Implementation of a European e-Infrastructure for the 21st Century

Executive Summary

This document proposes an implementation plan for the vision of an e-infrastructure as described in "*A Vision for a European e-Infrastructure for the 21st Century*". The objective of the implementation plan is to put in place the *e-infrastructure commons* that will enable *digital science* by introducing *IT as a service* to the public research sector in Europe.

The rationale calls for a hybrid model that brings together public and commercial service suppliers to build a network of *Research Accelerator Hubs* offering a range of services to a wide user base. The exploitation platform will make use of and cooperate with existing European e-infrastructures by jointly offering integrated services to the end-user. This hybrid model represents a significant change from the status-quo and will bring benefits for the stakeholders: end-users, research organisations, service providers (public and commercial) and funding agencies. *Research Accelerator Hubs* can be owned and operated by a mixture of commercial companies and public organisations. Their portfolio of services, using as a starting point those listed by e-IRG and the High Level Expert Group on Scientific Data, will be made available under a set of terms and conditions that are compliant with European jurisdiction and legislation with service definitions implementing recognised policies for trust, security and privacy notably for data protection. A funding model engaging all stakeholder groups is described.

The ability to fully exploit the potential for knowledge and job creation that is locked-up in the datasets and algorithms to be hosted by the *Research Accelerator Hubs* will require the nurturing of a new generation of *data scientists* with a core set of ICT skills.

A management board where all the *Research Accelerator Hubs* operating organisations are represented will provide strategic and financial oversight and a user forum, through which the end-users themselves, in a cross‐disciplinary body collaborate to define requirements and policies for the services.

A pilot service is proposed that can be rapidly established by building on the existing investments. The pilot service will demonstrate the feasibility of the network of *Research Accelerator Hubs* model for a range of scientific disciplines and evaluate the suitability for the ESFRI Research Infrastructures, that are currently under-development and represent Europe's future "*big data factories*". Implementation will start in 2014, initially offering a limited set of services at several prototype *Research Accelerator Hubs*.

CERN-OPEN-2013-019
25/10/2013

Contents

This document has been prepared by CERN's IT department on behalf of the EIROforum IT working group.

Introduction

This document proposes an implementation plan for the vision of an e-infrastructure as described in a separate document[1] "*A Vision for a European e-Infrastructure for the 21st Century*". The objective of the implementation plan is to put in place the *e-infrastructure commons* that will enable *digital science* by introducing *IT as a service* to the public research sector in Europe.
EIROforum[3] partners are intergovernmental research organisations – CERN, ESA, EMBL, EFDA, ESO, European XFEL, ILL and ESRF – covering disciplines ranging from particle physics, space science and biology to fusion research, astronomy, and neutron and photon sciences. The partner organisations have a truly European governance, funding and remit, and in many cases share a global engagement. They are world leaders in basic research, as well as in managing and operating large research infrastructures and facilities. The EIROforum collaboration is helping European science reach its full potential through exploiting its unparalleled resources, facilities and expertise. By combining international facilities and human resources, EIROforum exceeds the research potential of the individual organisations, achieving world-class scientific and technological excellence in interdisciplinary fields. EIROforum works closely with industry to foster innovation and to stimulate the transfer of technology.

This implementation plan establishes a pilot service aimed at addressing the challenges ahead to capture, manage, and process the vast amounts of data to be generated, not only by the established fundamental research domains but in a growing range of scientific disciplines, large and small. This pilot service is seen by the research organisations as an exploratory activity to be operated and funded in addition to their baseline plans for the production services they offer to their user communities.

[1] http://cds.cern.ch/record/1550136/files/CERN-OPEN-2013-018.pdf
[3] http://www.eiroforum.org/

Rationale

E-infrastructures have made remarkable progress over the last few years. Moving from bespoke installations for single applications to distributed structures serving many user communities. In order to fulfil the vision for the future these e-infrastructures must be brought together to provide integrated solutions to a wider range of users during all stages of the research life-cycle. The justification, or rationale, to push for the implementation of this vision is that the impact of combining the e-infrastructures to form an *e-infrastructure commons* will generate far greater added value for all the stakeholders. This rationale has been inspired by the Helix Nebula initiative's publication "*Helix Nebula – The Science Cloud: A catalyst for change in Europe*"[4] from which extracts have been taken.

Helix Nebula was conceived as a way of bringing coherence to a highly fragmented IT services industry through the vision of a federated 'science cloud' integrated with publicly-funded scientific e-Infrastructures. Two fundamental steps remain if the vision of Helix Nebula is to become reality. One is to bring together a critical mass of supply-side interests and the other is to do the same for the demand-side.
The work of Helix Nebula on an architecture model[5] has produced a number of scenarios exploring the means by which publicly funded infrastructures can interoperate with commercial cloud services[6]. Such hybrid systems are in the interest of the users of publicly funded infrastructures and funding agencies because they will provide '*freedom and choice*' over the source of resources to be consumed and the manner in which they can be obtained, facilitating the goal of *'making every researcher digital'*. Hybrid systems serve the interests of suppliers by encouraging the adoption of cloud services by more researchers and thus creating a larger and more vibrant marketplace in which they can offer their services.
This integration will allow the public infrastructure users to strike their own balance between publicly funded resources and commercial cloud services while taking into account aspects of policy and cost. Private sector service providers can supply resources/services and additional capacity not available in the public funded infrastructures. Research and innovation activities that have the potential for commercial exploitation can work with the private sector to unlock that potential, demonstrate feasibility and market acceptance. It is expected that services will migrate from publicly funded infrastructures to commercial suppliers as suitable offers, addressing cost and policy aspects, become available on the market. In this manner, Helix Nebula is implementing what the e-Infrastructrue Reflection Group (eIRG) refers to as the *e-Infrastructure Commons* in its 2012 Roadmap paper[7].

