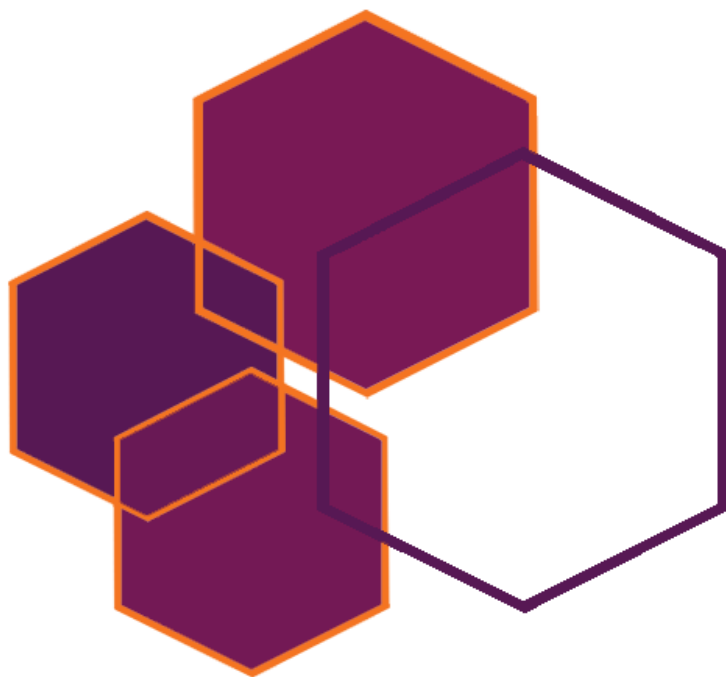


2014

STATEMENT OF INTENT



EU-T0
DATA RESEARCH AND INNOVATION HUB

Peter CLARKE;Manuel DELFINO;David
FOSTER;Volker GUELZOW;Giovanni
LAMANNA;Donatella LUCCHESI
<http://www.eu-t0.org/>
16/09/2014



STATEMENT OF INTENT

EXECUTIVE SUMMARY

The EU-T0 vision is to create a hub of knowledge and expertise to coordinate technological development, and integrate data and computing centres into a virtual Tier-0 centre, which can be more easily and transparently used by the different Scientific Communities. It will extend and simplify access to the resources, through a common approach to authorization, authentication and accounting frameworks. It will innovate new services and software tools as required to benefit the Scientific Communities and in particular the means to store, move, archive, preserve, and provide open access for very large scale data sets.

The EU-T0 consortium was initiated on 11th February 2014 by CERN, CIEMAT, DESY, IFAE, IN2P3, INFN, KIT and STFC, funding agencies and organisations which between them fund and provide very large scale computing and data processing resources for a set of scientific communities which already have Exabyte scale requirements. Participation in the consortium by similar funding agencies and organisations is encouraged.

The data centres of the consortium, and the communities they support, have been highly successful scientifically; they have innovated the development of distributed computing on a global scale, and represent an important part of the European e-Infrastructure for research. The forecast growth in both the rates and complexity of the data arising from the scientific communities is enormous, and will rise from Exabytes to Zettabytes within a decade. This gives rise to many technical and financial challenges in the future in order to meet these demands.

The EU-T0 partners already work closely together and now see the opportunity to cooperate even further in order to address the challenges and opportunities, with the expectation that this will yield scientific, technological and cost-efficiency benefits.

EU-T0 is multi-domain, its Scientific Communities include astronomy, astrophysics, astro-particle physics, cosmology, nuclear physics and particle physics and in some cases national and international photon and particle beam facilities as well as engineering and life sciences. EU-T0 seeks synergies with other research domains, and will collaborate with other projects and initiatives. EU-T0 will make an important contribution to structuring the European Research Area (ERA) through cooperation with ESFRI RIs.

EU-T0 will also seek to integrate with other e-Infrastructures and broaden existing collaborations beyond Europe.

EU-T0 will explore and propose a new model of international coordination for joint procurement and joint services and resources provision.

