The Grid: An information infrastructure requiring a Knowledge Infrastructure

Dr. Will Venters
Information Systems and Innovation Group
Department of Management
The London School of Economics
w.venters@lse.ac.uk

This research was undertaken as part of Pegasus
EPSRC: Grant No: EP/D049954/1
www.pegasus.lse.ac.uk
➢ The world's most powerful particle accelerator... the Large Hadron Collider (LHC) (from tomorrow!)

➢ Searching for Higgs Boson – "1 person in 1000 worlds, or a needle in 20 million haystacks"

➢ 12-14 PB per year.

➢ 40PB disk, 40PB tape.
Data Collection

- Qualitative longitudinal research through studies of work practices – ethnographic focus.
- Interviews: over 60 with GridPP members, technical experts and users;
- Participant Observation:
  - GridPP meetings
  - WLCG workshops, EGEE user forum, All Hands Meetings.
  - Visits to CERN
- Survey
- Secondary data
Pegasus Overview

- Technical Grid
- People and their Practices
- Knowledge Infrastructure

Socio-technical Grid

www.pegasus.lse.ac.uk
Pegasus Overview

People and their Practices

Technical Grid

Knowledge Infrastructure

Socio-technical Grid

www.pegasus.lse.ac.uk
Pegasus Overview

- Technical Grid
- People and their Practices
- Knowledge Infrastructure

Socio-technical Grid
Structure of this talk

Knowledge Infrastructure

Theoretical Framework

GridPP Case Using Theoretical Framework

Conclusions

Tentative Conclusions
Knowledge Infrastructure

- **Infrastructure**: sewers, telephone, railways – transparent, invisible in use.

- **Information Infrastructure**: standards, technology, networks. Socio-technical, path-dependent, emergent (Hanseth, Dahlbom).

- **Working definition of Knowledge Infrastructure**: ‘The taken for granted social and material structures which enable knowledgeable practice’.

- Knowing is a human trait, but humans are ‘beings-in-the-world’ (Heidegger 1962) – the material around us is both ‘because-of’ and ‘in-order-to’ (Heidegger 1962; Ciborra 2002)
Theoretical Framework

‘Stuff’/Matter
Social Structures
Knowledge
History & Culture

Knowledge Infrastructure

Based on (Venters & Wood 2007 – ISJ)
Knowledge

- Reject notion that knowledge is information++
- Practice perspective (Swan 2005; Schatzki, 2001): Knowledge is not a ‘thing’ to be objectified or captured since it has a tacit dimension (Polanyi 1962).
- Knowledge is social and ‘sticky’ (Brown & Duguid 2001) within social structures.
- Linked to ‘stuff’ (including Boundary Objects (Boland and Tenkasi 1995; Carlile 2002))
- Socially and culturally situated.
Stuff / Matter

- Materiality is an important component of knowing (Orlikowski 2006)
- We live in a world of matter/stuff.
- Scaffolding?
- Stuff ‘affords’ use – inscribed by the designer, and shaped in use.
- That the object of construction shapes and is shaped by the ‘tools’ of its construction.
Social Structures

- Social structures form part of the knowledge infrastructure
- Hierarchy, network, federation.

Communities of Practice (Wenger 1998)

- Emergent
- Context for learning and innovation.
- Trajectory of participation
- Identity – negotiated within a community.
- Boundary Spanners
Knowledge is socially situated and institutional. (Alavi and Leidner 2001).

History drawn upon in making sense (Weick 1995) of the present.

Epistemic cultures (Knorr-cetina 1999)

Maps the path-dependency of knowledge infrastructure
Structure of this talk

Theoretical Framework

GridPP Case using theoretical framework

Tentative conclusions
Knowledge

- Understood Purpose; Higgs boson, and physics aspirations
- Knowledge of action and how to act.
- ‘Humming’ ‘with itself, about itself’ (Knorr-cetina 1999)
- Key ‘knowledgeable’ individuals.
- Language of the project – acronym soup!
- Informal discussions “in the pub”.
- Knowledge of working in a distributed and collaborative fashion
- “you sort of hope that by attending [GridPP meetings] and listening and talking that gradually people start to understand”
- “we are obliged to know a bit about everything in our group just because everyone has short term contracts.”
GridPP2 Goal: To develop and deploy a large scale production quality grid in the UK for the use of the Particle Physics community.
# UK Grid Status at 26 Aug 2008 09:02:07


Blog, XML Version

---

**Resource Broker Summary**

|---|---|---|---|---|---|---|---|

**BDII Summary**

<table>
<thead>
<tr>
<th>RAL: Good</th>
<th>Scot: Good</th>
</tr>
</thead>
</table>

**LFC (Info)**

| RAL: Good | LHCb: 74% | MB/s: 46.0 |

---

Click below on an institute name for a summary for that institute.

