Sustainability in Grid-Computing

Christian Baun

Sustainability in Grid-Computing

- Important topics in Grid-Computing during GridKa-School 2007:
  - Grid applications
  - Grid middleware systems
  - Grid business models
  - Usability
  - Involvement of industry
  - …
- All these topics play an important role while archiving sustainability!

The D-Grid Project

- The aim of D-Grid is to build and run a reliable and **sustainable** Grid Infrastructure for e-Science in Germany
- 19 Community projects
  - Different scientific fields
  - Variety in manpower and financial possibilities
- 1 Integration project
  - Builds up the infrastructure
  - Integrates the developments from the different community projects in one common D-Grid platform
- Design parameters:
  - D-Grid 1: 2005 – 2008
  - D-Grid 2: 2007 – 2010
  - 24 Sites
  - Funding: 60 M€

Why do we need Sustainability? – Motivation (1)

- Typical progression in a scientific project:

Start of the project → **Duration of the project** → **End of the project (out of money)** → **bye bye**




➔ Not sustainable!

Why do we need Sustainability? – Motivation (2)

- Our goal:

Start of the project

Duration of the project

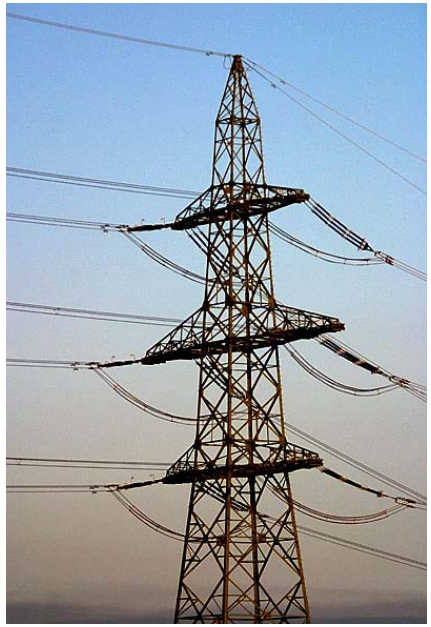
Long-time availability of

- infrastructure and
- services

➔ sustainable!

Why do we need Sustainability? – Motivation (3)

- Examples for sustainable infrastructures



- **The sustainability of an infrastructure with scientific purposes depends on how it becomes a normal element of the scientific process!**

A Grid Infrastructures can be seen as a Stock Market

User of services

Grid infrastructure

Resource provider

Properties of Sustainability in Grid-Computing (1)

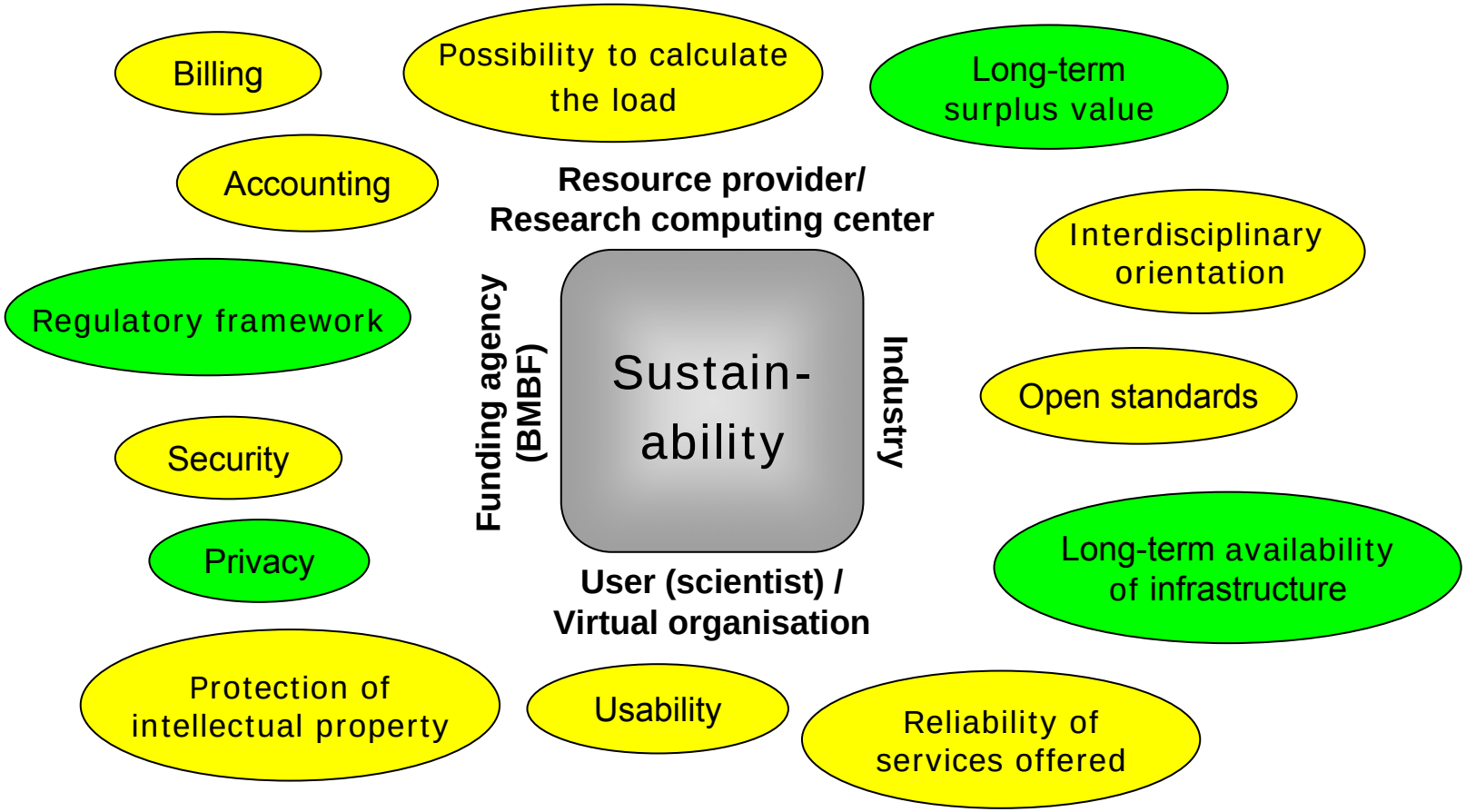
- Long-term availability of infrastructure and services
  - Users (scientists) need long-term available tools for their work
  - Industry demands for long-term available standards (markets)
- An surplus value for all participants exists
  - Surplus value does not only mean financial profit (renting resources)
  - Surplus value means also new knowledge is gained and scientific collaborations are enhanced
  - Surplus value means the daily work of users (scientists) get easier and more effective
- Long-term cost-covering operation with permanent re-investments in the infrastructure
  - Existence of a realistic business plan
  - Long-term and secure financing from more than one source