[4] http://cds.cern.ch/record/1537032/files/HelixNebula-NOTE-2013-003.pdf
[5] http://cdsweb.cern.ch/record/1478364/files/HelixNebula-NOTE-2012-001.pdf
[6] http://cds.cern.ch/record/1548323/files/HelixNebula-D6_1.pdf
[7] http://www.e-irg.eu/images/stories/publ/e-irg_roadmap_2012-final.pdf

This hybrid cloud has the potential to bring the public and private sectors together so that the public sector has something to offer the private sector more than a simple demand for cloud services. The creation of a continuum of services across the public and private sector, instead of a simple supply-demand model will alleviate the current situation which is inhibiting the establishment of hybrid cloud infrastructures. It is also inhibiting the establishment of big-data services since there are a number of sensitivities, explained below, that need to be addressed before a model can be envisage where all datasets are hosted on commercial cloud services.

Business models need to be developed that ensure the Data Stewards[8], which are frequently publicly funded research centres, retain control over the curation and usage of the datasets and the associated intellectual property. Researchers also need guarantees about the long-term availability of the cloud services and the data. Such a hybrid model would address these points and allow the progressive migration of services between public and private providers. Furthermore, it is of upmost importance that data catalogue services, such as those being developed by the EUDAT project[9], are available to all research communities' datasets hosted by both public and private cloud services.

There exist several institutes in the biomedical sciences with well-established data infrastructures, such as the European Bioinformatics Institute and the Swiss Institute of Bioinformatics. The new, distributed infrastructure ELIXIR will take data management, curation and interoperability to the next, Europe-wide level. ELIXIR in turn is coordinating the FP7-funded BioMedBridges project, which aims to integrate data and thus construct computational 'data and service' bridges between the new biomedical sciences infrastructures on the ESFRI roadmap, including BBMRI, INSTRUCT and ECRIN. Together these span a wide range of disciplines from the biological, medical, translational and clinical domains. Their diverse data management needs can be best catered for by a combination of domain-specific services that are integrated into a wider ecosystem of domain-agnostic services. Other community specific research data infrastructures have been established in recent years, such as The Catalog of Life indexing the world's known species[10], iMARINE[11] and GENESI-DEC[12], which have produced valuable data curation tools and expertise, along with data sharing policies. Being able to interface these new data e-infrastructures into a hybrid cloud model as proposed here will allow a larger user base to exploit the data, bigger opportunities to contribute scientific data to multi-disciplinary research, and provide sustainability models for their continued existence.

[8] http://en.wikipedia.org/wiki/Data_governance
[9] http://www.eudat.eu/
[10] http://www.catalogueoflife.org/
[11] http://www.i-marine.eu/Pages/Home.aspx
[12] http://www.genesi-dec.eu/

Initial investigations on the potential impact of cloud services in the research community suggest that the commercial public cloud services are likely to be adopted initially for the *'long tail of science'* conducted by researchers that do not have access to significant in-house computing resources and skills. Efforts must be made to simplify access to commercial cloud services for such groups that generally have straightforward requirements and frequently do not have sufficient in-house IT expertise to manage and operate their own computing resources.

Conversely, it will require further reduction in the costs of commercial services for large research users with important in-house computing capacity to find commercial cloud services as financially attractive as is the case for small scale users. As one industry representative put it '*why would I hire a car on a daily basis if I know in advance I will use it every day for 3 years – it will be cheaper to buy my own*'[13].

It is recognised that not all publicly funded research centres are in a position to make accurate estimations of the total cost of ownership of in-house IT services since some contributing costs are borne by different departments. But in order for the demand-side users to be encouraged to purchase cloud computing services, the services offered must be economically advantageous compared to other means of procuring IT services.

These alternatives include purchasing and operating IT equipment internally which requires capital investment and IT expertise but remains economically attractive for IT-intensive applications with a sustained and predictable usage.

