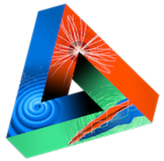


JENAA
Joint ECFA-NuPECC-APPEC Activities

eosc | EVERSE

SPECTRUM CoP

HSF
HEP Software Foundation

HSF, JENA, EVERSE and other European Inputs towards sustainable HEP Software

HEP-CCE AHM

Graeme Stewart, CERN EP-SFT
(Particular thanks to Torre Wenaus, Caterina Doglioni, Stefan Roiser, Sanje Fenkart for material)

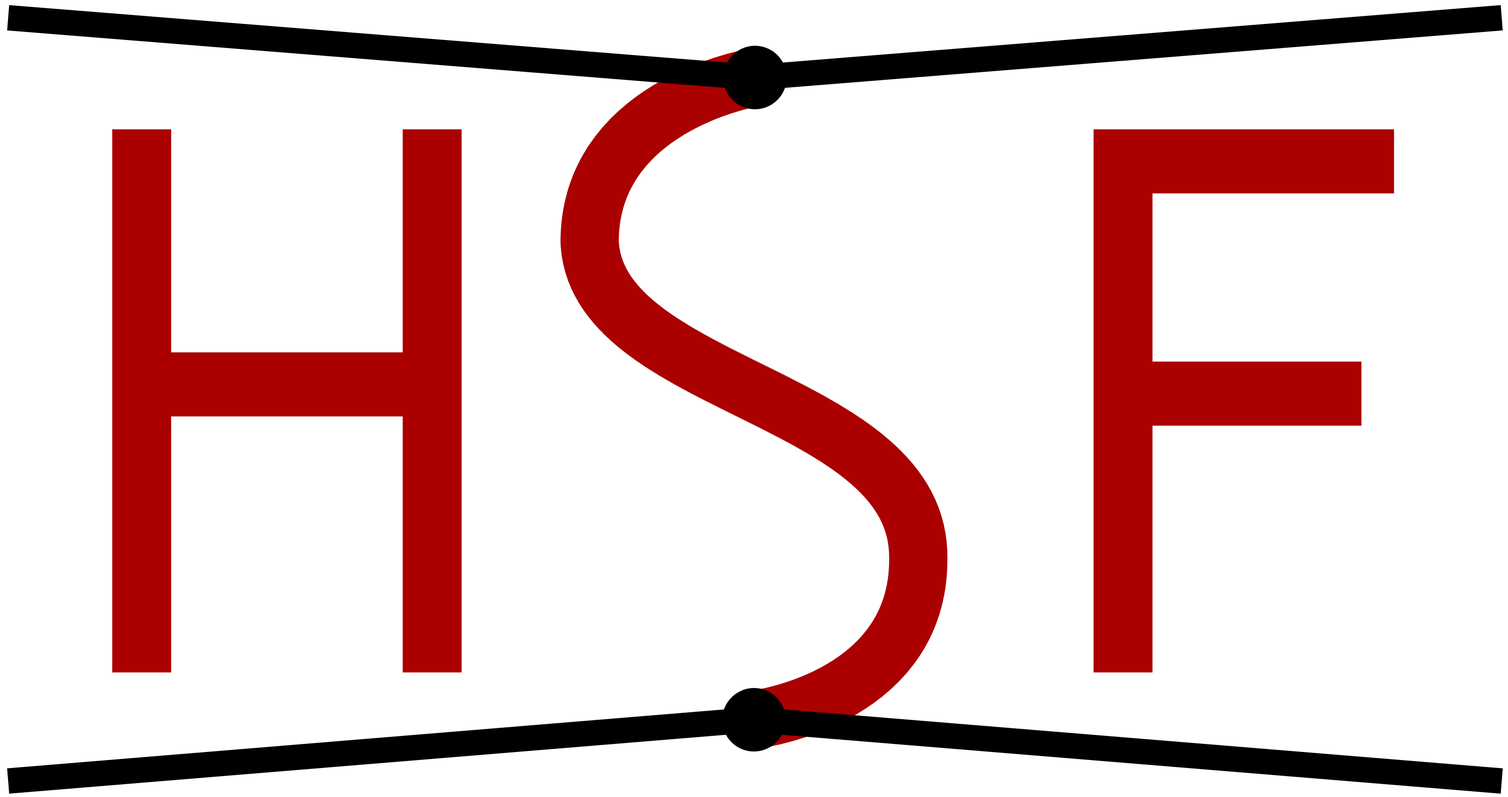
2024-07-23

CERN

Overview

- HEP Software Foundation
- JENA and SPECTRUM
- Software Sustainability
- EVERSE
- Conclusions for Collaboration

HSF



HEP Software Foundation

HEP Software Foundation

Happy Birthday HSF, we are 10 years old!

- Motivations today are very much the same as we had a decade ago
  - *Changing hardware landscape making efficient software more difficult to implement*
    - Today multi-core processing is very well established in all experiments
    - Accelerator programming has advanced significantly
      - But still far from being a solved problem: APIs and heterogeneous scheduling still have no general solutions
  - *An exciting physics programme ahead*, e.g., High-Luminosity LHC, but this will put tremendous pressure on software and computing
    - Event complexity and rate will jump, physics precision is growing across the board - but the budget is ~fixed
  - We need to *train a new generation of software experts*
    - And still give them recognition and a career path
    - RSEs are a new model from the same era
  - Try to *avoid duplicated solutions* and multiple rediscoveries of the same things, given the severe lack of experts

DESY Workshop

- We celebrated our birthday at DESY, Joint WLCG-HSF Workshop in May
