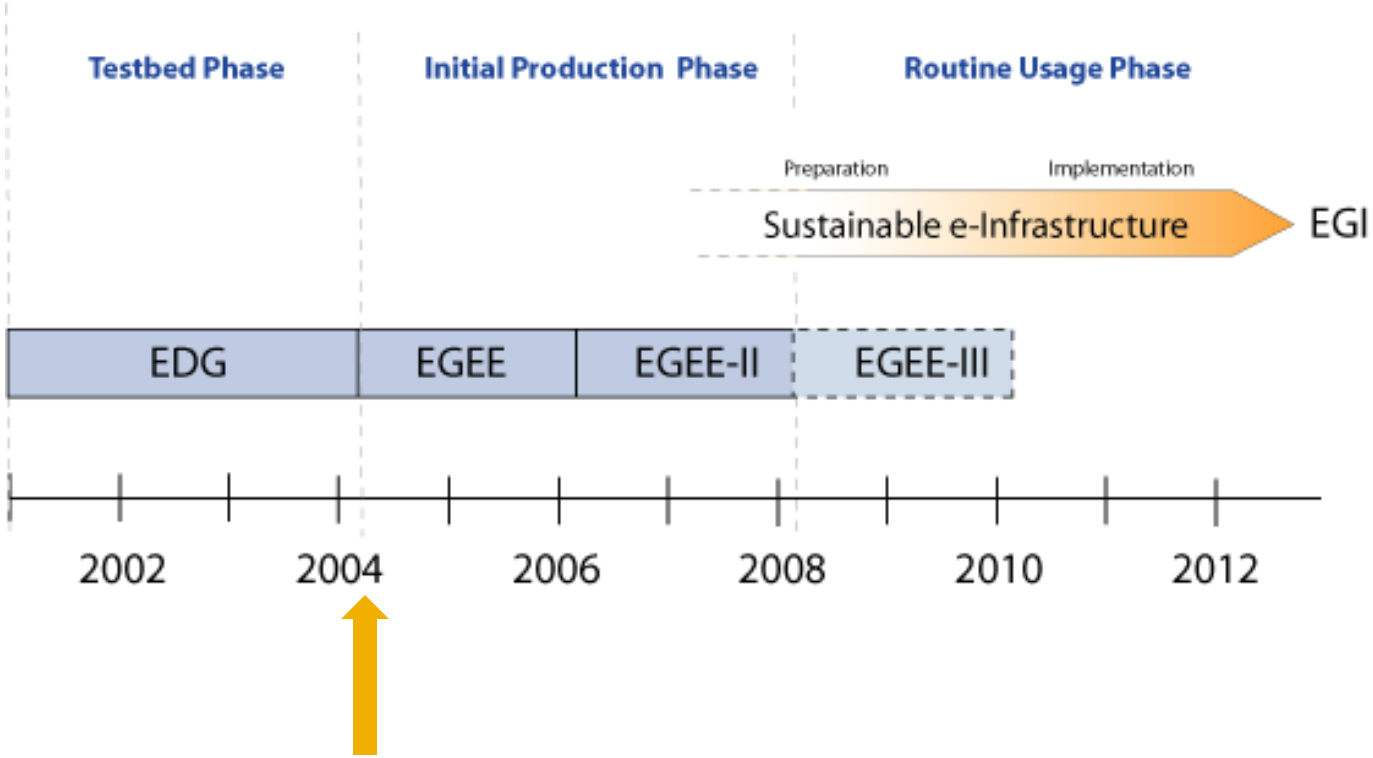
Impact assessment of e-infrastructures

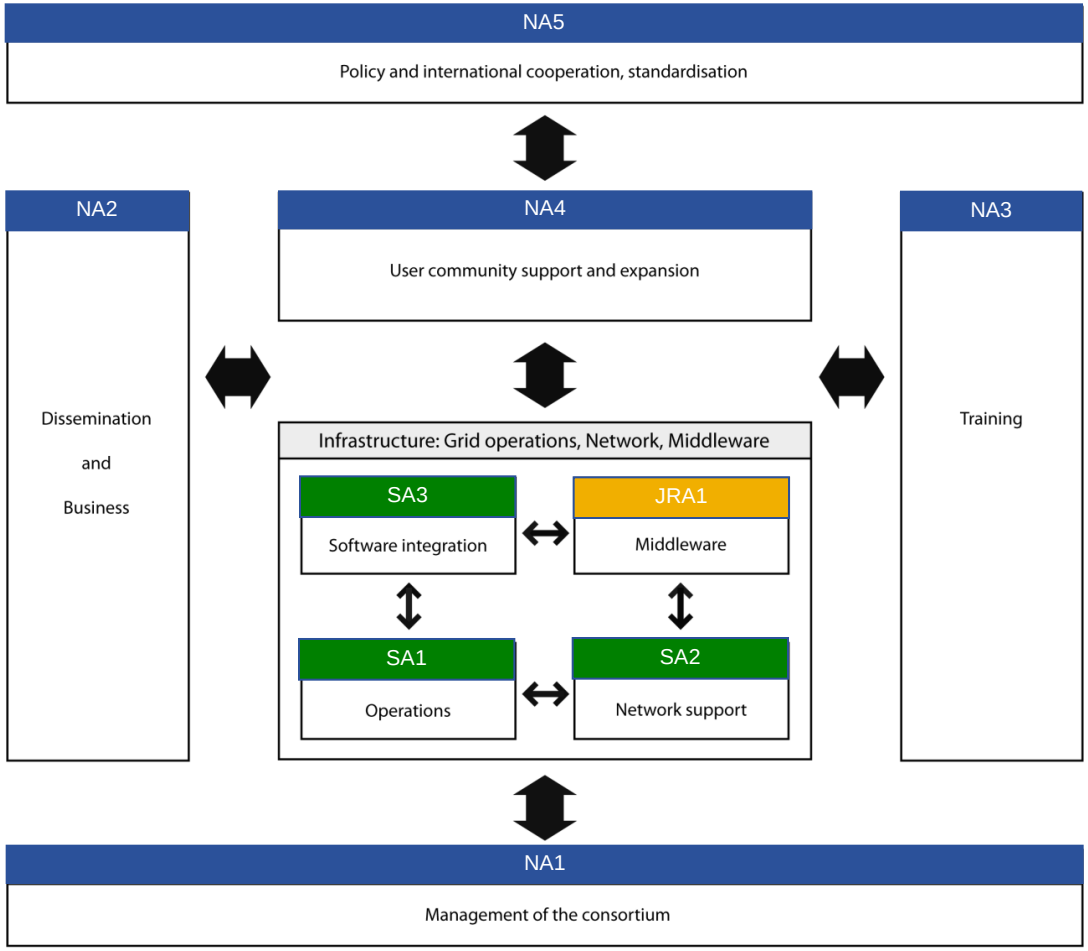
lessons learnt from EGEE

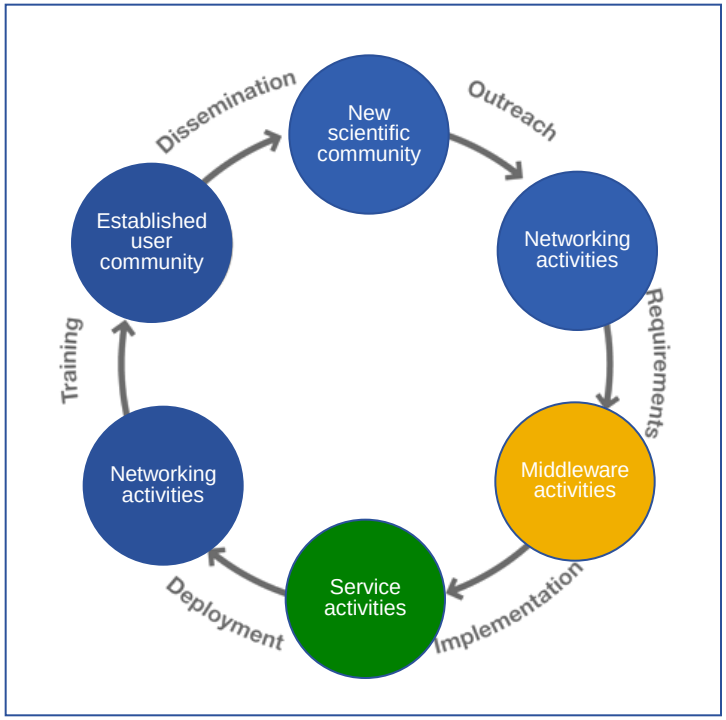
Bob Jones
CERN
IT Department



Metrics programme introduced then reviewed/refined every year

EGEE activities



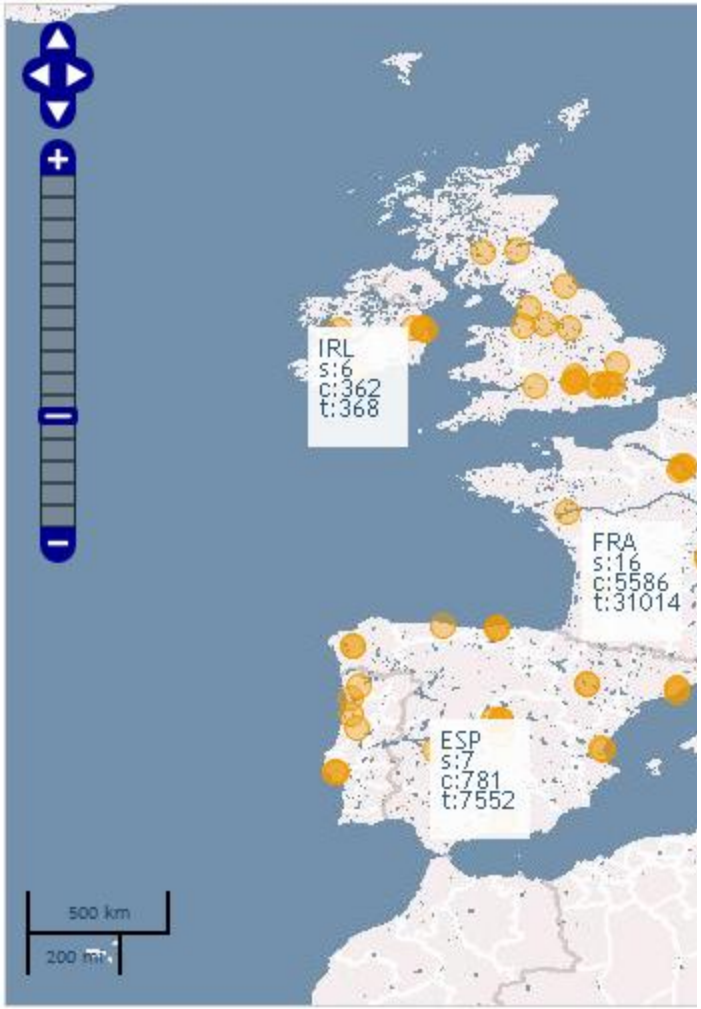


EGEE project metrics

Objectives	Metric	Targets at end of Year 1	Targets at end of Year 2	Q1	Q2	Q3	Q4	Y1 status	Q5	Q6	Q7	Q8	Y2 status
Expand scale of the infrastructure	Number of sites connected to the infrastructure	>250	>270	260	261	267	268	268	289	291	302	317	317
	CPU core made available to the user communities	76000	85000	72000	87000	110300	117500	117500	142783	146330	155066	235024	235024
	Number of countries connected to the EGEE infrastructure	~45	~46	45	45	48	48	48	49	50	50	52	52
