

HEP Software & Computing Knowledge Base

hepsoftware.org

Torre Wenaus, BNL CHEP 2016, San Francisco October 13, 2016





Why a HEP S&C knowledge base?

Number one on the original HEP Software Foundation (HSF) task list: "provide a system for facilitating information exchange"

Proposition accepted at HSF's formative workshop at SLAC in 2015: HEP S&C should have a **grass roots information gathering and exchange hub**

- Facilitate collaboration by increasing awareness of resources, projects, activities
- Provide a place to look for solutions, learn from the choices made by others
- Promote awareness of software and tools from outside our community, e.g.
 open source

A resource built and maintained by the community

- Must overcome the Catch 22: It's only useful if it's populated, and people will only help populate it if it's useful
- Entice people to contribute by making it fun (or at least pleasant) to do so



hepsoftware.org development

Content today: 603 entries, 963 relations created by 24 people

- Key design precepts
 - Entity-relation model: relations are as important as entries, much of the useful knowledge is in the interconnections (e.g. who uses what)
 - Make it fast, highly navigable, easy transiton between browsing and editing
 - Users are also creators
- Goals reached in the third prototype generation
 - 2014: Django, MySQL, xml data, based on ATLAS PanDA monitor. Clunky
 - 2015: Drupal + extensions, part of HSF website. Slow and clunky
 - 2016: Current version based on ATLAS 'data knowledge base' prototyping
 - 2017: No plans for a fourth, just clean up the code
- Data content carried forward and preserved generation to generation
- Implementation: a javascript app in the browser served by a node.js server
 - MySQL data repository
 - REDIS memory store for server side sessions
 - CouchDB/PocketDB for server/client json data channel
 - Works beautifully
- In github private repo, move to public when purged of secrets and housecleaned





Content today: 603 entries, 963 relations created by 24 people

- Two knowledge base usage modes:
 - Data subset at the client with queries to server for the rest
 - or, Fully client-resident data
 - Currently uses the latter; data volume is small
 - Content is all human typed text
- Non-negligible initial load time pays off in responsiveness
 - plus it works offline
 - 5 sec load time in my hotel room this morning
- Has operated very stably for close to a year
 - Server in Amazon EC2, provided by BNL
 - Backed up every 2 hours to EBS and S3, daily backups kept
 - Full revertible version histories for entries kept





The remainder of the talk is a lightning tour until my time runs out...

...what I don't get to is supplementary



Getting help



^f: search ^e: edit

esc: escape

 ■ Sections ■ **↑** Top Feedback How to use hepsoftware.org Scanning entries Entry content Keyboard shortcuts How to contribute: adding and editing entries Authenticating The editing panel Active edits Reverting to previous versions Listing new and updated entries Policy License About the knowledge base Issues **◆** Bottom

Click the info button for a guide, which is itself an entry in the KB

All KB content is in markdown

Navigable table of contents for large entries

Licensed, with its content, under Creative Commons Attribution-ShareAlike

Resources/Services

The High Energy Physics (HEP) Software & Computing Knowledge Base at hepsoftware.org is a collection point for HEP related software projects and information on HEP software and computing. You can use it to look for existing software, to make your own project known to the community, and to inform the community what software you and your experiment use.

The hepsoftware.org knowledge base (KB) is a project of the HEP Software Foundation. Everyone in the HEP computing community is encouraged to contribute to the content, by adding software, adding experiments, adding interrelationships describing who uses what, and so on. The more complete and reflective of community knowledge, experience and opinion the content is, the more useful it will be.

This knowledge base replaces and improves upon (in speed, flexibility, ease of use, extensibility) the version integrated with the HSF's Drupal-based website. The content there has been transferred here and will shortly be removed there.

Feedback

Feedback is encouraged! Send it to hepsoftware@gmail.com or to the author, Torre Wenaus wenaus@gmail.com.

How to use hepsoftware.org

The KB presently catalogs software packages, experiments, science fields, organizations, insitutes and resources. Also links can be added -- web links dressed with comments. tags, relations to KB entries etc. Under these buttons different views are available, e.g. successive clicks on the software button cycle through lists of software categories, categories with their associated software, and a list of all software. Click the clock to order, by latest changes rather than alphabetic. Similarly for Experiments, they can be listed together with the software they use or conversely software can be listed indicating the experiments that use it. And so on for Sciences, Organizations, Institutes, Resources etc.

Scanning entries

Scanning these listings with the mouse gives an instantaneous scan of the associated entries, until an entry is selected with a click (highlighting its name in red). The center panel then remains fixed on the selected entry. It can be de-selected by clicking the highlighted name in the center header or left list, or with ctrl-x.

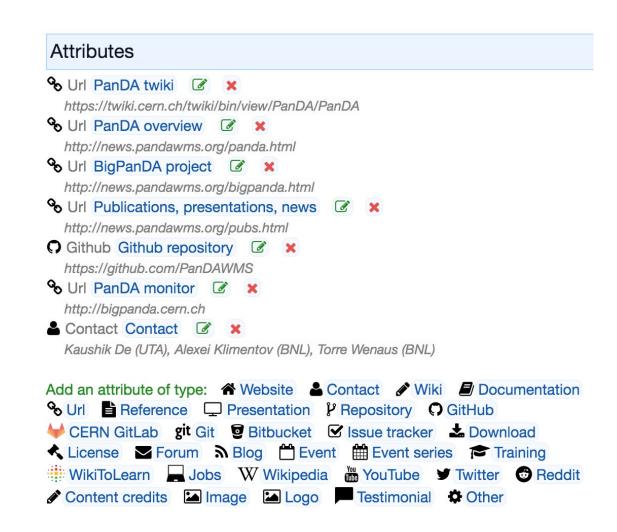
Entry content

The body of the entry is written in Markdown. Where it is long and structured enough to have sections (like this one), the Show Sections button in the header can be used to show a table of contents in the left window. Besides the body text describing the entry, it



Entries have attributes

- Contact people
- Website, twiki, wikipedia
- Email fora, blog
- Social media
- Events, event series
- Training
- WikiToLearn
- Reference links
- Documentation
- Presentations
- Downloads
- Repository
- Issue tracker
- Testimonials
- Jobs
- ...





Entries have relations

Key to the entity-relationship organization of the knowledge base

Ascribe relations between entries of contextually appropriate types

Builds up the knowledge part of the KB

Convey connections, enable discovery

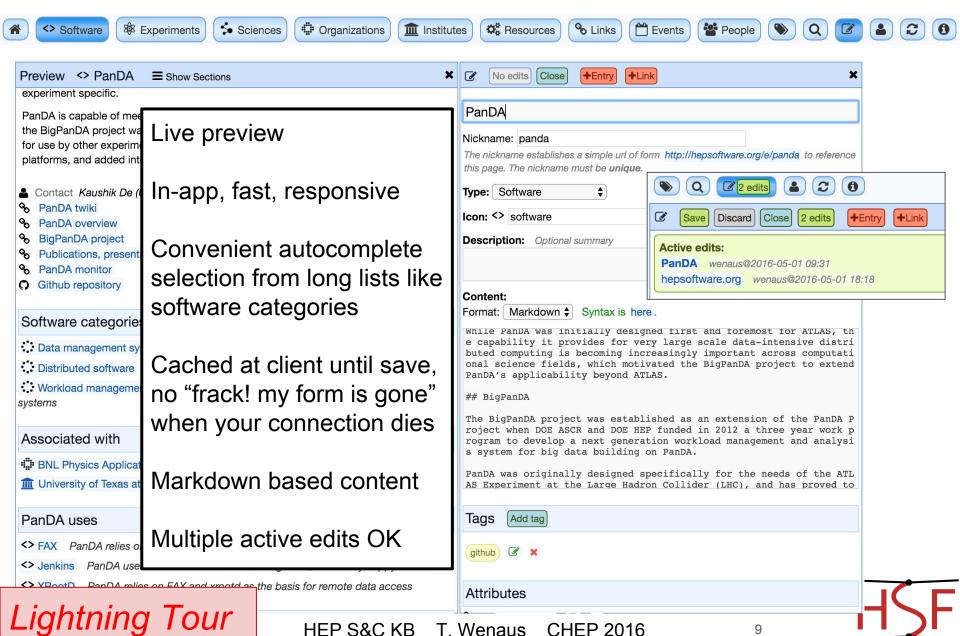
Share the wisdom of your experiment or project in what you use

