Some Thoughts on

Open Source

Arfon Smith (arfon@stsci.edu)

Creative Commons Attribution 4.0 International License.
2 minutes about me…
TL;DR: AS INTERESTED IN ANALYSIS SOFTWARE AS SCIENTIFIC RESULTS
SHAPE
Is the galaxy simply smooth and rounded, with no sign of a disk?

- Smooth
- Features or disk
- Star or artifact
predicting coronal mass ejections using machine learning methods

In this notebook, we will be predicting whether or not a flaring active region will also emit a Coronal Mass Ejection (CME). The analysis that follows is published in Bobra & Ilonidis, 2016, *Astrophysical Journal*. If you use any of this code, we ask that you cite our paper.

Generally, active regions associated with large flares produce coronal mass ejections, but there have been some notable exceptions - for example, the largest active region in the last 24 years, which appeared in October 2014, produced many large flares yet not a single CME. Here is the active region:

```python
c
```
Citing Matplotlib

If Matplotlib contributes to a project that leads to a scientific publication, please acknowledge this fact by citing Hunter et al (2007) using this ready-made BibTeX entry:

```bibtex
@Article{Hunter:2007,
  Author    = {Hunter, J. D.},
  Title     = {Matplotlib: A 2D graphics environment},
  Journal   = {Computing in Science \& Engineering},
  Volume    = {9},
  Number    = {3},
  Pages      = {90--95},
  abstract  = {Matplotlib is a 2D graphics package used for Python for application development, interactive scripting, and publication-quality image generation across user interfaces and operating systems.},
  publisher = {IEEE COMPUTER SOC},
  doi       = {10.1109/MCSE.2007.55},
  year      = {2007}
}
```

DOIs

- v2.1.0: DOI 10.5281/zenodo.1004650
- v2.0.2: DOI 10.5281/zenodo.579577
- v2.0.1: DOI 10.5281/zenodo.573577
Software citation principles

Arfon M. Smith¹ *, Daniel S. Katz² *, Kyle E. Niemeyer³ *, and
FORCE11 Software Citation Working Group

¹ GitHub, Inc., San Francisco, California, United States
² National Center for Supercomputing Applications & Electrical and Computer Engineering
Department & School of Information Sciences, University of Illinois at Urbana-Champaign,
Urbana, Illinois, United States
³ School of Mechanical, Industrial, and Manufacturing Engineering, Oregon State University,
Corvallis, Oregon, United States
* These authors contributed equally to this work.

ABSTRACT

Software is a critical part of modern research and yet there is little support across the scholarly ecosystem for its acknowledgement and citation. Inspired by the activities of the FORCE11 working group focused on data citation, this document summarizes the recommendations of the FORCE11 Software Citation Working Group and its activities between June 2015 and April 2016. Based on a review of existing community practices, the goal of the working group was to produce a consolidated set of citation principles that may encourage broad adoption of a consistent policy for software citation across disciplines and venues. Our work is presented here as a set of software citation principles, a discussion of the motivations for developing the principles, reviews of existing community practice, and a discussion of the requirements these principles would place upon different stakeholders. Working examples and possible technical solutions for how these principles can be implemented will be discussed in a separate paper.
Software citation principles

Aron M. Smith**, Daniel S. Katz**, Kyle E. Niemeyer** and POREC11 Software Citation Working Group

1 ORCID Inc., San Francisco, California, United States
2 National Center for Supercomputing Applications & Electrical and Computer Engineering
   Department & School of Information Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois, United States
3 School of Mechanical, Industrial, and Manufacturing Engineering, Oregon State University, Corvallis, Oregon, United States

* These authors contributed equally to this work.

