

# Updates on Software and the HEP Software Foundation

Benedikt Hegner  
(EP-SFT, CERN)

LHCC Referee Meeting  
1.3.2016

Using input from LHC experiments and HSF

# What's ahead of us?

- **Increasing complexity and multiplicity of physics events** a challenge for simulation and reconstruction performance
- The free lunch is over (no news!) and we have to **actively address problems of SW performance**
- Have to re-learn many things due to **evolution of CPUs** and **new architectures arriving**
- Do not have the resources to evaluate every **new technology** in each experiment independently
- Have to **seek commonality and collaboration** in new developments resulting from the evaluations

⇒ **Improve both on SW efficiency and efficient usage of people's time**

# Current Status and Activities

- **[HEP Software Foundation \(HSF\)](#) as the umbrella for addressing these problems together!**
  - Collection of ideas and proposals in 2014 and startup-team formed
  - Kick-off workshop Jan 2015 at SLAC
- **Sharing expertise**
  - Schools, trainings and courses (not always easy to find), [wikiToLearn](#)
  - [HEP S&C Knowledge Base](#)
  - HSF Technical notes
  - Topical fora and working groups in the HSF
- **New hardware architectures and technologies**
  - [Concurrency forum](#)
    - Evolving into a general software technology forum
  - Usage of resources provided on best-effort basis by e.g. CERN's TechLab / Openlab
  - Porting to new architectures efforts within the LHC experiments

# Current Status and Activities II

- **Software performance**
  - Simulation: parallelisation of Geant4; [GeantV](#)
  - Reconstruction: HSF common tracking SW forum + [Machine Learning Forum](#)
  - I/O: parallel ROOT I/O, key-value-store evaluations
  - Mathematics: MetaLibm
  - Ad-hoc improvements and parallelization in various SW projects
  - Performance tools (e.g. [igprof](#), [FOM tools](#))
- **Supporting developers and participating projects**
  - Providing best practices to facilitate integration into HEP ecosystem
  - Project templates for bootstrapping new projects
  - Development services
  - Help in selecting the proper SW license
- **Quite some activity in HSF, even though participation in the startup-team is on volunteer/best-effort level**

# HSF Activities and Working Groups

Working Group	Objectives	Forum - Mailing list
<a href="#">Communication and information exchange</a>	Address communication issues and building the knowledge base Technical notes	<a href="#">hep-sf-tech-forum</a>
<a href="#">Training</a>	Organization of training and education, learning from similar initiatives	<a href="#">hep-sf-training-wg</a>
<a href="#">Software Packaging</a>	Package building and deployment, runtime and virtual environments	<a href="#">hep-sf-packaging-wg</a>
<a href="#">Software Licensing</a>	Recommendation for HSF licence(s)	<a href="#">hep-sf-tech-forum</a>
<a href="#">Software Projects</a>	Define incubator and other project membership or association levels. Easy-start project templates	<a href="#">hep-sf-tech-forum</a>
<a href="#">Development tools and services</a>	Access to build, test, integration services and development tools	<a href="#">hep-sf-tech-forum</a>

# HSF Topical Fora

- **Software Technology Forum**
  - Technical issues to embrace new technology in our software
  - Ongoing activity
- **Reconstruction Algorithms Forum**
  - All matters of event reconstruction and pattern recognition software; 3 in-person meetings
- **Machine Learning Forum**
  - ML discussions and code development in the context of HEP
  - Development of relevant tools, methodology and applications

# Cross-experiment Collaborations

- **There are quite a few (more or less) new cross-experiment collaborations, with involvement or moderation of the HSF - going beyond WLCG**
- **Experiment frameworks**
  - Gaudi (ATLAS, LHCb, FCC)
  - FAIRRoot (FAIR, ALICE)
- **Common Conditions Data Project**
  - Discussion/cooperation between ATLAS, Belle II, CMS and LHCb
- **Common Software Build and Packaging Tool efforts**
  - [Working group of HSF](#) comparing HEP and non-HEP solutions
  - Starting point was LCG's Librarians and Integrators Meeting
- **Cooperation on Reconstruction Software**
  - "Connecting the Dots" tracking workshop extended by HSF session about common tracking implementations
- **AIDA2020** (EU funded)
  - DD4hep for detector description (LCD, FCC, potentially LHCb)
  - PODIO data model library (FCC, LCD, potentially LHCb)
- **DIANA (Data Intensive ANALysis)** (NSF funded)
  - 4-year project on on analysis software, including ROOT and its ecosystem

