The need for Research Software Engineering

College of RSEs proposal + other things

Heek 10

PPAP Meeting Birmingham Sept 12th 2019 75t month

Heek og

Pete Clarke + Davide Costanzo



Something's Gotta Give



The case for RSEs is not new

- Research Software Engineers (RSE, DevOps, ResOps, SRE...)
 - Professionals who are highly experienced in software engineering / techniques
 - Long term focus on software engineering
 - Same status as a physicists or hardware engineer
- RSEs needed for
 - Software Infrastructure for large scale distributed computing
 - Common physics code e.g. Geant, Data management,
 - Experiment specific code: Reconstruction, Simulation
 - Novel techniques development e.g. image processing
- RSEs needed across PPAN
 - LHC of course as we approach Run-III and the GPD upgrades
 - Other HEP experiments (DUNE, HyperK,...)
 - SKA
 - Other large and growing Astronomy projects (LSST, EUCLID, GAIA,..)
 - CTA, LZ, LIGO
 - Nuclear physics at JLAB (and FAIR)

New is that we have reached limit of "pretense"

- As any activity scales up \rightarrow computing requirements become large and distributed
 - Investment in a proper computing/software team for each large Activity is mandatory
 - This is not avoidable just because people would like to wish otherwise.
 - Head in sand approach no longer works
- Software construction should be treated with same respect as hardware construction
 - Can't be done with cottage industry
 - Nor transient people doing it in spare time
- We need to start recognising the "slnfrastructure"
 - Software infrastructure is part of the elnfrastructure
 - It needs to be constructed

New is that we have reached limit of "pretense"

- As any activity scales up \rightarrow computing requirements become large and distributed
 - Investment in a proper computing/software team for each large Activity is mandatory
 - This is not avoidable just because people would like to wish otherwise.
 - Head in sand approach no longer works
- Software construction should be treated with same respect as hardware construction
 - Can't be done with cottage industry
 - Nor transient people doing it in spare time
- We need to start recognising the "slnfrastructure"
 - Software infrastructure is part of the elnfrastructure
 - It needs to be constructed

Example: take the case of GPUs

•These could reduce hardware costs substantially

- •But only if major CPU hungry codes can be re-written to use them
- •This means simulation (generators, Geant)
- •This means a re-write from scratch you cant just "adapt" x86 code

→It is futile for the left hand to call for use of GPUs to reduce costs whilst the right hand effectively never funds the RSEs needed to get there

→A proposal to initiate this was rejected at recent opportunities call

Software levels in HEP terms

Layer	Responsible	Experiment 1	Exp. 2	Exp. 3	
6	Experiment physicist end users	Selecting data, running analysis code.			Community Projects
5	Experiment physics programmers and software engineers	Analysis frameworks, reconstruction code, calibration code			(e.g. HSF and WLCG working Groups)
4	Experiment computing teams	'Production' computing operations and software			\rightarrow Moving code
3	WLCG/GridPP	↑ Middleware interface to experiments, and experiment 'customer' support			down the stack to the common layer ←
2	WLCG/GridPP	Software infrastructure running on physical hardware infrastructure and WLCG Federal responsibilities ↓		↔ Essential Development of WLCG for the future (
		Physical Ha	ardware		

Software levels in HEP terms

Physicists (many very talented at

software)



One of problems is confusion of these





Something's Gotta Give



A Chronic problem is becoming acute

We can go on pretending that this can be magically be done with no effort

or

We can be serious and make the need for RSEs a first class item (as EPSRC have done)

We == everyone involved in making HEP wheels turn: STFC, peer review bodies, PIs at institutes

