Bridging between infrastructures in the polder

Collaboration amongst e-Infrastructures at the technical and policy level – a (biased) personal view

David Groep,

Nikhef and the Dutch National e-Infrastructure

TTC meeting 2015-11-23

Foto: Reinier Sierag, CC-BY-NC, https://www.flickr.com/photos/sierag/

‘Operational aspects of service provisioning across e-Infrastructures’

The term **e-Infrastructure** is used to indicate the **integrated ICT-based Research Infrastructure in Europe**.

Of course such an infrastructure builds on many ICT components that have been around for quite a while, such as networks, supercomputers and storage. There are many interdependencies between these components, so their future should be planned coherently. The e-Infrastructure viewpoint allows to join and fit all interrelated infrastructures together and start think of them as a system – and optimise not for each individual part, but for the whole.

e-Infrastructures Roadmap 2005

e-IRG inaugural meeting, Dublin 2004 – a wide range of policy makers from across many infrastruc

Cross-infrastructure services are not new

… even if success at times is incidental rather than by construc

… it may be two steps forward, then one step back …

*… since shortcuts are frequent as **It Has To Work**™ – right now*

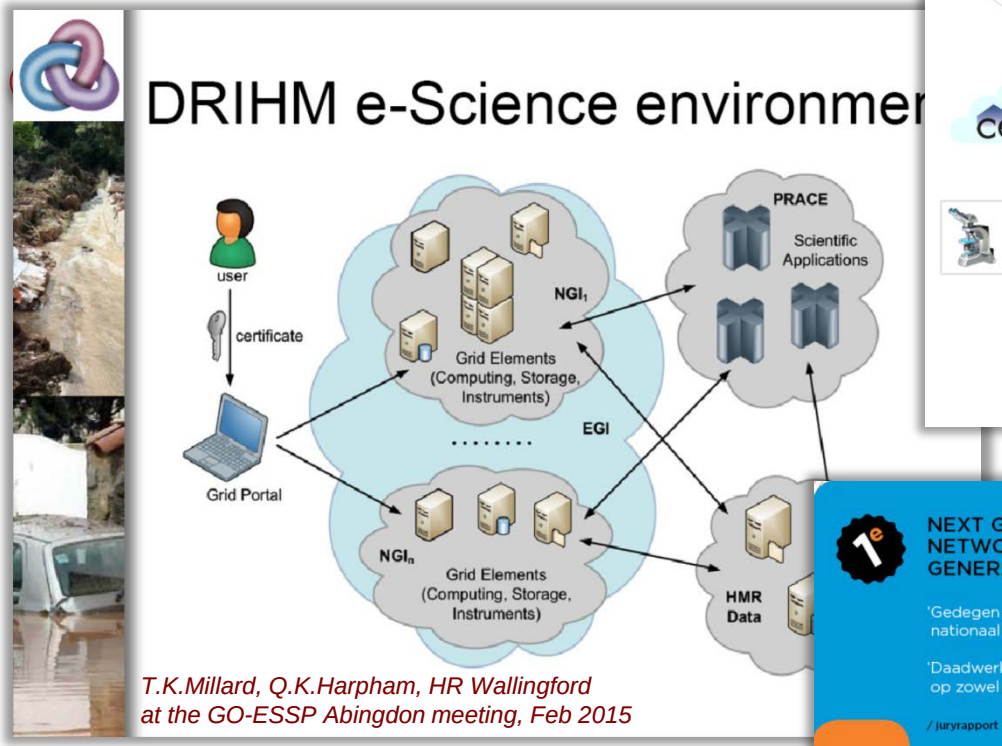
interconnection of complex heterogeneous High Performance Computing systems in a cross-organizational and cross-domain manner.

'designing middleware to support a particular kind of distributed application, known as a close-coupled application, and at assessing the middleware using industrial applications across the European GEANT network.'

Nikhef

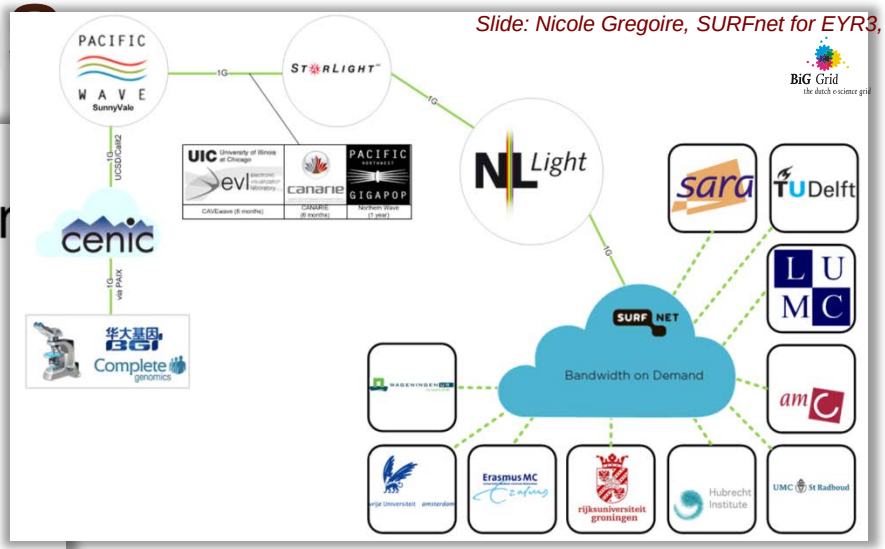
Background picture: Willibrord's processing, Echternach, LU
Photo: David Edgar, CC-BY-SA, from WikiMedia

