, its_bugs_changed, …
  - Figures, visualisation
  - Recommendations

[EDA] [http://dashboard.eclipse.org](http://dashboard.eclipse.org)

[BMG] [https://metricsgrimoire.github.io](https://metricsgrimoire.github.io)
A set of plugins for community metrics.

- **Software Configuration Management (SCM)**
  - Metrics: `scm_commits_7`, `scm_commits_30`, `scm_commits_365`, `scm_authors_7`, `scm_committers`, `scm_files`, …
  - Figures, visualisation
  - Recommendations

- **Mailing Lists (MLS)**
  - Metrics: `mls_senders_7`, `mls_senders_30`, `mls_senders_365`, `mls_sent_7`, `mls_threads`, `mls_senders_response`, …
  - Figures, visualisation
  - Recommendations
Alambic plugins – PMD Analysis & Configuration

A plugin to help users understand the output of PMD and configure the tool.

- **PMD Analysis**
  - Get a summary of violations found by PMD
  - Identify the most important violations to resolve.

- **PMD Configuration:**
  - Use only most important rules, with fewer violations.
  - Target *conscious* zero-default.
**Use Cases: The PolarSys dashboard**

**PolarSys** is an Eclipse Industry Working Group that provides an open-source tool chain for embedded systems.

A maturity assessment task force was initiated to ensure the overall consistency of projects regarding quality:

- Define a quality model with specific quality concerns
- Define means to measure quality attributes
- Define presentation of data (dashboard)
Use Cases: The PolarSys dashboard
Use Cases: The PolarSys dashboard
Use Cases: The PolarSys dashboard
Use Cases: The PolarSys dashboard
Use Cases: The Eclipse forge

Use the Eclipse Foundation's infrastructure to retrieve information from the various data sources and tools of the Eclipse forge:

- The PMI for project details (description, web site, downloads, etc.)
- Issue tracking (Bugzilla)
- Configuration Management (Git)
- Communication: Mailing lists and forums
- Hudson CI, PMD..

Also use external sources:

- Stack Overflow questions about the project
Use Cases: The Eclipse forge
Use Cases: The Eclipse forge

Alamblc dashboard for Eclipse - v2.0 (Beta)

StackOverflow Questions
This section shows information about 50 posts concerning the tag `egit` during the last 5 years (i.e. from 2011-05-05 to 2016-05-19).

Quick Summary
- There is a total of 1813 questions and 2438 posts related to `egit`.
- Last activity registered was on 2016-05-39 15:32.
- 712 of them (79%) have an accepted answer.
- 175 actually have answers but none of them is accepted.
- 126 have no answer at all.
- There is an average of 1.41 answers to questions.

The wordcloud on the right details frequent words used in the question titles. The size and colour of the words depend on their number of appearances in questions. It shows what are the concerns and issues of people when dealing with the project.

Recent questions waiting for an answer
The following recent questions (less than 1 month old) have been asked more than 2 days ago and still have no answer. Would you like to check them?

<table>
<thead>
<tr>
<th>Question</th>
<th>Created on</th>
<th>Score</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eclipse cdt Egit commit run configuration</td>
<td>2016-05-05 10:12:57</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Export deployable plugin-ins and fragments for projects stored in Git repository from Eclipse Kepler</td>
<td>2016-05-06 20:40:53</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Eclipse egit cannot not detecting changed files or add new file</td>
<td>2016-05-12 16:39:01</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Eclipse Git - Making changes at the same time</td>
<td>2016-05-14 13:33:27</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>Commit and Push keepa making new branch (Eclipse)</td>
<td>2016-05-16 01:51:15</td>
<td>0</td>
<td>16</td>
</tr>
</tbody>
</table>
Alambic 3.0 adds many features:

- **Architecture / Use Cases**
  - More data-centric, positioned as a central data management platform.
  - Address new use cases: DevOps, system engineering.
  - Address new forges: GitHub, Apache.

- **Plugins**
  - New types of plugins: post and global triggers for reports
  - Wizards to create new projects: Eclipse PMI, GitHub, Apache
  - Tools: easily use external tools (e.g. R for analysis and reporting)
Conclusion & Perspective

Alambic project:

• Home:  
  http://alambic.castalia.camp

• Source, issues, documentation:  
  http://bitbucket.org/BorisBaldassari/alambic

Check the live instances:

• Polarsys dashboard: http://dashboard.polarsys.org
• Eclipse forge dashboard: http://eclipse.castalia.camp

Check the full article!  

Castalia Solutions