There are many more relations in the database than there are entries

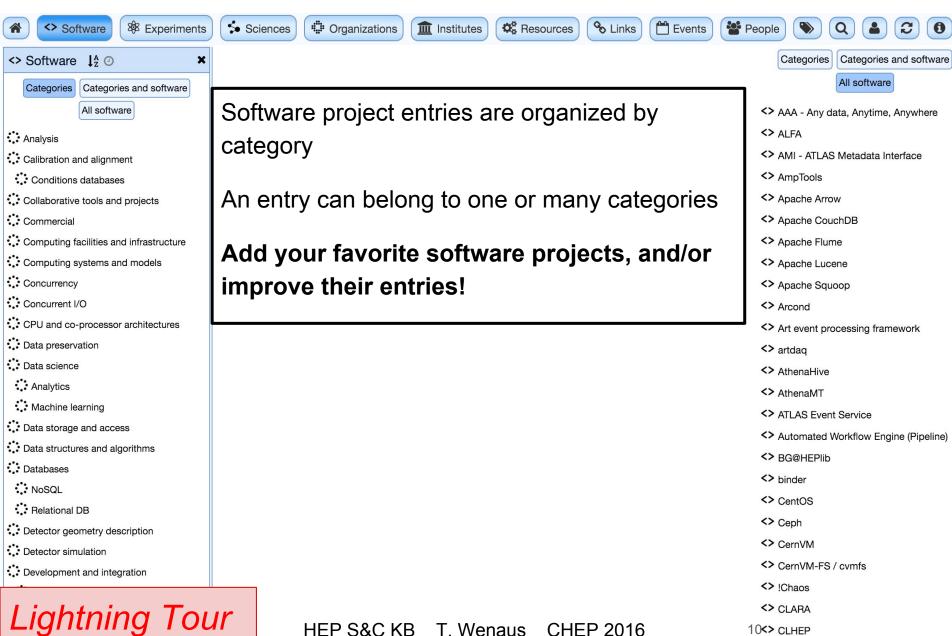
Relations Relations: Software category Distributed software PanDA software category Distributed software 2015-09-20/admin Software category Data management systems & PanDA software category Data management systems 2015-09-20/admin Software category Workload management systems & PanDA software category Workload management systems 2015-09-20/admin Uses Jenkins 3 X PanDA uses Jenkins for continuous integration, we're very happy with it 2015-10-25/wenaus Uses FAX 📝 🗙 PanDA relies on FAX and xrootd as the basis for remote data access 2015-10-25/wenaus Uses XRootD 🗷 🗙 PanDA relies on FAX and xrootd as the basis for remote data access 2015-10-25/wenaus UTA is a lead contributor to BonDA 2015-11-05/w Add a relation to another entry: Associated with in Institutes Associated with "Organizations Associated with ** Resources Associated with <> Software Associated with Groups More information Notes Related to <> Software Science field Sciences Software category :: Software categories Used by Experiments Used by institutes Used by " Organizations

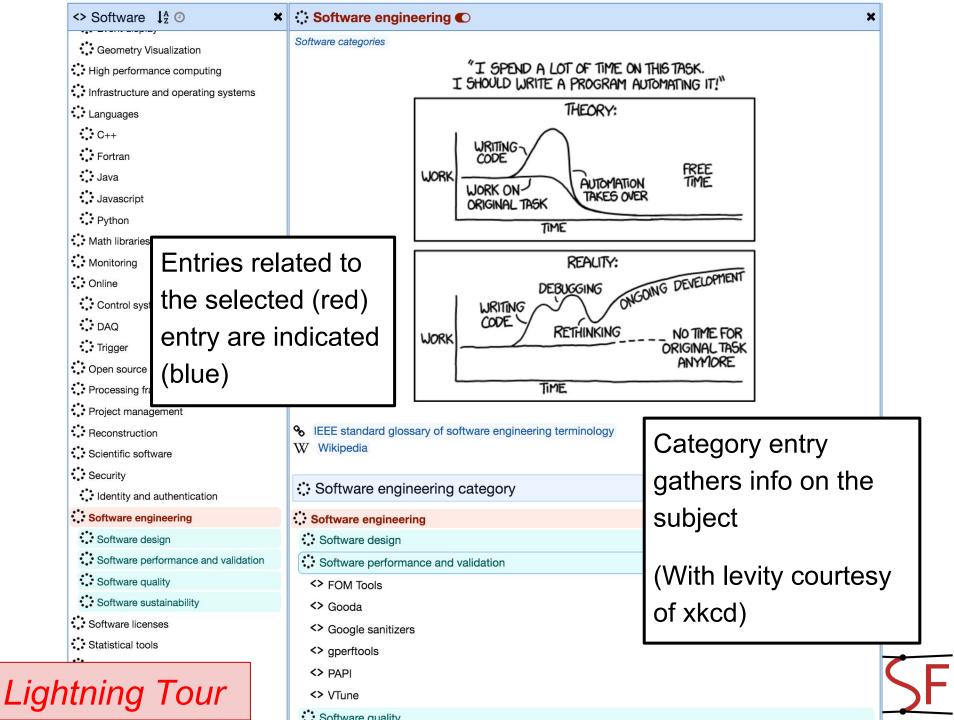
Used by Q Resources

Adding and editing material



Software projects, categories







:: Software quality

:: Control systems

DAQ.

Trigger

Open source

Processing frameworks

Project management

Reconstruction

Scientific software

Security :

:: Identity and authentication

Software engineering

Software design

Software performance and validation

Software quality

Software sustainability

Software licenses

Statistical tools

Supercomputing

Training software

Trigger/DAQ

:.. User

Environment Management

Virtualization and clouds

:.. Web

:: Web app frameworks

: Web based tools and services

Software categories



...WOW. THIS IS LIKE BEING IN A HOUSE BUILT BY A CHILD USING NOTHING BUT A HATCHET AND A PICTURE OF A HOUSE.



IT'S LIKE A SALAD RECIPE WRITTEN BY A CORPORATE LAWYER USING A PHONE AUTOCORRECT THAT ONLY KNEW EXCEL FORMULAS.



IT'S LIKE SOMEONE TOOK A TRANSCRIPT OF A COUPLE ARGUING AT IKEA AND MADE RANDOM EDITS UNTIL IT COMPILED WITHOUT ERRORS. OKAY I'LL READ A STYLE GUIDE.