Table of contents

EXECUTIVE SUMMARY	2
1. The EU-T0 Motivation and Vision.....	4
2. Objectives.....	6
3. Initial Activities.....	7
Appendix A: The Partners and their Scientific Communities.....	8

1. The EU-T0 Motivation and Vision

The EU-T0 initiative was launched on the 11th February 2014 by some major research funding agencies and organisations: CERN, CIEMAT, DESY, IFAE, IN2P3, INFN, KIT and STFC. The participation of similar agencies and organisations is encouraged. The above listed agencies and organisations are hereafter referred to as the *Partners*. A description of each partner is given in Appendix A.

The Partners have considerable overlap in their roles and responsibilities in support of scientific research. Those *Scientific Communities* which are supported by each Partner are listed in Appendix A. In most cases these include astronomy, astrophysics, astro-particle physics, cosmology, nuclear physics and particle physics. In some cases photon and particle beam facilities as well as engineering and life sciences interests are also included.

Between them, the Partners own or provide a very large amount of the physical resources used by their Scientific Communities. Their data centres, which rank as some of the largest globally, act as the primary international repositories for communities already dealing with data volumes on the scale of Exabytes, They have amassed a considerable amount of expertise in their operation. EU-T0 is therefore already de-facto a major, successful, and existing component of the European e-Infrastructure.

The forecast growth in both the rates and complexity of the data arising from the Scientific Communities is enormous, i.e. from Exabytes to Zettabytes. A commensurate increase in the computing power is required and the need to move, manage and access data globally will be even greater. In the light of such growing demands, the Partners see several common opportunities and challenges for the future, which include:

- a. How to innovate shared computing and data processing services, which can be more easily and transparently used by their Scientific Communities.
- b. How to provide for the increasing resource demands within challenging financial constraints, including realising economies of scale.
- c. How to provide greater and simpler accessibility, through a common approach to integrating Authentication, Authorization and Accounting frameworks.
- d. How to efficiently utilise the next generation of computer architectures.
- e. How to store, move, archive, preserve, and provide open access for Zettabyte scale data sets.
- f. How to engender information exchange in respect of technology, services and standards between scientific communities, and thereby enable new communities to capitalise on the existing investment.
- g. How to provide stability on the scale of decades which is required for the largest and longest lived scientific endeavours, and the ensuing requirements of data stewardship.
- h. How to share benefit and experiences from technological development programmes.

It is therefore natural and appropriate that the Partners continue to work together to address the future opportunities and challenges presented above, with the expectation that this will yield scientific, technological and financial benefits. The Partners will work towards integrating their data and computing centres, into a virtual Tier-0¹ centre accessible to a wide set of communities, and coordinate the development of future computing and data processing services. This is the rationale for the formation of EU-T0 Consortium and its vision:

“The EU-T0 vision is to create a hub of knowledge and expertise to coordinate technological development, and optimise the investment of the funding agencies in their existing data centres by broadening, simplifying, and harmonising access, driven by well-defined user requirements.”

The EU-T0 vision embodies the idea of a hub focussed on science domains, which will primarily undertake activities that follow directly, or indirectly, from the requirements of its Scientific Communities. Many years of experience have demonstrated that the most successful e-Infrastructures are those that are closely coupled to, and driven by, the Scientific Communities. Therefore EU-T0 recognises that its Scientific Communities must be paramount in defining its strategy and activities.

In the longer term EU-T0 will be open to incorporating additional scientific communities where this leads to benefits to all communities concerned, and furthers the long term strategic objectives set out in this document.

EU-T0 is consistent with the recommendations presented by the IT working group of the EIROforum in a recent document². EU-T0 is an example of a “research accelerator hub - Reach” focussed upon its Scientific Communities.

EU-T0 addresses the main recommendations provided by the e-Infrastructure Reflection Group (e-IRG). EU-T0 will be a contributor to structuring the European Research Area (ERA) through cooperation with ESFRI RIs and other domain focussed infrastructures.

¹ The term Tier-0 is used to mean a top-Tier data processing and storage system capable of archiving, processing and providing access to data arising from experimental and simulation facilities. Referring to the top-Tier as Tier-0 is consistent with the terminology used in the Worldwide LHC Computing Grid (WLCG) and the Partnership for Advanced Computing in Europe (PRACE).