ABSTRACT

Software is a critical part of modern research and yet there is little support across the scholarly ecosystem for its acknowledgement and citation. Inspired by the activities of the POREC11 working group focused on data citation, this document summarizes the recommendations of the POREC11 Software Citation Working Group and its activities between June 2015 and April 2016. Based on a review of existing community practice, the goal of the working group was to produce a consolidated set of citation principles that may encourage broad adoption of a consistent policy for software citation across disciplines and venues. Our work is presented here as a set of software citation principles, a discussion of the motivations for developing the principles, reviews of existing community practice, and a discussion of the requirements these principles would place upon different stakeholders. Working examples and possible technical solutions for how these principles can be implemented will be discussed in a separate paper.
GitHub releases data on 2.8 million open source repositories through Google BigQuery
Non-profit, based in Baltimore
Established in 1982 to operate HST for NASA
Science operations center for HST & JWST
Develop software to support community
Mixed staff (astronomers & engineers)
Archive for > 20 different missions (including Kepler)
Government contractor (funding via NASA SMD)
Three things...

1. Insights from GitHub
2. The work we do at STScI
3. Recommendations
Insights from GitHub
24M users
1.5M organizations
67M repositories
100M pull requests
1B commits in 2016

GitHub

Source: https://octoverse.github.com
Repositories (144) People (31)

Search repositories...

**ominas**
Forked from pontando/ominas
Opensource Multi-Instrument Analysis Software
- 1 Star
- Updated 22 minutes ago

**openmct**
A web based mission control framework.
- 23 Star
- Updated an hour ago

**utm-apis**
The collection of APIs for NASA's UTM project in the form of OpenAPI documents.
- 4 Star
- Updated 2 hours ago

Top languages:
- C
- Python
- C++
- JavaScript
- Java

People (31)
Every company is a software company.
Why Software Is Eating The World

By Marc Andreessen
August 20, 2011

This week, Hewlett-Packard (where I am on the board) announced that it is exploring jettisoning its struggling PC business in favor of investing more heavily in software, where it sees better potential for growth. Meanwhile, Google plans to buy up the cellphone handset maker Motorola Mobility. Both moves surprised the tech world. But both moves are also in line with a trend I've observed, one that makes me optimistic about the future growth...

Groupon Investor Marc Andreessen: 'No Tech Bubble'
In an interview with WSJ's Kevin Delaney, Groupon and LinkedIn investor Marc Andreessen insists that the recent popularity of tech companies does not constitute a bubble. He also stressed that both Apple and Google are...
Open source is in every company. (Even if they don’t know it)
Maturity model

- Releasing new open source software
- Contributing to open source software
- Consuming open source software
Open is the new normal.
Open source as a strategic advantage.
### Ten most-forked projects

<table>
<thead>
<tr>
<th>Project</th>
<th>Forks</th>
</tr>
</thead>
<tbody>
<tr>
<td>TENSORFLOW/TENSORFLOW</td>
<td>24K</td>
</tr>
<tr>
<td>TWBS/BOOTSTRAP</td>
<td>15K</td>
</tr>
<tr>
<td>GITHUB/GITIGNORE</td>
<td>10.6K</td>
</tr>
<tr>
<td>BARRYCLARK/JEKYL-NOW</td>
<td>10.7K</td>
</tr>
<tr>
<td>TENSORFLOW/MODELS</td>
<td>9.3K</td>
</tr>
<tr>
<td>VUE.JS/VUE</td>
<td>8.1K</td>
</tr>
<tr>
<td>FACEBOOK/REACT</td>
<td>8K</td>
</tr>
<tr>
<td>JLORD/PATCHWORK</td>
<td>7.4K</td>
</tr>
<tr>
<td>SPRING-PROJECTS/SPRING-BOOT</td>
<td>7.4K</td>
</tr>
<tr>
<td>TORVALDS/LINUX</td>
<td>6.1K</td>
</tr>
</tbody>
</table>