	Number of supported platform	4	5	3	5	5	4	4	4	4	4	4	4
Expand the User communities	Number of VOs registered	150	170	130	140	154	162	162	166	175	186	192	192 (*)
	Number of registered users	13500	>14000	13 800	15 000	16 759	18 000	18000	11552 (Clean-up)	12000	12235	12583	12583(*)
Expand scale of the usage of the infrastructure	Average jobs/day during the quarter	250000	300000	250 000	151 500	276 700	343 600	255 450	450500	441879	404717	527456	527456
	Data transfer: Site to site Data Movement with FTS per Month	2,3 Pbytes/M	2,6 Pbytes/M	2,3	2,1	1	1.5	1.7	1.7	1.4	1.03	1.75	1.75
	Computing resources consumed by the user communities per quarter	30800cpu (KSI2K.quarter)	36000cpu (KSI2K.quarter)	19750	23850	32600	50100	31800	55000	51682	80412	74290	74290
Further improve QoS of the Infrastructure	Resource center reliability	85%	90%	83.06%	86.15%	92.3%	89.3%	87.70%	93.98	92.33	94.88	93.33	93.33
Expand regional training autonomy	Number of accredited trainers	100	110	93	96	103	104	104	104	133	133	134	134
	Number of training events	60	120	21	+19	+43	+18	101	+21	+19	+21	+23	185
	Number of attendees trained	600	1200	332	+358	+400	+286	1376	+216	+215	+271	+234	2312
Dissemination & Outreach	Events attended	70	140	17	+26	+35	+9	87	+18	+19	+25	+12	161
	Events organised	80	160	13	+26	+16	+26	81	+24	+20	+9	+12	146
	Number of business events organised	10	20	1	+3	+1	+12	17	+2	+4	+1		24

http://project-egee-iii-na1-qa.web.cern.ch/project-EGEE-III-NA1-QA/EGEE-III/QoS/Follow-up/Overall.htm

Usage of infrastructure measures



EGI Availability and Reliability Report for VO OPS

Region Summary - Sorted by Name — **September 2010**

Data from Nagios and Gridview

https://twiki.cern.ch/twiki/pub/LCG/GridView/Gridview_Service_Availability_Computation.pdf

Availability = Uptime / (Total time - Time_status_was_UNKNOWN)
Reliability = Uptime / (Total time - Scheduled Downtime - Time_status_was_UNKNOWN)
HS06 : Installed capacity of the site measured in HEPSPEC06 (HS06)
Reliability and Availability for Region - Weighted average of sites in the Region (supporting this VO) based on installed capacity

Colour coding : N/A | < 50% | < Target | >= Target

EGI SLA Availability Target is 70 % and Reliability Target is 75 %

Region	Availability	Reliability
AsiaPacific	78 %	78 %
CERN	96 %	96 %
Italy	91 %	93 %
NGI_AEGIS	99 %	99 %
NGI_BY	86 %	86 %
NGI_CH	37 %	39 %
NGI_CYGRID	99 %	99 %
NGI_CZ	99 %	99 %
NGI_DE	87 %	87 %
NGI_FRANCE	85 %	88 %
NGI_GRNET	94 %	96 %
NGI_HR	100 %	100 %
NGI_HU	99 %	99 %
NGI_IBERGRID	96 %	97 %
NGI_NDGF	96 %	97 %
NGI_NL	95 %	95 %