[13] Similar cost arguments are made in this article: http://www.theregister.co.uk/2013/04/26/cloud_magic_number/

Benefits of the hybrid model

The hybrid model outlined in the rationale above represents a change from the status-quo but is necessary to kick-start new and emerging sciences use of e-infrastructures. The identified benefits of this model for the stakeholders include:

Users

- A wide range of integrated services and data resources providing a more complete working environment supporting the full lifecycle of research activities
- Easy means to publish their own research results and data
- Freedom and choice over which services to use according to preferences for functionality, policy and budget
- An environment into which their own bespoke services can be integrated
- Readily available training material and support
- Time to focus on their research objectives rather than having to learn about and dedicate time to operate IT infrastructure
- Keep control over their results and data with the assurance that they will be available over the long-term
- Have a voice on how the services they use are governed

Research Organisations

- A means to expose their research products to a wider community
- An exploitation platform through which research organisations can be more closely connected to their users and react more quickly to their changing needs
- Ability to procure additional resources on-demand
- Grouped negotiation of service procurement with other research organisations
- A ready-made channel for improving knowledge and technology transfer with industry

Service providers (public and commercial)

- Creation of a market with an extensive user base in to which providers can offer their services
- Privileged communication channel with end-users and purchasing organisations
- A means of estimating the scale and sophistication of services that would be suitable for the public research sector
- Reduced risk and up-front investment during the development and introduction of innovative services

Funding agencies

- A means of encouraging the uptake of cloud services by the public sector
- Creation of an public-private innovation platform for ICT
- A simple means for implementing open access policies of publications and data
- Democratisation of access to data and compute resources so they are available to any users, laboratory or project, regardless of size and location
- A channel for increasing the impact and exploitation of publicly funded research
- A means to monitor the impact of public investments

Architecture of the e-infrastructure commons

The e-infrastructure vision document proposes a common platform for the future that builds on the experience of the last decade and is flexible enough to adapt to technological and service innovations. Such a platform must provide the underlying layers of common services, but must be adaptable to the very different and evolving needs of the research communities. The proposal has 3 distinct layers of services:

1. European and international networks; services for identity management and federation across all European research and education institutions and integrated with other regions of the world;
2. A small number of facilities to provide cloud and data services of general and widespread usage.
3. Software services and tools to provide value-added abilities to the research communities, in a managed repository:
    a. The tools to provide those research communities that have access to large sets of resources the ability to federate and integrate those resources and to operate them for their community, potentially sharing with other communities;
    b. Tools to help build applications: e.g. tools to manage data, storage, workflows, visualisation and analysis libraries, etc.
    c. Tools and services to allow researchers to integrate everyday activities with the e-Infrastructure;
    d. Tools to help research communities engage the general public as citizen scientists.

Figure 1 below shows how the layers of services are brought together into an overall architecture. A key element is the *Research Accelerator Hub (*ReAcH) which is a centre that offers a set of managed IT services tailored to the needs of their users. Together the *Research Accelerator Hubs* form a network of interoperating centres that can support workflows addressing the full research life-cycle. The underlying high-performance network brings together the *Research Accelerator Hubs* with the other elements of the *e-infrastructure commons*. The high-performance network extends to scientific instruments which act as big-data factories and additional data providers hosting important datasets around the world. Super-computing sites connected to the network have a dual role: like scientific instruments they can be seen as big-data factories but they also offer capability-style data analysis facilities for datasets held at the *Research Accelerator Hubs* or provided by external data providers. Volunteer computing structures are an integral part of the *e-infrastructure commons* providing significant computing capacity as well as channels for engaging citizen scientists and the general public. All the elements of the *e-infrastructure commons* share a federated identity management system enabling single-sign-on access for users to any resource for which they have authorisation. A common digital object identifier service promotes the citation and re-use of any publication or dataset in the *e-infrastructure commons*.

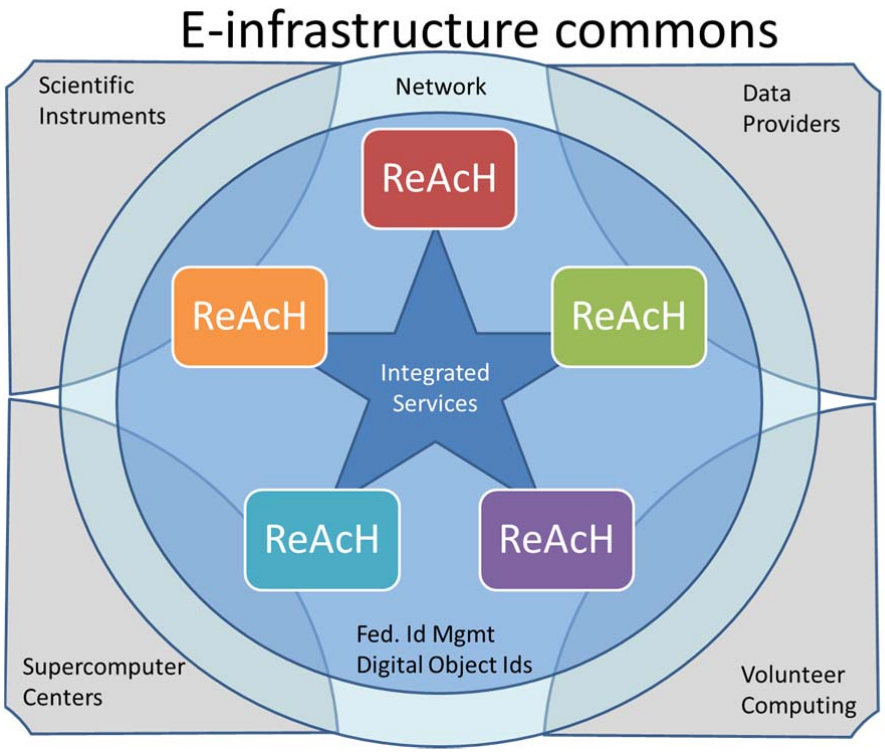

Figure 1 e-infrastructure commons architecture

Pilot Service

In order to verify the feasibility of this vision and rationale, a pilot service is proposed that can be rapidly established by building on the existing investments to demonstrate added-value to the research community and the sustainability of the approach.

