The four elements of achieving research software sustainability for long tail projects

Stephan Druskat[1,2,4] https://orcid.org/0000-0003-4925-7248

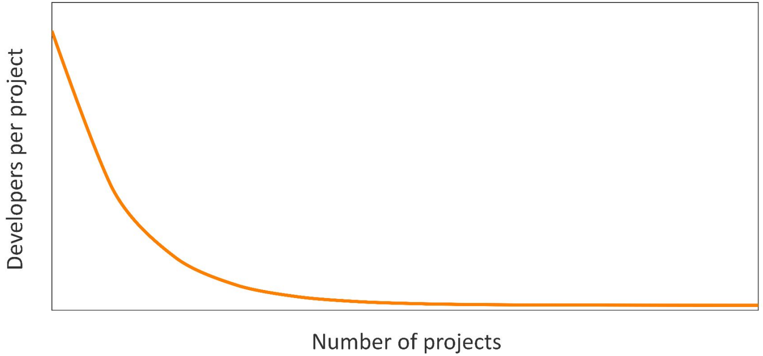
Thomas Krause[3] https://orcid.org/0000-0003-3731-2422

[1]German Aerospace Center (DLR), Institute for Software Technology; stephan.druskat@dlr.de
[2]Humboldt-Universität zu Berlin, Department of Computer Science
[3]Humboldt-Universität zu Berlin, Department of German Studies and Linguistics; thomas.krause@hu-berlin.de
[4]Friedrich Schiller University Jena, Department of English Studies

DOI 10.5281/zenodo.7654778

Q: How to sustain small research software projects?

“The long tail of science”



Our domain: projects with a bus factor → 1.

Approach: potential for technical sustainability + documentation + open infrastructures + maintenance strategy

Druskat, Stephan; Krause, Thomas; Lüdeling, Anke; Gast, Volker (2019): Infrastrukturstrategien für nachhaltige Forschungssoftware in befristeten Projekten. Poster. **deRSE19**. https://doi.org/10.6084/m9.figshare.11277764.v1

H1: “A minimal infrastructure for the sustainable development and provision of [small research software] consists of four elements”:

1. Develop for **technical sustainability**
2. **Documentation** as first class output
3. Use **existing infrastructures**
4. **Maintenance** strategy

Technical sustainability (a.k.a. follow good practice)

issues

pull-requests

mature-technologies

tdd

user-documentation

modularization

portability

continuous-integration

testing

developer-documentation

licensing

automated-builds

reproducible-builds

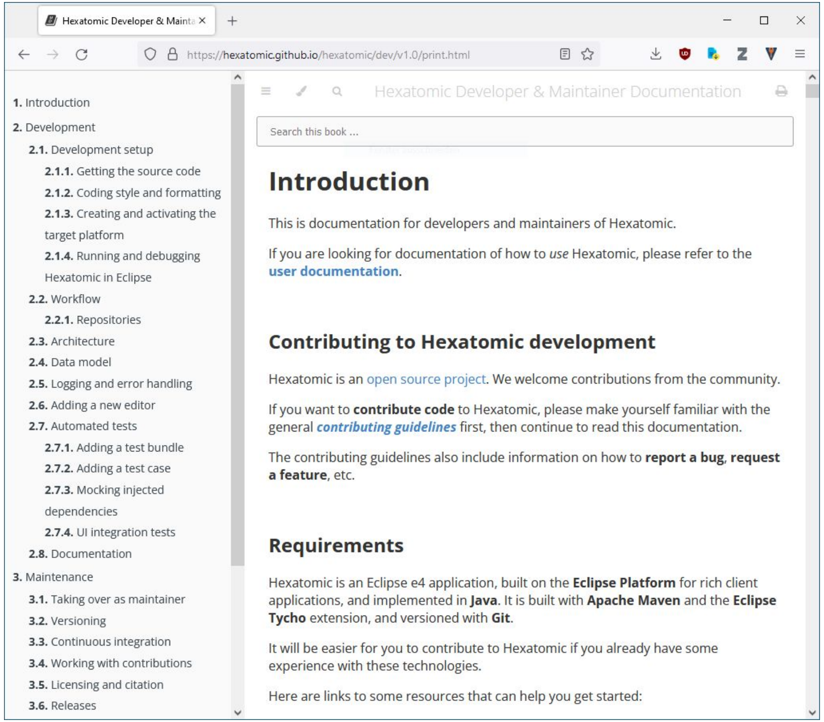
code-analysis

code-review

Documentation

- Documentation for all target groups:
  - users
  - developers and maintainers
  - interested parties (e.g, funders)

- No integration of undocumented changes



Hexatomic Developer & Mainta
https://hexatomic.github.io/hexatomic/dev/v1.0/print.html
Hexatomic Developer & Maintainer Documentation
Search this book ...

