EIROforum

Serving European Science

An e-Infrastructure for the 21st Century

eIRG workshop

Vilnius

4th November 2013

Bob Jones, CERN

The Goal

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

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

The Vision

Sustainable - RIs currently in construction (FAIR, XFEL, ELIXIR, EPOS, ESS, HiLHC, 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

Consolidation of Services

- Fragmentation of users (big science vs. long tail)
- Fragmentation of infrastructure (not integrated services)
- Common platform (*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**
- A *data continuum* - linking the different stages of the data lifecycle, from raw data to publication, and compute services to process this data

Governance by the Users

- Create a 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
- First meeting of the user forum is scheduled for 19-20 November
  - Initial members: EIROforum labs, ESFRI cluster projects, ERF, LIBER, LERU

Building Research Accelerator Hubs

- Build a hybrid model of public and commercial service suppliers into a network of *Research Accelerator Hubs (ReAcH)*
- Work with 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 EMBL-EBI

- This *ReAcH* will serve broad life science community based on successful Embassy cloud piloted since 2011
- Use 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 public & private)
  - 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

Example from CERN

- This *ReAcH* 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
  - Open access repository for publications and supporting data allowing users to create and control their own digital libraries (see www.zenodo.org)
  - Long-term archiving service
  - 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)

Beyond the prototype Research Accelerator Hubs

- Learn from the ReAcH prototypes to establish a network of similar structures around Europe
  - Not identical: each has its own portfolio of services and funding model
  - All interconnected: to offer a networked 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 Distributed Computing Infrastructures

- Fragility of Funding
  - National structures typically are funded on a 1 or 2 year horizon
- Lack of Control
  - Bodies such as EGI.eu and NGIs do not have ownership of the grid sites they coordinate
- High Operational Costs
  - Large number of small-scale sites offering identical services
- Interoperability and flexibility
  - Project structure means users can’t easily combine & swap services

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

Changing the DCI model

- Consolidate DCI sites into a reduced number of ReAcH with sufficient capacity to host a larger and more diverse portfolio of services
- Integrate volunteer computing infrastructures into the e-infrastructure commons
- Introduce a pay-per-usage business model
- Interoperate publicly funded DCI sites with commercial cloud services providers in a hybrid platform

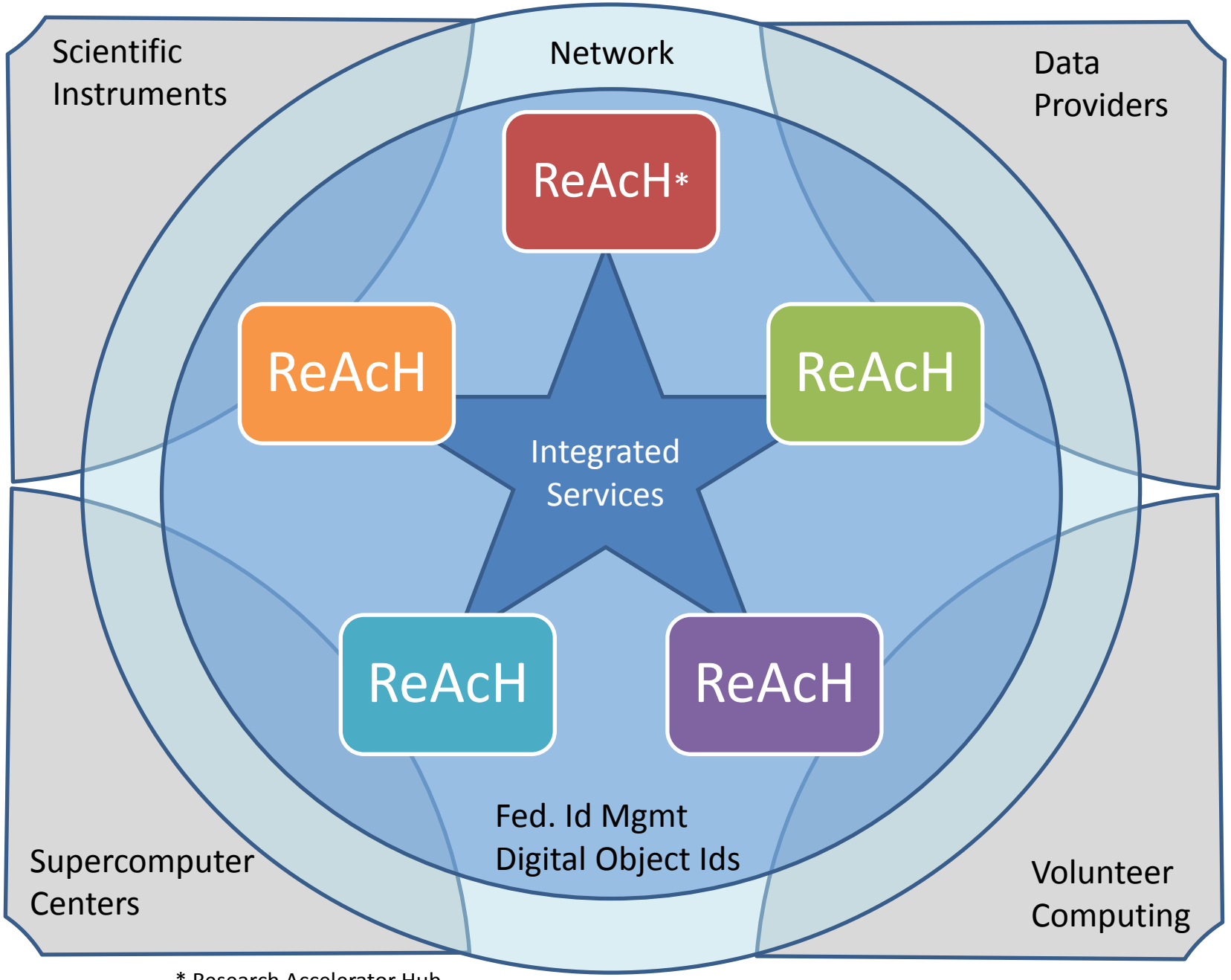
Introducing a pay-per-usage business model

