

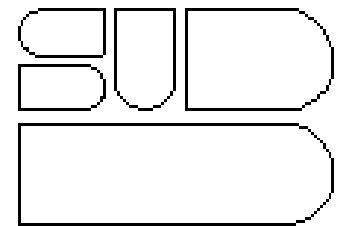
nestor II: e-Science and Preservation – A Perfect Match?

iPRES

International Conference on Preservation of Digital Objects



Heike Neuroth
Göttingen State and University Library (SUB)
Head Research & Development Department
neuroth@sub.uni-goettingen.de





ToC

- **nestor II**
- e-Science
- Grid
- Perfect Match?



nestor II

➤ **Network of Expertise in Long-Term Storage of Digital Resources:**

- Network of expertise
- Center of competence in Germany in digital long-term preservation (digital curation)
- Information platform in all activities in German speaking countries (DE, AT, CH)
- Liason with international developments, projects ...
- Expert database
- ...





nestor II

- Partners from libraries, museums, archives, research institutions, technology providers ...
 - nestor I: 2003 – 2006 (ca. 1 Mio Euro)
 - nestor II: 2006 – 2009 (ca. 1.8 Mio Euro)
- Focus on sustainable organizational model
 - Thematic concentration on education/teaching, **e-Science/Grid**, standards/standardization activities on national and international level (certification of trusted repositories ...) ...





ToC

- nestor II
- **e-Science**
- Grid
- Perfect Match?





e-Science

"e-Science is about ***global collaboration*** in key areas of science and the next generation of ***infrastructure*** that will enable it." *(John Taylor)*

- **Infrastructure** for research, science, and education
- **Technical Integration** (Grid, Middleware)
- **Knowledge Management** (cooperation, collaboration)





The e-Science Journey ...

➤ *... starts with*

- local and isolated applications (silo)
- static environments
- lots of manual and redundant effort

➤ *... aims to deal with environments that are*

- information-/data-intensive
- distributed and dynamic
- multi-disciplinary



e-Science Tasks

(Christine Borgman, e-Research)

➤ e-Science

- manage the “data deluge”
- facilitate communication and collaboration
- enable reuse (and remixing) in an open environment

➤ Information infrastructure

- establish a value chain of information
(relationships, context; early reuse and collaboration)
- infrastructure FOR information
rather than OF information
- ...



e-Science Opportunities

- Shared responsibilities,
Web of contributors (external parties),
→ harmonized workflows to orchestrate
- Open infrastructure
→ remixable services
- Collaborative environment
→ reusable resources



e-Science & Preservation

- Harmonized workflows
e.g. NDIIPP
- Remixable services, preservation infrastructure
e.g. PANIC (AU), PLANETS (EU), CASPAR (EU)
- Interoperability
e.g. Shibboleth, Grid Application Toolkit, Web Services
- Manage data deluge
e.g. LOCKSS





ToC

- nestor II
- e-Science
- **Grid**
- Perfect Match?