W Wikipedia

Programming style - Wikipedia

Indent style - Wikipedia

Naming convention - Wikipedia

W Coding conventions - Wikipedia

:: Software quality category

Software quality

Parent

Software engineering

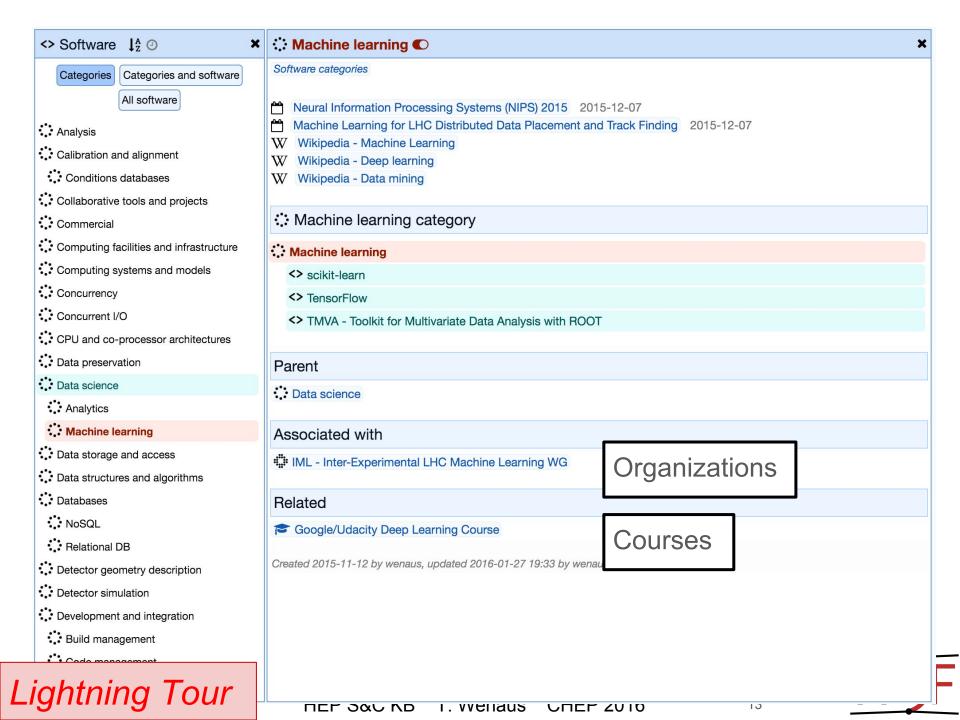
Related

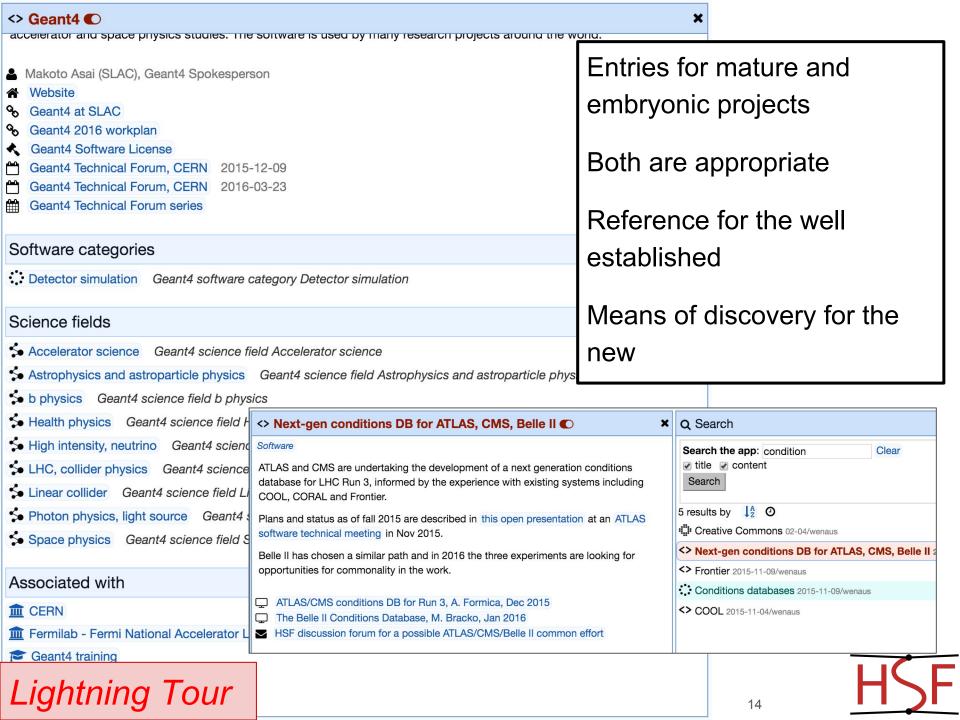
Developing maintainable software | Software Sustainability Institute