Implementation will start in 2014, initially offering a limited set of services at prototype *Research Accelerator Hubs*. This ramp-up phase will include a roadmap for how to extend the e-infrastructure to a group of independently operated and cooperating *Research Accelerator Hubs* offering a richer set of layered services and higher total capacity to serve the European Research Area. Such facilities will be *Research Accelerator Hubs* providing capacity style High Performance Computing (HPC)[14] services for data centric applications and will be federated into a network of such centres across Europe. The *Research Accelerator Hubs* will not be duplicates but rather offer a set of services tailored to their users' specific needs and business models.

[14] For a definition of HPC capacity computing refer to the ETP4HPC Strategic Research Agenda document http://www.etp4hpc.eu/wp-content/uploads/2013/06/ETP4HPC_book_singlePage.pdf

Portfolio of Services

The *Research Accelerator Hubs* will offer an initial portfolio of services taken from the list documented by eIRG in its blue paper of 2010[15] with the technical characteristics identified by the High Level Expert Group on Scientific Data in their "Riding the Wave" report from the same year[16].

While each *Research Accelerator Hub* will offer their own set of services, all such services will incorporate state-of-the-art security measures, including host-based protections, system logging and access traceability. All services will make use of a federated identity management system addressing authentication and authorization to ensure seamless and secure access to the services and data. While the *Research Accelerator Hubs* will take the role and responsibilities of Data Custodians[17], the Stewards of the data-sets hosted by the *Research Accelerator Hubs* will be responsible for defining and managing the access policy for their data. To ensure a trustful relationship is established with the users, it will always be possible for Stewards to extract or remove any material that has been deposited with the *Research Accelerator Hubs.*

Brokering services will be deployed so that the full portfolio of services from all *Research Accelerator Hubs* is visible to all users and can be compared and accessed.

Training and Education

The design, creation and operation of e-infrastructure services are essential tools in the development of skills and competencies for the European market. The services offered by the network of *Research Accelerator Hubs* will provide a focus for the creation of public-private teams of skilled personnel across the cooperating research organisations, service suppliers and users, enabling the sharing of knowledge to develop a co-design approach[18] to services and applications. The ability to fully exploit the potential for knowledge and job creation that is locked-up in the datasets and algorithms to be hosted by the centre will require the nurturing of a new generation of *data scientists* with a core set of ICT skills. To ensure the training and education services impact as large an audience as possible while controlling the costs of offering such services, massive open online course (MOOC[19]) techniques will be employed which will be supported by the *Research Accelerator Hubs* model itself via the digital library services. The EIROforum organisations have core competences in training and education[20] which can form the basis of this activity. Since its formation in 2002, EIROforum joint initiatives have already resulted in an important impact in the areas of Outreach and Education. High-profile activities have involved thousands of European science teachers and students, thus emphatically supporting Europe-wide efforts to raise interest in science and technology and to secure a sound recruitment base for European R&D efforts in the future. Such collaboration in the area of human resources also have a bearing on the European Research Area: by attracting and retaining world-class researchers in Europe,

[15] http://www.e-irg.eu/images/stories/eirg_bluepaper2010_final.pdf
[16] http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf
[17] http://en.wikipedia.org/wiki/Data_governance
[18] http://echallenges.org/e2010/outbox/eChallenges_e2007_ref_195_doc_3562.pdf
[19] http://en.wikipedia.org/wiki/Massive_open_online_course
[20] http://www.scienceinschool.org/

through technological exchange and common development, and by organising multi-disciplinary scientific conferences.

Integration of public and commercial service suppliers into a hybrid model

The investigations within Helix Nebula have revealed that with current pricing of commercial services many public organisations will continue to find it financially attractive to provide the services themselves in-house. Hence this implementation plan foresees that *Research Accelerator Hubs* can be owned and operated by a mixture of commercial companies and public organisations. Each *Research Accelerator Hub* will include a pay-per-usage element in its funding model but the operating organisation is free to define its own pricing strategy and implementation. The portfolio of services will be made available under a set of terms and conditions that are compliant with European jurisdiction and legislation with service definitions implementing recognised policies for trust, security and privacy notably for data protection.
By integrating commercial and publicly operated *Research Accelerator Hubs* into a hybrid model it will be possible to migrate workloads to commercial suppliers once acceptable pricing and terms and conditions have been negotiated. In this way the *Research Accelerator Hub* model will bolster the creation of a market to which the commercial suppliers can offer their services.