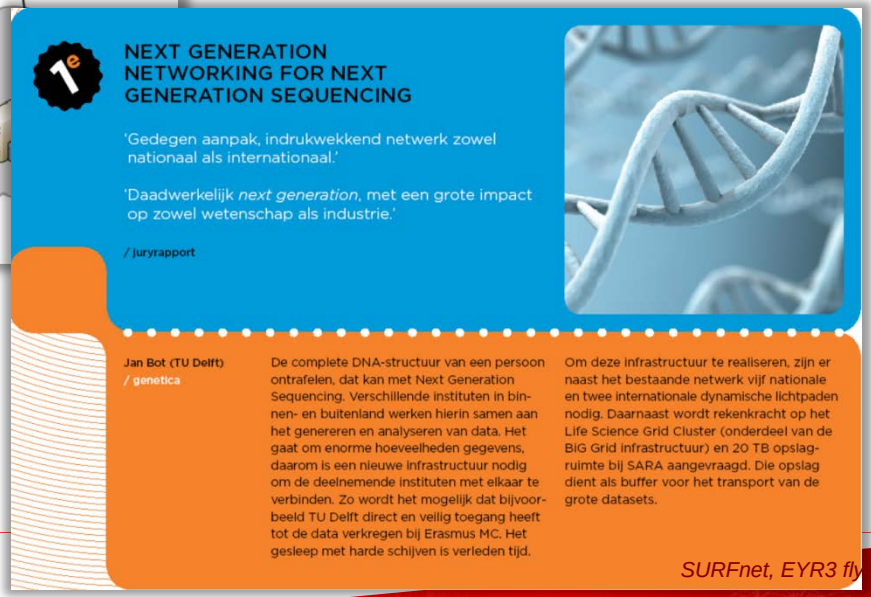
Shared applications



T.K.Millard, Q.K.Harpham, HR Wallingford
at the GO-ESSP Abingdon meeting, Feb 2015

Slide: Nicole Gregoire, SURFnet for EYR3, CC



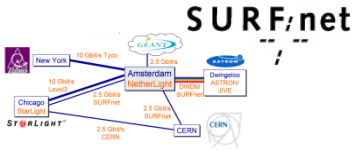


SURFnet, EYR3 flyer

David Groep
Nikhef
Amsterdam
PDP & Grid

Common technology and direction?

- Where needed & obvious, common solutions were applied



- But specific use cases and models resulted in divergent interfaces ar wares for (higher-level) se

David Groep
Nikhef
Amsterdam
PDP & Grid

- Driven by the need to Make It Work Now
  - the e-Infrastructures today are the result of an emergent system
  - there was not the luxury to ‘play around’ for a decade
  - operational production infrastructure for research (like LCG)

Shared users – shared concerns

Emerging *user-fronting* multi-domain infrastructures faced similar issues in Authentication, Policy, and Authorization

Authentication and identification

Accounting and metering

Application and Granting

Middleware Protocols

Acceptable Use

Authorization, access control, and provisioning

Coordinated Resource Sharing

‘EUGridPMA’

‘UR-WG’

‘IPG’

‘JSPG’

‘IGTF’

‘OGF’

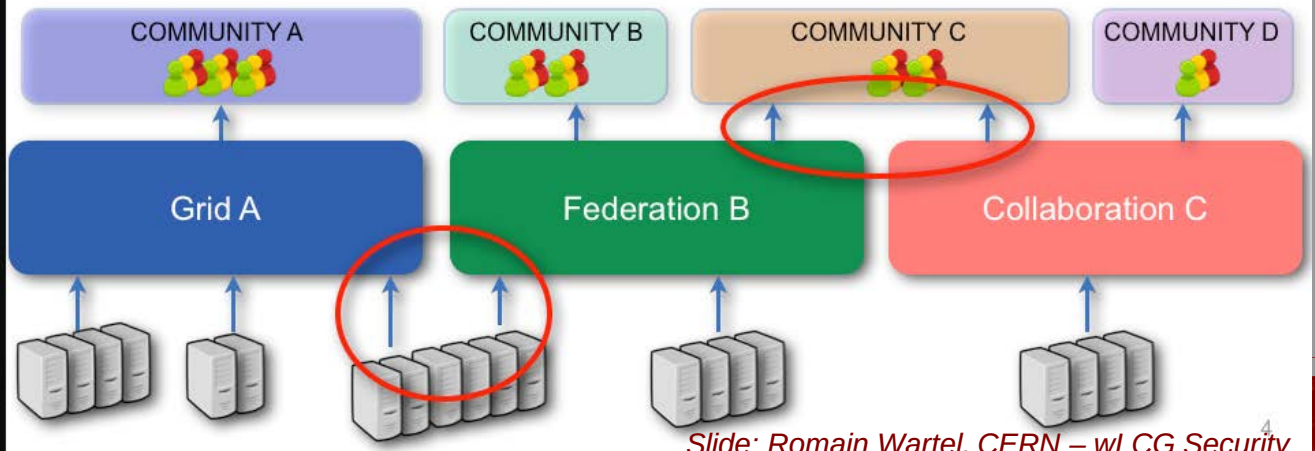
‘SCI’

