



LHC, EGI-InSPIRE & the “Digital Agenda”

Jamie.Shiers@cern.ch

~ ~ ~

Fifth Workshop on Data Preservation & Long Term Analysis in HEP, May 2011



Overview

- LHC as an example of long-term (multi-decade) data preservation
 - 1990s, 2000s, 2010s and beyond
- EGI-InSPIRE: broadening to other disciplines
 - Life Science, Earth Science, Astronomy & Astrophysics etc.
- The EU “Digital Agenda” and Framework Programme 8
 - Funding from 2014 – now is the time to influence call(s)!
- Wrap-up

LHC Computing

- We are now in the 3rd decade of LHC Computing
 1. Started around CHEP '92 in Annecy – “OO” & R&D projects; MONARC; Hoffmann
 2. Began at [CHEP 2000](#) in Padua – “grid-itus”;
 3. (IMHO) coincided with the beginning of serious LHC data taking end March 2010.
- There have been *many* major (and minor) migrations during this period – who can remember them all?
 - SHIFT, HPSS, CASTOR1, CASTOR2, EOS, grid interfaces, non-stop media migration and repack, server retirement and disk pool migration, searches for “dark data” & catalog consistency;
 - Objectivity/DB migration – 300TB of COMPASS + HARP (ATLAS & CMS “dropped” the old MC data once sufficient new generated).
- How many more decades ahead? How many more migrations?
- We “know” the effort involved, both on service and “user” sides...

- European Grid Initiative: Integrated Sustainable Pan-European Infrastructure for Researchers in Europe
- The 48 month EGI-InSPIRE project (from May 2010) will continue the transition to a sustainable pan-European e-Infrastructure started in EGEE-III
- It will sustain support for Grids of high-performance and high-throughput computing resources, while seeking to integrate new Distributed Computing Infrastructures (DCIs), i.e. Clouds, SuperComputing, Desktop Grids, etc., as they are required by the European scientific community

- This activity provides continued support for activities currently supported by EGEE while they transition to a sustainable support model within their own community or within the production infrastructure by:
 - Supporting the tools, services and capabilities required by different heavy user communities (HUCs)
 - Identifying the tools, services and capabilities currently used by the HUCs that can benefit all user communities and to promote their adoption
 - Migrating the tools, services and capabilities that could benefit all user communities into a sustainable support model as part of the core EGI infrastructure
 - Establishing e a sustainable support model for the tools, services and capabilities that will remain relevant to single HUCs
- **Unlike the other Work packages it runs for 36 months**

- Heavy User Communities (HUCs) are Virtual Research Communities (VRCs) that have been using EGEE and EGI routinely and thus have become more structured and advanced in terms of grid usage
- The communities identified as Heavy Users Communities (HUCs) are:
 - High Energy Physics (HEP)
 - Life Sciences (LS)
 - Astronomy and Astrophysics (A&A)
 - Computational Chemistry and Materials Sciences and Technologies (CCMST)
 - Earth Sciences (ES)
 - Fusion (F)
- HEP today represents around 40% of the registered “EGI” users but provides and consumes by far the most resources
- Quote from EGI User Forum: ***“WLCG delivered 80 CPU-millennia in 2010”***

Data Volumes & Lifetimes

- HEP might (today) have the largest data volumes but
 - a) This is changing fast;
 - b) The lifetime for which the data needs to be preserved is perhaps decades for HEP, whilst much (much) longer in LS, ES
- Life Science:
 - E.g. patient data (individual patient, collective)
 - 100 years +
- Earth Science:
 - Millennia?
- EGI-InSPIRE encourages us (as HEP) to work with these communities – we potentially have much to gain (and give)

Requirements vs Priorities

- Discussions on requirements – e.g. as performed at EGI Technical Forum in September 2010 – suggest that there is significant overlap
- What differs is more the priority of the various features
 - E.g. patient data needs to have much more stringent access control than event data but that is not the same as saying we neither want nor need *any* level of control
 - If cost and granularity are / become affordable, likely that convergence / overlap will increase...
- EGI-InSPIRE SA3 is in the process of building a “requirements matrix” across the supported communities

PARTNER 2

- A follow-on hadron-therapy project for which funding is hoped as from late 2011 / early 2012
- Asked for funding for two Marie-Curie fellows to work on specific aspects of data preservation
- With or without this funding we still have to face and solve the issues involved in keeping LHC data alive for the next two decades
- And there is recent interest in late LEP data too
 - LEP experiments now registered as Virtual Organizations

FP8 and the Digital Agenda

- The EU has been exposing its plans for FP8 since ~the start of EGI-InSPIRE
 - Talks at both the Technical Forum September 2010 in Amsterdam as well as the User Forum April 2011 in Vilnius
- They definitely want the focus to be “multi-disciplinary” – from science to humanities
- This year and next is the time to prepare
- Proposals in 2013, funding from 2014
- With our current involvement not only in EGI-InSPIRE but also other ES and LS projects we are (well?) positioned
- What should we do?

