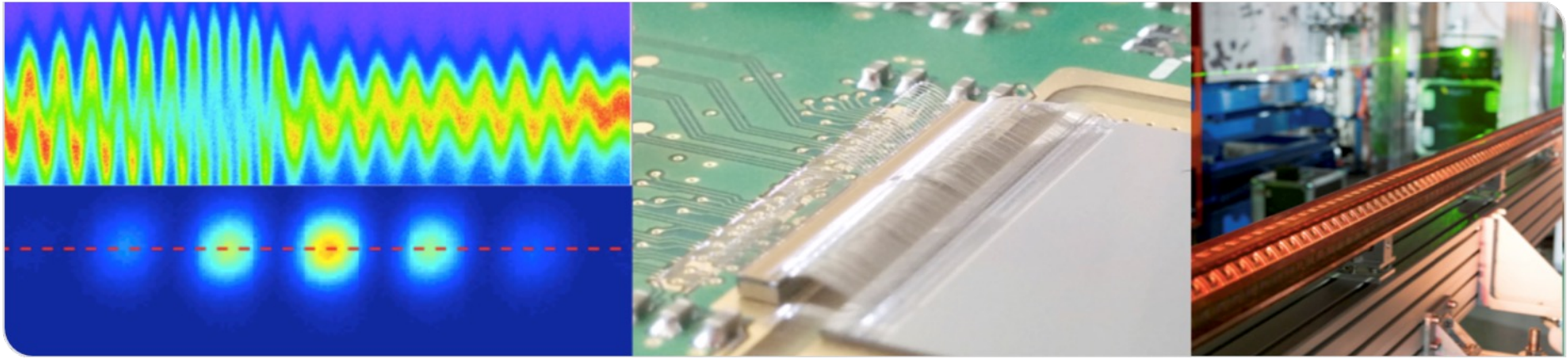
Karlsruhe Institute of Technology

HELMHOLTZ MTDTS

Kennzahl: Forschungsdaten und Software

Andreas Kopmann



Should KIT publish open source software?

- Helmholtz introduces open source software as performance indicator starting in 2022
  - Funding agencies have an increasing interest, that our research results are accessible and can be re-used

- Regulation in KIT (in preperation):
  - ”The head of the institute decides in consultation with the creator of the software, how / if it should be be published”
  - A common widely-used, known license should be selected

Kennzahlen

1. Anzahl der WoS-, SCOPUS oder Open Research Europe indexierten Publikationen
2. ... davon Open-Access-Publikationen – *neu ab 2021*
3. Drittmittelerträge in TEUR
4. Anzahl der abgeschlossenen betreuten Promotionen
5. Anzahl der Postdocs – *neu ab 2021*
6. Anzahl der Nachwuchsgruppen
7. Anzahl der ausgewählten koordinierten, nationalen und internationalen Förderprogramme
8. Anzahl der Kooperationen mit der Wirtschaft und externen nichtwissenschaftlichen Institutionen, öffentlich oder privat finanziert – *neu ab 2022*
9. Anzahl der Ausgründungen (Spin-offs) und kompetenzbasierten Gründungen (Start- ups) – *neu ab 2022*
10. Wissenstransferaktivitäten in FTE, gemessen am Programmbudget – *neu ab 2022*
11. Anzahl der Zitierbar publizierten Forschungsdaten- und Forschungssoftware- Publikationen – *neu ab 2022, zunächst Probeerhebung für 2 Jahre*

Indikator: Forschungsdaten und -software

- Anzahl von zitierbar publizierten Forschungsdaten-Publikationen und Forschungssoftware-Publikationen sind je separat zu erfassen.
- Die zitierbar publizierten Forschungsdaten- und Forschungssoftware-Publikationen müssen in einem **Repositorium mit Metadaten** gespeichert und mit einem persistenten Identifikator (insb. Digital Object Identifier – **DOI**) versehen sein.
- Erlaubte Repositorien sind solche, die entweder
  - unter Beteiligung eines Helmholtz-Zentrums betrieben werden oder
  - extern sind und in **re3data.org** gelistet oder zertifiziert sind.
  - Für die Publikation von Forschungssoftware können etablierte Software-Repositorien genutzt werden.
- Bei Softwarepublikationen wird jeder Release als Publikation gezählt.
- Die Erfassung erfolgt als Einstiegsindikator:
  - erstmals für das **Berichtsjahr 2022**,
  - nur als Helmholtz-interne Probeerhebung mit jährlicher Evaluierung der Erhebung;
  - ab Berichtsjahr 2024 ist die Einführung eines verpflichtenden Qualitätsindikators geplant.