‘e-IRG White Papers’

Shared users spurring practical, *global*, collaboration …

Federation = BIG attack surface

- Increase in collaboration means
  - Shared users
  - Shared resources
- Collaboration => incident propagation vector



Slide: Romain Wartel, CERN – wLCG Security

David Groep
Nikhef
Amsterdam
PDP & Grid

Common authentication of users … a ‘bottom-up’ approach may just work

* First meeting March 2003 at GGF 7 in Tokyo
* Will continue to meet at GGF conferences. Next meeting Seattle.
* Will work on forming the Grid Policy Management Authority: GRID PMA.org
  - Develop Minimum operational requirements - based on EDG work
  - Develop a Grid Policy Management Authority Charter
* Representatives from Major Grid PMA
  - European Data Grid and Cross Grid PMA
    * 16 countries, 19 organizations
  - NCSA Alliance
  - Grid Canada
  - DOEGrids PMA
  - NASA Information Power Grid
  - TERENA
  - Asian Pacific PMA
    * AIST, Japan
    * SDSC, USA
    * KISTI, Korea
    * Bll, Singapore
    * Kasetsart Univ., Thailand
    * CAS, China

‘TeraGrid/XSEDE US HPC’

You know them ☺

David Groep
Nikhef
Amsterdam
PDP & Grid

Common AUP to enable collaboration

E-Infrastructures, being 'end-user facing' and independent of 'home' administrative domains, need an AUP to inform and bind users – and to foster collaboration, all major international 'e-' and 'cyber' infrastructures agreed on a single common basis: the 'Taipei Accord' (2005)

5. Logged information, including information provided by you for registration purposes, shall be used for administrative, operational, accounting, monitoring and security purposes only. This information may be disclosed to other organizations anywhere in the world for these purposes. Although efforts are made to maintain confidentiality, no guarantees are given.

"The e-IRG notes the timely operation of an EGEE/LCG/OSG group working on a common Acceptable Usage Policy for multidisciplinary Grid infrastructures and it expresses its satisfaction and support for the current draft AUP proposed in this white paper and would like to encourage the group to consolidate it asap. It is felt that such an effort would greatly promote pan-European resource sharing for eScience."

e-IRG
e-Infrastructure Reflection Group

David Groep
Nikhef
Amsterdam
PDP & Grid

EGI, wLCG, OSG, PRACE, XSEDE

e-IRG White Paper created a fruitful collaboration

tacar

eugridpma

Authentication policy development

- The e-IRG notes the timely operation of the EUGridPMA in conjunction with the TACAR CA Repository and it expresses its satisfaction for a European initiative that serves eScience Grid projects.

Authorization policy development

- The e-IRG endorses the principle of the EUGridPMA and TACAR. The e-IRG welcomes this development which positions Europe in the forefront of Grid and eScience interoperability. The e-IRG strongly encourages the EUGridPMA / TACAR to continue their valuable work and recommends that they be supported by the relevant EU / national projects and agencies.

e-IRG white paper 2004 (IE)

IGTF
Interoperable Global Trust Federation
AP|EU|TAG

Infrastructure 'Relying Party' members in 2005 and l

eGee
Enabling Grids for E-sciencE

XSEDE
Extreme Science and Engineering Discovery Environment

PRACE

SEE-GRID

TeraGrid

Distributed European Infrastructure for Supercomputing Applications

egi

Open Science Grid

PRAGMA

TERENA

David Groep
Nikhef
Amsterdam
PDP & Grid

A mandate that has been there all along

[In the] Council of the European Union … a substantial majority of delegations supported the following Presidency conclusions: …

COUNCIL OF
THE EUROPEAN UNION

EN

E.15. UNDERLINES that the present actions to support existing research infrastructures, integrating activities and trans-national access to facilities should be continued and reinforced …
Including e-Infrastructures, i.e. GEANT infrastructure projects for the interconnection of electronic research networks and GRID architecture empowered infrastructures.