### Projects with the most contributors

<table>
<thead>
<tr>
<th>Project</th>
<th>Forks</th>
</tr>
</thead>
<tbody>
<tr>
<td>MICROSOFT/VSCODE</td>
<td>15K</td>
</tr>
<tr>
<td>FACEBOOK/REACT-NATIVE</td>
<td>8.8K</td>
</tr>
<tr>
<td>NPM/NPM</td>
<td>7.6K</td>
</tr>
<tr>
<td>ANGULAR/ANGULAR-CLI</td>
<td>7.4K</td>
</tr>
<tr>
<td>TENSORFLOW/TENSORFLOW</td>
<td>7.3K</td>
</tr>
<tr>
<td>FORTAWESOME/FONT-AWESOME</td>
<td>6.8K</td>
</tr>
<tr>
<td>ANGULAR/ANGULAR</td>
<td>6K</td>
</tr>
<tr>
<td>DOCKER/DOCKER</td>
<td>6K</td>
</tr>
<tr>
<td>JLORD/PATCHWORK</td>
<td>5.9K</td>
</tr>
</tbody>
</table>

This list doesn’t include MOOC courses, but an R programming assignment on Coursera has forks in the thousands, too. If forks are any indication, more than 100,000 students have started this assignment.
Sustainability is an increasing concern.
Roads and Bridges: The Unseen Labor Behind Our Digital Infrastructure

14 July 2016
Libraries.io

Search open source libraries, frameworks and tools...

Libraries.io monitors 2,501,647 open source libraries across 34 different package managers, so you don't have to. Find out more

- **Discover new software**
  Search 2.5M projects by licence, language or keyword, or explore new, trending or popular projects.
  Explore

- **Monitor your dependencies**
  Stay up to date with notifications of updates, licence incompatibilities or deleted dependencies.
  Login

- **Maintain your OSS project**
  Understand your users and make informed decisions about features with usage and version data.
  Login

- **Use Libraries.io data**
  Use Libraries.io data in your applications, services or research. Use our API to stay up to date.
  Documentation

Supported Package Managers

- Go
  811K Projects
- npm
  914K Projects
- Rubygems
  162K Projects
- Pypi
  114K Projects
- Packagist
  179K Projects
- Maven
  108K Projects
- NuGet
  135K Projects
- Bower
  55.5K Projects
- WordPress
  53.5K Projects
‘Data-oriented’ software is particularly fertile open source territory.
Institutional Partners

Institutional Partners are organizations that support the project by employing Jupyter Steering Council members. Current Institutional Partners include:

- Continuum Analytics
- Bloomberg
- Netflix
- Cal Poly
- Berkeley University of California
- QuantStack

The Partners contribute to the development of Jupyter via patronage and by employing Jupyter contributors as part of their business.
Data science driving much of this.
### Pinned repositories

<table>
<thead>
<tr>
<th>Repository</th>
<th>Description</th>
<th>Language(s)</th>
<th>Stars</th>
<th>Forks</th>
<th>Updated</th>
</tr>
</thead>
<tbody>
<tr>
<td>javascript</td>
<td>JavaScript Style Guide</td>
<td>JavaScript</td>
<td>61.9k</td>
<td>11.9k</td>
<td></td>
</tr>
<tr>
<td>enzyme</td>
<td>JavaScript Testing utilities for React</td>
<td>JavaScript</td>
<td>11.5k</td>
<td>1.2k</td>
<td></td>
</tr>
<tr>
<td>aerosolve</td>
<td>A machine learning package built for humans</td>
<td>Scala</td>
<td>4.2k</td>
<td>542</td>
<td></td>
</tr>
<tr>
<td>react-dates</td>
<td>An easily internationalized, mobile-friendly datepicker library for the web</td>
<td>JavaScript</td>
<td>6.7k</td>
<td>551</td>
<td></td>
</tr>
<tr>
<td>lottie-android</td>
<td>Render After Effects animations natively on Android and iOS, Web, and React Native</td>
<td>Java</td>
<td>17k</td>
<td>2.8k</td>
<td></td>
</tr>
<tr>
<td>lottie-ios</td>
<td>An iOS library to natively render After Effects vector animations</td>
<td>Objective-C</td>
<td>11.3k</td>
<td>1.4k</td>
<td></td>
</tr>
</tbody>
</table>

