

Ian Bird

WLCG Workshop, Copenhagen

12th November 2013

e-Infrastructure for 21st Century

WLCG today

- Successfully supported LHC run 1
- Many lessons have been learned – already several significant changes to the computing models
- Experiments pushing to higher and higher data rates
- Funding for future computing is a problem
 - Flat budgets are the (optimistic) working assumption

WLCG Strategies

- Reduce operational effort so that WLCG Tiers can be self supporting (no need for external funds for operations)
- Position ourselves so that the experiments can use pledged and opportunistic resources with ~zero configuration
 - (Grid) clusters, clouds, HPC, ...
- Collaborate with other science communities
 - Share expertise, experience

External funding?

- WLCG benefitted greatly from funding from EC, US (DoE/NSF), and other national initiatives
- This funding has largely stopped now
- Prospects for future funding exist – but the boundary conditions will be very different:
 - Must demonstrate how we benefit other sciences and society at large
 - Must engage with Industry (e.g. via PPP)
 - HEP-only proposals unlikely to succeed
 - Also it is essential that any future proposal is fully engaged in by CERN (IT+PH) and experiments and other partners

HEP value?

- Building and operation of the world's largest globally federated, distributed, infrastructure
- Management of multi-petabyte data sets and facilities
- Record of collaborating with other scientific domains (EGEE, OSG), and industry (openlab, Helix Nebula, ...)
- And more...
- Other sciences now need to address some of the same problems as HEP: we must collaborate
 - This is one reason why we must avoid HEP-specific solutions as much as possible, we don't have a good record of building broadly useful tools

e-Infrastructure vision

The Changing Times

- E-Infrastructures created to aid collaboration between domain specific scientists
 - But creates “silo’s” of competencies and activities.
 - Much duplication as collaborations solve similar problems
- Leverage existing investments in computing infrastructure
 - But new sciences treat computing as an operational (Opex) cost, not investment (Capex).
 - Leads to different models, e.g Grids and Clouds
- Large investments made in Distributed Computing Infrastructures (DCI)
 - But has not created sustainable infrastructures due to resources being “free”
- Commercial e-infrastructure services have gained in popularity (Google docs, dropbox etc)
 - But have limitations when applied to large science (cost models, trustworthiness etc)
 - But there is no coherent engagement with industry
- Large research infrastructures have benefitted by being able to leverage the infrastructure investments
 - But what about the “long tail” of science

Evolution

- The goal is to transform existing Distributed Computing Infrastructures (DCIs) based on a range of technologies into service-oriented platforms for the global research community that can be sustained through innovative business models

What exists today?

- Existing European e-infrastructure *long-term* projects
 - GEANT, EGI, PRACE
- Many “pathfinder” initiatives have prototyped aspects of what will be needed in the future
 - Includes much of the work in the existing e-Infrastructure projects but also projects such as EUDAT, Helix Nebula, OpenAIRE+, etc
 - Thematic projects such as WLCG, BioMedBridges/ CRISP/ DASISH/ ENVRI, as well as Transplant, VERCE, Genesi-DEC and many others

E-Infrastructure Commons – key ideas

- Bring together public funded infrastructure and commercial partners into a hybrid model
 - Innovation for emerging science needs focused through *Research Accelerator Hubs* (ReAcH)
 - Commercial partnerships commoditise the services
- Encourage consolidation and commercial engagement
 - Create consolidated innovative services for the broad science domain through less centers with broader reach
 - Engage with industry to offer commodity services in a competitive and consistent way
- Ensure sustainability
 - Innovate business models based on a paid service model
- Provide legal frameworks
 - Define legal models that will allow for the rapid uptake of services

EIROForum papers published

- EIROforum is a partnership between eight of Europe's largest inter-governmental scientific research organisations that are responsible for infrastructures and laboratories:
 - CERN, EFDA-JET, EMBL, ESA, ESO, ESRF, European XFEL and ILL.
- 3 EIROforum e-infrastructure papers published in 2013
 - A Vision for a European e-Infrastructure for the 21st Century:
 - <https://cds.cern.ch/record/1550136/files/CERN-OPEN-2013-018.pdf>
 - Implementation of a European e-Infrastructure for the 21st Century:
 - <https://cds.cern.ch/record/1562865/files/CERN-OPEN-2013-019.pdf>
 - Science, Strategy and Sustainable Solutions, a Collaboration on the Directions of E-Infrastructure for Science:
 - <https://cds.cern.ch/record/1545615/files/CERN-OPEN-2013-017.pdf>

Vision

Sustainable - RIs currently in construction (FAIR, XFEL, ELIXIR, EPOS, ESS, SKA, ITER and upgrades to ILL and ESRF etc.), need to be convinced that e-Infrastructure will exist and continue to evolve throughout their construction and operation phases if they are to take the risk and invest in its creation & exploitation

Inclusive - Need an e-Infrastructure that supports the needs of the whole European research community, including the “*long tail of science*”, and interoperate with other regions

Flexible - Cannot be a one-size-fits-all solution

Integrated - Coherent set of services and tools must be available to meet the specific needs of each community

Innovative - Essential that European industry engages with the scientific community to build and provide such services

User driven - The user community should have a strong voice in the governance of the e-Infrastructure

Business of research

- Publicly funded research communities make significant investments in e-infrastructure that must be justified
 - To justify these investments the e-infrastructures must show a clear impact for the research communities
 - To gauge the impact, this market of end-users must be well understood by funding agencies and e-infrastructure services providers
- User communities must have a strong voice in the governance of the e-infrastructures to ensure they remain relevant and upto-date
 - Has been a strong recommendation in many reports.
- E-infrastructure providers have governance models limited to their infrastructure collaborators (their “customers”).
 - The “user” has been hard to identify and include

Governance by the Users

- Create a pan-European forum for organisations and projects that operate at an international level
 - Present to the policy makers and the infrastructure providers the common needs, opinions and identify where there is divergence
 - Independent of any supplier and engage across research domains
 - Supplements but does not replace existing e-infrastructure user engagement channel
 - Engages with the “long tail” of science
- Provides the essential “market” information to E-Infrastructure providers
 - Market research deliverable including analysis and trends.