# (Physics) Code Optimization

- During LS1 all experiments invested heavily in improving their software
  - Required detailed expert knowledge rare
    - Important to invest on the right topic
  - Significant gains achieved

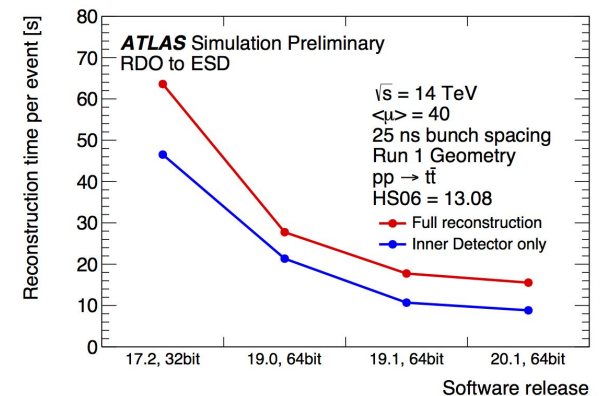
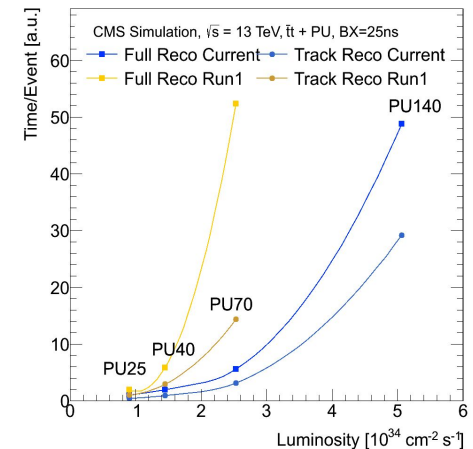
- **One example - ATLAS and CMS Tracking**

- Better/vectorized implementations of calculations
- Addressing hot-spots, e.g. magnetic field in ATLAS
- Tuned reconstruction strategy for higher pileup, e.g. new seeding algorithms in CMS
- Up to x4 improvements in speed

- **Gained expertise fed back into community**

- Ongoing efforts to make improved tracking code available to the wider community

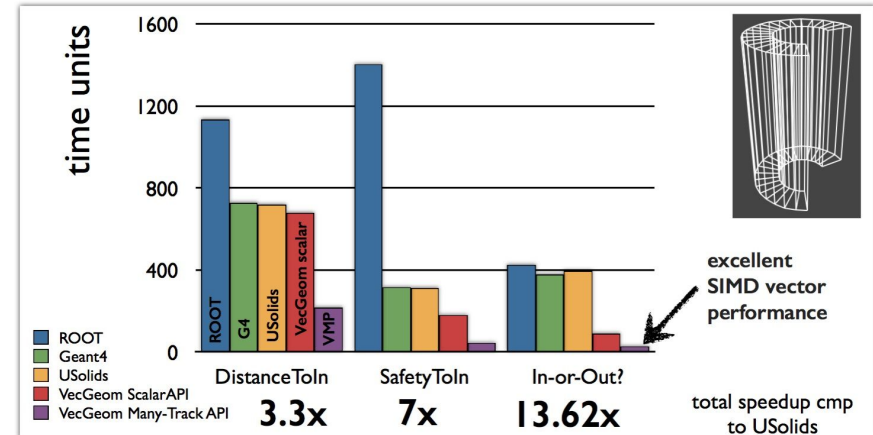
- **Tracking performance will stay an important topic for HL-LHC**