### lottie-web

after effects to html library

- Language: JavaScript
- Stars: 8,203
- Forks: 623
- License: MIT
- Updated 2 minutes ago

### Top languages

- JavaScript
- Ruby
- Java
- Python
- HTML
Featured Projects

ReAir
Easy-to-use tools for replicating tables and partitions between Hive data warehouses.
View Project

Apache Superset (incubating)
A data exploration platform designed to be visual, intuitive, and interactive.
View Project

Enzyme
JavaScript Testing utilities for React.
View Project

Aerosolve
A machine learning package built for humans.
View Project

Apache Airflow (incubating)
Use Apache Airflow (incubating) to author workflows as directed acyclic graphs (DAGs) of tasks.
View Project

Airpal
A web-based, query execution tool for Facebook's PrestoDB.
View Project
**hollow**
Hollow is a Java library and toolset for disseminating in-memory datasets from a single producer to many consumers for high performance read-only access.

- **Type:** Java
- **Stars:** 1870
- **Forks:** 52
- **Language:** Java
- **Updated:** 8 minutes ago

**genie**
Distributed Big Data Orchestration Service

- **Language:** Java
- **Stars:** 788
- **Forks:** 198
- **Updated:** an hour ago

**msi**
Message Security Layer

- **Language:** JavaScript
- **Stars:** 132
- **Forks:** 68
- **Updated:** an hour ago
Netflix Open Source Software Center

Netflix is committed to open source. Netflix both consumes and provides open source technology focused on providing the leading Internet television network. Our technology focuses on providing immersive experiences across all Internet-connected screens. Netflix deployment technology allows for continuous build and integration into our worldwide deployments serving members in over 50 countries. Our focus on reliability defined the bar for cloud-based elastic deployments with several layers of failover. Netflix also provides the technology to operate services responsibly with operational insight, peak performance, and security. We provide technologies for data (persistent & semi-persistent) that serve the real-time needs of our 82 million members, as well as power the big data analytics that allow us to make informed decisions on how to improve our service. If you want to learn more, jump into any of the functional areas below to learn more.

Big Data
Tools and services to get the most out of your (big) data

Data is invaluable in making Netflix such an exceptional service for our customers. Behind the scenes, we have a rich ecosystem of big data technologies facilitating our algorithms and analytics. We use and contribute to broadly-accepted open source technologies including Hadoop, Hive, Pig, Parquet, Presto, and Spark. In addition, we’ve developed and contributed some additional tools and services, which have further elevated our data platform. Serve is a powerful, REST-based abstraction to our various data processing frameworks, notably Hadoop. Stitch provides detailed insights into the performance of our Hadoop jobs and clusters. Libflick shows the workflow of Pig jobs in a clean, visual format. And Astethus enables the bulk abstraction of data out of Cassandra for downstream analytic processing.

Build and Delivery Tools
Taking code from desktop to the cloud

Netflix has open sourced many of our Gradle plugins under the name Nebula. Nebula started off as a set of strong opinions to make Gradle simple to use for our developers. But we quickly learned that we could use the same assumptions on our open source projects and on other Gradle plugins to make them easy to build, test and deploy. By standardizing on Gradle development, we’ve lowered the barrier to operating them. If you want to learn more, jump into any of the functional areas below to learn more.
**docker-forensics**
Tools to assist in forensicing docker
- Python
- Star count: 21
- Fork count: 4
- BSD-3-Clause license
- Updated 2 hours ago

**rcloud**
Collaborative data analysis and visualization
- JavaScript
- Star count: 341
- Fork count: 118
- MIT license
- Updated 3 hours ago

**ast**
This is the AT&T Software Technology ast software download site from AT&T Research. The AT&T AST OpenSource Software Collection provides an overview and Practical Reusable UNIX Software provides a historical perspective. You can join the ast and unix mailing lists, or monitor the mail archives. There is also an AT&T internal talk. The man page index...
The Social Coding Environment for Big Data Analytics & Visualization

**Exploration**
Interactive, web-based environment to explore data on an interactive command line, including visualization and interactive graphics.