PRESS RELEASE

2624th Council Meeting

But working together takes time – let's look at the Dutch example

Photo: Teun Spaans – WikiMedia commons CC-BY-

Dutch National e-Infrastructure today

operational partners

SURF SARA SURF NET

university of groningen

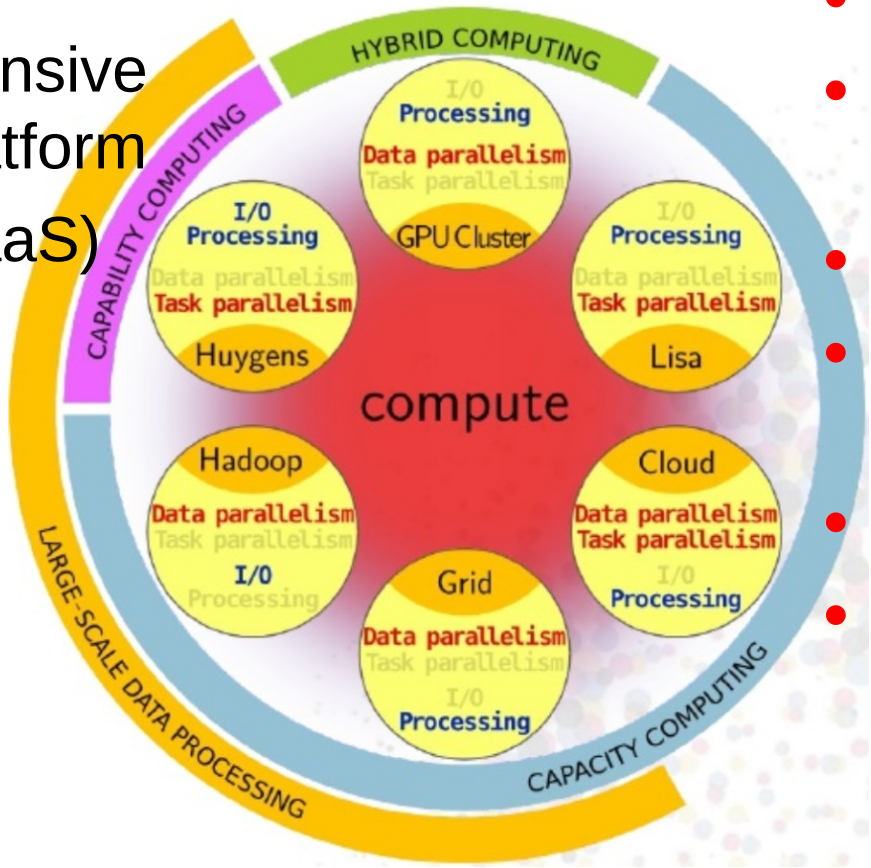
Nikhef

coordinated by SURF

David Groep
Nikhef
Amsterdam
PDP & Grid

DNI data and compute services

- ‘Grid’ data intensive processing platform
- HPC Cloud (IaaS)
- Hadoop
- GPU
- BeeHub
- SURFdrive



- Data Ingest Service
- High-throughput storage
- Long-term archival
- “Lisa” low-latency compute cluster
- Carthesius T1 HPC
- HTC IaaS data processing Cloud (several, est: 2016)