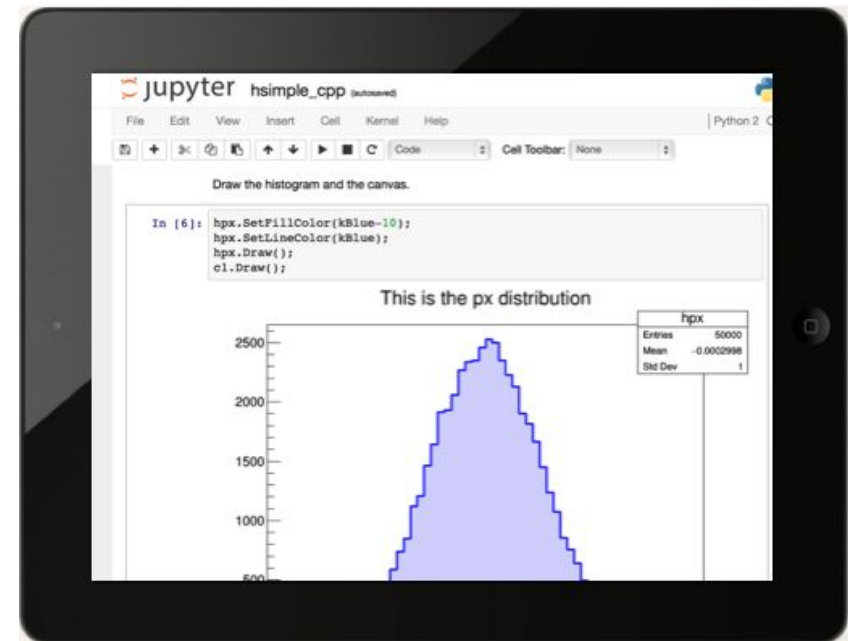
# Preparing Simulation for the Future

- **Detector simulation with Geant4 is one of the essential ingredients**
  - It is the largest (50%) single consumer of the WLCG cycles.
- **Geant4 has been (naively) parallelized with Geant4 Version 10**
- **The potential of the current Geant4 design for long-term adjustment is limited**
  - Software design more than 20 years old (but serves the community well!)
  - Triggered new project - *GeantV*(ectorized)
    - Currently in R&D phase
    - 2016 review milestone is a CMS-like detector
- **Necessary revolution of software can even be backported**
  - Geometry navigation takes 30-40% of CPU time in simulation
  - Addressed with new “VecGeom” library in GeantV
  - Substantial performance gains also in scalar mode & Geant4



# New Paradigms arriving

- **Software needs to operate well in a wider set of environments**
  - Clouds and supercomputers require more flexibility on the application side
- **Cloud services not only limited to providing resources and storages**
  - Software-as-a-service a very successful model (e.g. webmail, Google Doc)
  - Could extend this to the physics analysis use-case
- **ROOT investigating new paradigms**
  - Notebook interfaces (Jupyter) for browser-based analysis
  - Declarative instead of imperative data analysis (“the what and not the how”)
- **Impact and chances for physics analysis workflows yet unclear**