**Open Source**
Code is available on Github – collaborators and contributors welcome!

**Collaboration**
Social coding integrated development environment (IDE): search, run, comment, fork, modify and share code and analyses – deployed as web services, web pages or dashboards.

**In the Cloud**
Accessible from any connected device or computer anywhere, scalable power of distributed computing.

“Recently, the American Business Awards recognized RCloud as one of the Technical...”
License is important.
Open source license usage on GitHub.com

March 9, 2015  benbaltar  Watercooler

Open source simply isn’t open source without a proper license. Unless you’ve explicitly told others that they can modify and reuse your work, you’ve only showed others your code; you haven’t shared it. Here at GitHub, we’re big fans of open source, so we set out to better understand how our users approached licensing their code by looking at license usage across public, non-forked repositories, in hopes of encouraging more users to share their work with others.

**Percentage of repositories licensed**

[Graph showing percentage of repositories licensed over years from 2008 to 2015]
Breakdown of license usage

We also wanted to look at the relative breakdown of the most popular open source licenses. You can see their popularity expressed below as a percentage of licensed projects on GitHub.com:

<table>
<thead>
<tr>
<th>Rank</th>
<th>License</th>
<th>% of projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MIT</td>
<td>44.68%</td>
</tr>
<tr>
<td>2</td>
<td>Other</td>
<td>16.68%</td>
</tr>
<tr>
<td>3</td>
<td>GPLv2</td>
<td>12.96%</td>
</tr>
<tr>
<td>4</td>
<td>Apache</td>
<td>11.19%</td>
</tr>
<tr>
<td>5</td>
<td>GPLv3</td>
<td>8.88%</td>
</tr>
<tr>
<td>6</td>
<td>BSD 3-clause</td>
<td>4.53%</td>
</tr>
<tr>
<td>7</td>
<td>Unlicense</td>
<td>1.87%</td>
</tr>
<tr>
<td>8</td>
<td>BSD 2-clause</td>
<td>1.70%</td>
</tr>
<tr>
<td>9</td>
<td>LGPLv3</td>
<td>1.30%</td>
</tr>
<tr>
<td>10</td>
<td>AGPLv3</td>
<td>1.05%</td>
</tr>
</tbody>
</table>

Unsurprisingly, MIT, Apache, and GPL are the clear front runners, with some 15% of licensed projects opting for a non-standard license or standard license not among those listed on choosealicense.com.

License breakdown by repository creation date

![License breakdown chart](https://via.placeholder.com/150)
Last, we looked at how license usage has changed over time. Again, you see a swift uptick of the three featured license (MIT, Apache, GPL) in mid-2013, with the relative percentages remaining otherwise steady over the past six years.

Developers use GitHub because they want to share their code with the world, and the data suggests that when the tools we use make it a little bit easier, developers do just that. When presented with the option, they choose to license, and they license very permissively.

Under the hood

To detect what license, if any, a project is licensed under, we used an open source Ruby gem called **Licensee** to compare the repository’s LICENSE file to a short list of known licenses. However, it’s important to note that this approach doesn’t count a project as licensed if the README indicates a specific license or if individual project files contain a
License is important.
Business friendly, ‘permissive’ licenses are dominant in data-oriented software.
Academia can learn from open source.
Insights from GitHub: Ideas Academia should copy
Idea 1: Open collaborations
Open **Source** vs Open **Collaborations**
Open Source: the right to modify
Open Collaborations: a highly collaborative development process and are receptive to contributions of code, documentation, discussion, etc from anyone who shows competent interest.
The pull request
Code first, permission later
Capture context
Expanded About Section

No one is assigned

Milestone: About 2.0

Starting to work on https://team.githubapo.com/projects/15

At this point I've just applied the repo admin styling to give us some about sections and shuffled the existing content around. It's rough, but you get the idea.