- Good time to look back at our accomplishments
  - Community White Paper (CWP) process that mapped out the roadmap towards HL-LHC
  - Provided a voice for software and its importance in high-level strategic discussions
    - LHCC, European Strategy Update, Snowmass, P5, ECFA, JENA, …
  - Advice on practical matters such as licensing and tooling
  - Input from experts on GeantV, GPUs, DUNE frameworks
  - Plus community forums, from working groups in topical areas to, e.g., Compute and Accelerator Forum
  - Reference implementation for conditions databases
  - Google Summer of Code
  - HSF Training Initiatives with close to 3000 students trained - new Training Centre online

The activity of the HSF has helped create a positive atmosphere where more software projects are successfully funded: IRIS-HEP, SWIFT-HEP, HEP-CCE, ErUM-Data-Hub

This is very different from 10 years ago!

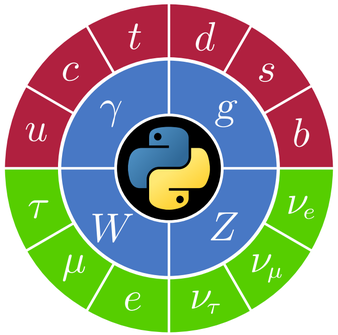
But how can we do better…?

- We asked the question about how we could be even more effective in the next years
  - How can we sustain, grow and prioritise?
  - Strengthening our connections to the community, to projects and the experiments
- Reorient the HSF organising team to a **Steering Group**
  - Take more executive oversight of the HSF
    - Though not discouraging our do-ocratic nature!
  - Very open to new members who would like to contribute to steering the HSF
- In the process of setting up an **Advisory Group**, primarily from the experiments (LHC, DUNE, Belle II, EIC/ePIC, MCnet, …)
  - More regular way to gather external feedback on the HSF's activities and how these are helping the community