See also: LondonGrid, NorthGrid, ScotGrid, SouthGrid

---

### Table: Resource Broker Summary

| Institute | CPU Tot | CPU Free | Jobs Cur | Jobs Wait | Disk Tot | Disk Free | CE | SE | SRM | 24 Hrs | Week | 24 Hrs | Week | SE | CE | ATLAS Tests | CMS | LHCb |
| Trinity Dublin | 692 | 368 | 326 | 159 | 35.5 | 33.1 | W | P | P | 100% | 99% | P | 100% | 100% | S | Any | | |
| Brunel | 409 | 8 | 190 | 1152 | 10.7 | 2.2 | P | P | P | 100% | 100% | P | 100% | 94% | S | dje-grid-40 | | |
| Imperial HEP | 245 | 308 | 64 | 391 | 47.0 | 14.0 | W | P | P | 100% | 78% | P | 100% | 95% | | | | |
| Imperial LeSC | 200 | 303 | 40 | 0 | 0.0 | 0.0 | P | P | P | 100% | 93% | | 0% | 0% | | | |
| QMUL | 1480 | 1709 | 18 | 33236.1 | 29935.0 | P | P | P | 50% | 79% | P | 100% | 100% | | | | |
| RHUL | 400 | 1 | 375 | 305 | 149.0 | 72.2 | P | P | P | 100% | 99% | P | 100% | 96% | | | |
| UCL CCC | 2560 | 1026 | 938 | 0 | 309.1 | 304.0 | M | P | P | 0% | 0% | | 0% | 0% | | | | |
| UCL HEP | 115 | 110 | 5 | 0 | 1.0 | 0.3 | M | P | P | 0% | 12% | | 0% | 0% | | | | |
| Lancaster | 380 | 333 | 47 | 1 | 85.4 | 70.4 | P | P | P | 23% | 78% | P | 100% | 100% | | | | |
| Liverpool | 479 | 407 | 73 | 0 | 34.6 | 30.1 | F | P | P | 100% | 100% | P | 100% | 94% | | | | |
| Manchester | 1740 | 795 | 797 | 2 | 1038.1 | 41.4 | P | P | P | 100% | 100% | P | 100% | 94% | | | | |
| Sheffield | 200 | 188 | 13 | 0 | 11.8 | 8.6 | P | P | P | 100% | 98% | P | 100% | 94% | | | | |

---

**Steve Lloyd's ATLAS Grid Tests**

#### Network Tests Updated

- I have readded the network tests as agreed at a recent meeting and added a new one for tunning the test. This new test is based on the previous one and involves copying files from the Tier 1, to the Tier 2, then to the Tier 3. It is a way to test the reliability of the ATLAS tests for the ATLAS tests. I have also added a new test for the ATLAS tests which is to test the reliability of the ATLAS tests for the ATLAS tests. It is a way to test the reliability of the ATLAS tests for the ATLAS tests.

#### LHCb Tests

- I have added the LHCb tests to the collection. These tests are actually SAM tests for the LHCb tests. They are used to test the LHCb tests for the LHCb tests. It is a way to test the reliability of the LHCb tests for the LHCb tests.

#### Local problems

- The bad news is that we have a problem with the ATLAS tests. The tests are not running as expected and the problems are being investigated.

---

Steve Lloyd
Social Structures

- “…the original proposal for GridPP was set up deliberately to make it look like an experiment. This whole idea of a collaboration board for instance comes out of the idea of what an experiment does.”

- Meaning, identity located in particle physics practices.

- Network rather than hierarchy: Virtual, federated, overlapping and inter-connected.

- Long tradition of global collaborations (Traweek 1988).

- Computer Scientists will “Physicists will come up with the most hacked solution in the world… but it will work”

- Natural selection: “the cream comes to the top. Things that work win out, and that’s how we worked it”
History and Culture

- Fluid architectures; EGEE middleware is modularized and released gradually.
- Down-to-earth and creative approaches embedded in particle physics tradition (Traweek, 1988).
- Stories of past experiments.
- Academic publication as retrospective narrative.
- Long history of success in computing; Cray X-MP; WWW; Open source; linux; Gigabit networks...
- Mood –reflective; jovial; supportive ...
Structure of this talk

Knowledge Infrastructure

Theoretical Framework

GridPP Case

Analysis using theoretical framework

Conclusions

Tentative conclusions
Very Tentative Contributions

1) Usability of the Grid requires an accommodation of knowledge infrastructures between various relevant groups:
   1) ‘Users’
   2) ‘Maintainers’
   3) ‘Developers’

2) Knowledge Infrastructure – emergent and path dependent and hence resistant to ‘design’

3) Stuff is complex and political - Provision of technical infrastructure of Wiki, Blog and web servers may be beneficial but...

4) Consideration of **Boundary Objects** between communities in terms of social structures, history, knowledge.