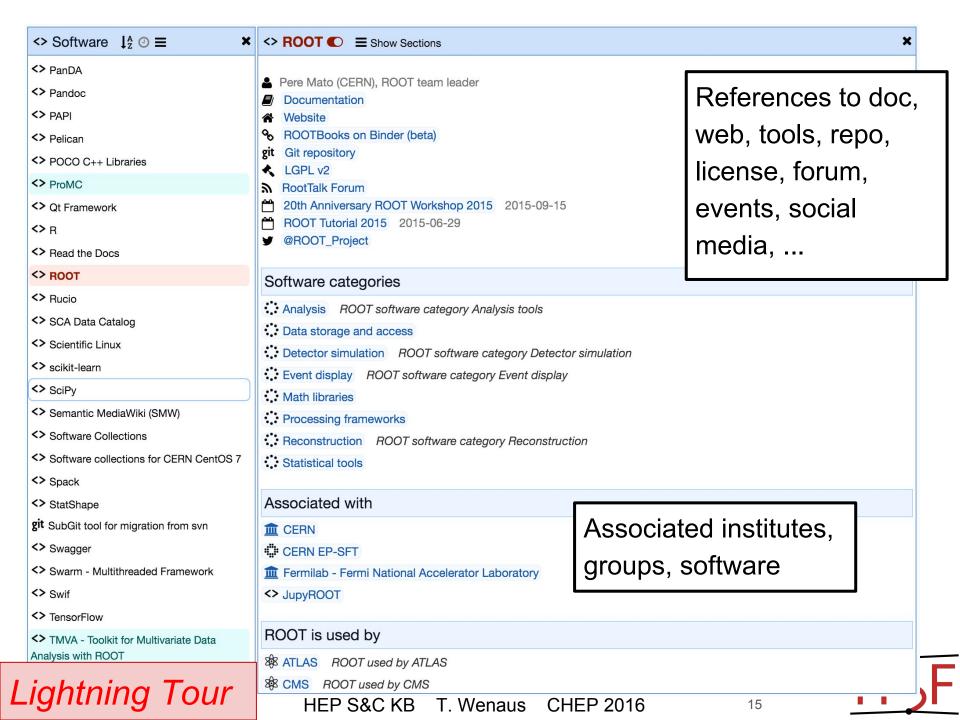
The Beauty of Code - Paris Review

Writing readable source code | Software Sustainability Institute

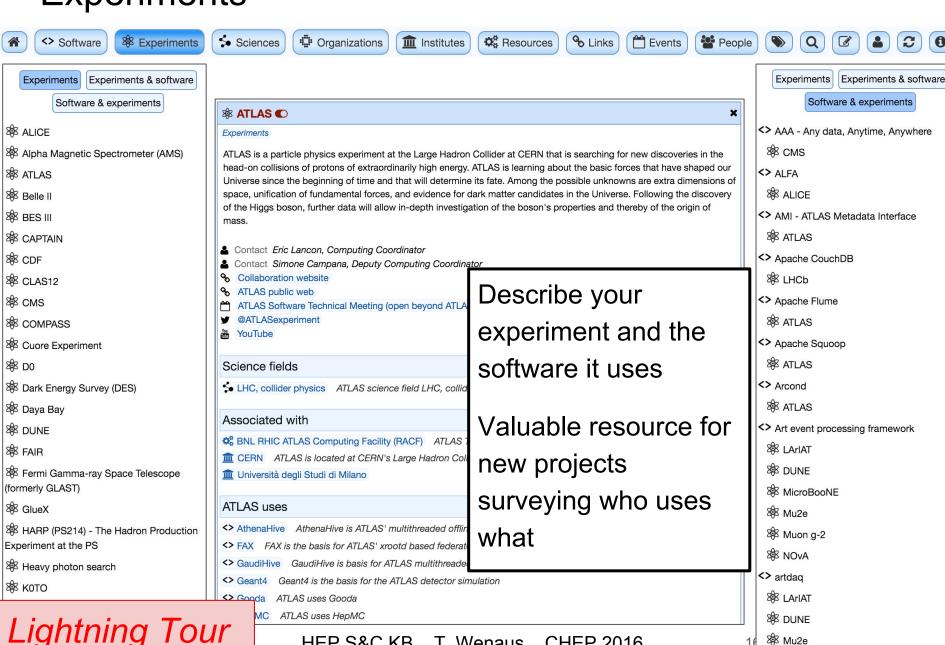
Linked articles







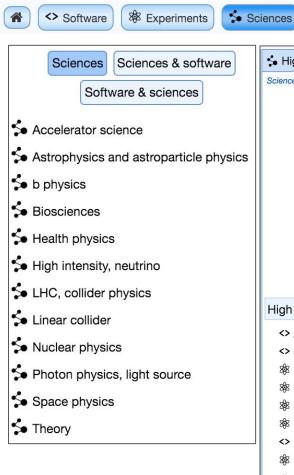
Experiments



HEP S&C KB T. Wenaus CHEP 2016

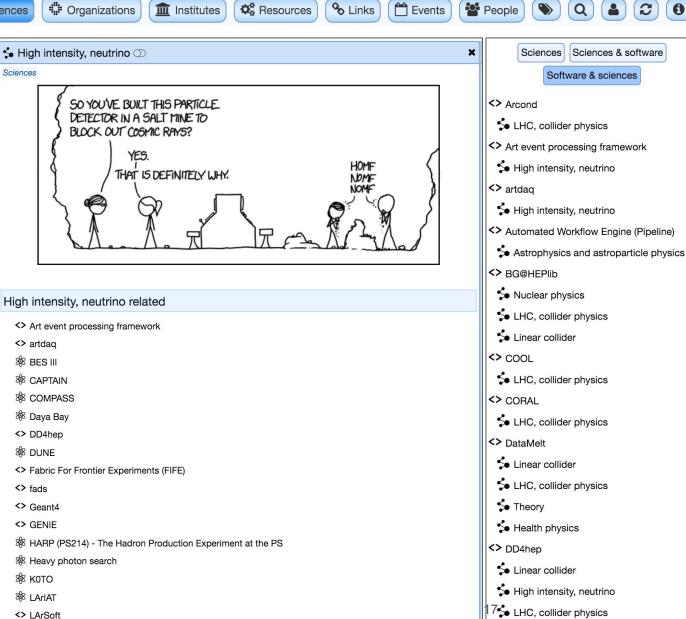
₩ Mu2e

Sciences

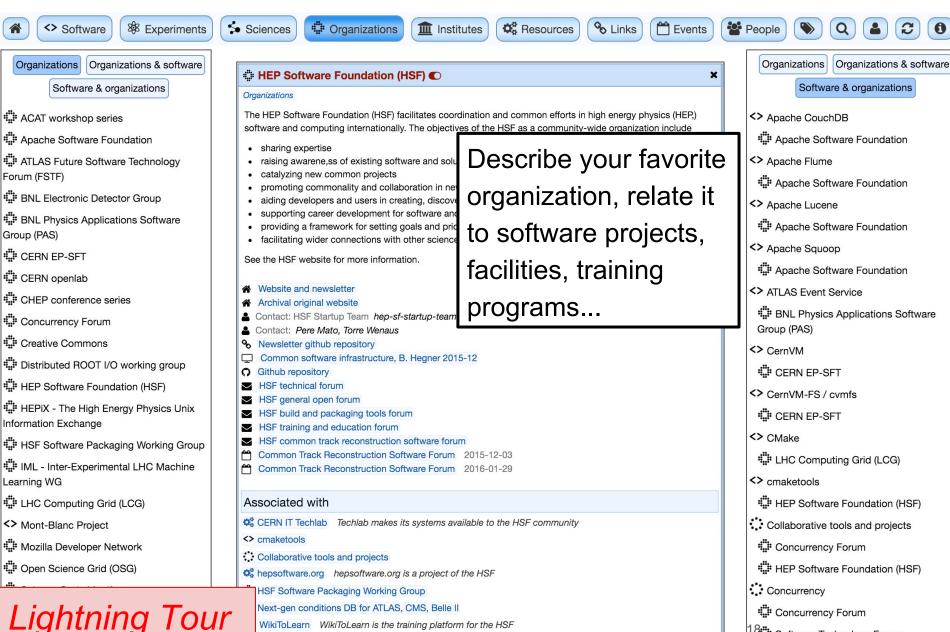


Lightning Tour

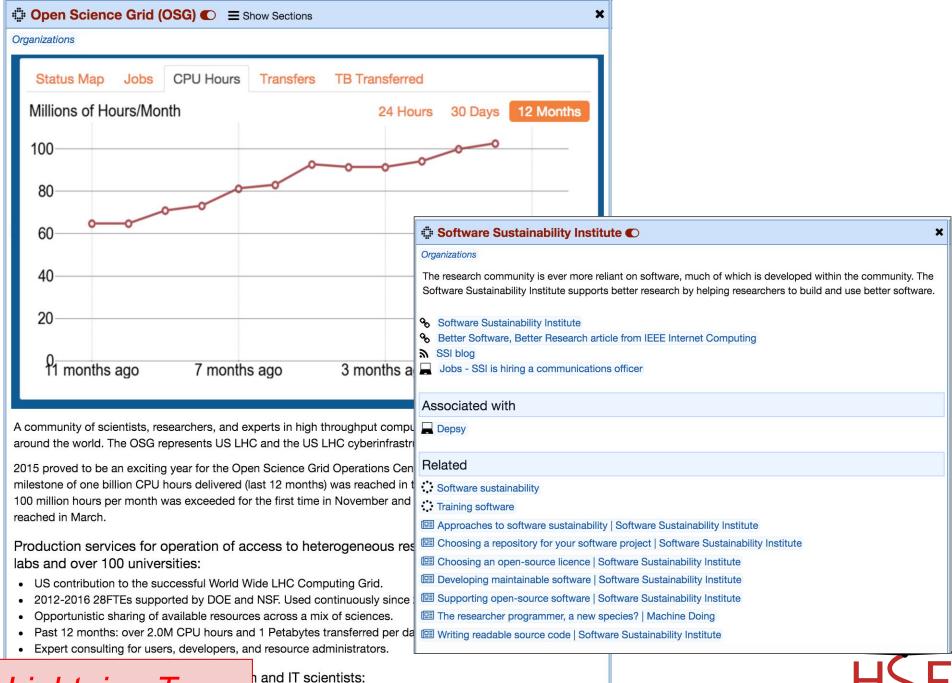
<> LArSoft



Organizations



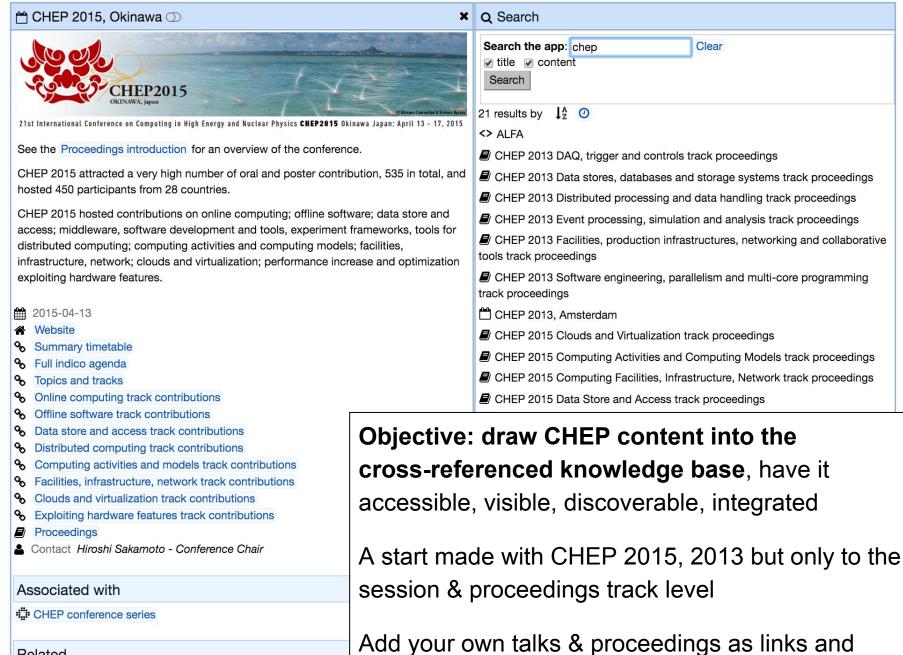
Software Technology Forum



Lightning Tour

e driven by research community.

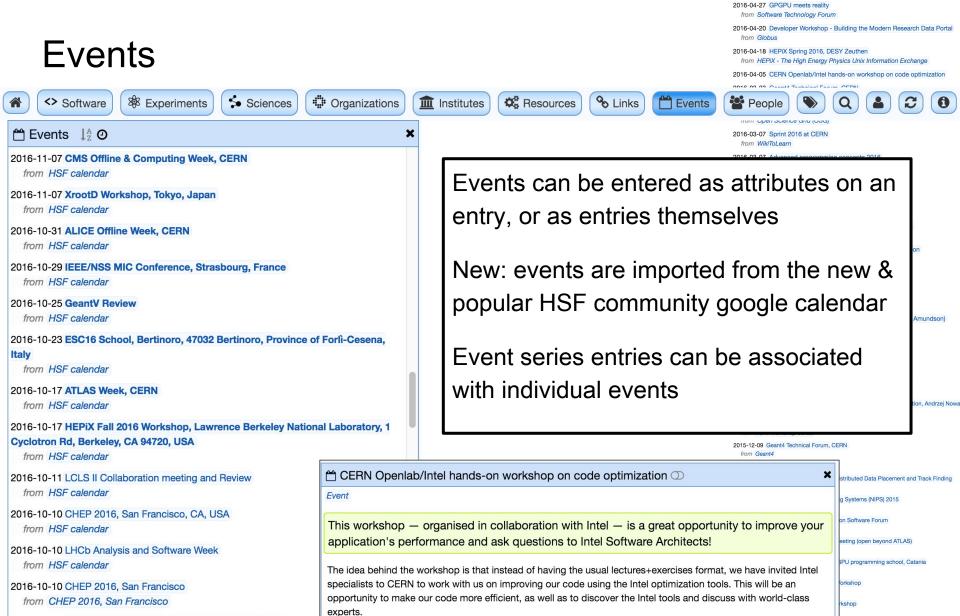




Lightning Tour

Related

ack proceedings omputing Models track associate them with their session, sw category etc



£ 2016-04-05

Lightning Tour

2016-08-31 Higgs Hunting 2016, Orsav

2016-09-12 4th Workshop on Sustainable Software for Science

from 4th Workshop on Sustainable Software for Science

2015-10-12 HEPIX Fall 2015, BNL
from HEPIX - The High Energy Physics Unix Information Exchange

2015-06-29 ROOT Tutorial 2015 from ROOT

2015-09-15 20th Anniversary ROOT Workshop 2015 from ROOT

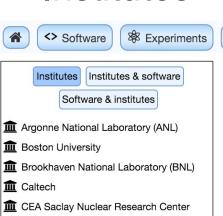
Andrzej Nowak

T. Wenaus CHEP 2016

Institutes

Sciences

Organizations



- **III** DESY Deutsches Elektronen-Synchrotron
- Fermilab Fermi National Accelerator Laboratory
- m Harvard University

Cornell University

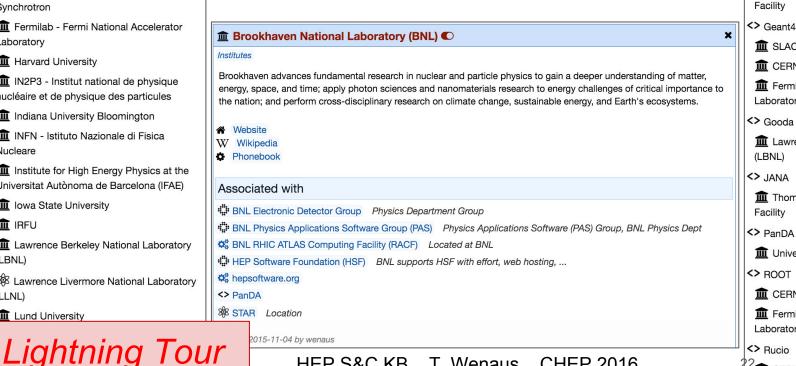
CERN

- i IN2P3 Institut national de physique nucléaire et de physique des particules
- Indiana University Bloomington
- INFN Istituto Nazionale di Fisica Nucleare
- Institute for High Energy Physics at the Universitat Autònoma de Barcelona (IFAE)