**Backup**

# Support for multiple Architectures

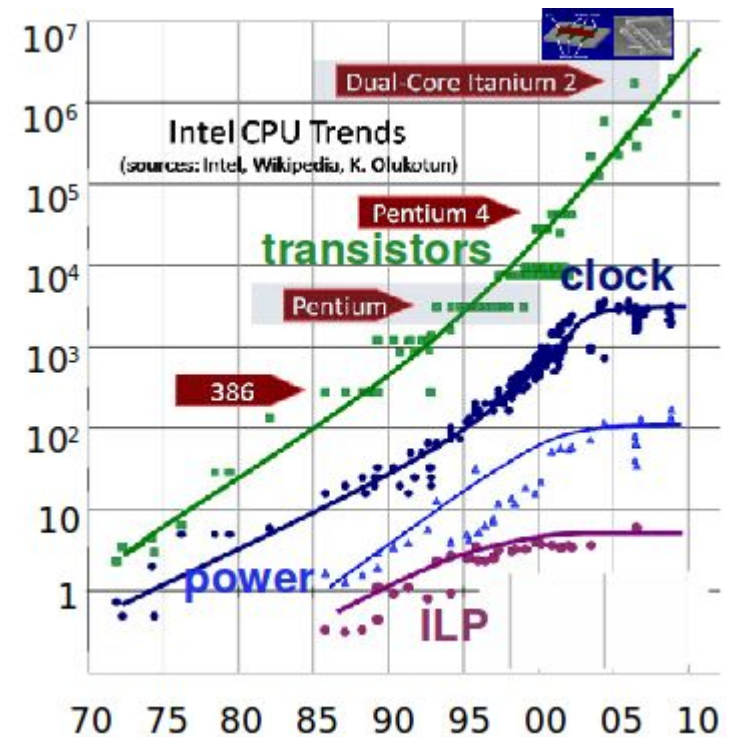
- **For many years we have been developing for/running on x86 (Intel) architectures**
  - Other platforms mainly for code-checking
- **The landscape is much more complex now**
  - ARM coming from the low-power consumption market
  - Power8 a common platform in super-computing/HPC environment
  - Dedicated vector processors (e.g. GPUs)
- **Ongoing activities**
  - Projects like Geant4 / ROOT already being tested on a variety of platforms
  - CERN IT providing multiple platforms on best-effort basis (TechLab)
  - Porting efforts within the experiments
  - Concurrency forum

# HSF Timeline

- Jan 2014: [HEP software collaboration proposed](#)
- Apr 2014: [HEP software collaboration meeting](#)
- Spring/Summer 2014: gathering White Papers from the community.
- Oct 1 2014: [Startup plan approved and startup team established](#). Agreement communities and software domains to focus on initially.
- Nov 11 2014: [White Paper Analysis and Proposed Startup Plan](#) released, followed by discussions with many parts of the community prior to the SLAC workshop.
- Jan 20-22 2015: [SLAC HSF workshop](#) established concrete activities and next steps
- Apr 17, 2015: [HSF meeting at CHEP 2015, Okinawa](#) to present progress, assess opportunities emerging from CHEP, and discuss next steps.
- June/July 2015: Intensive discussions in Packaging Working Group
- Sep 2015: [Technical Notes](#) policies published - more in the queue
- Sep 2015: HSF on WikiToLearn
- Nov 2015: HEP Knowledge Base finished
- May 2016: HSF Workshop in Paris

# Stagnation in Speedup

- Performance of our code scales with clock cycle (hence is stagnant!)
- Needs will increase more than tenfold and the budget will be constant at best
- HEP code needs to exploit new architectures and to team with other disciplines to share the optimization effort
  - Data & instruction locality and vectorisation
- Portability, better physics and optimization will be the targets
- Simulation can lead the way to show how to exploit today's CPU's resources more effectively in complex applications



# Knowledge Base

- **Software catalog, software categories, science fields, community, and events**
  - implementation is a browser-based app ( javascript client, node.js server, json in between, MySQL)
  - authentication is via github, google, facebook etc.
  - **emphasizes easy adding/editing of content, and extensibility. Adding content should be fun.**
- Available at <http://hepsoftware.org>
  - Comments/feedback are welcome!
  - Just start filling it!
- *Implementation based on ATLAS' DKB (data knowledge base)*

The screenshot displays the AthenaHive portal interface for the ATLAS experiment. At the top, there is a navigation bar with various icons. Below it, the 'Experiments' section is active, showing a list of experiments on the left and detailed information for 'ATLAS' on the right.