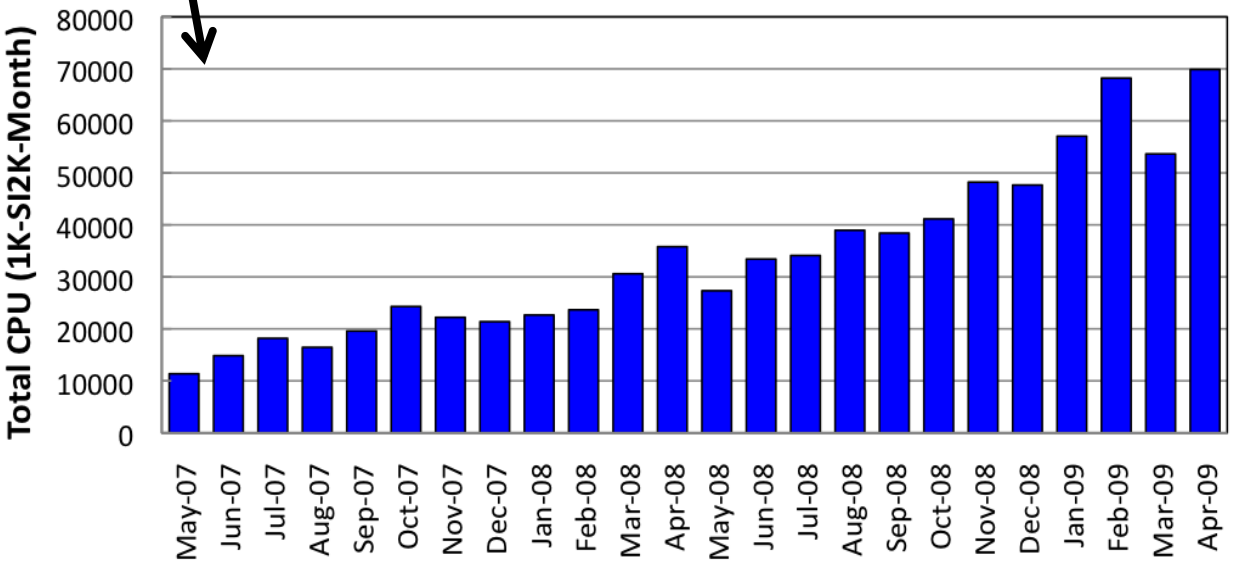
Region	Availability	Reliability
SouthEasternEurope	87 %	90 %
UKI	88 %	89 %

http://gstat-prod.cern.ch/gstat/geo/openlayers#/Country/ALL

https://documents.egi.eu/public/RetrieveFile?docid=219&version=1&filename=EGI_Sep2010.pdf

Use of User measures

Consistent doubling every 12-18 months.
HEP largest users / contributors
AA/ES/other show strong increase



Domain	VOs	Users
AstroPhy & Astronomy	20	373
Comp Chem	4	347
Comp Sci	4	21
Earth Sci	7	142
Fusion	2	68
High Energy Phys	36	8577
Life Sci	9	379
"Regional"	26	1658
Other	28	1816
TOTAL	136	13381