- i lowa State University
- **m** IRFU
- Lawrence Berkeley National Laboratory (LBNL)
- Lawrence Livermore National Laboratory (LLNL)
- in Lund University

Universities and labs, with relations to the software, S&C groups, services, facilities, experiments etc. that are associated with them

III Institutes

Resources



Institutes Institutes & software Software & institutes AMI - ATLAS Metadata Interface IN2P3 - Institut national de physique nucléaire et de physique des particules <> Ceph **III** CERN Rutherford Appleton Laboratory (RAL) <> CLARA Thomas Jefferson National Accelerator Facility <> Geant4 III SLAC CERN Fermilab - Fermi National Accelerator Laboratory <> Gooda **III** Lawrence Berkeley National Laboratory (LBNL) <> JANA Thomas Jefferson National Accelerator Facility

university of Texas at Arlington (UTA)

Fermilab - Fermi National Accelerator

CERN

Laboratory

CERN

Q

Figure 1

People

S Links

HEP S&C KB T. Wenaus CHEP 2016

Describe your group as an organization, relate it to your software projects

© CERN EP-SFT €

×

Organizations

SoFTware Development for Experiments

The SFT group in the EP department develops and maintains common scientific software for the physics experiments in close collaboration with the EP experimental groups, the IT department and external HEP institutes. The majority of the group is involved in projects organised as part of the Applications Area of the LHC Computing Grid (LCG) project. In addition, several group members have direct responsibilities in the software projects of the LHC experiments. SFT projects are organised in four core areas: Simulation, Libraries and Frameworks, Distributed Systems and Collaborations with Experiments.