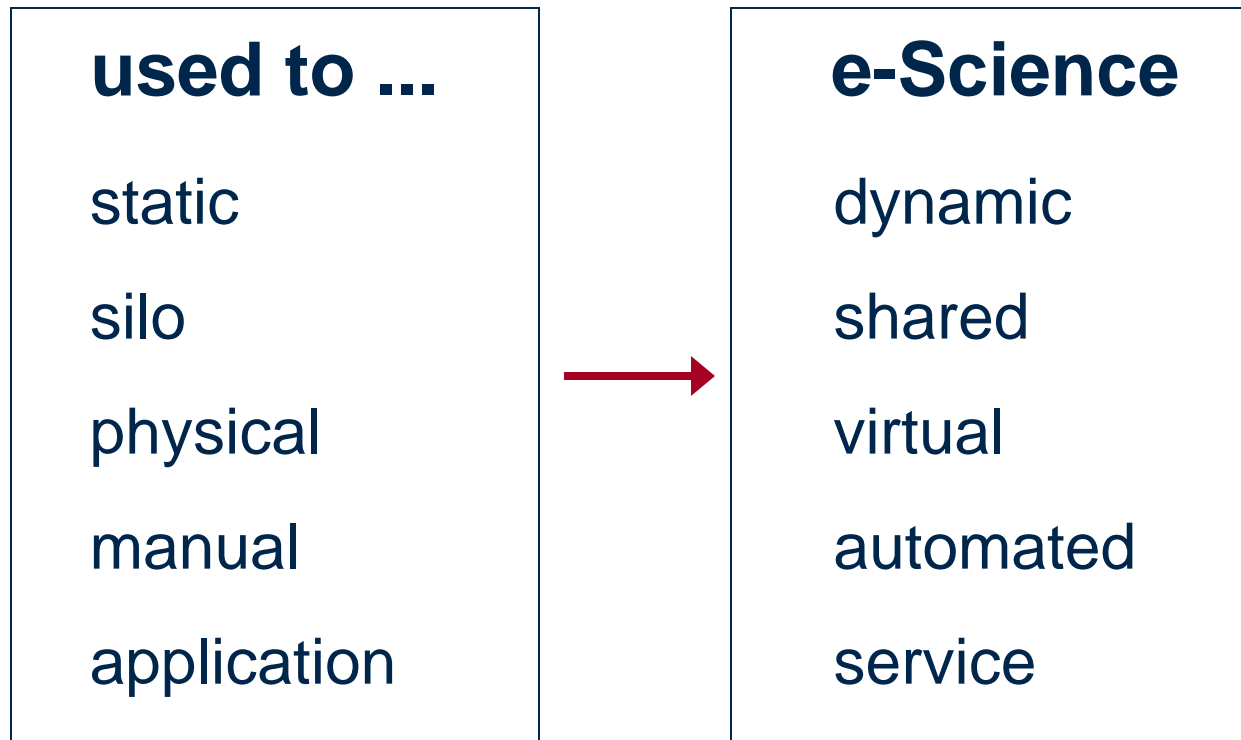
Grid (-Technology)

Enable communities (“virtual organizations”) to share geographically distributed resources as they pursue common goals -- *assuming the absence of ...*

- central location
- central control
- ...

(The Anatomy of the Grid: Enabling Scalable Virtual Organizations. I. Foster, C. Kesselman, S. Tuecke. International J. Supercomputer Applications, 15(3), 2001)