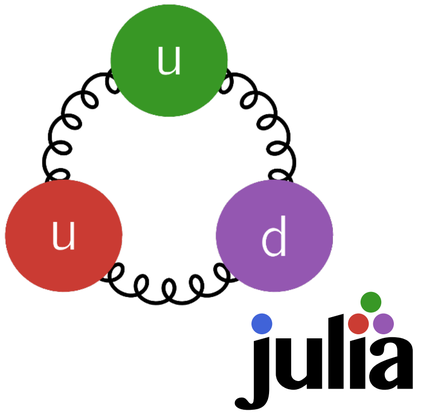
Working Groups, Activities, Seminars

- After the CWP we setup a number of Working Groups
  - e.g., Simulation, Reconstruction, Event Generators, Data Analysis
- These have been quite successful, but more as forums for discussions than places where work is actually organised
  - We decided that in many cases these groups would be better described as activity forums
  - Contribute to a new ***HSF Seminar Series***
    - Regular time slot (Wednesdays 16h30 CE(S)T/CERN time)
    - Alternating with the Compute and Accelerator Forum
      - e.g., planning a meeting on 4D tracking at the LHC experiments
  - Notwithstanding that there will still be focused non-seminar activities/meetings, e.g., the generator group would like to organise some discussions on the evolution of HepMC3

Tools, Languages, Training, …



- Other HSF areas really are helping to organise work, or doing it themselves
  - Language evolution has been a successful activity
    - PyHEP was a success to organise both coherent discussions between developers, as well as didactic and outreach events for physicists
      - PyHEP online workshops most recent edition 1-4 July
      - PyHEP.dev (2nd edition) will be in Aachen, 26-30 August
    - JuliaHEP helping marshall activity related to exploring Julia for high-energy physics
      - Next JuliaHEP workshop at CERN, 30 September - 4 October
    - Training Group very active in organising regular training events for HEP
      - Core software skills, Python analysis and CI/CD are very popular
      - Contributions to the training programme always welcome - material or helping to teach
      - Pre-CHEP workshop on Training in HEP, 19-20 October



HSF helped these activities achieve a certain critical mass beyond their funding (if they had any)

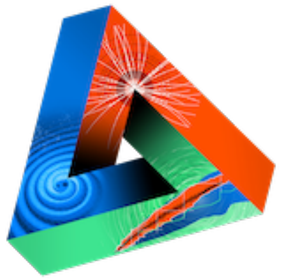
Community Software

- At the workshop we had a great series of topics promoted and developed by the community
  - Particularly strong series of talks from the generators teams
    - Nice to have the neutrino experiments discussing their needs and how they align/contrast with respect to HEP
  - Also talks on tracking (ACTS), simulation (Gaussino), L1 triggers (Allen)
  - And community software projects like SciKit-HEP, ROOT, HSF India, Key4hep, Phoenix
- Shows a great strength of the HSF as a forum for the community to share knowledge and discuss common approaches and solutions
- Our “do-ocracy” approach has been a success
  - Thanks to the efforts put in by a large number of people

Software Projects

- The HSF itself didn't become the home of software projects
  - A few projects were very closely aligned (prmon, phoenix, harvester)
  - We did not control funding (deliberately!) and development services evolved greatly (GitHub and GitHub actions)
- We propose to revisit the idea of HSF supported projects
  - So that the HSF can **endorse projects** as recognition of their role in the community
    - Considering a differentiated way to do this, depending on the criteria fulfilled (open development model, usage in the community, etc.)
  - Projects always retain full control of their development and management
- We are also happy to arrange for expert reviews or to help work towards reference implementations and white papers for common software solutions

JENA and SPECTRUM

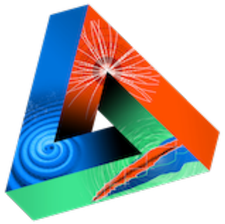


JENAA

Joint ECFA-NuPECC-APPEC Activities

SPECTRUM CoP

JENA Computing Initiative



JENAA
Joint ECFA-NuPECC-APPEC Activities

- JENA: Joint ECFA - NuPECC - APPEC
  - ECFA - European Committee for Future Accelerators
  - NuPPEC - Nuclear Physics European Consortium
  - APPEC - AstroParticle Physics European Consortium
- Three major scientific communities working together on science and infrastructure themes of common interest
- JENA Symposia in 2022 in Madrid
  - Plenary presentations and the closed sessions discussed computing
  - Funding agency representatives identified an increased need for discussions on the strategy and implementation of European federated computing at future large-scale research facilities