- Website
- % CERN SFT Google Summer of Code
- CERN SFT GitHub repository

BNL Physics Applications Software Group (PAS)

Organizations

The Physics Applications Software (PAS) group in BNL's Physics Department dev software in support of BNL's HEP program. The largest PAS effort goes to the ATI members also contribute to the DUNE neutrino physics collaboration and the LSS

- Website
- Contact Alexei Klimentov, PAS Group Leader

Associated with

- ATLAS Event Service BNL's PAS group is a lead contributor to the Event Ser
- m Brookhaven National Laboratory (BNL) Physics Applications Software (PAS)
- hepsoftware.org PAS developed and hosts hepsoftware.org
- NICOS nightly build system NICOS was developed by BNL's PAS Group
- PanDA PAS is a lead contributor to PanDA

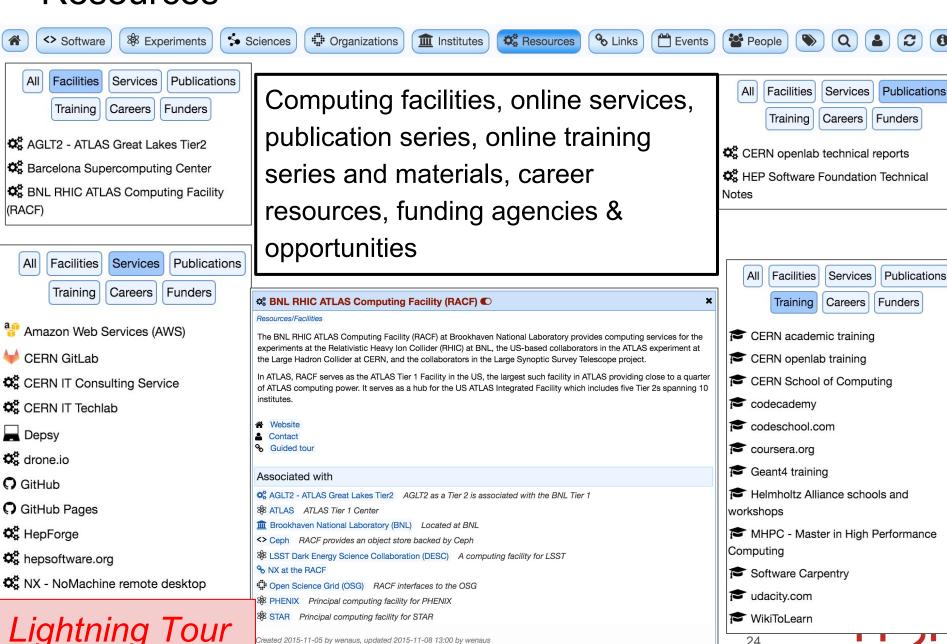
Associated with

- **CERN**
- <> CernVM
- Concurrency Forum
- GENSER Generator Service Project
- <> HepMC3
- Jupyter@CERN
- LHC Computing Grid (LCG)
- <> ROOT
- Software Technology Forum

Reporting they are used by CERN EP-SFT

<> CernVM-FS / cvmfs

Resources



24

Created 2015-11-05 by wenaus, updated 2015-11-08 13:00 by wenaus

and its projects

AGLT2 - ATLAS Great Lakes Tier2 Show Sections

ATLAS Great Lakes Tier 2 provides computing and storage capacity for US ATLAS physicists running ATLAS simulations and data analysis. We currently provide more than 4000 total job slots (6600 cores) and 3.7 Petabytes of storage capacity interfaced to the Open Science Grid. Our job scheduling system is Condor and we utilize dCache, Lustre, AFS, and NFS as

AGLT2 is a federated Tier 2, comprising facilities at U Michigan and Michigan State, operating as a single unified facility.

AGLT2 is one of the top ATLAS Tier 2 sites worldwide in terms of CPU hours provided. In any given 24 hour period we average more than 8000 completed ATLAS jobs. We transfer an average of 1-2TB of data per day to and from other sites.

AGLT2 is also an ATLAS Muon Spectrometer Calibration Center.

Networking

storage systems.

AGLT2 is a leader in advancing networking, including advancing perfSONAR as a basis for uniform network performance monitoring across LHC computing resources.

PunDIT

AGLT2 is a collaborator on the PuNDIT project which will integrate and enhance several software tools to provide an infrastructure for identifying, diagnosing and localizing network problems. In particular, the core of PuNDIT is the Pythia tool that uses perfSONAR data to detect, identify and locate network performance problems.

OSIRIS

AGLT2 is currently developing OSIRIS, a distributed, multi-institutional storage infrastructure that will allow researchers at any of our campuses to read, write, manage and share their data directly from their computing facility locations.

Our goal is to provide transparent, high-performance access to the same storage infrastructure from well-connected locations on any of our campuses. We intend to enable this via a combination of network discovery, monitoring and management tools and through the creative use of CEPH features.

By providing a single data infrastructure that supports computational access on the data "in-place", we can meet many of the data-intensive and collaboration challenges faced by our research communities and enable these communities to easily undertake research collaborations beyond the border of their own Universities.

- Website
- Contact
- Wiki
- **OSIRIS**
- & PunDIT

Associated with

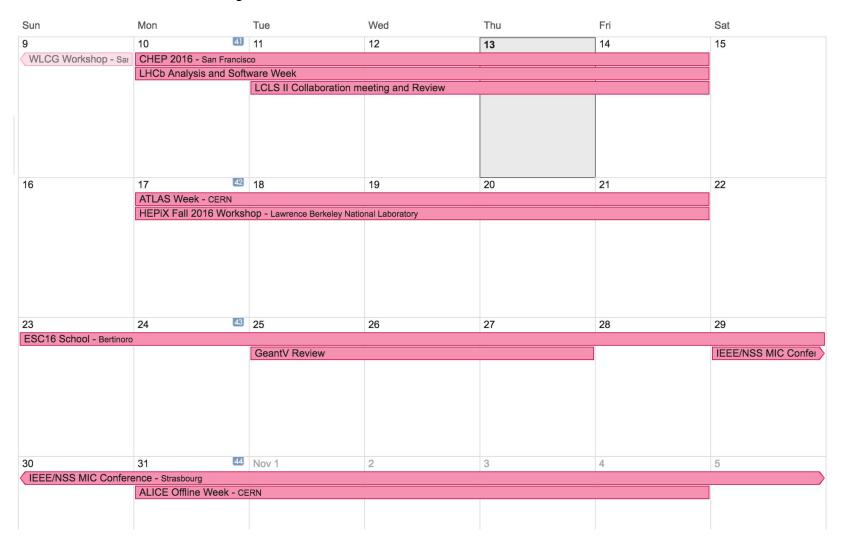
S BNL RHIC ATLAS Computing Facility (RACF) AGLT2 as a Tier 2 is associated with the BNL Tier 1

Open Science Grid (OSG) AGLT2 interfaces to the OSG





HSF community calendar



https://calendar.google.com/calendar/embed?src=e4v33e1a1drbncdle1n03ahpcs%40group.calendar.google.com