Re3data.org

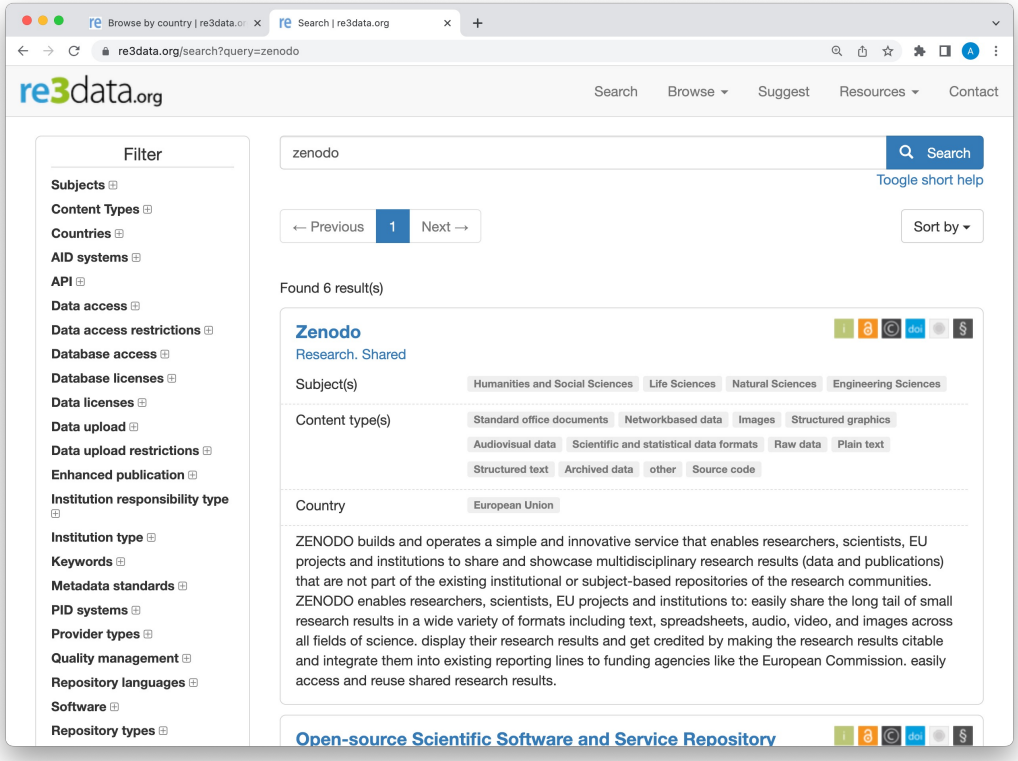
Re3data is a global registry of research data repositories