David Groep
Nikhef
Amsterdam
PDP & Grid

Graphic: SURFsara, Evert Lamm

An Interesting Journey: 2000…2015…2018+

2000

vl·e

virtual laboratory for e-science

SURFnet

NCF

BiG Grid

the dutch e-science grid

2008

ICTRegie

Nationaal regieorgaan voor ICT-onderzoek en -Innovatie

Ministerie van Onderwijs, Cultuur en Wetenschap

2013

NWO

Nederlandse Organisatie voor Wetenschappelijk Onderzoek

SURF

2015

netherlands eScience center

SURF SARA

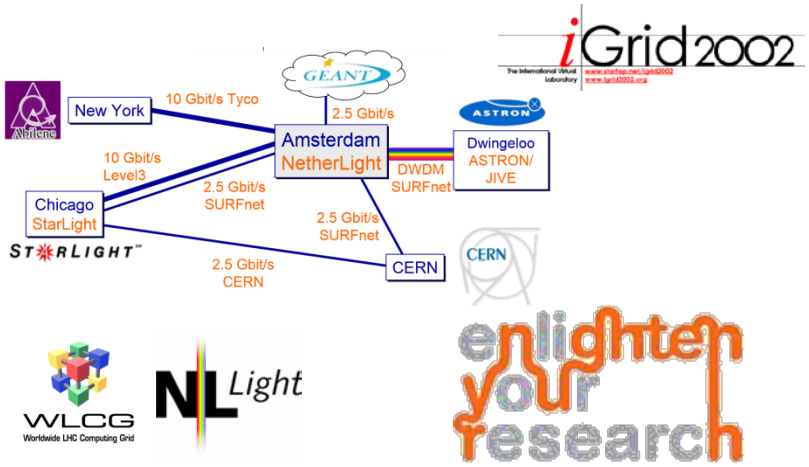
SURF NET

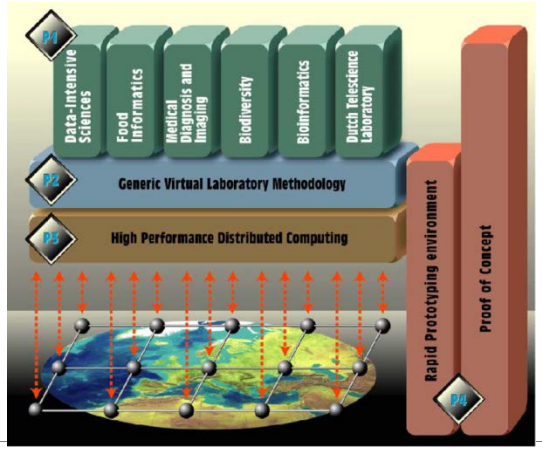
SURF MARKET

Nikhef

Convergence needs concrete use cases

Especially in the ’90s and early millenium SURFnet was a highly ‘researchy’ NREN supporting basic research through GigaPort – but engaged real end-users at the same time





Virtual Laboratory for e-Science: a joint experiment driven by users, but set up to establish a scalable, reliable infrastructure

vl·e virtual laboratory for e·science

food informatics, medical imaging, biodiversity, bio-informatics, ‘tele-science’ lab, data-intensive sci

David Groep
Nikhef
Amsterdam
PDP & Grid

Development of an Infrastructure

In the Netherlands, throughout VL-e, EU DataGrid NL, and NCF 'NL-Grid', although driven by (mainly) Nikhef and UvA use cases, the target was explicitly to build a generic infrastructure for research and innovation.

NWO
Nederlandse Organisatie voor Wetenschappelijk Onderzoek

BiG Grid
the dutch e-science grid

It yielded BiG Grid – the Dutch e-Science Grid – following a intense (but messy) episode, a foundational project *driven by actual end-users* but giving a prominent role to operational infrastructure partners (esp. SARA, but also RUG-CIT and at the Philips Research HTC)

David Groep
Nikhef
Amsterdam
PDP & Grid

BiG Grid – the Dutch e-Science Grid

~ 30M€ data and compute infrastructure ‘pathfinder’ investment project

- “Best project I’ve ever been involved with” (and yielding ~ 1MEur/page)
- Combining HTC/Grid, HPC co-funding, data services, and infra innovation (cloud)

- Hardware (80%)
- **Application support** (20%) with a focus on LS + humanities
- Almost entirely application driven: Nikhef + NBIC + NCF (on behalf of the HPC users) but pushing for a service fostering a national e-Infrastructure
  We had seen what happened elsewhere: if we want to keep getting funding for HEP and LS to come from generic sources, we have to present and run it as a generic service. For a single domain one will never get generic funding – so you need to diversify, and SARA was the obvious nucleus
- With industrial R&D involvement for both applications and operations

But VL-e, GigaPort, and BiG Grid were all impulse-financed – and not sustainable. A radically new approach was needed …

David Groep
Nikhef
Amsterdam
PDP & Grid

ICTRegie rapport

- In 2008 we – application domain, infrastructure providers, networks – collectively convinced the government that "*maintaining the recognized high level of scientific output in the Netherlands requires an excellent ICT infrastructure. Even more, the Netherlands has great opportunities in the field of the development of e-science. These opportunities are originating from directed investments in the network (SURFnet6/GigaPort), in extension and renewal of supercomputers (Huygens, Blue Gene), and in projects like LOFAR, VL-e and BIG Grid, next to NWO research programs like STARE, GLANCE and VIEW.*"
- Everyone collaborated because the push was to move from project to sustainable funding – a luring perspective!

ICTRegie

Nikhef

You then get some …

‘users don’t care who provides the service, as long as they get all of it in one spot, without hassle’