Technical Challenges



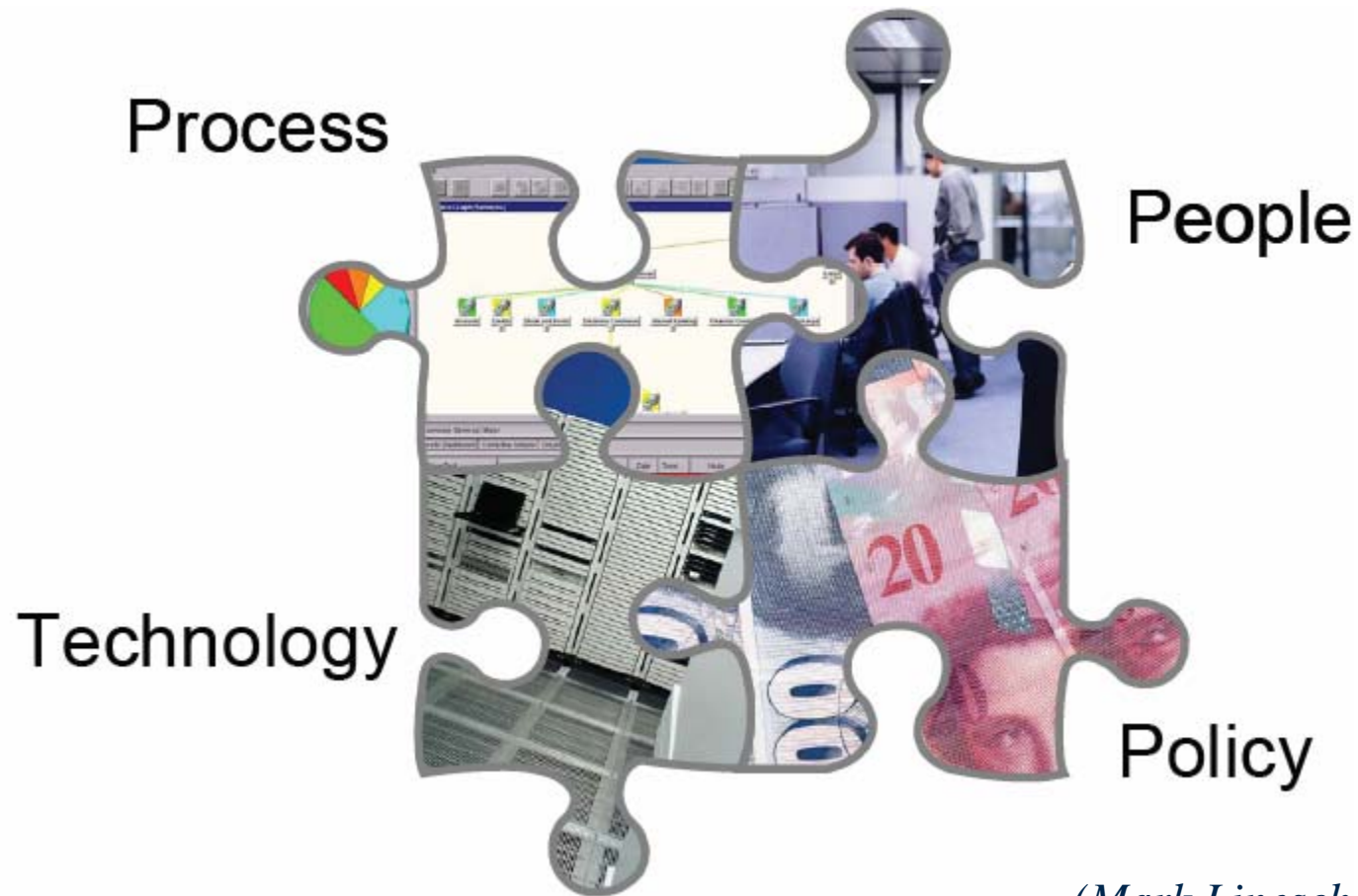
Silo Architecture



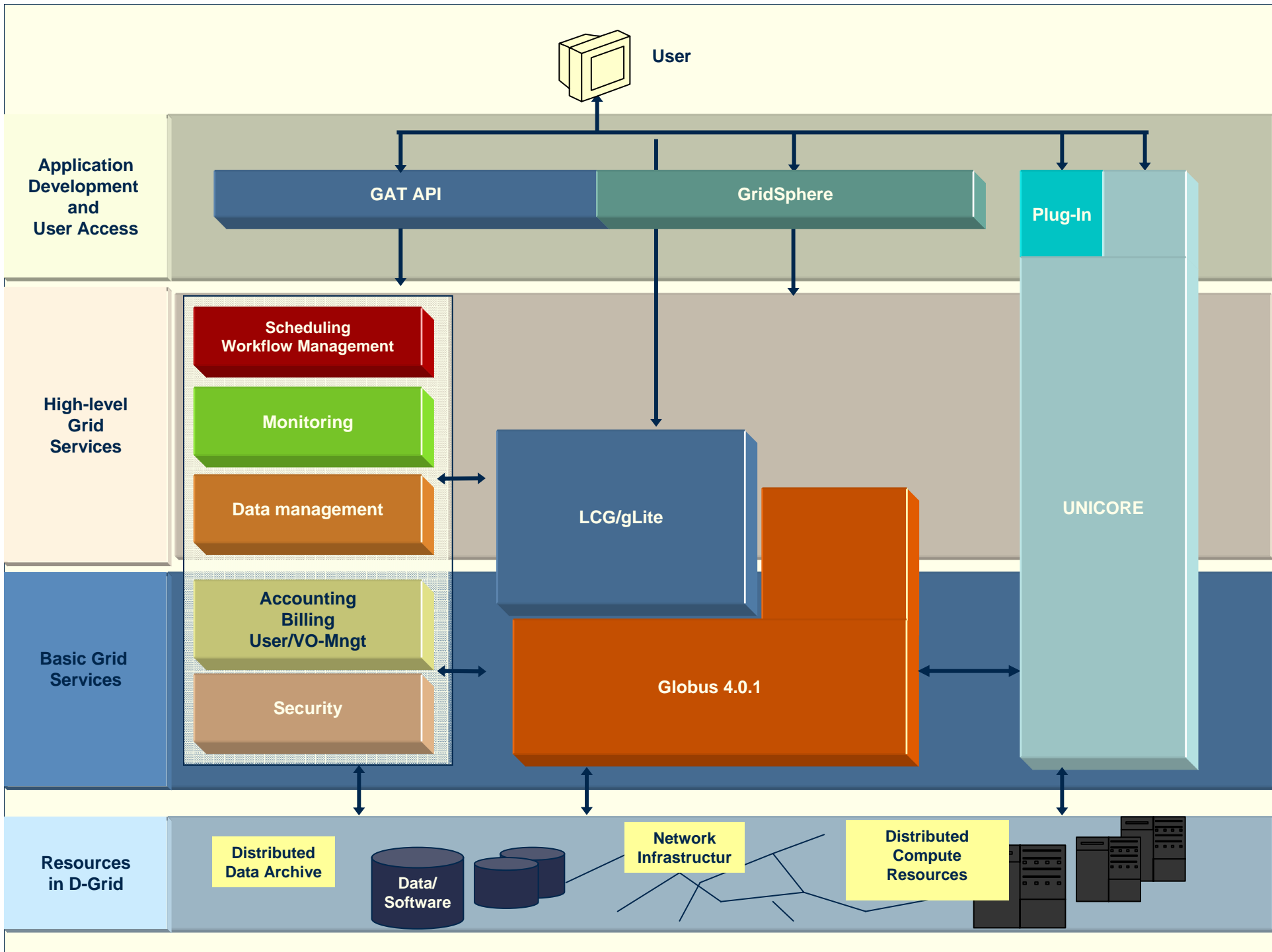
**Service Oriented
Architecture (SOA)**



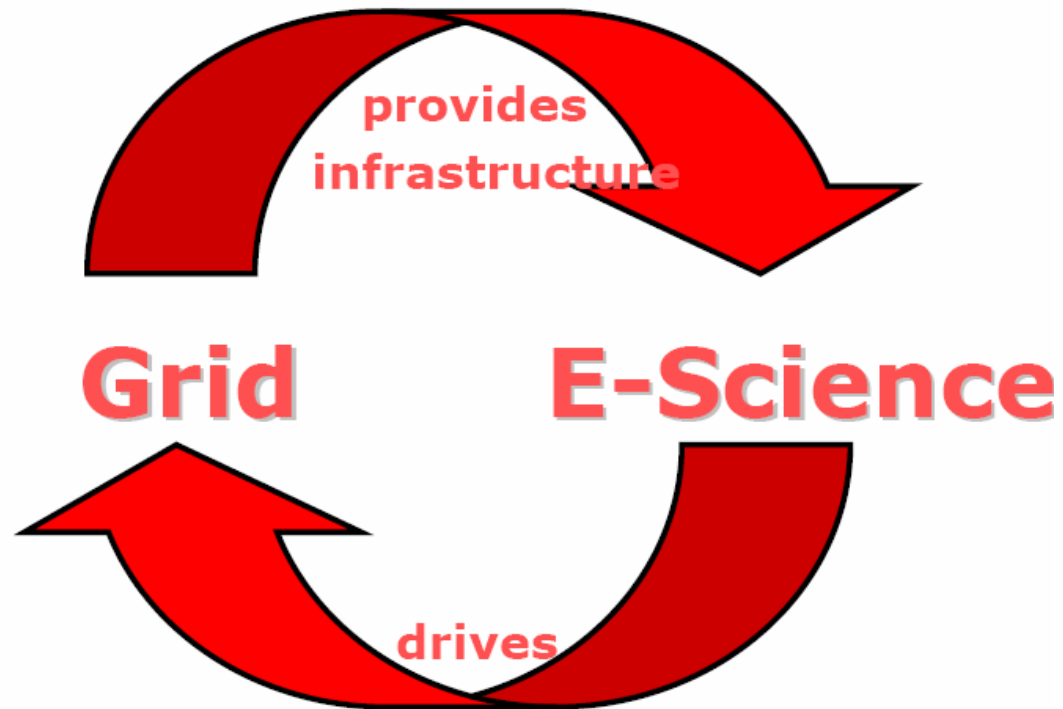
Challenge of Integration



(Mark Linesch; GGF, HP)



e-Science & Grid



(Alexander Reinefeld; ZIB Berlin)



e-Science Initiatives

- UK e-Science Programme
<http://www.rcuk.ac.uk/escience/>
- US Cyberinfrastructure programme
<http://www.nsf.gov/od/oci/reports/toc.jsp>
- AustrianGrid
<http://www.austriangrid.at/>
- BIG GRID, the Dutch e-Science Grid
<http://www.dutchgrid.nl/>
- German D-Grid Initiative
<http://www.d-Grid.de/>



German D-Grid Initiative

- Announcement of German Ministry for Research and Education (BMBF): Call for proposals in the areas of:

- **Grid Computing**
 - E-Learning
 - **Knowledge Management**
- e-Science*
- 

- 2005 – 2009: ca. 100 Mio Euro
- Over 100 institutions (including World Data Centers in Germany) are involved

Vision of D-Grid