We make it easier to collaborate with others and share your projects with the universe.
Exposed process
Every time this happens the **community** learns
Idea 2: Culture of reuse
Software composed of many components
Your software is the thing that is *different*
Current academic credit model rewards novelty
Idea 3: Verification
“open source is... reproducible by necessity”

Fernando Perez

Pull requests

1. Public Key as a language with .asc and .pub extensions
   - #1844 opened 17 minutes ago by pchen

2. Add support for Turtle and SPARQL
   - #1941 opened 18 hours ago by demx

3. Add Lisp and PicoLisp language.
   - #1837 opened 2 days ago by zwarnick

4. Optimization of generated?
   - #1832 opened 2 days ago by pchau

5. Set "missing" ace modes to `text`
   - #1831 opened 2 days ago by gltrkian

6. Fix bug in #1731; add .xml.fr and .html.fr
   - #1830 opened 2 days ago by lanbar-khoff

7. Test folder for files without a language
   - #1829 opened 2 days ago by pchau

8. Support for Rc from Plan9
   - #1828 opened 3 days ago by pchau

9. SaltStack state files use the same extension as old scheme library source files
   - #1825 opened 3 days ago by iggy
Remove `.script!` hack #1786

bkeepers wants to merge 3 commits into strategies from remove-script-art

bkeepers commented 9 days ago

It is no longer necessary.

/cc @aron @pchaingo

bkeepers added some commits 9 days ago

- Remove `.script!` hack
- Don't need a separate class for filename lookup
- Remove puts

bkeepers commented 9 days ago

I should note that this is also dependent on #1633.
makes mistakes. I don’t think it would be able to choose among all script languages...
If we continue with this behavior we should make sure first that the classifier will do a good job.

bkeipers commented 9 days ago
Fair point.

Add more commits by pushing to the remove-script-ext branch on github/linguist.

Merge with caution!
You can also merge branches on the command line.

ProTip! Add .patch or .diff to the end of URL for Git’s plaintext views.
GitHub/Linguist

Language Savant. If your repository's language is being reported incorrectly, send us a pull request!

Build Matrix

<table>
<thead>
<tr>
<th>Job</th>
<th>Duration</th>
<th>Finished</th>
<th>Ruby</th>
<th>OS</th>
</tr>
</thead>
<tbody>
<tr>
<td>3869.1</td>
<td>1 min 51 sec</td>
<td>8 days ago</td>
<td>1.93</td>
<td>Linux</td>
</tr>
<tr>
<td>3860.2</td>
<td>1 min 52 sec</td>
<td>8 days ago</td>
<td>2.00</td>
<td>Linux</td>
</tr>
<tr>
<td>3860.3</td>
<td>2 min 10 sec</td>
<td>8 days ago</td>
<td>2.1</td>
<td>Linux</td>
</tr>
<tr>
<td>3860.4</td>
<td>2 min 10 sec</td>
<td>8 days ago</td>
<td>2.2</td>
<td>Linux</td>
</tr>
</tbody>
</table>

Commit: b123b34 #1734: Remove script hack
strategies - Remove puts

Brandon Keepers authored and committed

Commit b122959 #1766 Remove script hack

using worker: worker-linux-6-2.bb.travis-ci.org:travis-linux-19

Build system information

$ git clone --depth=50 git://github.com/github/linguist.git github/linguist
$ rvm use 2.1 --install --binary --fuzzy
$ export BUNDLE_GEMFILE=~/Gemfile
$ ruby --version
$ rvm --version
$ rvm 1.26.0 (master) by Wayne K. Seguin <waynesegui3@gmail.com>, Michal Papis <apapiss@gmail.com>
$ bundle --version
$ gem --version
$ 2.4.2