Included are

- Github
- KITopen
- Zenodo

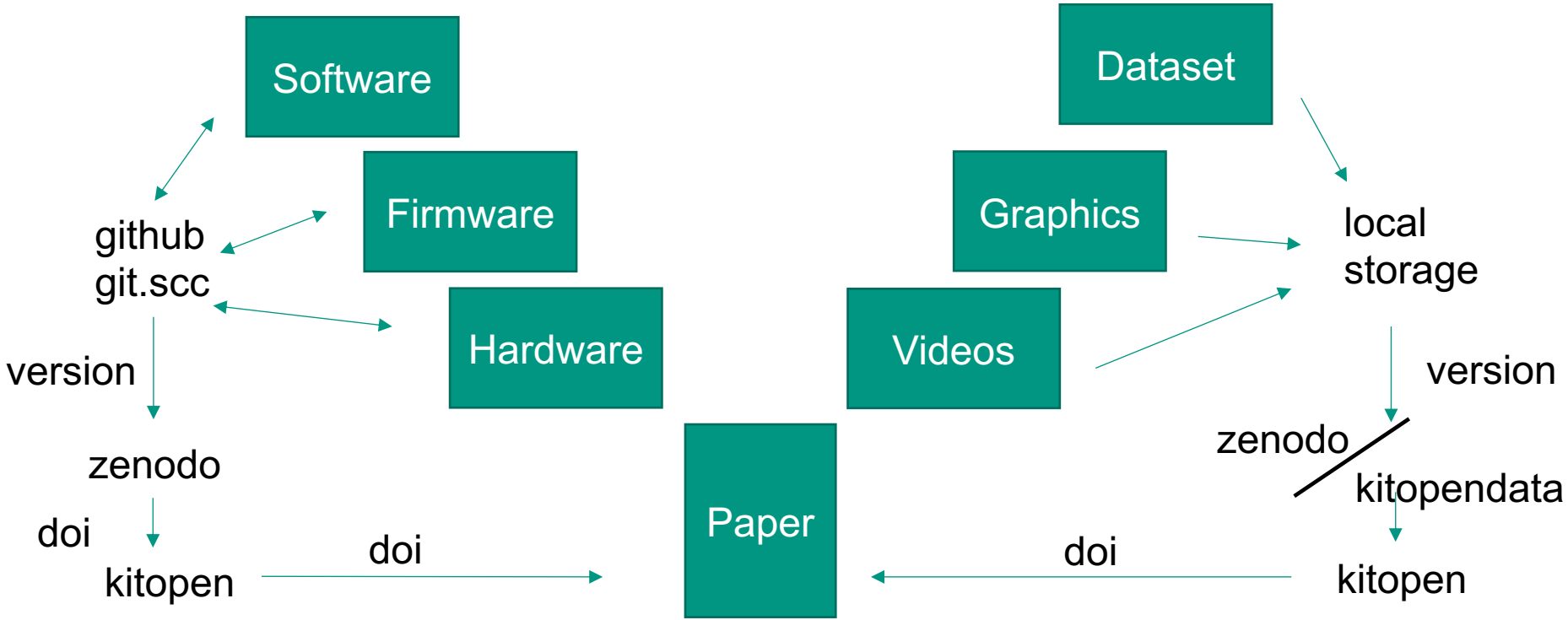
Not listed:

- Git.scc.kit.edu
- KITopenData (bwdatadiss)



Repository	Data	Metadata	DOI	Citations	Limitations
GitHub	Line-based	README, Version history, …	No (possible with Zenodo)	-	None
Git.scc.kit.edu	Line-based	README, Version history, …	No	-	None
KITopen	Publications	Extensive, Helmholtz measures	Yes	Yes	
KITopenData (part of KITopen)	Research data		Yes	No	180 GB/dataset Embargo < 2a On tape
Zenodo		Yes Github via .zenodo.json	Yes, reserve DOI option, versioning	Yes	50 GB/dataset

Workflow of research data



Workflow

- Write paper, prepare suplemental material, e.g. datasets
- Create a repository of the used software / firmware / hardware in a GIT repository
- Generate a dedicated version for the paper in zenodo
- Register software / firmware / hardware and suplemental material in KITopen
- Add reference (DOI) of software / firmware / hardware and suplemental material to the paper
- Upload the paper to a preprint server
- Submit the paper to conference / journal
- After aceptance register paper in KITopen
- Add relations

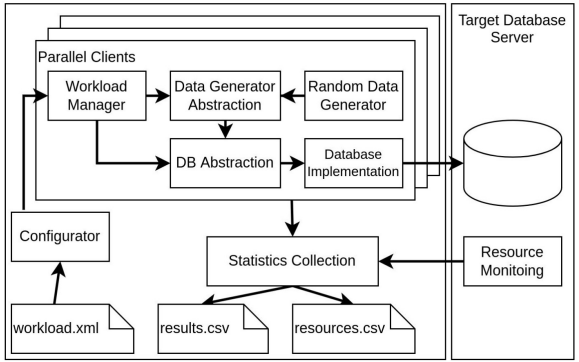
Real-time data storage

SciTS – a tool to evaluation time-series databases (TSDB)