JENA Computing Workshop

- We started this process organising a computing workshop in Bologna, June 2023
  - Reviewed the major challenges for each field
    - Particle Physics
      - High Luminosity LHC - enormous stress on software and computing from higher trigger rates and higher pileup
    - Nuclear Physics
      - GSI-FAIR experiments will increase data rate and computing needs
      - Need to support many experiments, including small ones
      - Try to make the software and the computing as common a problem as possible - with common solutions (example: FAIR-ALICE collaboration)
    - Astro(patricle) Physics
      - Observatory model - running and upcoming projects with data piped into existing computer centres
      - Need to generalise access and interfaces to give scientists access to data products and alerts
      - Vera Rubin, SKA - very high data rates

Known Challenges

- Hardware evolution continues with manufacturers adapting to smaller scales
  - Chiplet architectures and 3D lithography
- Heterogeneity is inevitable
  - Many software challenges arise from this to make algorithmic changes and implement efficient solutions
- Significant European investment in HPCs (see Maria's talk)
  - We need to engage and find common ground to allow HPCs to be more effectively used by data intensive sciences
  - Has to integrate into our established distributed model
- Machine Learning (AI) increasingly important in workflows
  - Resource implications not yet really well understood
    - Hope is for higher fidelity and reduced costs for inference cf. other methods
    - But how much training will be required?

Computing Initiative Working Groups

- HPCs
  - Augmenting the computing capacity available and facilitating the federation with existing data facilities
- Software and Heterogeneous Architectures
  - Find common solutions to using heterogeneous architectures, lowering the costs of computing and addressing sustainability
- Federated Data Management, Virtual Research Environments, FAIR Principles and Open Data
  - Managed by the ESCAPE Consortium, strengthen data management synergies and how workflows interact with FAIR principles
- Machine Learning and Artificial Intelligence
  - Quantify the resource needs and to define the interfaces and services that are needed by physicists to run ML workloads
- Training, Dissemination, Education
  - Leverage the experience in the HSF training initiative and find common ground with other sciences; explore how to develop a curriculum with universities

Goal is to write a whitepaper, drafted for November 2024, to be finalised at the next JENA Symposium

Software and Heterogeneous Architectures

- We have surveyed the *prior art* material from all three communities
  - Now in the process of synthesising these inputs into summary documents
- Meanwhile we have been preparing a **survey** that looks at different aspects of software and computing in the community
- We have done this in conjunction with the SPECTRUM project
  - Strategic Research, Innovation and Deployment Agenda (SRIDA) and a Technical Blueprint for a European compute and data continuum
    - Concentrating on HEP and Radio Astronomy
- JENA/SPECTRUM Survey (Announcement)
  - Answers from all international partners also very welcome!
  - The overall survey is pretty huge, but answering the questions on software and training should take < 15 minutes

Sustainable Software

Image Credit: Kitware

FAIR Software

FAIR Principles for Research Software
https://doi.org/10.15497/RDA00068

- Public repository with version control
- A licence
- Code is catalogued
- Enable citations
- Use a software quality checklist

https://fair-software.eu/

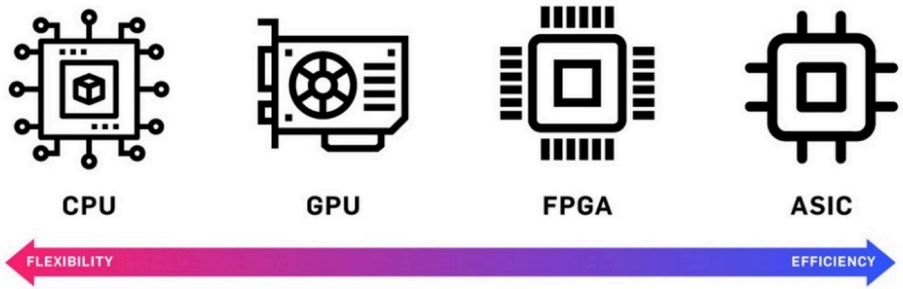
The FAIR4RS Principles are:

F: Software, and its associated metadata, is easy for both humans and machines to find.
F1. Software is assigned a globally unique and persistent identifier. • F1.1. Components of the software representing levels of granularity are assigned distinct identifiers. • F1.2. Different versions of the software are assigned distinct identifiers. F2. Software is described with rich metadata. F3. Metadata clearly and explicitly include the identifier of the software they describe. F4. Metadata are FAIR, searchable and indexable.
A: Software, and its metadata, is retrievable via standardized protocols.
A1. Software is retrievable by its identifier using a standardized communications protocol. • A1.1. The protocol is open, free, and universally implementable. • A1.2. The protocol allows for an authentication and authorization procedure, where necessary. A2. Metadata are accessible, even when the software is no longer available.
I: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.
I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards. I2. Software includes qualified references to other objects.
R: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).
R1. Software is described with a plurality of accurate and relevant attributes. • R1.1. Software is given a clear and accessible license. • R1.2. Software is associated with detailed provenance. R2. Software includes qualified references to other software. R3. Software meets domain-relevant community standards.

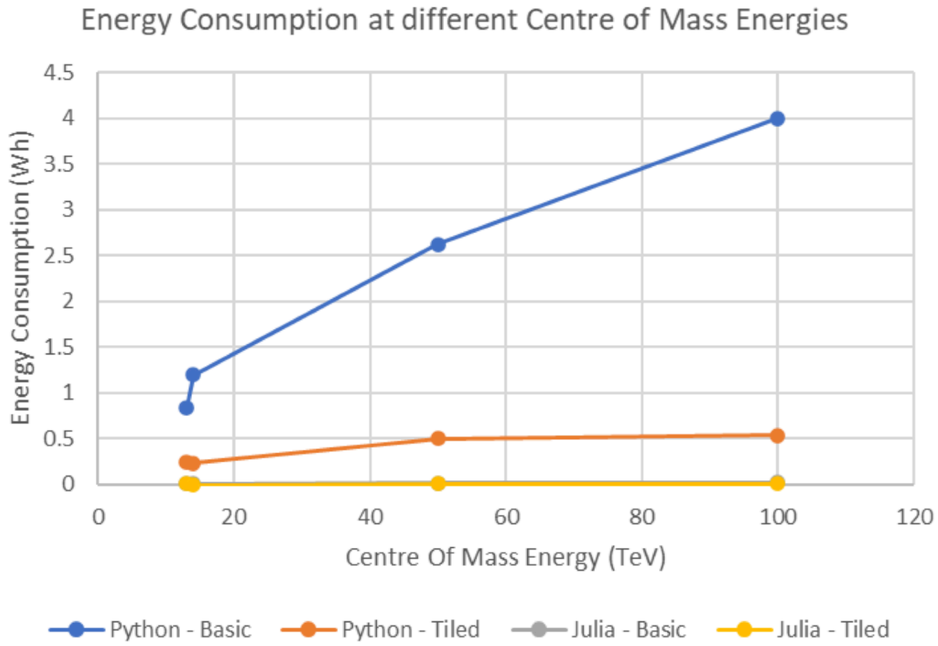
Table 1: The FAIR Principles for Research Software

Green Software

- From https://learn.greensoftware.foundation/
- Concrete example from SwiftHEP where x50 improvements were gained in ATLAS's use of the Sherpa event generator's multi-leg NLO calculations
  - Also supported in HEP-CCE!