Introduction

This is documentation for developers and maintainers of Hexatomic.

If you are looking for documentation of how to *use* Hexatomic, please refer to the **user documentation**.

Contributing to Hexatomic development

Hexatomic is an open source project. We welcome contributions from the community.

If you want to **contribute code** to Hexatomic, please make yourself familiar with the general ***contributing guidelines*** first, then continue to read this documentation.

The contributing guidelines also include information on how to **report a bug**, **request a feature**, etc.

Requirements

Hexatomic is an Eclipse e4 application, built on the **Eclipse Platform** for rich client applications, and implemented in **Java**. It is built with **Apache Maven** and the **Eclipse Tycho** extension, and versioned with **Git**.

It will be easier for you to contribute to Hexatomic if you already have some experience with these technologies.

Here are links to some resources that can help you get started:

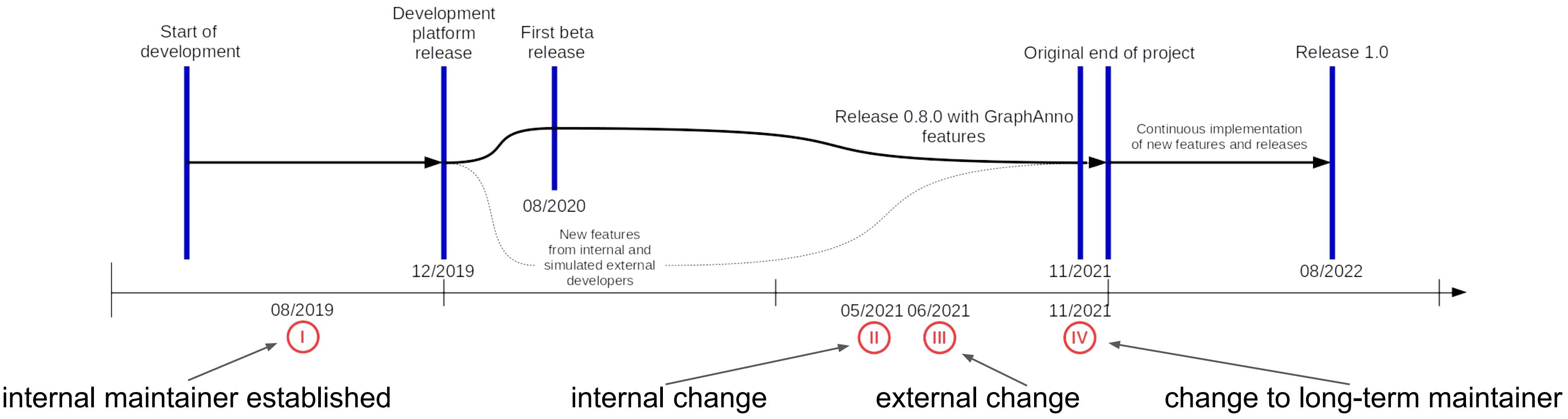
Existing open infrastructures

H_2: To sustainably provision small research software, four infrastructure elements are needed.

1. Source code repository platform (here: **GitHub**)
2. Repository for build artifacts (here: **Zenodo**)
3. Repository for dependency artifacts (here: **Maven Central**, **Eclipse P2**)
4. Maintainer

Maintenance strategy

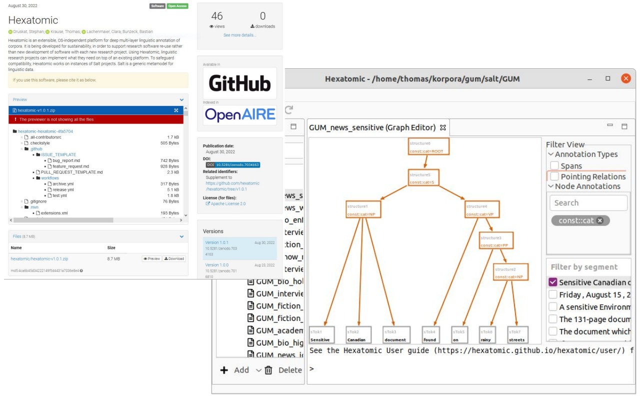
- Maintainer is the central role in the project
- Documentation for sunsetting and maintainer off-/onboarding
- Maintenance does not need experience (student assistants)
- Experiments: documented maintainer changes