- Majority of DCI sites are supported by national funding agencies based on the set-up & operational costs
- 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 DCI sites that do not become ReAcH?

- Many sites joined DCI projects in order to contribute to scientific challenges, get training and international exposure
- Volunteer computing structures offer an avenue by which they can continue to contribute but with reduced operational costs
  - DEGISCO project and International Desktop Grid Federation
- Integrate volunteer computing into the overall e-infrastructure commons
  - EDGI developed bridge between volunteer computing & grids & clouds
  - Offer a channel for engaging the general public and citizen scientists
- ReAcH will offer training/secondment opportunities

E-infrastructure commons



* Research Accelerator Hub

Addresses SIENA recommendations

- **Expand support for DCI efforts to provide mechanism to federate across multiple cloud suppliers**
  - *link independently operated ReAcH into a network offering a continuum of interoperable services*
- **Introduce business models for use of clouds by research**
  - *introduce pay-per-usage model and seek funding from multiple stakeholders*
- **Re-use tangible and intangible assets produced by DCIs**
  - *encourage existing DCI sites to become ReAcH*
  - *channel additional contributions to volunteer computing*
  - *migrate existing Virtual Research Communities*

I - Addresses "Cloud for science and public authorities" recommendations

1. Use EC Funding and initiatives to promote the integration and federation of clouds and enable the migration from e-infrastructures towards a European marketplace of connectivity and cloud services for e-Research – ***Create a network of ReAcH and encourage stakeholders to adopt a pay-per-usage model***
2. Promote and extend the use of clouds across multiple scientific domains and the development of a cloud services ecosystem, in order to narrow the gap between the supply and user communities and overcome cultural and resistance barriers – ***Open to all scientific domains and seeds the innovation of new services***
3. Support the consistent, comprehensive and business-case oriented analysis of cloud computing costs compared to other computing resources, requiring full cost assessment in all public funded projects – ***Makes explicit the funding models and price of services***
4. Promote the transformation of the business models and organizational structure of e-infrastructure providers – ***Introduces a pay-per-usage business model and consolidates the structure of e-infrastructures***

II - Addresses "Cloud for science and public authorities" recommendations

5. Create the next-generation of cloud enthusiast, supporting the change of mindsets and the development of the new skills sets needed for new cloud services and e-infrastructures – ***the best practices guides produced by the project will be valuable training material***
6. Promote innovative SMEs developing cloud-based services, also leveraging spin-offs and start-ups – ***market information will provide the basis of the business case for innovation of new services through start-ups and SMEs***
7. PRACE to start offering cloud services with a pay-as-you-go model – ***PRACE centres are encouraged to become Research Accelerator Hubs***

Summary

- A new model is 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