Governance Model

The governance model foreseen for the pilot service will be based on two bodies:

- A management board where all the *Research Accelerator Hubs* operators are represented and the strategic and financial aspects of the *Research Accelerator Hub* model are addressed. Policy aspects impacting the operation of the network of *Research Accelerator Hubs* will be addressed by the management board. The introduction of a new *Research Accelerator Hub* into the network will be approved by the management board subject to verification of the policies (see Integration of public and commercial service suppliers) under which it operates and the quality of the services provided.
- A user forum[21], through which the end-users themselves, in a cross-disciplinary body collaborate on requirements and policies concerning the e-infrastructure services. The User Forum will also act as a consultation body for management board.

The pilot service will demonstrate the feasibility of the e-infrastructure *Research Accelerator Hubs* model for a range of scientific disciplines. The pilot service will also

[21] David Foster, Bob Jones, "Science Strategy and Sustainable Solutions; A Collaboration on the Directions of e-Infrastructure for Science", CERN-OPEN-2013-017, http://cds.cern.ch/record/1545615/files/CERN-OPEN-2013-017.pdf

provide a means for the ESFRI Research Infrastructures that are currently under-development and represent Europe's future "*big data factories*", to evaluate the suitability of the services provided by *Research Accelerator Hubs*. The integrated service portfolio offered by the network of *Research Accelerator Hubs* will enable multi-disciplinary research across the Research Infrastructures and contribute to ensuring excellence in science.

Relationship to existing European e-Infrastructures

The *Research Accelerator Hubs* will make use of and cooperate with existing European e-infrastructures to build the e-infrastructure commons by jointly offering integrated services to the end-user. The relationship with these existing e-infrastructures will be defined via written agreements and is described below.

GEANT

The *Research Accelerator Hubs* will be connected to the GEANT network in order to provide high performance access for the whole European Research Area and ensure integration with other regions of the world so Europe can be a leading participant in global research challenges. Building on the experience gathered through the Helix Nebula initiative which is extending GEANT connectivity to commercial data centres will greatly simplify the establishment of a hybrid model engaging both public and commercial service providers.

PRACE

The capacity style HPC services for data centric applications offered by the *Research Accelerator Hubs* will be complimentary to the capability style HPC services provided by PRACE. The expertise developed by PRACE and related projects in efficient parallel programming paradigms and optimising software for a range of architectures is directly relevant to the e-Infrastructure *Research Accelerator Hubs* and application/service developers. The European Technology Platform for High Performance Computing project[22] recently published a Strategic Research Agenda for achieving HPC leadership in Europe[23] which specifically highlights the upcoming big-data challenges for leading research activities and the relevance of cloud services:

> "*Europe is in a unique position to excel in the area of* ***HPC Usage and Big Data*** *owing to the experience level of current and potential users (and the recognition of the importance of data by such users as CERN, ESA, and biological data banks) and the presence of leading ISVs for large-scale business applications. Europe should exploit that knowledge to create competitive solutions for big-data business applications, by providing easier access to data and to leading-edge HPC platforms, by broaden the user base (e.g., through Cloud Computing and Software as a Service (SaaS), and by responding to new and challenging technologies.*"

[22] http://www.etp4hpc.eu/
[23] http://www.etp4hpc.eu/wp-content/uploads/2013/06/ETP4HPC_book_singlePage.pdf

So as well as sharing expertise, the services offered by the *Research Accelerator Hubs* and the PRACE Tier-0 and Tier-1 centres should be integrated to form part of the overall e-infrastructure ecosystem. This will require the PRACE HPC centres to participate in the federated identity management scheme and data sharing services if the PRACE centres are to be fully integrated as service providers within this model.

EGI

The experience gathered by EGI in managing a federated grid infrastructure will be directly relevant to the network of *Research Accelerator Hubs* model. EGI has also been evaluating cloud technologies via the EGI federated cloud[24]. It is proposed that a consolidated set of EGI sites become *Research Accelerator Hubs*. This will give the EGI distributed computing infrastructure a clear direction for how to contribute its experience and make a larger portfolio of services accessible to its existing user-base while introducing the innovation potential created by the uptake of cloud computing in research and business sectors.

Volunteer Computing

Volunteer computing initiatives across Europe have established production structures which serve a range of research communities. Such structures allow research and education organisations as a well as individuals and citizen scientists to contribute and participate in research activities. Significant computing resources are assembled by structures such as the International Desktop Grid Federation[26] that can support a growing range of application types with very modest operational and coordination overheads. It is important that such structures become an integral part of the e-infrastructure commons.

EUDAT data services

It is expected that data services currently under development by various projects, notably EUDAT, will provide candidates for future services and potentially additional e-infrastructure *Research Accelerator Hubs*. A goal will be to introduce services that can profit from the co-location of data and compute services to support multi-disciplinary research. Metadata and indexing facilities across the set of services in all the e-infrastructure *Research Accelerator Hubs* are seen as being particularly relevant. It is essential that new services are fully integrated with existing services to preserve the data and compute continuum of the exploitation platform and support the *e-infrastructure commons*.