**Experiments List (Left Panel):**

- ALICE
- Alpha Magnetic Spectrometer (AMS)
- ATLAS** (highlighted)
- Belle II
- BES III
- CAPTAIN
- CDF
- CLAS12
- CMS
- COMPASS
- Cuore Experiment
- D0
- Dark Energy Survey (DES)
- Daya Bay
- DUNE
- FAIR
- Fermi Gamma-ray Space Telescope (formerly GLAST)
- GlueX
- HARP (PS214) - The Hadron Production Experiment at the PS

**ATLAS Page Content (Right Panel):**

**ATLAS**

*Experiments* <http://www.hepsoftware.org/e/atlas>

ATLAS is a particle physics experiment at the Large Hadron Collider at CERN that is searching for new discoveries in the head-on collisions of protons of extraordinarily high energy. ATLAS is learning about the basic forces that have shaped our Universe since the beginning of time and that will determine its fate. Among the possible unknowns are extra dimensions of space, unification of fundamental forces, and evidence for dark matter candidates in the Universe. Following the discovery of the Higgs boson, further data will allow in-depth investigation of the boson's properties and thereby of the origin of mass.

**Contact:**

- Contact *Eric Lancon, Computing Coordinator*
- Contact *Simone Campana, Deputy Computing Coordinator*

**Links:**

- [Collaboration website](#)
- [ATLAS public web](#)
- [ATLAS Software Technical Meeting \(open beyond ATLAS\)](#) 2015-11-09
- [@ATLASexperiment](#)
- [YouTube](#)

**Science fields**

[LHC, collider physics](#) *ATLAS science field LHC, collider physics*

**Associated with**

- [BNL RHIC ATLAS Computing Facility \(RACF\)](#) *ATLAS Tier 1 Center*
- [CERN](#) *ATLAS is located at CERN's Large Hadron Collider (LHC)*
- [Università degli Studi di Milano](#)

**ATLAS uses**

- [AthenaHive](#) *AthenaHive is ATLAS' multithreaded offline framework*



# Software Packaging

- Topics
  - package building, deployment, runtime environment, new technologies like Docker, cmake best practices
- Organized a series of discussions/presentations on packaging and build tools (8 meetings)
  - Current practices inside and outside HEP
  - Document to summarize findings being prepared
- Trying a hands-on approach to increase share of actual code even if existing experiments and projects locked-in to a certain packaging solution
  - Common “build recipes” protocol

join the [hep-sf-packaging-wg](#)

# Build and Packaging Software Review

Looked at many tools, in particular

- worch, cmsBuild, aliBuild, LCGCMake, SciSoft, contractor (HEP), homebrew, Nix, conda

Main problems in HEP software

- reinvention of the square wheel
- non share even within the community

	Multi-Rel	Multi-BuildVar	MultiShell-RTE	Relocation
cmsBuild	+	+	+	+
Contractor	+	+	?	?
Homebrew	-	-	NA	-
LCGCMake	+	+	+	+
Nix	+	-	+	-
SciSoft	+	+	+	+
Worch	+	+	+	+

Main problems in non-HEP software:

- non HEP-tools prefer rolling releases / care less about reproducibility
- little support for multi-environment setups

Evolving document available at <https://github.com/HEP-SF/documents/tree/master/HSF-TN/draft-2015-PKG>

# Fostering Collaboration

HSF may have the role of foresting and publicising common software development initiatives. Some examples:

- Next-generation conditions data
  - Belle II / ATLAS / CMS / LHCb
  - 2 meetings: [Dec 10](#), [Jan 21](#)
- Track reconstruction
  - Huge challenges ahead
  - Should try to not only share ideas, but concrete code
- Frameworks: Gaudi
  - ATLAS / FCC / LHCb

Satellite projects in the HSF constellation. Examples:

- DIANA (Data Intensive ANALysis), 4-year NSF funded  
Focus on analysis software, including ROOT and its ecosystem

# Technical Notes

- Technical Notes can be proposals, ideas, whatever people want to add
- First TN with the TN policy has been published
- Some more in preparation:
  - Licence Guidelines, Naming conventions, packaging tools landscape, ...
- Repository and version control in GitHub