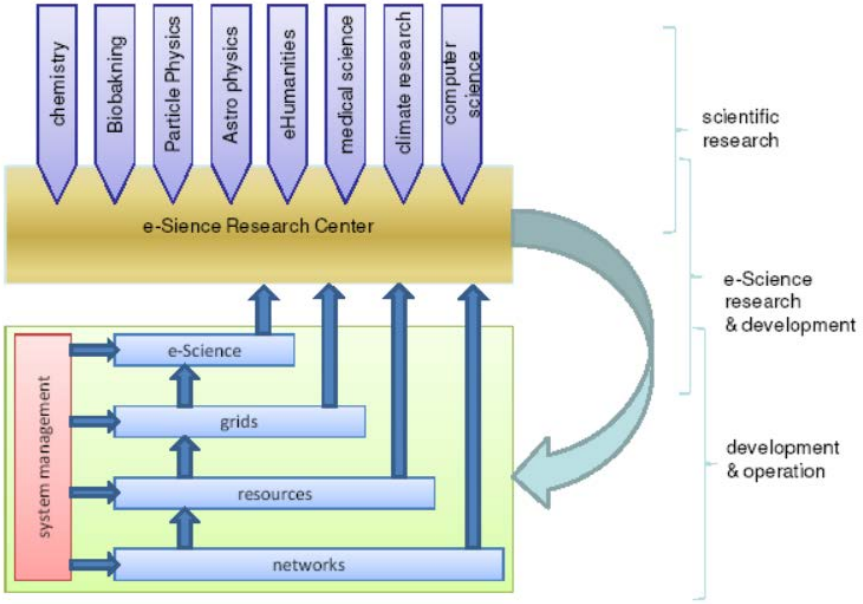


Figure 4 The E-science Research Center and ICT infrastructure for scientific research

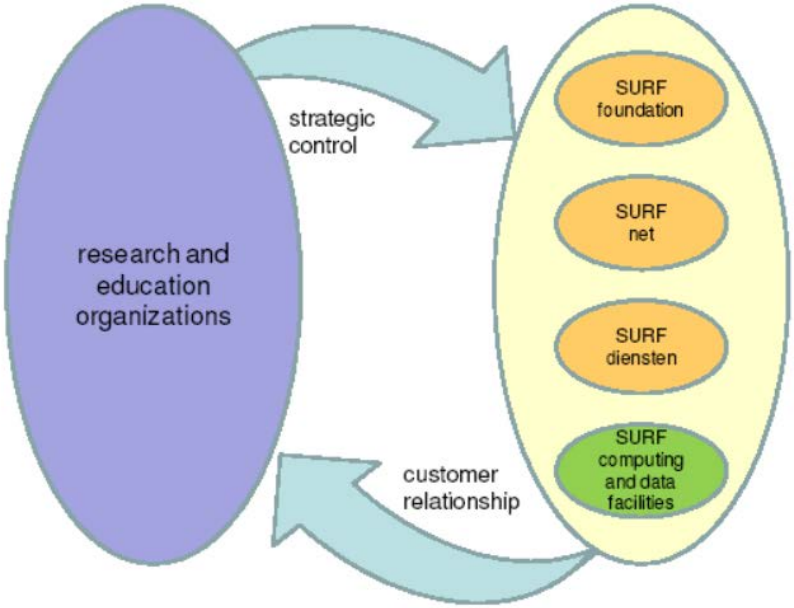


Figure 3 Proposed SURF organization

science domain input through both SURF stakeholder engagement and via the NLeSC

after the dust has settled, which takes >1-2 y

… but not all

	R&D	innovation	operation
e-science	1.5	2.0	0.5
grid	1.5	1.5	0.5
storage		3.0	1.5
capacity		2.5	2.5
capability		6.0	4.0
network	1.0	9.0	
system management	0.5	0.5	2.5
support			2.0
	4.5	24.5	13.5

	R&D	innovation	operation
European key position		12.0	8.0
		12.0	8.0

Table 1 Annual budget for the ICT infrastructure (in € million)

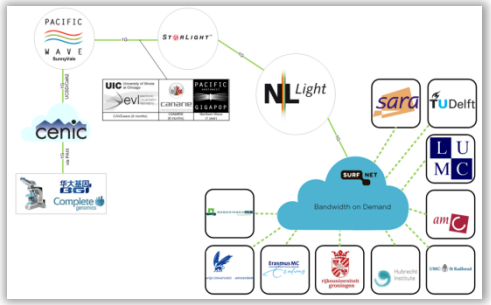
- First of all, you get less mor
  - "SuperNode" (PRACE-T0) ambition had to be abandoned
  - Not all the money was forthcoming as intended, so we got both sustained funding as well as a sustained funding gap

- Granting bodies (NWO)did not want to relinquish control over the (super)computing granting process which they traditionally controlled
  - government followed their advice
  - and that anomaly hunts us to this day …

David Groep
Nikhef
Amsterdam
PDP & Grid

But once you think of it …

Hardware
Support people
Money



Jan Bot – ‘Next generation networking for next generation sequencing’
linking data sources at BGI through lightpaths and GLORIAD to a distributed grid of clusters at UMCs and with central HPC Cloud @SARA

Frank Seinstra – ‘Jungle Computing’
Dynamic allocation of HTC, HPC, and cloud resources connected by lightpaths

David Groep
Nikhef
Amsterdam
PDP & Grid

Nikhef

SURFdrive

SURF

ownCloud → GÉANT → SURF NET

DNI CCDP @ SURFsara

→ SURF SARA

David Groep
Nikhef
Amsterdam
PDP & Grid

SURF

Compute

- Super computer (Caratesius)
- Beowulf cluster (LISA)
- GRID (GINA + LSG)
- HPC Cloud
- HADOOP