Goal: direct storage of DAQ data in database engines

- An evolution of TSDBs motivated by time-series special characteristics:
  - Indexed by its corresponding timestamps
  - Continuously expanding in size
  - Usually aggregated, down-sampled, and/or queried in ranges
  - has very write-intensive requirements
- **SciTS** is a benchmark for TSDBs in large-scale experiments and industrial IoT
  - Motivated by experiences in the KATRIN experiment
  - Parameterizable ingestion workloads & 5 practical query workloads based on our study of data management tools in scientific experiments

https://github.com/jalalmostafa/scits



jalalmostafa / SciTS Public

Notifications | Fork 2 | Star 8

Code | Issues 1 | Pull requests | Actions | Projects | Security | Insights

v1.0 | 1 branch | 1 tag | Go to file | Code

jalalmostafa add bibtex citation	b834de3 on Sep 6	9 commits
.vscode	Initial commit	7 months ago
BenchmarkTool	fix default commandline	2 months ago
DDL	Initial commit	7 months ago
Jupyter	update readme	7 months ago
Scripts	Initial commit	7 months ago
Workloads	Initial commit	7 months ago
.gitignore	Initial commit	7 months ago
BenchmarkTool.sln	Initial commit	7 months ago
README.md	add bibtex citation	last month

README.md

SciTS

A tool to benchmark Time-series on different databases

Requires .NET 6.x cross-platform framework.

Citation

Please cite our work:

> Jalal Mostafa, Sara Wehbi, Suren Chilingaryan, and Andreas Kopmann. 2022. SciTS: A Benchmark for Time-Series Databases in Scientific Experiments and Industrial Internet of Things. In 34th International Conference on Scientific and Statistical Database Management (SSDBM 2022). Association for Computing Machinery, New York, NY, USA, Article 12, 1–11. https://doi.org/10.1145/3538712.3538723

Bibtex

About

A tool to benchmark Time-series on different databases

postgres | benchmarking | benchmark | database | timeseries | influxdb | time-series | metrics | clickhouse | postgresql | databases | scientific-computing | benchmarks | database-management | benchmarking-suite | benchmark-framework | scientific-publications | timescale | timescaledb

Readme

8 stars

1 watching

2 forks

Releases 1

v1.0 Latest
6 days ago

Packages

No packages published

Languages

Jupyter Notebook 87.0% | C# 12.7% | Other 0.3%

September 30, 2022

Software Open Access

jalalmostafa/SciTS: v1.0

Jalal Mostafa

Supporting version of SSDBM 2022 publication

Preview

SciTS-v1.0.zip

- jalalmostafa-SciTS-b834de3
  - .gitignore 6.7 kB
  - .vscode
    - launch.json 2.3 kB
    - settings.json 53 Bytes
    - tasks.json 1.3 kB
  - BenchmarkTool
    - BenchmarkTool.csproj 957 Bytes
    - ClientRead.cs 4.9 kB
    - ClientWrite.cs 2.8 kB
    - Config.cs 13.5 kB
    - ConfigurationKeys.cs 2.7 kB
    - Constants.cs 615 Bytes
    - CsvLogger.cs 1.4 kB
    - Database
      - ClickhouseDB.cs 11.5 kB
      - DatabaseFactory.cs 886 Bytes
      - IDatabase.cs 690 Bytes

Files (712.7 kB)

Name	Size	
jalalmostafa/SciTS-v1.0.zip	712.7 kB	Preview Download

md5:a1b85f2f7a2bbccc443bde9f6e0ca1b0

Beta Citations 0

Show only: Literature (0) Dataset (0) Software (0) Unknown (0) Citations to this version

Search

16 views

1 downloads

See more details...