Consolidation of Services

- Avoid fragmentation of users (big science vs. long tail)
- Avoid fragmentation of infrastructure (not integrated and duplicated services)
- Provide Common platforms (*e-infrastructure commons*) with 3 integrated areas
 - **International network, authorization & authentication, persistent digital identifiers**
 - **small number of facilities to provide cloud and data services of general and widespread usage**
 - **Software services and tools to provide value-added abilities to the research communities, in a managed repository (SciencePAD)**
- Provide for *data continuum* - linking the different stages of the data lifecycle, from raw data to publication, and compute services to process this data

Research Accelerator Hubs

- Build a hybrid model of public and commercial service suppliers into a network of *Research Accelerator Hubs*
- Make use of existing European e-infrastructures to jointly offer integrated services to the end-user
- *ReAcH* can be owned and operated by a mixture of commercial companies and public organisations offering a portfolio of services
 - Services made available under a set of terms & conditions compliant with European jurisdiction & legislation and service definitions implementing recognised policies for trust, security and privacy notably for data protection
- A management board where the *ReAcH* operators are represented to provide strategic and financial oversight - coupled with the user forum
- A pilot service (2014) initially offering a limited set of services at prototype *ReAcH*

Example from CERN

- This *prototype* will focus on data-centric services representing a platform on which more sophisticated services can be developed
- Use the resources installed by CERN at the Wigner Research Centre for Physics in Budapest, Hungary
- Services will be accessible via single sign-on through a fed id. mgmt system
 - Multi-tenant compute environment to provision/manage networks of VMs on-demand
 - ‘dropbox’ style service for secure file sharing over the internet
 - Point-to-point reliable, automated file transfer service for bulk data transfers
 - Long-term archiving service
 - Open access repository for publications and supporting data allowing users to create and control their own digital libraries (see www.zenodo.org)
 - Integrated Digital Conferencing tools allowing users to manage their conferences, workshops and meetings
 - Online training material for the services

Sustainability of CERN's ReAcH

- Partners will
 - curate their data-sets
 - connect their identity federations
 - deploy their community specific services & portals
 - manage the interaction with their registered users and associated support activities
- Beyond this first year, partners engage to fund the cost of the services their users consume according to a pay-per-usage model (to be jointly-developed with CERN during the first year)

Prototype ReAch – Example from EMBL-EBI

- This *prototype* will serve broad life science community based on successful Embassy cloud piloted since 2011
- Use the resources installed by EMBL-EBI in its tier-3 data centres in London
- Services
 - Well known resources and datasets: UniProtKB, Emsembl, PDBe, ENA
 - IaaS to other organisation (tenants – currently 8 from public & private sector)
 - Private sector “pay at cost”
 - In 2014 will expand scale of resources
 - Support large-scale analysis of genomic data via partnership with International Cancer Genome Consortium
 - Integrate with other centres and technologies resulting from Helix Nebula to serve ELIXIR

Beyond the initial prototypes

- Learn from the prototype ReAcH to establish similar structures around Europe
 - Not identical: each has its own portfolio of services and funding model
 - All interconnected: to offer a continuum of services
 - All integrated with public e-infrastructures:
 - GEANT network
 - commercial networks are not excluded!
 - PRACE capability HPC centres
 - EGI

Sustainability challenges for existing DCIs

- Short term funding from national funding agencies
 - National Grid Infrastructures (NGIs) typically are funded on a 1 or 2 year horizon
- Coordinating bodies do not control the resources
 - EGI.eu and NGIs do not have ownership of the grid sites they coordinate
- Operational costs are too high
 - 300+ sites in EGI offering similar services
 - NGIs in EGI are using >40% of their effort for operational tasks
- Interoperability and flexibility
 - Project structure means users can't easily combine and swap services

Changing the Approach

- Comparing DCIs to ESFRI RIs
 - ESFRI RIs are brokering long-term commitments from member states
 - RIs are selecting individual centres in member states to provide specific services
- Proposed approach is to
 - Rationalise the DCIs into a smaller number of ReAcH that have sufficient capacity and can offer a diverse portfolio of services
 - Introduce a pay-per-usage business model
 - Interoperate publicly funded DCI sites with commercial cloud services providers in a hybrid platform

Introducing a pay-per-usage business model

- Majority of DCI sites are supported by national funding agencies based on the set-up & operational costs
- Propose to introduce a pay-per-usage model so funding is linked to level of usage
 - Funding agencies can see the impact of a service hence have justification for their investment
- Give financial control to the users
 - Encourage existing Virtual Research Communities to adopt this model
 - They will choose services that offer better value-propositions
- Total cost of service provisioning will be reduced
- Services will continue to be free at the point of use

What happens to existing DCI sites that are not equipped to become ReAcH?

- Many sites joined DCI projects in order to contribute to scientific challenges
- Volunteer computing structures offer an avenue by which they can continue to contribute but with reduced operational costs
 - DEGISCO project
 - International Desktop Grid Federation
- Integrate volunteer computing into the overall e-infrastructure commons
 - EDGI project has developed bridges between volunteer computing and grids and clouds

Conclusion

- A new model has been proposed
 - Combining commercial and public funded e-infrastructures
 - Addressing large and small science
 - Managing Governance and Sustainability
 - Transitioning to an integrated service model
 - Evolving existing e-infrastructures
- Prototype ReAcH and the business models will be tested in 2014