Start of development
Development platform release
First beta release
Original end of project
Release 1.0
Release 0.8.0 with GraphAnno features
Continuous implementation of new features and releases
08/2020
New features from internal and simulated external developers
12/2019
11/2021
08/2022
08/2019
05/2021
06/2021
11/2021
I
II
III
IV
internal maintainer established
internal change
external change
change to long-term maintainer

Results

Software: Hexatomic

- **Successful:** development and provision through minimal infrastructure
- **Successful:** external contribution integrated



August 30, 2022

Software | Open Access

Hexatomic

Druskat, Stephan; Krause, Thomas; Lachenmaier, Clara; Bunzeck, Bastian

Hexatomic is an extensible, OS-independent platform for deep multi-layer linguistic annotation of corpora. It is being developed for sustainability, in order to support research software re-use rather than new development of software with each new research project. Using Hexatomic, linguistic research projects can implement what they need on top of an existing platform. To safeguard compatibility, Hexatomic works on instances of Salt projects. Salt is a generic metamodel for linguistic data.

If you use this software, please cite it as below.

Preview

hexatomic-v1.0.1.zip

The previewer is not showing all the files

```
hexatomic-hexatomic-4fa5704
  .all-contributorsrc                 1.7 kB
  .checkstyle                      505 Bytes
  .github
    ISSUE_TEMPLATE
      bug_report.md                742 Bytes
      feature_request.md           928 Bytes
    PULL_REQUEST_TEMPLATE.md          2.3 kB
    workflows
      archive.yml                  317 Bytes
      release.yml                     5.1 kB
      test.yml                        1.8 kB
  .gitignore                        76 Bytes
  .mvn
    extensions.xml                 193 Bytes
```

Files (8.7 MB)

Name	Size	
hexatomic/hexatomic-v1.0.1.zip	8.7 MB	Preview Download

md5:4ca6b45d34222149f5d4431a7036e8e4

46 views

0 downloads

See more details...

Available in: GitHub

Indexed in: OpenAIRE

Publication date: August 30, 2022

DOI: 10.5281/zenodo.7034163

Related identifiers: Supplement to https://github.com/hexatomic/hexatomic/tree/v1.0.1

License (for files): Apache License 2.0

Versions

Version 1.0.1 — 10.5281/zenodo.7034163 — Aug 30, 2022

Version 1.0.0 — 10.5281/zenodo.7016810 — Aug 23, 2022

Hexatomic - /home/thomas/korpora/gum/salt/GUM

GUM_news_sensitive (Graph Editor)

Filter View

Annotation Types

☐ Spans

☐ Pointing Relations

Node Annotations

Search

const::cat

Filter by segment

☑ Sensitive Canadian c

☐ Friday, August 15, 2

☐ A sensitive Environm

☐ The 131-page docum

☐ The document which

```
See the Hexatomic User guide (https://hexatomic.github.io/hexatomic/user/) f
>
```

✅

Maintainer changes

- **Successful:** changes to internal maintainer (milestones II/IV)
- **Semi-successful:** change to external maintainer
    - Lack of experience led to **stalling integrations**
    - **Steep learning curve** due to previously unknown complex **technology** (Java, OSGi, Eclipse RCP)
- Incremental improvements in maintenance documentation

😐

Lessons learned

- Minimal infrastructure works, but minimal != small
  - Setting up a **sustainable** infrastructure should be part of the project planning, even if the main benefit might happen after the project ends
  - Templates for specific Programming Languages and types of research software would be helpful
- Maintenance may be non-trivial, depending on technological experience
  - **We need RSEs** to do maintenance
- Processes are important:
  - Code review
    - Supported by a review checklist and quality metrics (static code analysis via CI)
    - Manual testing, building, running
    - Periodic triage of unreviewed code as workaround for urgent bug fixes by the maintainer
  - Only integrate high quality changes:
    - Static code analysis quality metrics
    - Complete documentation
    - Only tested changes
- Infrastructure and processes improve RSE practice even in solo projects

Thank you

- **Hexatomic:** https://github.com/hexatomic/hexatomic | hexatomic.github.io
  > Druskat, Stephan, Krause, Thomas, Lachenmaier, Clara, & Bunzeck, Bastian. (2022). Hexatomic (1.0.1). Zenodo. doi:10.5281/zenodo.7034163

- **Paper** forthcoming

- **Contact:** stephan.druskat@dlr.de, thomas.krause@hu-berlin.de, mailing list

- **Team:** Volker Gast, Anke Lüdeling, Bastian Bunzeck, Clara Lachenmaier 🙏

- **Funded** under the call *Research Software Sustainability* issued by Deutsche Forschungsgemeinschaft, grant number 391160252. 🙏
  T. Krause as **current maintainer** is funded by Deutsche Forschungsgemeinschaft SFB 1412, grant number 416591334