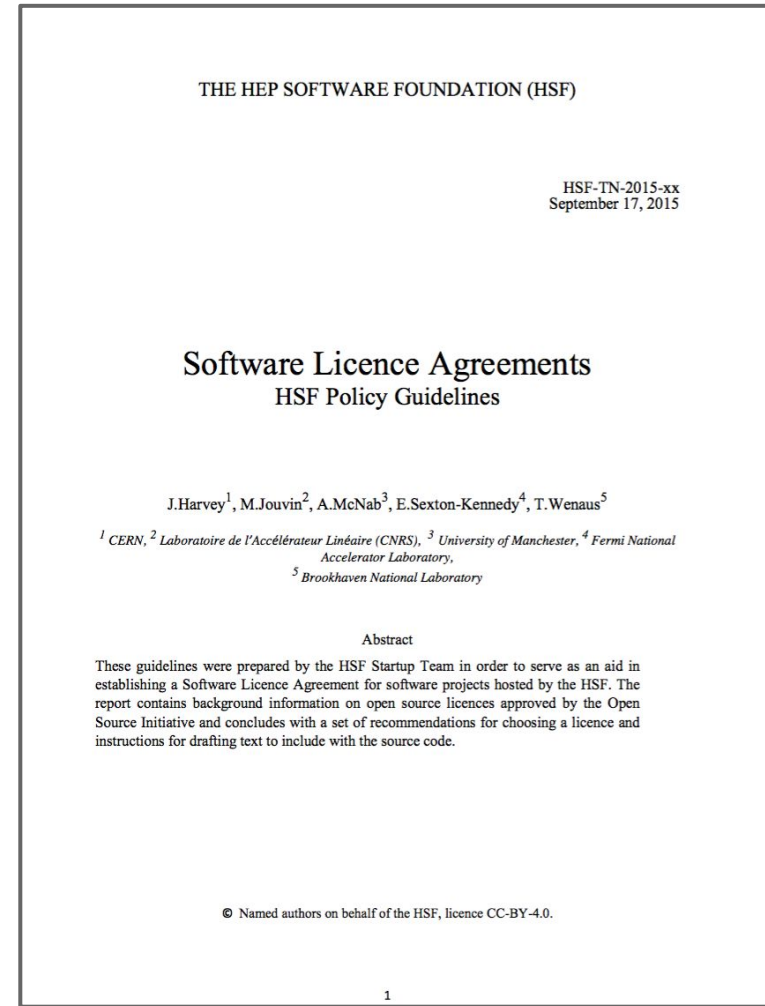
TN Number	Title	Authors	Download
HSF-TN-2015-01	HSF Technical Notes policy	A. McNab	<a href="#">PDF</a> <a href="#">GitHub</a>

## Drafts in the acceptance process:

Draft TN Reference	Title	Authors	Download
HSF-TN-2015-LIC	(Draft) Software Licence Agreements HSF Policy Guidelines	J. Harvey et al.	<a href="#">GitHub</a>
HSF-TN-2015-NAM	(Draft) HSF Platform Naming Conventions - A Proposal	B. Hegner	<a href="#">GitHub</a>
HSF-TN-2015-PKG	(Draft) HSF Packaging Working Group Report	B. Hegner, L. Sexton-Kennedy	<a href="#">GitHub</a>

# Software Licensing

- TN for the HSF Licensing Guidelines is being finalized in the technical forum
  - Plan to publish it this week
- Background information on open source licences approved by the Open Source Initiative
- Set of recommendations for choosing a license and instructions for drafting text to include with the source code



# WikiToLearn

- **WikiToLearn** is a wiki-based platform tailored at training and teaching
- Initiated in the context of italian universities
  - Basic idea was that students can improve and extend the material of their professors, while still being quality-controlled
- HSF jumped onto that to see whether we can take advantage of it
  - Started adding material to this site
- Now investing in providing interactive tutorials
  - think of the combination of jupyter style notebooks and a privately owned sandbox - start tutorial now, resume later (this even triggered a new collaboration w/ the ROOT team)
- This is only the shell, content has to come by the community