Prospects for e-Infrastructures in the ICT and innovation policy context

Overview: Kostas Glinos will present an overview of the current policy developments at EU-level and the role of e-Infrastructures in this evolving context. He will focus on two Europe 2020 flagship initiatives, the Digital Agenda for Europe and the Innovation Union, and on the steps towards the construction of the European Research Area. He will also refer to the ongoing consultation on the successor of the current Framework Programme for Research.

Discussion...

Conclusions

- EGI-InSPIRE requires us (HEP, CERN IT-ES) to work with multiple disciplines, including Life Science & Earth Science – we are involved in other projects with these communities too
 - Hadron therapy: PARTNER, ULICE, ...
 - Environmental: EnviroGRIDS, UNOSAT, ...
- FP8 is being prepared now – and will have a significant focus on “the Digital Agenda” (EGI Technical Forum talk)
- Caveat – there can be significant reporting overhead in EU projects!
- But can we afford *not* to be involved?

ALLAROUND US IT IS THERE WHEN YOU WATCH TELEVISION
美と字印 び抜す 国出のシ品 致最ま ゴ図ンは証 メ密万

TRIXIT IS ALLAROUND US IT IS THERE WHEN YOU WATCH
THE MATRIX HEIS THE ONE DREAMWORLD NEO
YOUR WALPAPER.COM

の補 及術文写て 感ザ絵しオ会観美イ 力版もレ 保の 文精なフト社明 をに美と字印 び抜す 国出のシ品 致最ま

と字印 び抜す 国出のシ品 致最ま
IS ALLAROUND US IT IS THERE WHEN YOU WATCH TELEVISION

IT IS THERE WHEN YOU WATCH TELEVISION

及術文写て 感ザ絵しオ会観美イ 力版もレ 保の 文精なフト社明 をに美と字印 び抜す 国出のシ

TRIX HEIS THE ONE DREAMWORLD NEO ANAGENT TRINITY
をに美と字印 び抜す 国出のシ品 致最ま ゴ図ンは証 メ密万

90+54 17199 > H5801K P>>R+4

刷の補 及術文写て 感ザ絵しオ会観美イ 力版もレ 保の 文精なフト社明 をに美と字印 び抜す 国出のシ品 致

出す 国出のシ品 致最ま ゴ図ンは証 メ密
DREAMWORLD NEO ANAGENT TRINITY WHAT IS YHE MAT

ND US IT IS THERE WHEN YOU WATCH TELEVISION

感ザ絵しオ会観美イ 力版もレ 保の 文精なフト社明 をに美と字印 び抜す 國
IS THE MATRIXIT IS ALLAROUND US IT IS THERE WHEN
をに美と字印 び抜す 国出のシ品 致最ま ゴ図ンは証 メ密万

17199 > H5801K P>>R+4

レ 保の 文精 刷の補 及術文写て 感ザ絵しオ会観美イ 力版もレ 保の 文精なフト社明 をに美と字印 び抜す

THEIS THE ONE DREAMWORLD NEO ANAGENT TRINITY WHA
THE MATRIXIT IS ALLAROUND US IT IS THERE WHEN YOU

00+54 17199 > H5801K P>>R+4

感ザ絵しオ会観美イ 力版もレ 保の 文精なフト社明 をに美と字印 び抜す 国出のシ品 致最ま
致最ま ゴ図ンは証 メ密万

NEO ANAGENT TRINITY WHAT IS YHE MAT
をに美と字印 び抜す 国出のシ品 致最ま ゴ図ンは証 メ密万

278 20307 17199 > H5801K P>>R+4

刷の補 及術文写て 感ザ絵しオ会観美イ 力版もレ 保の 文精なフト社明 をに美と字印 び抜す 国出のシ品 致最ま

ATIS THE MATRIXIT IS ALLAROUND US IT IS THERE WHEN
HE MATRIX HEIS THE ONE DREAMWORLD NEO ANAGENT TR

をに美と字印 び抜す 国出のシ品 致最ま
MATRIXIT IS ALLAROUND US IT IS THERE WHEN YOU WATCH

90+54 17199 > H5801K P>>R+4

刷の補 及術文写て 感ザ絵しオ会観美イ 力版もレ 保の 文精なフト社明 をに美と字印 び抜す 国出
IS THE ONE DREAMWORLD NEO ANAGENT TRINITY WHAT
IT IS THERE WHEN YOU WATCH TELEVISION

90+54 17199 > H5801K P>>R+4