[24] https://wiki.egi.eu/wiki/Fedcloud-tf:UserCommunities
[26] http://desktopgridfederation.org

Engagement of funding agencies

The implementation plan and associated funding model have been designed so that the e-infrastructure *Research Accelerator Hubs* can be sustained by their operating organisations according to a continuum of financial models ranging from sponsored resources for peer-reviewed scientific cases to communities who would pay for the services they receive. Additional resources will be required in order for these services to be expanded and to serve a wider range of users. The European Commission and national funding agencies will be invited to become stakeholders and contribute to the expansion of the e-infrastructure *Research Accelerator Hub* model. The guiding principle is that funding from such stakeholders will be focused on innovation of services and uptake by new user communities and business actors while the operational costs will be borne by the operating organisations and the users themselves. Below is a non–exhaustive list of areas where funding agencies may contribute to the expansion of the *Research Accelerator Hub* model:

- Development of new services to be deployed on the e-infrastructure. Significant effort will be required to co-develop scalable services that can operate in a distributed environment and serve a wide range of users.
- Financial incentive scheme to increase adoption of services (both public and private) by users including 'long-tail of science' research groups and SMEs.
- Engaging the use of the services by new research communities (e.g. curation of data-sets, connection of identity federations, deployment of community specific services, training for new users, etc.)
- Develop training and educational activities building on the e-infrastructure services to maximise their impact. This can also include expansion of services to support for volunteer computing so that researchers can build citizen-cyberscience communities and further engage the general public in science.
- Organisation of user forum events as well as outreach and dissemination to a range of audiences and production of material for policy related activities.
- International collaboration (beyond Europe) through interoperation with equivalent structures in other regions of the world.
- Expansion of the network of the *Research Accelerator Hubs* across the European member states to address national and thematic needs.
- Many research organisations that operate research infrastructures do not have the mandate to provide e-infrastructure services to their users for the management and processing of their experimental data. This represents a gap in the scientific lifecycle and a missed opportunity to highlight the results and impact of public funded research. These research organisations will require assistance to bridge this gap by supporting their users so they can make use of e-infrastructures services to manage and process their experimental data.

The recently completed eInfraNet project[27] brought together national funding agencies[28] from across Europe to explore the economics of e-Infrastructures and align European

[27] http://e-infranet.eu/
[28] http://e-infranet.eu/partners/

policies with the needs of the member states. In its final report[29] the project recommended:

> "*The use of standards-based private cloud infrastructures may also provide the needed flexibility in case of jobs that exceed available resources through cloud bursting (the use of external public clouds for the exceeding capacity) or by pooling individual institutional private clouds into a federated cloud that can provide the necessary resources in a federated way, following the extremely successful approach already in use for GRIDs.*"

The model for a network of *Research Accelerator Hubs* described in this document can serve as a blueprint to be expanded across Europe. The establishment of *Research Accelerator Hubs* could be supported by national funding agencies to address local or thematic needs and provide a focus for innovation in e-Infrastructures. Such national or thematic *Research Accelerator Hubs* should be integrated with existing European e-Infrastructures (see
Relationship to existing European e-Infrastructures) and will be in a position to host data, information and services that for policy reasons need to remain within a national context. As these *Research Accelerator Hubs* come online they must be federated together to provide a European dimension to the *e-infrastructure commons*.

[29] Cloud Computing Economics: An evidence-based approach for Research Applications http://e-infranet.eu/wp-content/uploads/2013/02/Cloud-Computing-Economics-An-evidence-based-approach-for-Research-Applications1.pdf

Prototype Research Accelerator Hubs

Several *Research Accelerator Hubs* will be required to satisfy the needs of the public research sector. Here we outline two EIROforum prototypes that can act as "pathfinders" for the overall model and pilot service. It is proposed to use the resources installed by CERN at the Wigner Research Centre for Physics in Budapest, Hungary[30], and by EMBL-EBI at Telecity's Powergate site in London, UK, to develop prototype *Research Accelerator Hubs* for the pilot service.

EMBL-EBI as a Research Accelerator Hub

EMBL-EBI will operate a pilot *Research Accelerator Hub* to serve the broad life science community based on its successful Embassy cloud, which it has been piloting since 2011. EMBL-EBI is Europe's principal provider of life science services, and provides many well-used resources, including UniProtKB, Ensembl, PDBe and ENA. The largest datasets, such as the ENA's DNA sequence archive requires around 2 Petabytes of storage, doubling every year, and many other datasets are growing rapidly. The Embassy cloud offers Infrastructure as a Service to other organisations (described as "tenant" organisations) using the proven VMWare technology. The service is operated from EMBL-EBI's resources within its tier-3 London data centres.