² EIROforum IT working group (2013). e-Infrastructure for the 21st century. ZENODO DOI: [10.5281/zenodo.7592](https://doi.org/10.5281/zenodo.7592)

2. Objectives

The objectives of EU-T0 are:

- I. To build upon the existing, successful and well proven, computing and data processing centres that are owned or funded by the EU-T0 Partners, by enhancing cooperation and harmonisation in respect of all aspects of the infrastructure.
- II. To integrate the computing and data processing centres into a European Tier-0 centre by creating shared services which can be more easily and transparently used by the different Scientific Communities.
- III. To extend and simplify access to the EU-T0 data centres and their services, through a common approach to integrating Authentication, Authorization and Accounting frameworks.
- IV. To achieve a more cost effective provision arising from economies of scale, minimisation of duplication, and adoption of common solutions.
- V. To innovate new services and software tools as required for the benefit of the Scientific Communities. In particular to develop the means to store, move, archive, preserve, and provide open access for Zettabyte scale data sets.
- VI. To integrate with other e-Infrastructures and to work towards convergence of HPC and HTC services.
- VII. To engage with commercial infrastructure providers to develop a workable mixed model of provision.
- VIII. To broaden existing collaborations beyond Europe.

The EU-T0 consortium will act as a focus to address future opportunities targeted at supporting these objectives. It will support collaboration with other projects where appropriate and will also support the initiation of new projects and proposals where appropriate.

3. Initial Activities

A likely set of initial activities is described below. The nature and extent of activities undertaken will be determined by the resources that can be made available by partners or external funds which can be won through competition. In the immediate future EU-T0 members expect to present proposals to the Horizon 2020 framework programme for some of the activities identified below.

- I. Coordination and collaboration to develop and implement solutions which will allow ingestion, storage, movement, archival, preservation and access (including open-access when relevant) of increasingly large and complex datasets, with a horizon of reaching in a decade aggregate levels of Zettabytes with Exascale number of data objects. Improving the ease of data exchange with HPC installations. This will allow services to fulfill the increasing requirements of the Scientific Communities and the corresponding large-scale facilities.
- II. Coordination of the work to develop and deploy virtualised services, collaboration with other projects working in this area, and piloting of the incorporation of commercial cloud resources. This will bring the benefit of providing a simpler underlying computing fabric which can more easily be (virtually) tailored to different Scientific Communities as required, including those of long tail science.
- III. Explore joint procurement opportunities where these would be cost effective.
- IV. Coordination of deployment of common AAA systems to simplify access for diverse Scientific Communities. This may include international policy and trust agreements, and endorsing and supporting projects to develop European or global AAA infrastructures. This will bring the benefit of “single sign on” to multiple diverse resources.
- V. Coordination and collaboration in the development, provision and deployment of a portfolio of services, including new services and software tools, as required. In particular those which allow the building of customizable Virtual Research Environments, composed of services for data workflow management, user access to software, archives, and scientific analysis systems; digital platform relating publications, data and analysis software repositories.
- VI. Working closely as a single point of contact with initiatives from scientific domains, such as ESFRI RIs, to develop agreements to provide resources and deploy domain based services across the EU-T0 data centres.
- VII. Reduction of cost through the sharing of resources and expertise between various scientific communities and data centres.
- VIII. Further harmonisation of security coordination, policy development and incident handling.
- IX. Promotion of a training network to create and develop data scientists.