What's going on to seek support

College of RSEs



Priority projects – outline case

Project Name

College of Research Software and Infrastructure Engineers

Name	Where	Relevant Position
Peter Clarke*	Edinburgh	Director IRIS, Dep PL GridPP
Jeremy Yates	UCL	Dep Dir. IRIS, Innovation Dir DiRAC, ex-Director DiRAC
Dave Britton	Glasgow	Project Leader GridPP
Dave Colling	Imperial	Chair CAP
Andrew Sansum	STFC-RAL	Director IRIS, ex-Head SCD
Mark Wilkinson	Leicester	Director DiRAC
Clare Jenner	UCL	Project Scientist DiRAC
Anna Scaife	Manchester	Lead for SKA RDC
Andrew McNab	Manchester	Dep. Computing Coordinator LHCb
Debora <u>Sijacki</u>	Cambridge	Chair DiRAC Project Board
Mark Heron	Diamond	Head of Computing for Diamond
Robin Pinning	Hartree	Technical Director, Hartree
Stuart Martin-haugh	STFC-RAL	PPD
Catherine Jones	STFC-RAL	Software Engineering Group Leader, ALC Dep. Division Head
Roger Jones	Lancaster	Head of Computing and Head of School Lancaster
Tom Griffin	STFC-RAL	Head of SCD
Gordon Brown	STFC-RAL	Ada Lovelace Division Head and Project Manager for ALC
Jonathan Hays	QMUL	Academic Lead for Computing and GridPP in School of Physics and Astronomy a QMUL
Thomas Kitching	UCL/MSSL	Euclid Science Lead
lain Bethune	STFC-Hartree	Technical Programme Manager & Research Software Engineering Group Lead
Colin Morey	STFC-Hartree	Systems Architect
Bob Mann	LSST	LSST UK Project Lead
Davide Costanzo	Sheffield	ATLAS Computing Coordinator Designate
	+	

College of RSEs: Objectives

- To create a pool of RSEs able to be seconded to the national facilities and science programme activities for extended periods to carry out well defined software/infrastructure engineering tasks.
- To provide a focus to both attract some RSEs from industry and place others in industry for short periods for knowledge exchange.
- To improve the career structure of RSEs by developing a meaningful accord with any institute which STFC subsequently funds.
- To provide a flow of RSEs to UK industry job market.
- To enhance the software engineering experience and training of RSEs and PhD students through College activities in conjunction with the Data intensive Science CDTs¹.

[1: To be discussed and agreed with the CDTs]

College of RSEs

- Proposal includes:
 - A minimum cohort of ~ 25 RSEs to be seconded to STFC projects.
 - Short or long term secondments
 - To execute a project not ongoing support
 - Build up to ~12 "RSE Fellows" (existing)
 - Flexible ways to take these (e.g. people with existing posts)
 - Software engineering training for PhD students [with CDTs] ~ 50 PhDs per annum.
 - Associate scheme
 - To allow people in industry and elsewhere to be associated to College
 - CPD activities
 - Career concordat
 - Community building Annual meeting (s)
- Cost ~ 4M per annum

College of RSEs

- College of RSEs proposal was submitted as an exemplar (place holder) on behalf of all STFC communities
 - To set a clear marker for the requirement and scale of costs
- It clearly states that should funds become available, then an immediate exercise would be done to design it in consultation with stakeholders
 - Community (experiments)
 - CDTs
 - ATI
 - Higgs Centre
 - EPSRC RSE initiative
 - UK RSE Association
 -
- I.e. you would get to own the detailed shape of it



Proposal to Opportunities call

"Efficient Computing for Particle Physics"

- Cross-LHC request to fund a fraction of postdoc FTE (named and appropriately skilled) from ATLAS, CMS and LHCb
 - These three LHC experiments have a small number, in the UK, of *online* software experts skilled in code profiling and adaptation to new computing architectures specific to the online (trigger)
 - Aim of this project: Use those skills to profile the LHC experiments' offline code leading to a report detailing key bottlenecks in LHC computing performance
 - Proof-of-principle demonstrations of new computing hardware platforms (look broadly: GPU, FPGA etc)
 - A review of technical hardware approaches to improving efficiency
 - Formation of a new community of experts in the UK, with international engagement, aiming at providing a cost-efficient computing solution for the upgrade era
 - A detailed plan and impact assessment for a subsequent 3-year R&D project
- Support letters from nVidia, GridPP, IRIS-HEP (US project)

IRIS (STFC Consortium)

Some of you will have heard of IRIS (STFC wide consortium)

- > Coordinating computing across STFC
- > Funding common capital infrastructure
- > (no people)

Home		
STFC supports a diverse set of <i>Science Activities</i> . These activities require a substantial "eInfrastructure" to manage, preserve, analyse and simulate their data. Such eInfrastructure includes both the physical infrastructure (HPC and HTC computing resource, disk and tape storage), and the software infrastructure needed to enable the data to be processed.		
IRIS is the co-ordinating body for the provision of this eInfrastructure and is a collaboration between STFC, its eInfrastructure <i>Providers</i> and representatives from the <i>Science Activities</i> themselves.	SEARCH	
More details of what IRIS is (and indeed what it isn't) can be found under what is IRIS? and IRIS-FAQ.	Search Q	
	ABOUT IRIS	
Convright 2017-2019 the IRIS partners and contributors.	IRIS aims to create a common eInfrastructure for STFC science	