Licenses

Software

Website

Downloads

Tutorials

Jobs

Related

Solution Simple Sim

<> DD4hep

<> Geant4

<> FOM Tools

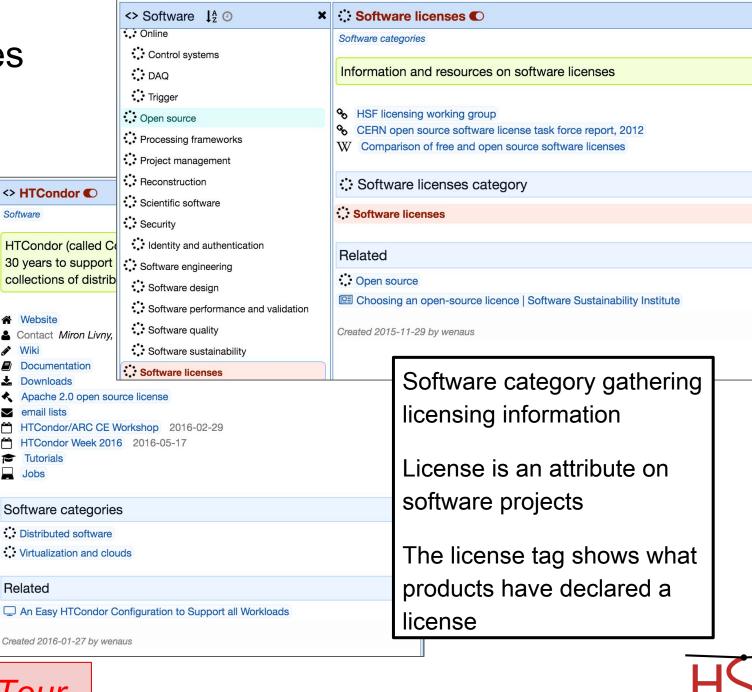
<> HTCondor

<> TensorFlow

<> ROOT

hepsoftware.org

Has designated a license

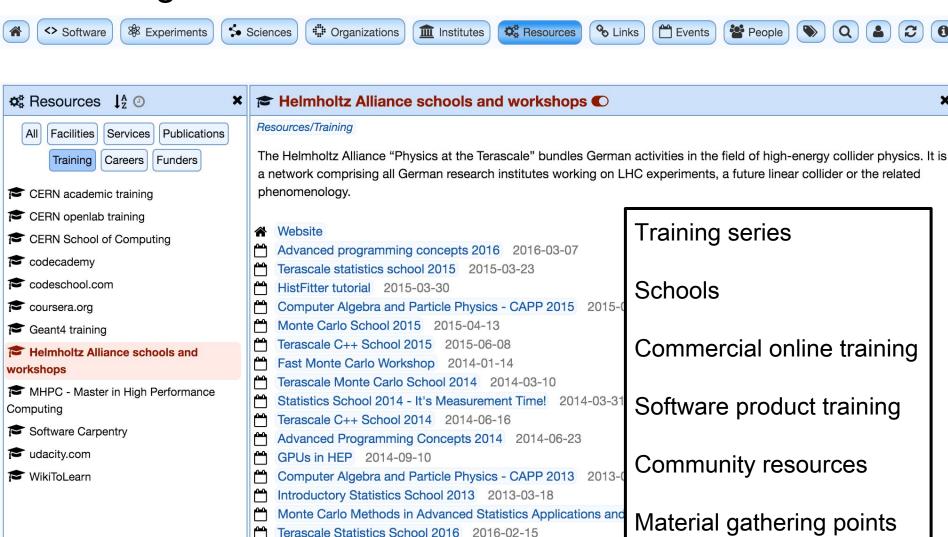




HEP S&C KB T. Wenaus

27

Training

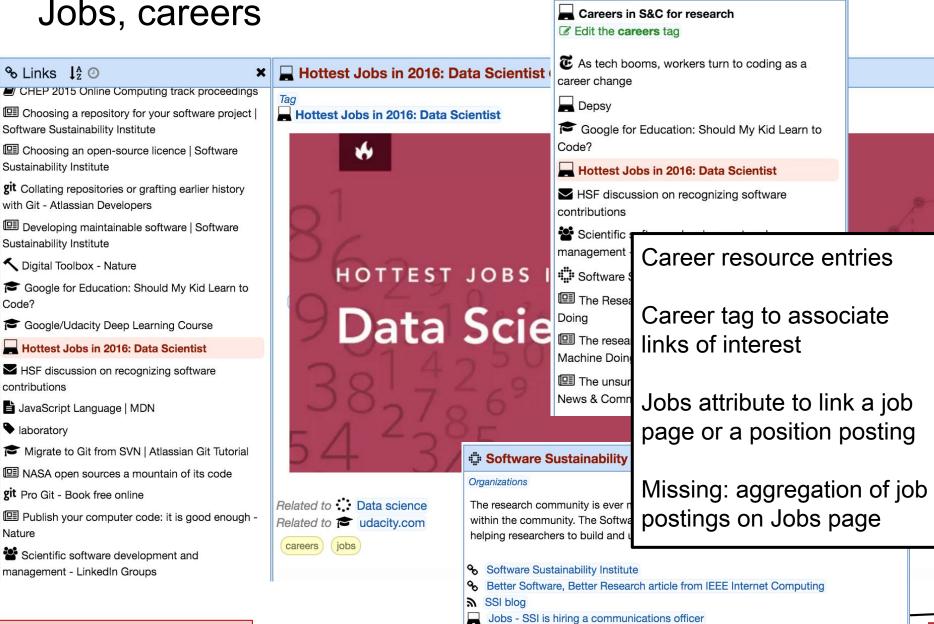






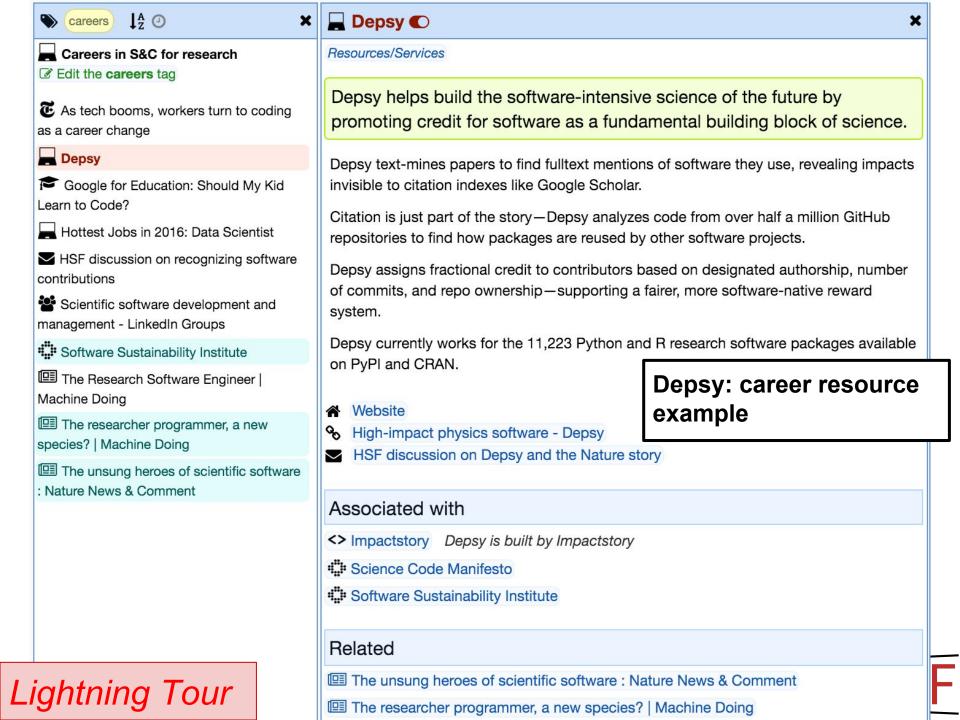
Created 2016-01-23 by wenaus

Jobs, careers



Lightning Tour

1A @



Tags



Lightning Tour

Recent addition not much used yet but powerful

& Links

Events

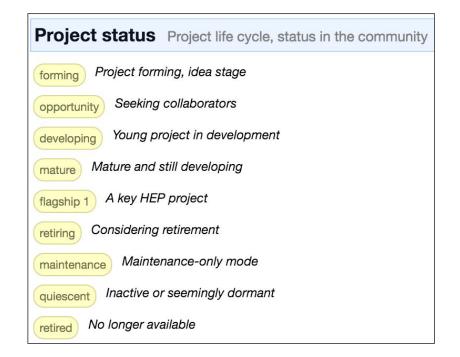
People P

Full flexibility to invent classifications

♥ Resources

iii Institutes

" Organizations

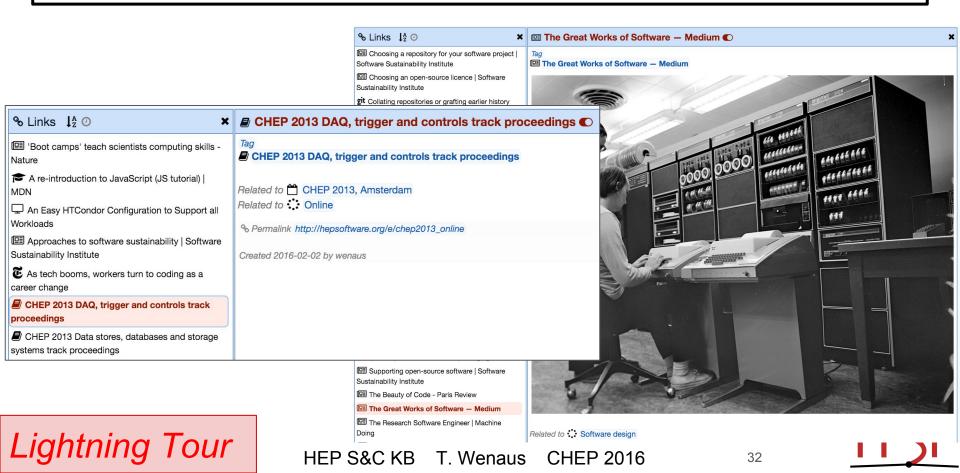


Links

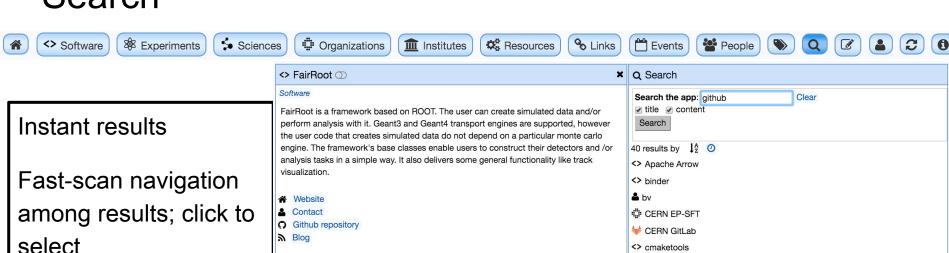
You can add anything as a link (as long as it's a link)

When you read a cool software design article, add it as a link

eg to relate a CHEP track's proceedings to its software category, add it as a link

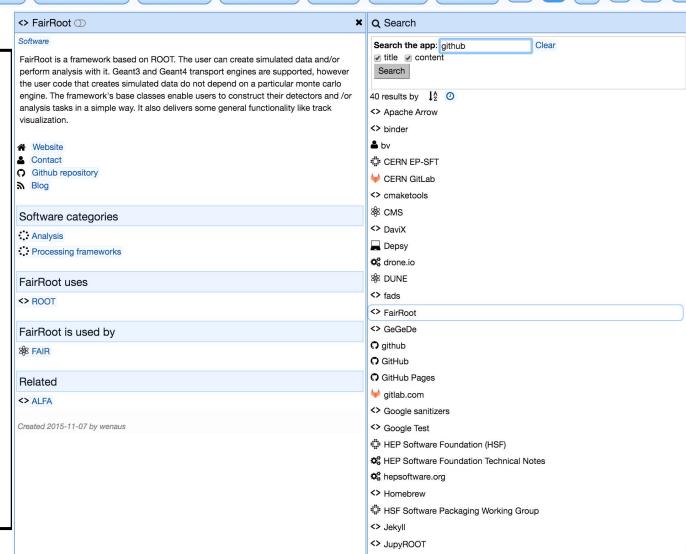


Search



Search by title and/or content

Empty search shows all entries, sorted by alpha or modification time



Accounts, authentication

Sciences



Experiments

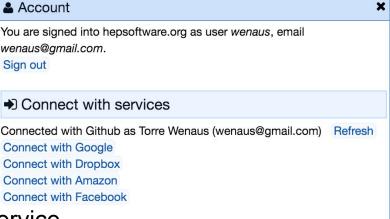
<> Software

- No internal user/pass database
- Authentication based on email address
- Obtained from your choice of third party
 - GutHub, Google, Dropbox, Amazon, ...
 - As you often see these days on websites
- Missing is CERN, they have a new OAuth2 service

Organizations

institutes

- REDIS based server-side session management
- Once authenticated, you have an identity in the KB, associating you with content you create and modify, testimonials you add, preferences you set
 - People not supported as entries properly yet
 - You should be able to relate people to other entries (software, experiments, institutes, ...) but can't yet
- Perhaps more personalizastion in the future
 - Experiment association, enabling experiment-specific and -internal content
 - Notifications of new content... RSS? email?



People

& Links

Resources

Events



Versions, visibility

Previous versions are archived and can be viewed and restored as the published version