"... New quality of digital scientific infrastructure which enable our globally connected scientists to **collaborate** on an international basis, and exchange information, documents, and publications about their research work in real time, and guaranteeing efficiency and stability even with huge amounts of data from measurements, laboratories, and computational results ..."



TextGrid

e-Humanities:

TextGrid - Modular platform for collaborative textual editing

A community Grid for the Humanities

<http://www.TextGrid.de/>





ToC

- nestor II
- e-Science
- Grid
- **Perfect Match?**



Perfect Match?

TextGrid and digital preservation:

- **Storage Grid:** about ?? TByte/PByte on data, e.g. digital editions, retrodigitized material, digital dictionaries ...
- **Service Grid:** tools for text processing like tokenizer or metadata annotation, text retrieval, link-editor ...
- **Match:** Preservation in Grid Environment – storage and services!
 - *Example: Fedora* – IR integrated into Grid environment – in a sense of *Storage* (inside Grid/Middleware or interface between both)

Perfect Match?

nestor II and expert studies on:

- Long-term preservation meets e-Science in the field of scientific raw/primary data ...
- Potential synergies between Grid/e-Science technologies and preservation technologies ...
- Common standards in the field of e-Science and preservation, need of activities ...
- ...
 - *Example: LOCKSS* – integrated into *Service Grid* (inside Grid/Middleware as a tool for making content redundant)



Perfect Match?

Fedora – Service Grid?

**Preservation Integrity Service
Preservation Alerting Service**

...

Sandy Payette
Cornell Information Science





Perfect Match?

Storage Grid?

**"Superman returns"
200 TByte !!!**

...

Carl Fleischhauer
LoC





Perfect Match?

Other ideas?



Thank you for listening!

iPRES

International Conference on Preservation of Digital Objects

Questions, comments??



Heike Neuroth
Göttingen State and University Library (SUB)
Research & Development Department
neuroth@sub.uni-goettingen.de