Available in GitHub

Indexed in OpenAIRE

Publication date:
September 30, 2022

DOI:
DOI 10.5281/zenodo.7129102

Related identifiers:
Supplement to
https://github.com/jalalmostafa/SciTS/tree/v1.0

License (for files):
Other (Open)

Versions

Version v1.0 Sep 30, 2022
10.5281/zenodo.7129102

Karlsruhe Institute of Technology

KITopen: link to software repositories

- SciTS: A benchmark for Time-Series Databases …
  - Conference proceedings – Github repository

Mitwirkende	Titel	Publikationstyp	POF-Struktur	Peer-reviewed	Jahr	KITopen-Status	Volltext
Mostafa, Jalal	jalalmostafa/SciTS: v1.0	Forschungsdaten	54.12.02		2022	Veröffentlicht	
Mostafa, Jalal	Linux Patch Fix for XDP sockets with XDP_SHARED_UMEM flag	Forschungsdaten	54.12.02		2022	Veröffentlicht	
Mostafa, Jalal Wehbi, Sara Chilingaryan, Suren **Kopmann, Andreas** Ich	SciTS: A Benchmark for Time-Series Databases in Scientific Experiments and Industrial Internet of Things	Proceedingsbeitrag	54.12.02	Ja	2022	Veröffentlicht	Hochladen

Verweis auf die Quelle der Forschungsdaten

DOI 10.5281/zenodo.7129102

URL

Use relations in KITopen

Relationen

Relationen in KITopen

Art der Relation	KITopen-Publikation	
Verweist auf	**jalalmostafa/SciTS: v1.0** Mostafa, Jalal (2022)	Forschungsdaten KITopen-ID: 1000151126