Appendix A: The Partners and their Scientific Communities

- CERN is the world's largest particle physics lab and home of the Large Hadron Collider (LHC), the world's most powerful particle accelerator. CERN provides research facilities for HEP researchers across the globe. LHC experiments will run up to 1 million computing tasks per day and generate around 15 petabytes of data per year over a 10- 15 year period. Analysis of the data relies on the combined resources of some 200 computer centres world- wide. CERN leads the World- wide LHC Computing Grid project (WLCG), to provide computing resource to store, distribute, analyse and access (in near real- time) LHC data for a community of more than 10,000 physicists worldwide.
- CIEMAT (<http://www.ciemat.es>) is a public research body assigned to the Spanish Ministry of Economy and Competitiveness under the Secretariat of State for Research, Development and Innovation. CIEMAT is involved in basic and applied science and technology projects in the fields of energy, environment and the technologies related to them, at national and international levels. CIEMAT collaborates as well with other institutions, universities and business in the sector, to transfer the knowledge and technology that it has generated. The Scientific Computing Unit works on advanced scientific computing techniques, providing infrastructure and support to the huge and complex requirements of the aforementioned projects, using distributed computing, grid and cloud technologies. The Unit participates in the Port d'Informació Científica (PIC), a joint undertaking with IFAE (see PIC description under IFAE). Past and current projects include WLCG, EU DataGrid, Enabling Grids for e-Science (EGEE), EGI-Inspire, EELA, EDGeS and the nuclear fusion related EUFORIA project.
- DESY, the “Stiftung Deutsches Elektronen-Synchrotron DESY” in Germany, is one of the world leading laboratories for particle physics and photon science and has long-standing experience in the design and operation of accelerators for high energy physics and synchrotron radiation research. DESY operates one of the most brilliant synchrotron radiation sources world wide (PETRA III), two Free Electron Lasers (FLASH and FLASH-II) and will operate the European XFEL accelerator, a project from the ESFRI list. DESY holds a 50% share of the European XFEL, an ESFRI-project. As a member of the Helmholtz Association in Germany, DESY is a non-profit research organization funded by public funds from the government and the states.
- IFAE, the Institut de Física d'Altes Energies (<http://www.ifae.es/>), is a consortium of the Universitat Autònoma de Barcelona (UAB) and the Government of Catalonia, part of CERCA, the public Catalan network of research Institutes (www.cerca.cat). IFAE conducts experimental and theoretical research at the frontier of fundamental physics, including Particle Physics, Astrophysics and Cosmology. IFAE also works at the cutting edge of detector technology, applying its know-how to Medical Imaging, Data-Intensive Computing and projects in other research and engineering domains. IFAE was granted in 2012 the Spanish Government's Severo Ochoa award. IFAE's Experimental Division currently participates in ATLAS at the Large Hadron Collider, in the T2K neutrino experiment, and in the MAGIC and CTA Cherenkov gamma-ray telescopes. In the field of sky-surveys searching for dark energy, IFAE leads the Physics of the Accelerating Universe (PAU) survey, and participates in DES and in ESA's EUCLID mission. IFAE manages Spain's largest scientific data centre, the Port d'Informació Científica (PIC). Created in 2003, PIC is a joint undertaking of the Spanish and Catalan governments through CIEMAT and

IFAE, with additional support from UAB. PIC has been designated by the Spanish government as its LHC Tier-1 centre, and it is the main (Tier-0) data centre for the MAGIC telescope and the PAU dark energy survey, as well as some of the largest sky-survey catalogue simulations performed on supercomputers. PIC maintains a transversal innovation activity with many significant results over the years, related to software, hardware, monitoring and energy efficiency. It also seeks collaborations to transfer the knowledge and technologies to other fields, including aerodynamic design and medical imaging analysis.