Carbon efficiency	Build applications that are carbon efficient. Minimise the amount of carbon emitted per unit of work
Energy efficiency	Build applications that are energy efficient. Write software that maximises/matches your hardware's energy efficiency
Carbon awareness	Consume electricity with the lowest carbon intensity. Choose your energy sources wisely, if you can
Hardware efficiency	Build applications that are hardware efficient. Software should be well matched to hardware, and optimise hardware lifecycles
Measurement	Improve sustainability through measurement. More information is better!
Climate commitments	Defining the exact mechanism of carbon reduction. Strive to improve, in a quantitative and well-defined way



Akshat Gupta, Caterina Doglioni, Graeme Stewart - compare energy consumption in different languages / algorithms

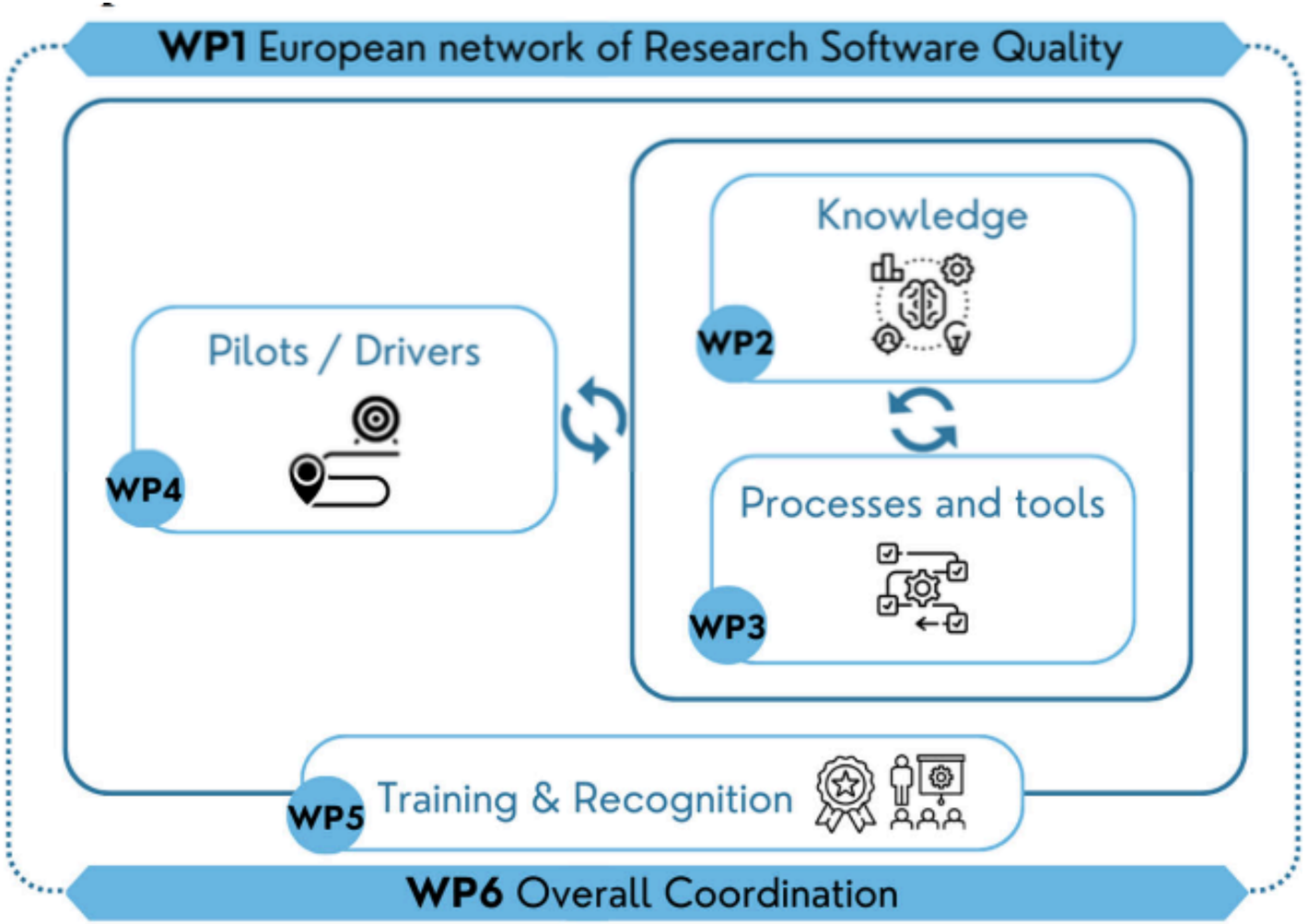
EVERSE

EVERSE

- EVERSE Project is 3 year EU funded project, €8M
  - Started 1 March 2024
- Build on the software quality aims of the five science clusters in the European Open Science Cloud
- Aims to
  - Create a framework for research software and code excellence
  - Build a network for software quality, based on existing community standards
  - Lay the foundations for a future virtual institute for research software quality

EVERSE Organisation

- Significant HEP connections
  - WP1 - Graeme Stewart
  - WP3 - Thomas Vuillaume
  - WP4 - Caterina Doglioni
  - WP5 - Stefan Roiser



EVERSE Activities

- The core activities:
  - *Assess current best practices* by researchers and make those accessible via the EVERSE “RSQkit”
    - The RSQkit is a meta-repository of documentation, practices, and examples (content still in flow)
    - Prototype for a different purpose/field: RDMKit (from Elixir)
- *Link tools and services* into common pipelines or frameworks for software quality and integrate them into the tools developed by the EOSC clusters
  - Guiding principles are *FAIRness* and *Open Science*
- Exercise these concepts via the science clusters’ use cases, e.g. these are some of the HEP/ESCAPE use cases:
  - Analysis code for high throughput data from the LHC based on “xAODAnaHelpers” [talk]
  - Prototyping of machine learning tools for data compression [talk]