Storage

- Archive
- Grid (disk + tape)
- OwnCloud
- NoSQL / ElasticSearch

Network

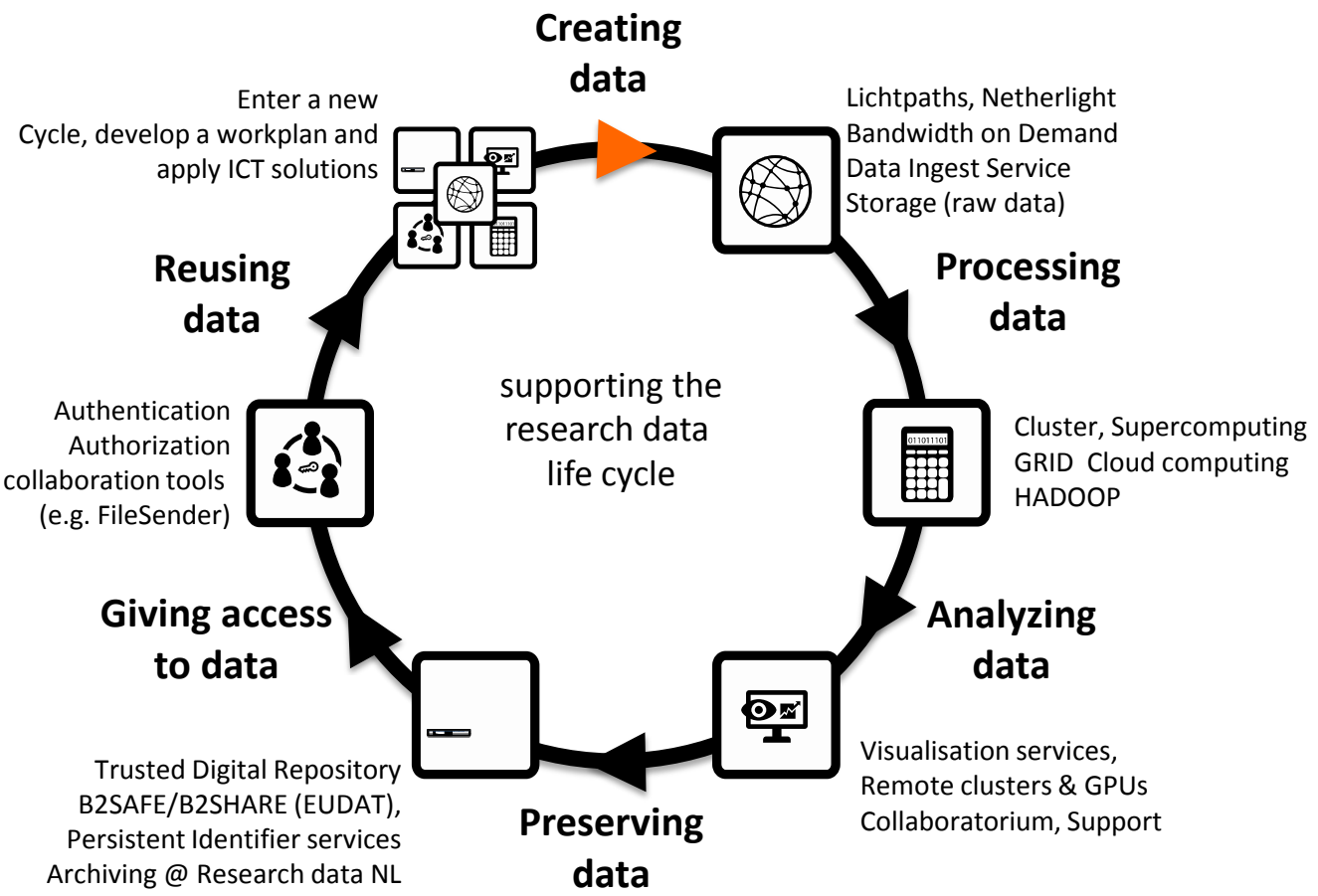
- Routed
- Lightpaths (MSPs)
- Wireless

- SURFsara: NGI + super computing center
- SURFnet: NREN
- SURFmarket: collaborative buying
- NLeSC: Netherlands eScience Center

- SURF is (as of 2015) a cooperative. Universities, University Medical Centers and research institutes are members.

Courtesy of Jan Bot, SURFsara

Research Data Life Cycle & SURF Portfolio

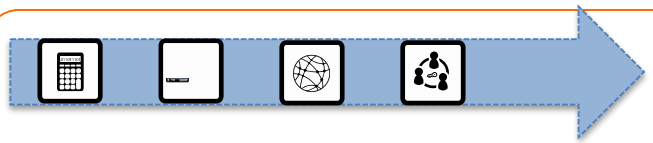


SURFmarket
licences/brokering

NLeScience center
integration support

Courtesy of Jan Bot, SURFsara

Research Services & Support Context



Researchers ask
Where to get facilities?
Where do I find expertise, training and support?
What regulation do I need to comply with?