- The CNRS - *Centre National de la Recherche Scientifique* (National Centre for Scientific Research) is a government funded research organization, under the administrative authority of France's Ministry of Research. As the largest fundamental research organization in Europe, CNRS carried out research in almost all fields of knowledge, through its eight research institutes among which the National Institute of Nuclear and Particle Physics (IN2P3).
Founded in 1971, the aim of IN2P3 is to promote and unify research activities in the fields of nuclear physics, particle and astroparticle physics. It coordinates programmes within these fields on behalf of the CNRS and universities, in partnership with CEA. Whilst these main fields represent the core of the discipline, IN2P3 also has several additional vocations: enabling other scientific domains to benefit from its competencies and solving certain problems posed by society, and accompanying universities in contributing to youngsters' training. IN2P3 is also responsible for running several major national facilities including particle accelerators. It also supports several international facilities. As the experiments require significant investments in terms of financing and personnel, from the outset the Institute took the structure of a limited number of large laboratories and Infrastructures or technological platforms in order to facilitate the pooling and optimization of its resources and competencies. IN2P3 also shares with the CNRS/INSU Institute for Astrophysics the support to some projects in Astrophysics and Cosmology. CNRS, together with CEA-IRFU, fund almost all national e-Infrastructures for these domains. These e-Infrastructures provide data archive and computing resources also to research projects in other fields such as biology and life science, humanity and social science.
- INFN, the National Institute for Nuclear Physics, is the Italian research agency dedicated to the study of the fundamental constituents of matter and the laws that govern them, under the supervision of the Ministry of Education, Universities and Research (MIUR). It conducts theoretical and experimental research in the fields of subnuclear, nuclear and astroparticle physics. All of the INFN's research activities are undertaken within a framework of international competition, in close collaboration with Italian universities on the basis of solid academic partnerships spanning decades. Fundamental research in these areas requires the use of cutting-edge technology and instruments, developed by the INFN at its own laboratories and in collaboration with industry. Groups from the Universities of Rome, Padua, Turin, and Milan founded INFN on 8th August 1951 to uphold and develop the scientific tradition established during the 1930s by Enrico Fermi and his school, with their theoretical and experimental research in nuclear physics. In the latter half of the 1950s, INFN designed and built the first Italian accelerator, the electron synchrotron developed in Frascati, where its first national laboratory was set up. During the same period, INFN began to participate in research into the construction and use of ever-more powerful accelerators being conducted by CERN, the European Organisation for Nuclear Research, in Geneva. Today, INFN employs some 5,000 scientists whose work is recognised internationally not only for their contribution to various European laboratories, but also to numerous research centres worldwide.

- KIT, the Karlsruhe Institute of Technology, was founded in 2009 by the merger of the Karlsruhe Research Centre (Forschungszentrum Karlsruhe, FZK) and the University of Karlsruhe (TH). It is a public corporation according to the legislation of the state of Baden-Württemberg and fulfils the parallel missions of a university and of a national research centre within the Helmholtz Association. With 9,400 employees (including nearly 6,000 employees working in research and education), 370 professors, 3,200 doctoral researchers, and an annual budget of about 790 million (2013), KIT is one of the biggest research and education institutions in Europe. At Steinbuch Centre for Computing (SCC) R&D focuses on computational science and engineering, data life cycle management of scientific data, IT management, web engineering, federated identity management and process integration. Furthermore, SCC is also the IT service provider of KIT. SCC operates very powerful HPC systems, among others the high-performance computer of the State of Baden-Württemberg, being used by regional and federal scientists from academia and industry for computationally intensive projects. In this context SCC's SimLabs form the interfaces between users and providers focusing on joint R&D in HPC and computational science. SCC is a member of the Gauß-Alliance that promotes and facilitates strategic HPC research in Germany. SCC is the location of GridKa – one of twelve WLCG Tier-1 centres worldwide. SCC is leading several national and state-wide projects related to big data management, archival and analysis. With the LSDF SCC has realised an innovative concept for storage, archival and analysis of scientific data. The LSDF supports the complete data life cycle and is in principle available to all scientific disciplines. In several state-funded projects the LSDF is enhanced with additional services for sync&share, block storage and archival available to all other universities in the state of Baden-Württemberg. A federated identity management system enables a seamless usage of the LSDF as well as powerful HPC and data analysis facilities in the state. In SCC's Data Life Cycle Labs (DLCLs), data experts perform joint research and development activities with scientists from several research fields to advance specific data life cycles; this is enhanced by technology research for generic, multi-community tools and services meeting the demands on IT security and trust.
- STFC, the Science and Technology Facilities Council in the UK, is the funding agency with responsibility for providing support for all staff and resources for Particle Physics, Particle Astrophysics, Astronomy, Astrophysics, Cosmology and Nuclear Physics. STFC funds all e-Infrastructure for these domains and in particular funds the GridPP project which provides all WLCG computing and the DiRAC project which provides HPC facilities for theoretical particle physics and astronomy. STFC also runs several major UK national laboratories including the Rutherford Appleton Laboratory, the Daresbury Laboratory, the Astronomy Technology Centre and facilities including neutron and light sources. STFC also hosts the Hartree HPC centre.