  - Software infrastructure for the tracking of charged particles based on “ACTS” [talk]

EVERSE Network

- Establish a community of practice in the quality of research software
- European based, but an international context
  - Provide standard and documented practices around tools and training for software developers
  - Define the meaning and principles of software quality for the research domain
  - Enhanced recognition for software developers contributing to research activities
- Value
  - EOSC Association expert group on software
  - Help shape software quality practices
  - Define the key aspects around recognition for research software activities
- Currently we are discussing with early adopters to help us shape the network
  - Contact Graeme and Sanje if you are interested to join

Conclusions

Conclusions for Collaboration

- Software and computing in high-energy physics remains as challenging as ever
  - A critical part of our science exploitation, with HL-LHC coming soon
- The landscape of large scale data intensive sciences is also growing
- HEP Software Foundation has helped our community exchange information, advocate for software at high levels and take practical steps to improve our software
  - We want to continue to work with all partners to keep doing that
  - We are utterly reliant on *members of our community contributing to the HSF* to do that!
- Broader European initiatives, such as JENA Computing, help align the interests of HEP with other sciences
  - Speak with a common voice to funding agencies and facilities
  - The JENA/SPECTRUM survey is a chance for many people to contribute their view
- EVERSE project works more broadly towards software quality criteria, training and recognition

In all of these areas possibilities to work usefully with HEP-CCE exist!

Stay in Touch

- HSF: hsf-steering@googlegroups.com
  - If you are not signed up to the HSF forum mailing list, please subscribe
- JENA: jena-computing-wp2-software@cern.ch
  - Or mail to Graeme Stewart, Paul Laycock and Adrien Matta (chairs)
- SPECTRUM: community@spectrumproject.eu
- ESCAPE: everse-contact@lists.certh.gr
  - Or contact Graeme Stewart, Stefan Roiser or Caterina Doglioni (HEP WP leaders)