# Software Project Templates

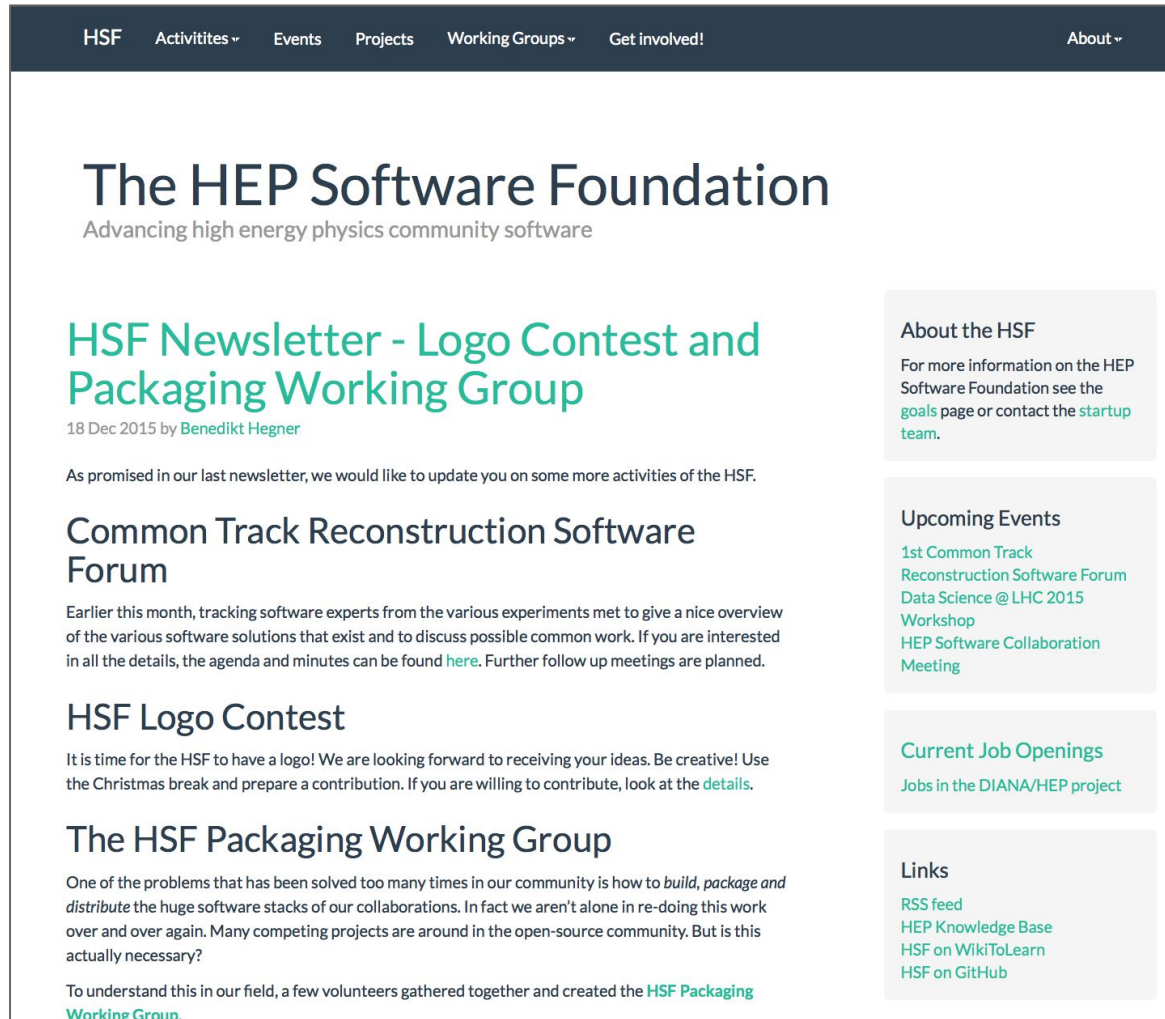
- The idea is to develop a **project template** implementing these guidelines and best practices
  - For example using the PODIO project (Toolkit for Event Data Models) as guinea pig developed under the AIDA 2020
  - A few more in the pipeline
- Prototype template available at <https://github.com/HEP-SF/tools>
- To support small projects that do not have a collaboration environment available
- To serve as example for shared projects across collaborations
  - reducing impedance mismatch