>13000
Registered Users

CIC Portal: http://cic.gridops.org/index.php?section=home&page=volist

Accounting Portal: http://www3.egee.cesga.es/

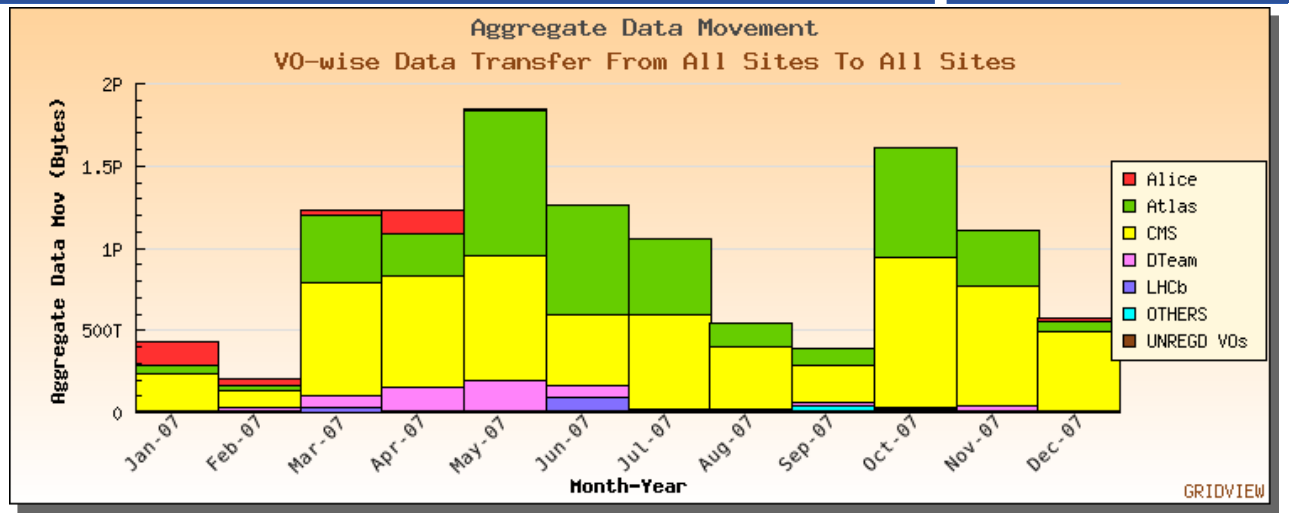
Measurements for cost comparison

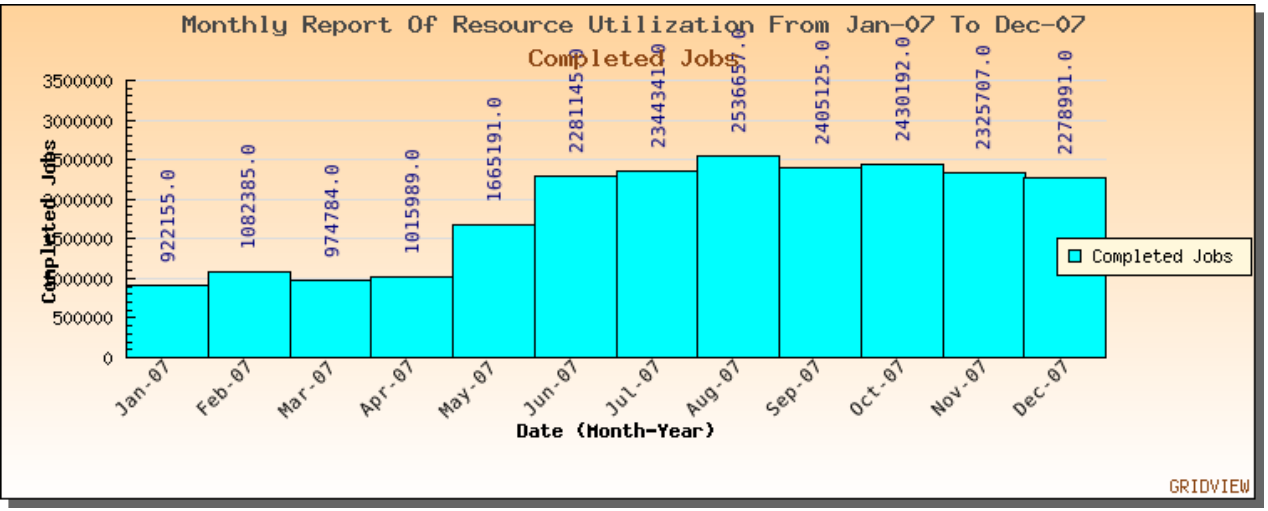
EGEE workload (2007)

Data:

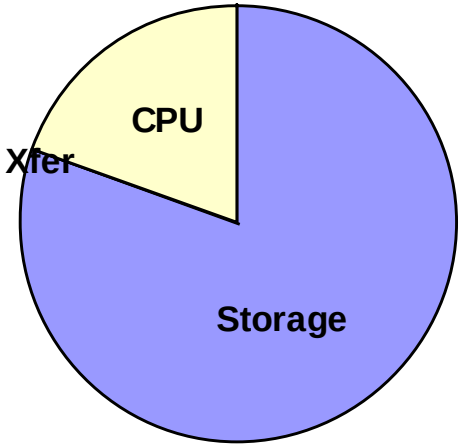
25Pb stored

11Pb transferred





CPU: 114 Million hours



Estimated cost if performed with Amazon's EC2 and S3: £ 29,977,003

http://gridview.cern.ch/GRIDVIEW/same_index.php http://calculator.s3.amazonaws.com/calc5.html? 17/05/08 $58688679.08