$ git fetch origin master:master
$ git fetch origin test/attributes:master/attributes
$ git fetch origin test/master:master/test/master
$ sudo apt-get install libhifi-dev-dev -y
$ bundle install --jobs=3 --retry=3
$ bundle exec rake

Run options:

# Running tests:

Finished tests in 25.317782s, 28.4559 tests/s, 308.6015 assertions/s.
249 tests, 0 assertions, 0 failures, 0 errors, 0 skips

ruby 2.1.4p265 [2014-10-27 revision 48166] [x86_64-linux]

The command "bundle exec rake" exited with 0.
makes mistakes. I don't think it would be able to choose among all script languages...
If we continue with this behavior we should make sure first that the classifier will do a good job.

bkeepers commented 9 days ago

Fair point.

Add more commits by pushing to the remove-script-ext branch on github/linguist.

All is well — The Travis CI build passed. Details

This pull request can be automatically merged.
You can also merge branches on the command line.

Write

Leave a comment

Attach images by dragging & dropping, selecting them, or pasting from the clipboard.

ProTip! Add .patch or .diff to the end of URL for Git's plaintext views.
<table>
<thead>
<tr>
<th>Merge pull request #6794 from m-ko/mm/new_trajectory_saving</th>
<th>latest commit 5909738f19</th>
</tr>
</thead>
<tbody>
<tr>
<td>omsbuild authored 5 days ago</td>
<td></td>
</tr>
<tr>
<td>- Alignment</td>
<td>Morgo pull request #5644 from eavdeova/NewScenario</td>
</tr>
<tr>
<td>- AnalysisAlignan</td>
<td>removed refractor from interface</td>
</tr>
<tr>
<td>- AnalysisDataFormats</td>
<td>2 months ago</td>
</tr>
<tr>
<td>- BigProducts/Simulation</td>
<td>Changes for root 5.34.22</td>
</tr>
<tr>
<td>- CallibCalorimetry</td>
<td>disable LTO for biglibs with gcc 4.9.x</td>
</tr>
<tr>
<td>- CallibFormats</td>
<td>Allow experiment type to be set</td>
</tr>
<tr>
<td>- CallibMuon</td>
<td>Make CaloTPGTransconductor interface thread safe</td>
</tr>
<tr>
<td>- CallibTracker</td>
<td>Morgo pull request #5777 from ramanjot/DTRecoConditions</td>
</tr>
<tr>
<td>- Calibration</td>
<td>Morgo pull request #5784 from hideapp/TrackerToBe</td>
</tr>
<tr>
<td>- CaloOnlineTools</td>
<td>Morgo pull request #5772 from mateosanftopic_HLTUdateFor7x</td>
</tr>
<tr>
<td>- CaloOnlineTools</td>
<td>Use thread-safe access to HCAL containers</td>
</tr>
<tr>
<td>- CommonTools</td>
<td>Morgo pull request #8604 from vander/73x-fxModule/Reconfiguration</td>
</tr>
<tr>
<td>- CondCore</td>
<td>Morgo remote branch origin/CMSSW_7_3_X into CMSSW_7_0_X</td>
</tr>
<tr>
<td>- CondFormats</td>
<td>fix build errors in PhysicsTools/Objects</td>
</tr>
<tr>
<td>- CondFormats</td>
<td>7 days ago</td>
</tr>
<tr>
<td>- CondFormats</td>
<td>5 days ago</td>
</tr>
<tr>
<td>- CondFormats</td>
<td>15 days ago</td>
</tr>
<tr>
<td>- CondFormats</td>
<td>11 days ago</td>
</tr>
</tbody>
</table>
added shifted beamspot #6790

- civanch wants to merge 1 commit into `cms-sw/cmssw / 4.x` from `civanch/shifted-beamspot`

**Conversation**

- civanch commented 2 days ago
  - HCAL requested special beam spot position to test HF phi asymmetry.