Nothing can be deleted by the user, you can hide items

Versions

Previous versions are accessible here. They can be loaded as the edited version (so long as doing so would not overwrite active edits), and then saved, in order to revert to previous versions.

Current 2014-10-20/wenaus, updated 2015-11-11 17:13/wenaus

Archive 2014-10-20/wenaus, updated 2015-11-11/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-11-11/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-11-11/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-11-05/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-11-04/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-10-27/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-10-27/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-10-26/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-10-25/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-10-25/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-10-15/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-10-15/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-10-07/wenaus	Load
Archive 2014-10-20/wenaus, updated 2015-10-07/wenaus	Load
Archive 2014-10-20/wenaus, updated 2014-10-20/wenaus	Load



Next steps

Add and improve content! An on-and-off hobby of mine but it needs you

 New entries, testimonials on software, dressing entries with tags and relations, adding that cool article on online training you found, adding your training event, ...

In terms of development, best effort and driven by interest

- Ease of use, aesthetics, performance
- Clean up the code and open it up in github
- People as full fledged entries with associations
- People-experiment association to enable experiment-specific (including private) material



Summary

The HEP S&C knowledge base at hepsoftware.org is there for you to try, as a user and as a content contributor

Stable, functional, zippy, even kind of fun to use and add stuff when you get in the swing of it

Benefits from a couple rounds of prototype cycles

Smart javascript client locally hosting the data makes for a very responsive dynamic browser app

Node.py + MySQL based back end in EC2 has been solid & stable

What it needs most now is content!