Benefits and cost of the metrics programme

CERN IT Department

- Benefits
  - Provided a dashboard for the project management
  - Highlighted where the project was doing well (and not so well)
  - Helped keep the activities focused
  - Provided good input for EC reporting and dissemination tasks
  - Helped demonstrate value to the project stakeholders
    - Funding agencies (EC and national agencies)
    - Users (and potential new users)
    - Resource providers
- Costs
  - Does not come for free so keep the set of measurements to a minimum
  - Be very specific about what you want to measure
  - Know how you intend to use the measurements
  - Think carefully about how often you need to measure

And there were some things we could not measure within the project….

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it

Impact we did not measure

"One week after the first collisions, the first papers were published electronically. And these were all peer-reviewed."

iSGTW INTERNATIONAL SCIENCE GRID THIS WEEK

About | Archive | Calendar | Learn | Interact | Press Room Search

Home > iSGTW - 28 April 2010 > Feature: Peer-reviewed physics, at the speed of light

Q&A: Peer-reviewed physics, at the speed of light

Sergio Bertolucci is the director for research and computing at CERN. Over the noise of nearby cathedral bells chiming the hour, iSGTW caught up with him on the steps of the University Building in Uppsala during a coffee break at the EGEE User Forum. We asked him about the spate of new papers coming out from the LHC, and what it all means for science.

iSGTW: We have heard that a lot of papers have already been published in the time since the start-up of the LHC. Is that right?

Bertolucci: Four papers on high-energy physics have already published, and 15 are in preparation as of today, April 14, all based on the collisions that just happened.

One week after the first collisions, the first papers were published electronically. And these were all peer-reviewed.

Sergio Bertolucci during an interview with a member of the Swedish press. *Image courtesy Corentin Chevalier, GridTalk*

iSGTW: That's very fast, compared to the sometimes years-long process of traditional print journals. How was the peer-review done?

Bertolucci: Each experiment is not a monolith, but consists of 3,000 people working in teams on separate projects within it. When a paper is about to be produced, it passes through different teams for internal review, then is reviewed by teams on other experiments, then reviewed by CERN. So, there have been five layers of review before it even goes to the external publication, which has its own, independent review system. The whole system sounds complex, but nonetheless is fast, as seen by one week between observation and publication.

iSGTW: Where have these papers appeared?

Bertolucci: Open-access journals, such as the Journal of High Energy Physics. Our paradigm has always been open access, and this is a good place to practice it.

iSGTW: What was the content of what was published?

Bertolucci: For one thing. The number of charged particles per unit angle at different

iSGTW 16 June 2010

- Feature - Grid makes drug discovery "crystal clear"
- Feature - Back to basics: Q&A on meshes
- Feature - Cloud infrastructure: OpenNebula becomes what you want
- Link of the week - Linux versus *E. coli*
- Image of the week - A child's-eye view of physicists

Announcements

- DEISA Digest now available
- "Why science is cool" video contest, 15 July deadline
- Abstracts now being accepted, 5 Sept deadline, CGW10, Krakow, Poland
- Jobs in grid, 1 new this week

Subscribe

Enter your email address to subscribe to *iSGTW*.

enter email...

Unsubscribe

iSGTW Blog Watch

Keep up with the grid's blogosphere

Mark your calendar

June 2010

http://www.isgtw.org/?pid=1002493

Generic impact measures for e-infrastructures

CERN IT Department

- Impact needs to be expressed differently for each stakeholder (users, funding agencies etc.) but can often be based on the same measures
- Scale
  - Total resources made available to users - significance
  - Geographical coverage – engagement of member states
- Usage
  - Size and diversity of the user communities
  - Frequency and scale of usage
  - Churn - ability to recruit and keep users
- Quality
  - Availability & reliability of specific services and support
- Cost
  - Total cost of services provided

CERN IT Department
CH-1211 Genève 23
Switzerland
www.cern.ch/it