From a HEP point of view IRIS is investing some small efforts into RSE work for PPAN

IDIS

- AAI infrastructure for single sign on (your institute credential, not a certificate)
- DIRAC job submission for multiple VOs (Virtual Organisaitons)
- Rucio data management for multiple VOs
- Common accounting across STFC
- This is a minimal contribution because IRIS has no resource for staff
- They are capitalised software creation ("digital assets") not posts
- This wil continue for a further 2 years at approx ~ 1FTE-year per annum

IRIS-HEP (USA software project)

(C) Iris

IRIS-HEP is a software institute funded by NSF

- Develop software cyberinfrastructure required for HL-LHC
- > 25 M\$ for 5 years, funding RSEs at Universities
- "Narrow" focus on ATLAS and CMS to tackle the SW challenge of the next decade
- Started in October 2018. Staff being recruited

Computational and data science research to enable discoveries in fundamental physics

Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP)

IRIS-HEP is a software institute funded by the National Science Foundation. It aims to develop the state-of-the-art software cyberinfrastructure required for the challenges of data intensive scientific research at the High Luminosity Large Hadron Collider (HL-LHC) at CERN, and other planned HEP experiments of the 2020's. These facilities are discovery machines which aim to understand the fundamental building blocks of nature and their interactions. Full Overview



Upcoming Events:

Sep 10–11, 2019 Fermilab					
Blueprint: Accelerated Machine Learning					
and Inference					
Sep 12–13, 2019 Fermilab					
IRIS-HEP - Institute Retreat					
Oct 16–18, 2019 The Cosener's House, Abingdon, UK					
PyHEP 2019 Workshop					
Oct 23–25, 2019 Catholic University of America, Washington DC					
Blueprint: A Coordinated Ecosystem for					
HL-LHC Computing R&D					
Dec 13–14, 2019 Vancouver Convention Centre					
Machine Learning and the Physical					
Sciences at NeurIPS 2019					
Jan 15–17, 2020 New York University					
ML4Jets2020 (in planning)					

About - Connect - Activities - Jobs

- Focus areas identified by the Scientific Software Innovation Institute
 - Data analysis systems
 - Reconstruction algorithms and software trigger
 - Application of machine learning
 - Data organisation, management and access (DOMA)
- FIRSTHEP funded to concentrate on training
 - 500k\$ over 3 years for tutorials, SW carpentries, and training events across the US
 - Train the next generation of physicists on SW tools

IRIS-HEP (USA software project)

(C) iris

IRIS-HEP is a software institute funded by NSF

- **Develop software** \geq
- 25 M\$ for 5 years, \succ
- "Narrow" focus o \triangleright challenge of the n
- Started in Octobe >

- **Focus areas**
 - Data ana •
 - Reconstr ٠
 - Applicati ٠
 - Data organisation, management and access (DOMA) ٠
- FIRSTHEP funded to concentrate on training •
 - 500k\$ over 3 years for tutorials, SW carpentries, and training events across the US
 - Train the next generation of physicists on SW tools ٠



Upcoming Events



About - Connect - Activities - Jobs

UKRI axis

- UKRI is working on UKRI infrastructure (M.T. in charge)
- As part of infrastructure is: elnfrastructure
- The elnfrastructure expert group recently worked on
 - Whitepapers
 - Roadmap for UKRI
- One of those white papers was on software and skills. Recommendations on:
 - > Improve funding policy and practices to support research software
 - > A coherent plan for funding research software throughout its lifetime
 - Identifying and filling identified gaps in provision of skills
 - Community building

Summary

Software development lies at the heart of UK research and innovation. In 2017, software development saw the largest growth in expenditure on R&D by UK businesses, increasing by 34.7% to £1.4 billion per year, meaning that it represented 6% of the total investment by UK business into R&D.¹ Software development is equally vital in academia. A recent survey² found that 92% of UK researchers use research software and 56% develop their own research software, whilst only 29% have received training in software development. Adoption of software is now occurring at a pace that far outstrips the UK research community's access to software skills. This is a significant risk to the reliability of results and, ultimately, the UK's reputation for research.

Summary

- RSEs have been needed for a long time
- There is a history of it being difficult/impossible to get these seriously via normal grant funding routes
- A chronic problem is now becoming acute across PPAN
- Its time to back all efforts to get a systematic injection into this sector
 - College of RSEs or equivalent
 - UKRI "sInfrastructure"" investment
 - Other opportunistic sources will help but not be systematic
- Meanwhile best efforts continues