Properties of Sustainability in Grid-Computing (2)

- Legal Security
  - Analysis of legal Framework needs to be done
  - Nobody wants to violate law and get sued
- Collaborations with similar objectives
  - No need for silo projects
  - Pushing the integration in international projects
- Helping the users with their daily work
  - Users (scientists) need good tools and infrastructures
  - Advancement of the research location

Roles and their Needs

- To achieve sustainability we need to identify the stakeholders involved and their needs and find ways to satisfy them:



Privacy (1)

- Most European nations:
  - **Privacy** is considered **highly important**
- USA, Japan and other important developed countries:
  - **Privacy** as something **less important**
- International collaborations in Grid computing increase
- Distributed IT-infrastructures are getting more complex
- Difficult to guarantee privacy

Privacy (2)

- Example: Personal data collected for statistical purpose should be processed in a Grid
  - Mostly incalculable how many nodes are storing that data
  - Difficult to say where the nodes are exactly located
  - Grid service providers have to guarantee that personal data, distributed over the Grid, are never duplicated on the nodes
  - Problem: Data on nodes are typically stored in backups
- Grid service providers have to make sure that all personal data is non-recoverably erased from the nodes after processing
- The users have the right to revoke their permission for collection and processing their personal data any time
  - Grid service providers need to have the capabilities to erase personal data of single users from the Grid resources at any time

Privacy (3)

- Sometimes personal data needs to be transferred to resources in other countries without equal or stronger privacy laws
  - Difficult agreements between resource providers and Grid service providers are required
  - See safe harbor acknowledgement between the European Union and USA
  - Such a proceeding is very complex and in a huge and dynamic Grid infrastructure neither realistic nor feasible

Privacy: Solutions (1)

- Making the personal data anonymous
  - Modifying the data it in a way that assigning it to allocatable persons is impossible or requires an extraordinary amount of time, cost and manpower
  - Nearly impossible to restore the original data
  - No more problems with privacy
  - But: The **quintessence** of the data is **lost**

Example:

Heike Hansen (Hamburg)	→	Mrs. A from X
Martin Müller (Mannheim)	→	Mr. B from Y
Helmut Haffner (Heidelberg)	→	Mr. C from Z

Privacy: Solutions (2)

- Altering the personal data
  - Data can only be assigned to allocatable persons with a code or a cryptographic method
  - After altering the data with a key or hash it is easy for the user to restore the personal data
  - If encryption is strong, no more problems with privacy
  - Problems: **Altering** the personal data is **not always possible** or useful

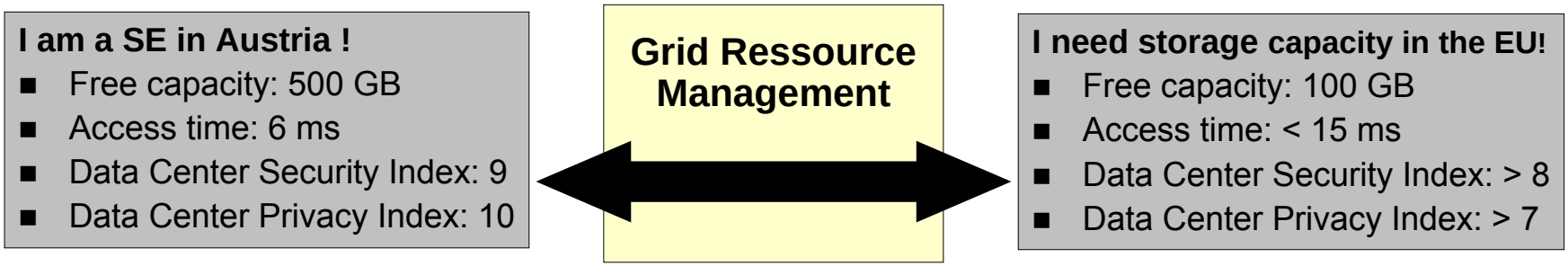
Example:

Heike Hansen (Hamburg)	→	Ifjlf Ibotfo from Ibncvsh
Martin Müller (Mannheim)	→	Nbsujo Nvmmfs from Nbooifjn
Helmut Haffner (Heidelberg)	→	Ifmnvu Ibggofs Ifjefmcfsh

Privacy: Solutions (3)

- Levels of privacy and security of resources and their environment
  - Periodically rating by transparent and standardized audits
  - Logging any access to user data
  - Disqualification:
    - Cannot guarantee privacy
    - Produces significant overhead

Example:



Privacy: Solutions (4)

- Privacy and the Grid:
  - No satisfactory solutions existing
  - Working solutions get more and more important with growing participation of industry in grid projects
  - Still much work to do

- In the contract between Grid service providers and resource providers it has to be made clear:
  - What kind of measures the parties are taking to ensure privacy
  - Who is to blame if personal data is getting stolen

Legal Topics: Much Work to do

- Legal topics of Grid-Computing
  - Currently not well-investigated
  - No court decisions exist
  - Many aspects of Grid-Computing are not new
  - Huge similarities with outsourcing of business processes and with web hosting offers of Internet service providers or Internet web hosting providers
- Problems arise in international projects
  - Who is to blame if a poor programmed grid-job causes a defect
  - Different legal systems have to be considered

What are the Costs for the Infrastructure?

- The biggest part of the total costs are personnel costs
  - Personnel is needed for running and improving infrastructure, user support and possibly software engineering
  - In Germany the employer of a scientist needs to budget € 80.000 through € 100.000 per employee per year
    - (Includes: salary, insurances, equipment, fees for training courses, …)
- Hardware for running an core-grid-installation for testing new software versions
  - The hardware needs to be **reinvested** every 3 years

What are the Costs for the Infrastructure?

- Additional costs
  - Costs for electrical power and cooling
    - With water cooling: approximately € 4 per watt per year
    - With air cooling: approximately € 5 per watt per year
  - Additional costs per server
    - Rack and storing position
    - Administration and batch licenses
- Helpdesk-Tool (Trouble Ticket System)
  - Open Source Tools: OneOrZero, XOOPS, Request Tracker, …
  - Proprietary Tools: Remedy, …
    - Purchase costs depend on product and number of users
    - Support contract: approximately 15% of purchase costs per year
- Marketing: flyer, poster, conference fees, hosting workshops, …

Software Licenses (1)

- Fear of complications because of incompatible software licenses in the beginning of D-Grid
- Questions asked:
  - Is it possible to mix software under different Open Source software licenses and proprietary software licenses?
  - Is it allowed to collect all needed Grid software und distribute it on one CD?
  - What software licenses give us the benefits of Open Source and leave the door open for industry?
  - What Open Source software license is suited best for developing Grid applications?
- The most popular Open Source software licenses were investigated for their appropriateness in Grid environments
  - GPL, LGPL, Apache License 2.0, Mozilla Public License, Q Public License, …
  - Result: Apache License 2.0 is best suited

Software Licenses (2)

- Apache License 2.0:
  - Related to the BSD license
  - Non-viral: derived software is not required to be redistributed as Open Source
  - Software linked to software under the terms of the BSD or Apache License 2.0 does not need to have the same software license
  - Securing the project sovereignty while protecting the project name
  - Short and easy to understand
- Most common Grid software uses the Apache License 2.0 or another BSD style license:
  - **Unicore**: BSD license
  - **Globus Toolkit** ≥ 4.0.1: Apache License 2.0
  - **gLite**: EGEE Software License. Switch to Apache License 2.0 is planned
  - **GridSphere**: Apache License 2.0
  - **Shibboleth**: Apache License 2.0
  - **VOMS**: EU DataGrid Software License (EDG). BSD style license
  - **iRODS**: BSD style license

Next Steps for D-Grid

- Actually D-Grid develops a business model
- A German Grid support facility will be installed. Its tasks are:
  - Running the core services (Monitoring, Security, Billing, …)
  - Support for users and resource providers (Helpdesk, Phone)
  - Consulting of developers and resource providers
  - Consulting of resource providers in legal topics
- Costs of German Grid support facility depend on the services and number of customers
  - Major part of costs are personnel costs
  - Also environmental costs and costs for running a core-grid-installation
- The customers (users and resource providers) have to pay for the services they consume and will finance the German Grid support facility

Thank you
for your attention!