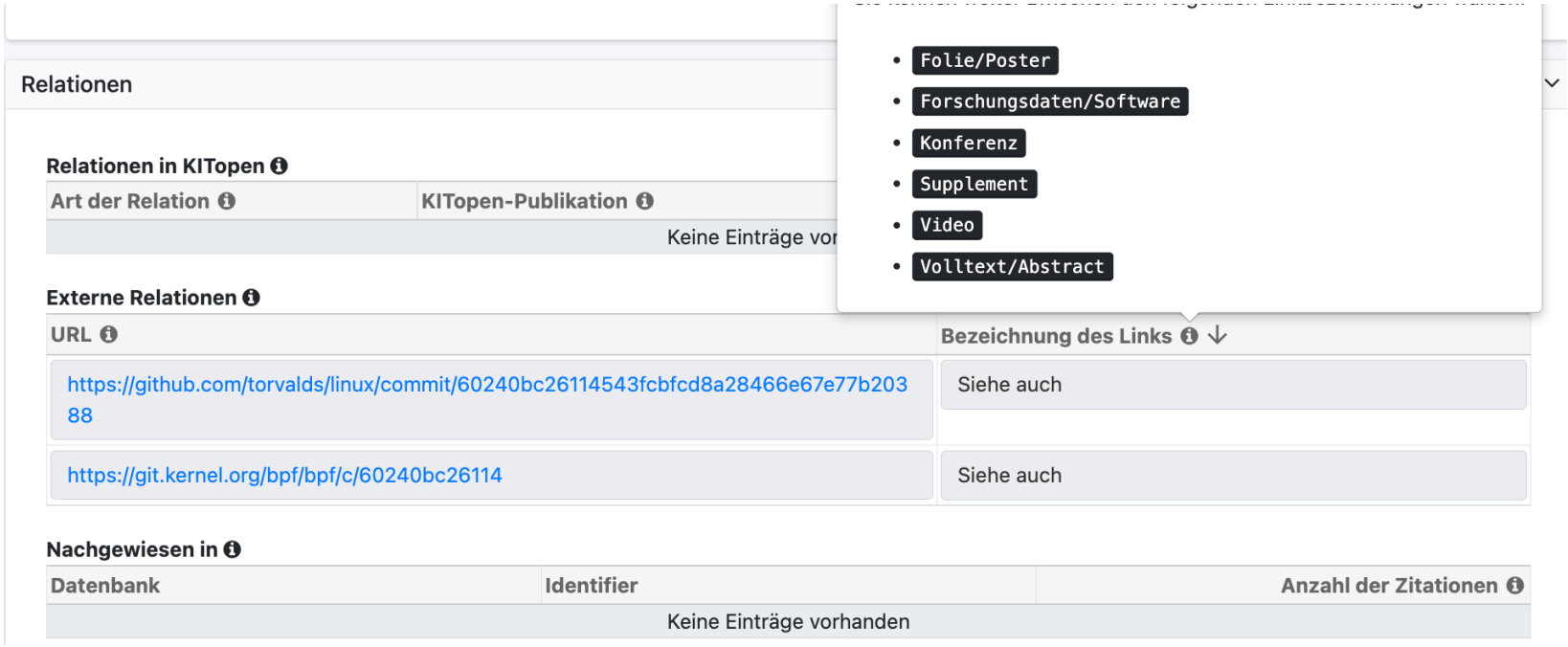
Externe Relationen

URL	Bezeichnung des Links
Keine Einträge vorhanden	

Nachgewiesen in

Datenbank	Identifier	Anzahl der Zitationen
Scopus	85137681496	0
Dimensions	pub.1150434125	0

KITopen: external relations



Research datasets

Forschungsdaten

Forschungsdaten in KITopenData

DOI

10.5445/IR/1000143

Relationen

Relationen in KITopen

Art der Relation	KITopen-Publikation	
Verweist auf	**High-Resolution Real-World Electricity Data from Three Microgrids in the Global South** Luh, Matthias; Phipps, Kaleb; Britto, Anthony; Wolf, Matthias; Lutz, Marek; Kraft, Johann (2022)	Forschungsbericht/Preprint KITopen-ID: 1000144567

Externe Relationen

URL	Bezeichnung des Links
Keine Einträge vorhanden	

Nachgewiesen in

Datenbank	Identifier	Anzahl der Zitationen
Keine Einträge vorhanden		

The Journal of Open-Source Software

- Requirements:
  - Open source as per the OSI definition
  - Obvious research application
  - Must not focus on new research results
- Review criteria:
  - Age of software (is this a well-established software project) / length of commit history.
  - Number of commits (not less than three months of work for an individual)
  - Number of authors.
  - Total lines of code (LOC); < 300 LOC will be rejected
  - cited in academic papers
  - Is software sufficiently useful; is it likely to be cited by your peer group.
- Efficient publishing and review procedure in Github

peer-reviewed
BUT currently
not in Scopus

The Journal of Open-Source Software

Paper in markdown:

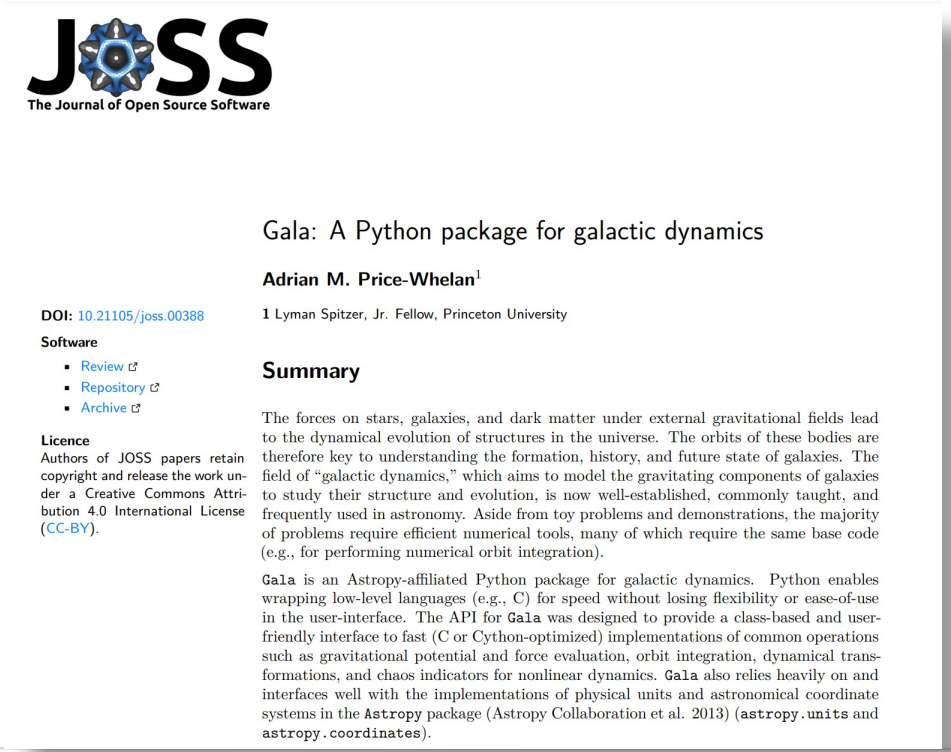
title: 'Gala: A Python..'
tags:
authors:
affiliations:
Date:
Bibliography