The principal benefit offered by the Embassy cloud is the potential of having virtual machines running on the same high performance network as the substantial public biology data resources held by EMBL-EBI. This allows high bandwidth, low latency operations to successfully occur, for example, mounting file systems directly to the virtual machines. This eliminates the requirement to replicate such data within the tenant's own remote infrastructure. Tenant organisations build and administer their own virtualised infrastructure, both machines and networks with access to EMBL-EBI's portfolio of datasets. There is no fundamental change of data access rights by being a tenant organization; the benefit is removing both the need to duplicate data and the complex data management of keeping a remote copy in sync with a EBI dataset.

As of June 2013 there are 8 tenant organisations participating in the Embassy Cloud trial, from both the academic and private sectors. The private sector will be paying "at cost" for the machine usage; currently academic users are selected by their use case. Use cases have included testing direct access to the 1000 Genomes data collection, as well as live public services operated by third parties such as a European mirror for OMIM.

EMBL-EBI plans to continue to evolve the Embassy model from 2014. This will involve a significant scale-up in capacity to include high-performance compute, providing a more efficient combined pool for internal and tenant organisations. Planned use cases include large-scale analysis of genomic data via partnerships with the International Cancer

[30] http://press.web.cern.ch/press-releases/2013/06/cern-and-wigner-research-centre-physics-inaugurate-cern-data-centres

Genome Consortium and the metagenomics dataset of the high profile Tara Oceans project. Further developments will target other life science use cases in medicine, agriculture and ecosystems, integration into a wider ecosystem of federated cloud infrastructures, and further adoption of open technologies as they become available. Many of these activities will be based upon the outcomes of EIROforum Helix Nebula initiative, as well as related ELIXIR Pilot Actions.

CERN and Wigner as a Research Accelerator Hub

The Wigner centre was contracted by CERN following a competitive call to tender to which approximately 20 sites around Europe responded. Wigner is operated as an extension to the CERN Data Centre (hosted on the CERN site in Switzerland) and consequently CERN can assure the business continuity for the critical systems in case of major problems. The Wigner data centre is connected to CERN via multiple 100Gbs network links. CERN is integrated with many of the world leaders in international networking, including hosting a GEANT Point of Presence (PoP) as well as being a major center for commercial networking for both telecoms suppliers as well as Internet Service Providers (ISPs). Additional network options provided by public and private bodies are welcome according to the needs expressed by the users, ranging from commodity IP services to dedicated circuits directly to Wigner or via CERN.

The following initial portfolio of services will be provided by this prototype *Research Accelerator Hub*:

- A virtual multi-tenant compute environment to provision and manage networks of virtual machines on-demand.
- A 'dropbox' style service for secure file sharing over the internet.
- A point-to-point reliable, automated file transfer service for bulk data transfers[31].
- An openAIRE compliant open access repository [32] for publications and supporting data allowing users to create and control their own digital libraries. Persistent digital object identifiers will be assigned to all publically available uploads so as to make them citable and permit the creation of a digital data continuum spanning from experimental data through to publications including links to commercial publishers.
- A long-term archiving service[33].
- Integrated Digital Conferencing [34] tools allowing users to manage their conferences, workshops and meetings.
- Training services: Online training material[35] will be made available and re-enforced with advanced training via brief residential programmes and secondments, such as CERN's openlab summer student programme[36], which has

[31] http://information-technology.web.cern.ch/services/file-transfer
[32] http://www.zenodo.org/features
[33] http://castor.web.cern.ch/
[34] http://indico-software.org/
[35] http://indico.cern.ch/categoryDisplay.py?categId=88
[36] http://openlab.web.cern.ch/news/cern-openlab-summer-student-programme-invites-2013-applications

recently been expanded with the engagement of EIROforum organisations and more companies.

The services will be accessible in a single sign-on (SSO) manner supported by a federated identity management system. These services will be implemented using the OpenStack[37] open source software suite accessible via the most popular interfaces (including EC2, S3 and potentially OCCI). OpenStack has been selected as the framework for the e-infrastructure services offered by the prototype *Research Accelerator Hub* because it has emerged as a clear leader in the vibrant cloud management market. CERN has extensively tested the software and has adopted it for the management of its production resources at its data centre in Geneva. OpenStack is fully open source and distributed under the Apache 2 license supported by a growing consortium of more than 180 private companies and public organisations with a well-defined and transparent governance structure representing a global community of more than 9,000 people across 87 countries. More than 800 developers are contributing to the software. All of these attributes means OpenStack does not require public funds to be maintained and offers one of the best environments for innovation.

Recent developments lead CERN to conclude that the open source software technology will shortly be available to make this federation of a network of *Research Accelerator Hubs* technically feasible. Federation is stated as an objective of the OpenStack Foundation[38]:

> "*The vision of federation across deployed OpenStack clouds is critical to the OpenStack vision. At the core OpenStack is a framework for managing and provisioning compute, network and storage. From that core many things are possible. But, we hope and will push for the OpenStack core to remain robust, consistent and stable. Yes, every deployment will have its own configuration and technical specifications, but as long as the core concepts and native APIs are robust and consistent, the possibility to drive real interoperability will exist and the value from this is a huge advantage of the open cloud concept*".

Long term sustainability, governance and collaboration with industry are achieved by building the prototype *Centre of Excellence* on the existing open source community and means national and European funding can be focused on innovation.