- cmsbuild added this to the Next CMSSW_7_4_X milestone 2 days ago
  - `operations-pending` `pending-signatures` `simulation-pending` `tests-pending`

- cmsbuild added `operations-pending` `pending-signatures` `simulatiom-pending` `tests-pending` labels 2 days ago

- cmsbuild commented 2 days ago
  - A new Pull Request was created by civanch (Vladimir Ivanchenko) for CMSSW_7_4_X.
  - added shifted beamspot
  - It involves the following packages:
    - Configuration/StandardSequences
    - CMS&EventVertexGenerators
A new Pull Request was created by @civanch (Vladimir Ivantchenko) for CMSSW_7_4_X.

It involves the following packages:

Configuration/StandardSequences
OMG/EventVertexGenerators

@civanch, @nclopezo, @mdhillreth, @cmsbuild, @franzoni, @davidlange6 can you please review it and eventually sign off? Thanks.

@ghellwig, @makortel, @GiacomoSquazzoni, @Grover, @EVinNn, @Appeltel, @cerati, @dghulhan this is something you requested to watch as well.

You can sign-off by replying to this message having '+1' in the first line of your reply.
You can reject by replying to this message having '-1' in the first line of your reply.

+1

Tests Pending

+1

Tested at: 64158fc

Tested Approved
Comparison is ready


This pull request can be automatically merged by project collaborators. Only those with write access to this repository can merge pull requests.
The work we do at STScI
The Data Science Mission Office is responsible for maximizing the scientific impact of data at STScI. We do this through a combination of: **Building and maintaining** world-leading data management infrastructure and services to support past, current and future missions. Being a **reliable and trusted** source of data, software, services and knowledge for the communities we support. **Developing institutional and community expertise** in combining data science and astronomy and **supporting the careers** of those individuals doing this work. **Exploring new technologies, services and infrastructure** designed to realize the potential of multi-wavelength, multi-mission astronomy.

**Data Science Mission Office**

[Logo: STScI | Space Telescope Science Institute]
Data Analysis Tools, Calibration Pipelines, Data Management Infrastructure & Archives
Most basic archive: A safe store
Better: Science-ready data archive
Archive
Archive, Services
Archive, Services, Software
MAST: Multi-mission archive
It works!
Archive, Services, Software
Status quo (until relatively recently)

http://mast.stsci.edu
http://archive.stsci.edu
VO services
JSON API
casjobs
TOPCAT

Traditionally thought of as science centers activities
Looking to the future.
Data-intensive missions

LSST

WFIRST
Hybrid missions

- Physics experiments, ‘planned’ science (e.g. surveys)
- ‘Guest observer’ mode (WFIRST)
- Vast data-rich archives
- Core set of services provided by science centers (data, storage, compute)
- Co-created software (both pipelines and data analysis tools)
- Centralization of projects means more contributed data products/re-analyses
Community contributions at all levels

- Community software (e.g., Astropy)
- Community Alert brokers/agents
- ‘L3’ data products

Data

Software

Services

Community software + L3/L4/L5 pipelines
Community-built Services
L3/L4/L5 data products (HLSPs)
One example

![Graph showing the number of code contributors over time from 2012 to 2017.](image-url)
Science center as a facilitator of these activities

“Design for community contributions at all stages”
Open source is critical for delivering on our mission to support the scientific community.
Recommendations
Business-friendly licenses are important.
MIT, BSD, Apache 2

(PUBLIC DOMAIN IS *NOT* CONSIDERED OPEN SOURCE)
GPL-compatibility is important too.
Standard NASA contracts are overly-restrictive (typically reserve all IP)
Boiler-plate language for contracts that would support the ability for any party to acquire copyright and release software developed under the contract.
Process described at code.nasa.gov is complex (and slow)
Kepler Science Data Processing Pipeline

The Kepler telescope launched into orbit in March 2009, initiating NASA’s first mission to discover Earth-size planets.
Great Observatories have been an exemplar for open data. Let’s make open source software a defining feature of NASA science too.
Thanks!

Arfon Smith (arfon@stsci.edu)