Summary

Statement of need

Acknowledgements

References



JOSS
The Journal of Open Source Software

Gala: A Python package for galactic dynamics

Adrian M. Price-Whelan[1]

1 Lyman Spitzer, Jr. Fellow, Princeton University

DOI: 10.21105/joss.00388

Software

- Review
- Repository
- Archive

Summary

The forces on stars, galaxies, and dark matter under external gravitational fields lead to the dynamical evolution of structures in the universe. The orbits of these bodies are therefore key to understanding the formation, history, and future state of galaxies. The field of "galactic dynamics," which aims to model the gravitating components of galaxies to study their structure and evolution, is now well-established, commonly taught, and frequently used in astronomy. Aside from toy problems and demonstrations, the majority of problems require efficient numerical tools, many of which require the same base code (e.g., for performing numerical orbit integration).

`Gala` is an Astropy-affiliated Python package for galactic dynamics. Python enables wrapping low-level languages (e.g., C) for speed without losing flexibility or ease-of-use in the user-interface. The API for `Gala` was designed to provide a class-based and user-friendly interface to fast (C or Cython-optimized) implementations of common operations such as gravitational potential and force evaluation, orbit integration, dynamical transformations, and chaos indicators for nonlinear dynamics. `Gala` also relies heavily on and interfaces well with the implementations of physical units and astronomical coordinate systems in the `Astropy` package (Astropy Collaboration et al. 2013) (`astropy.units` and `astropy.coordinates`).

Which license?

- **No license**
  - work is under exclusive copyright by default.
  - nobody else can copy, distribute, or modify
  - publishing on Github allow others to view and fork (BUT not still to use!)
  - **Don’t use the software** or ask the author to add a license

- **Use the license common in the community**
  - **MIT or BSD:** make people use your name when they use your work
  - **GPL:** prevent people from making changes to your work without providing them back to you

Popular licenses

- **GNU General Public License version 3 (GPL-3.0)**
- **GNU Affero General Public License version 3 (AGPL-3.0)**
- Mozilla Public License 2.0 (MPL-2.0)
- GNU Library or "Lesser" General Public License version 3 (LGPL-3.0)
- Eclipse Public License 1.0 (EPL-1.0)
- Apache License 2.0 (Apache-2.0)
- **MIT license (MIT)**
- BSD 2-Clause ("Simplified" or "FreeBSD") license (BSD-2-Clause

Further information: https://choosealicense.com/

Typical file structure in a software repository

README.md	Descripton of the sofware
LICENSE.txt	Standard license text, name of the author
.gitignore	Files not to be included in the repository
.zenodo.json	Metadata for publishing

Non-software licenses

- Data, media, etc.
  - CC0-1.0, CC-BY-4.0, and CC-BY-SA-4.0 are open licenses used for non-software material ranging from datasets to videos. Note that Creative Commons does not recommend its licenses be used for software or hardware.
- Hardware
  - Designs for open-source hardware ranging from furniture to FPGAs are covered by CERN Open Hardware licenses: CERN-OHL-P-2.0 (permissive), CERN-OHL-W-2.0 (weakly reciprocal), and CERN-OHL-S-2.0 (strongly reciprocal).
  - Open-Source Hardware Association - https://www.oshwa.org/definition/

Source: https://choosealicense.com/non-software/

Volunteers requested!

- Which software, datasets, firmware, hardware projects could serve as example?