The services provided by the prototype *Research Accelerator Hub* will not be made available on a commercial basis. CERN rents space in the Wigner data centre and commercial suppliers can discuss directly with the local authorities if they are interested in locating their own services in Wigner. This co-location model is already used in a number of centres around Europe offering services to the business and public research sectors, including an internet exchange point at CERN's data centre in Geneva[39] and the EMBL-EBI Embassy cloud that enables external organisations to perform secure data analyses on large datasets using virtual machines.

[37] http://www.openstack.org/
[38] http://www.rackspace.com/blog/an-open-letter-to-the-openstack-community/
[39] The CERN Internet eXchange Point, http://cixp.web.cern.ch/

The proposed funding model for the pilot service is that CERN will enter into formal agreements with partner research organisations that wish to jointly develop and use the services offered by the prototype *Research Accelerator Hub*. CERN will take responsibility for operating the services at the *Research Accelerator Hub* and cover the costs during the first year of operation. Beyond this first year, each partner research organisation engages to fund the cost of the services their registered users consume according to a pay-per-usage model where the detailed cost model will be jointly-developed during the first year. Should a partner research organisation withdraw from the pilot service then access to the services of the prototype *Research Accelerator Hub* for their registered users will be disabled following a grace period after which the host organisation will no longer guarantee the integrity of any material under its stewardship. The costs of operating the *Research Accelerator Hub* at the capacity foreseen for CERN's own usage will continue to be paid by CERN itself. CERN will charge to the partner research organisations the additional costs which are incurred when increasing the site's capacity for the purpose of the *Research Accelerator Hub*. These additional costs will include processors, storage, network switches, cabling, energy, software licensing costs and operational staff. CERN will operate the services on the same hardware configurations[40] and at the same level of quality as those offered to its high energy physics user community. As the Data Custodian, CERN will not exert any ownership or intellectual property rights over the material that has been deposited. Tools for monitoring the resources consumed and the quality of services provided will be made available. The services will be integrated in a bottom-up manner, starting with basic IaaS services and allowing each partner research organisation to integrate and manage their own higher-level services and portals. Any new proposed service will only be introduced if there is at least one identified partner research organisation that is prepared to contribute to the costs of operating the service. Support for the cloud services will be integrated into CERN existing support structures (e.g. the standard service portal[41]). Effort will be required from each partner research organisation to curate their data-sets, connect their identity federations, deploy their community specific services and portals as well as manage the interaction with their registered users and associated support activities.

[40] Technical Description Servers and Storage for Physics Data Processing, Acquisition and Control, CERN, MS-3903/IT, http://cds.cern.ch/record/1472671/files/MS-3903_Technical_Description.pdf
[41] https://cern.service-now.com/service-portal/

Funding Models for the Prototype Research Accelerator Hubs

To ensure the sustainability of the prototype *Research Accelerator Hubs* funding models are envisaged where all the stakeholders participate. Sustainability is seen as a process rather than a fixed solution and hence the funding models will evolve over time.

The rationalisation offered by the prototype *Research Accelerator Hubs* and their integration into a hybrid model will provide a number of advantages:

- Provide a clear example of the scale and nature of the services that would be required from commercial suppliers.
- Offer "user aggregation" of procurement since the prototype *Research Accelerator Hubs* will be in a position to negotiate with suppliers on behalf of all the partner research organisations. This will ensure that the most cost effective means of providing a service will be available at any moment in time. This will also simplify the contractual and administrative aspects of procuring services from commercial suppliers for the cooperating research organisations.
- The procurement and deployment cycle for the centre will take several months and so the integration with commercial cloud service providers will be used to provide additional capacity in the interim. This will bring added value to the cooperating research organisations and their users by making it simpler for them to use commercial cloud services in '*bursts*' to access additional resources to fulfil any short-term requirements.
- The technical aspects of migrating user workloads to commercial services will be simplified and accelerated.
- The prototype *Research Accelerator Hubs* will make it possible to host data-sets on a cost-recovery basis while offering open access so they can be exploited via commercial services. This will remove the need for commercial suppliers to make investments to host data-sets without a clear estimation of how much business they will generate.
- The archiving service would offer a means of ensuring that data-sets have a safe back-up copy.
- The prototype *Research Accelerator Hubs* will, in relations to Europe's HPC strategy, address the need for more flexible, easier-to-use, more productive and more cost-effective HPC systems.

Timeline

The timeline below shows the major milestones foreseen from conception through to the establishment of the pilot service.

Fourth Quarter 2013:

- Discuss implementation plan with relevant EC projects
- Refine implementation plans for the prototype *Research Accelerator Hubs* with further details of each service offered, SLA, terms and conditions of service
- First meeting of the user forum

First quarter 2014:

- First meeting of management board
- Initial portfolio of services (see Portfolio of Services) made available for testing by a small number of research partner organisation(s) via the prototype *Research Accelerator Hubs*

Second Quarter 2014:

- Management board meets to approve opening of the pilot service

Third Quarter 2014:

- Start of pilot service