# Startup Team

- Amber Boehnlein (SLAC)
- Peter Elmer (Princeton)
- Daniel Elvira (FNAL)
- Frank Gaede (DESY)
- Benedikt Hegner (CERN)
- Michel Jouvin (LAL, IN2P3)
- Pere Mato (CERN, co-lead)
- Dario Menasce (INFN)
- Elizabeth Sexton-Kennedy (FNAL)
- Graeme Stewart (Glasgow)
- Craig Tull (LBNL)
- Andrea Valassi (CERN)
- Brett Viren (BNL)
- Torre Wenaus (BNL, co-lead)



- <http://hepsoftwarefoundation.org>



The screenshot shows the homepage of the HSF website. The navigation bar at the top includes links for HSF, Activities, Events, Projects, Working Groups, Get involved!, and About. The main heading is 'The HSF Software Foundation' with the tagline 'Advancing high energy physics community software'. The featured article is 'HSF Newsletter - Logo Contest and Packaging Working Group' by Benedikt Hegner, dated 18 Dec 2015. The article text discusses updates on HSF activities, a forum for reconstruction software, a logo contest, and a packaging working group. The right sidebar contains sections for 'About the HSF', 'Upcoming Events' (listing the 1st Common Track Reconstruction Software Forum Data Science @ LHC 2015 Workshop and HEP Software Collaboration Meeting), 'Current Job Openings' (Jobs in the DIANA/HEP project), and 'Links' (RSS feed, HEP Knowledge Base, HSF on WikiToLearn, HSF on GitHub).

HSF   Activities ▾   Events   Projects   Working Groups ▾   Get involved!   About ▾

## The HSF Software Foundation

Advancing high energy physics community software

### HSF Newsletter - Logo Contest and Packaging Working Group

18 Dec 2015 by [Benedikt Hegner](#)

As promised in our last newsletter, we would like to update you on some more activities of the HSF.

#### Common Track Reconstruction Software Forum

Earlier this month, tracking software experts from the various experiments met to give a nice overview of the various software solutions that exist and to discuss possible common work. If you are interested in all the details, the agenda and minutes can be found [here](#). Further follow up meetings are planned.

#### HSF Logo Contest

It is time for the HSF to have a logo! We are looking forward to receiving your ideas. Be creative! Use the Christmas break and prepare a contribution. If you are willing to contribute, look at the [details](#).

#### The HSF Packaging Working Group

One of the problems that has been solved too many times in our community is how to *build, package and distribute* the huge software stacks of our collaborations. In fact we aren't alone in re-doing this work over and over again. Many competing projects are around in the open-source community. But is this actually necessary?

To understand this in our field, a few volunteers gathered together and created the [HSF Packaging Working Group](#).

#### About the HSF

For more information on the HEP Software Foundation see the [goals page](#) or contact the [startup team](#).

#### Upcoming Events

- [1st Common Track Reconstruction Software Forum Data Science @ LHC 2015 Workshop](#)
- [HEP Software Collaboration Meeting](#)

#### Current Job Openings

- [Jobs in the DIANA/HEP project](#)

#### Links

- [RSS feed](#)
- [HEP Knowledge Base](#)
- [HSF on WikiToLearn](#)
- [HSF on GitHub](#)

# Some Challenges

- **Status of Technology Tracking**

- Future HW architectures
- New technologies/trends worth looking at  
(cloud based analysis, e.g. Data Mining-as-a-Service? )
- Instrumentation and tools for measuring and improving SW performance
- What else?

- **Evolution vs. Revolution**

- Parallelism / vectorization implies revolutions in our SW
- Challenge to backwards compatibility
- Results of Revolution can still be included as evolution  
(GeantV's VecGeom as “preview” in Geant4 10.2)

- **Managing available manpower efficiently**

- build up more commonality in software projects and procedures
- prepare a common curriculum of development essentials
- provide an easy entry point for people to apply best practices  
(HSF project template + infrastructure to set up + tools)
- improve on quality and ease-of-use of the software we develop  
(less incentive on reinventing the wheel!)