Funding bodies
Regulations
Collaborations

European Commission
NWO
ZonMw
NFU
VSNU
University of Groningen
University of Twente
Radboud University
TU/e
TU Delft

Int. Research e-Infra
NL Research e-Infra
Instit. Research e-Infra

"Generic" Providers/Support
"LS&H" Support Providers
Support/Guidelines/ Good practices/ Networking

PRACE
EGI
EUDAT
Commercial providers

elixir
EMBL-EBI
BBMRI Biobanking and Biomolecular Resources Research Infrastructure

RDA Research Data Alliance

SURF
NIKHEF
rdnl research data netherlands
Commercial providers

SURF
trait a ctmm project
BBMRI-NL Biobanking and Biomolecular Resources Research Infrastructure

SURF
netherlands eScience center
DTL Dutch Techcentre for Life Sciences

Leiden Centre of Data Science
Amsterdam Data Science

rijksuniversiteit groningen
LUMC
UM
Universiteit Utrecht
umcg
VUmc

* Project based services

Courtesy of Jan Bot, SURFsara

SURF


SURF
SURF MARKET
SURF NET
SURF SARA
WWW.SURF.NL

Support4research (S4R)

- A four year program across all SURF subsidiaries to address the aforementioned challenges
  - 2015 - 2018
  - Work on visibility of SURF
  - Enhance knowledge exchange
- Started from another outreach activity: the Enlighten Your Research competition (Dutch, Global)
  - Combined support across all SURF subsidiaries highly appreciated by users
- Leverage the strengths of the individual subsidiaries
  - SURFnet (NREN): links with ICT directors at research institutes
  - SURFsara (NGI): links with researchers & research support
  - SURFmarket: knowledge of market services
  - NLeSC: software development & research involvement

Courtesy of Jan Bot, SURFsara

SURF

Goals Support4Research (S4R)

- Gather e-infrastructure requirements form research institutes

→ Facilitate interaction, discussions, pilots

- Increase knowledge (at the institutes) on compute, data and network services

→ **Integrated catalog for SURF and partner institutes**

- Strengthen bonds between research supporters

→ Community building & knowledge exchange

→ Improved **support flow between SURF subsidiaries and institutes**

- Increase knowledge about e-infrastructures

→ Knowledge exchange, seminars → Involve **local information managers**

- SURF and the Dutch research institutes are taking a ‘local first’ approach

→ Scale to national facilities when local e-infra is not sufficient or when explicitly requested

→ Scale to European infrastructure when necessary or in context of European projects: great for cross border collaborations, Innovation & technology provider, Knowledge exchange partner, Part of the service portfolio of SURF

Courtesy of Jan Bot, SURFsara

Access for Research programme

Remember that granting body for some infrastructures remained with NWO

Problems in applying for resources

- Network resources (SURFdashboard) via home institute to SURFnet
- Archive and OwnCloud via home institute to SURFsara or SURFnet, respectively (but it's a joint service!)
- HTC Grid, HPC Cloud, Hadoop, visualisation, Data Ingest, LSGrid, HTStorage via *researcher* self-service form to National e-Infra via SURFsara
- NRC "Lisa" beowulf compute and HPC "Carthesius" via NWO heavy-weight review process (and an incomprehensible form)

Takes between 2 weeks and 6+ months

Problems in reporting

- Institute-granted resources are typically co-funded by the home organisation and the researcher need not report anything
- National e-Infra (SURFsara) self-request resources request a yearly short PR report, but not much more
- "Lisa" and "Carthesius" need a formal yearly report-out

Each filed to respective granting body

David Groep
Nikhef
Amsterdam
PDP & Grid

Improvements and issues identified for NL

- Access is centered around the offerings, not the requested use cases
- No obvious relation between capacity request and financing or cost – apart from those for HPC and “Lisa”
  - makes for difficult capacity planning and funding thereof
  - does not match regular science funding for other investments
- Increase engagement of research groups in the access and granting process
  - explicitly *keeping* the link between individual researchers and the experts in the national e-Infrastructure
- Balance between rapid-reaction granting and efficient capacity planning is obvious, but both need to be there
- Offering, access method, ownership, and development of the infrastructure are not yet well coordinated
- Home organsiation IT “Support for Research” groups does not know the national capabilities and thus cannot effective collaborate or refer

David Groep
Nikhef
Amsterdam
PDP & Grid

Visit: surf.nl/support4research

Compute resources

Storage resources

Data transfer

Visualisation

Translation and integration support

Network infrastructure

Collaboration infrastructure

Services and support for research

SURF's operating companies work together to offer a world-class communication and computing infrastructure to facilitate scientific and scholarly research. Learn about the high-end services we offer, that will power up your research. Whether you need solutions for data transfer, compute, collaboration or visualisation, we offer and support it all.

Services overview

- Compute resources: high-end solutions thousands of times more powerful than your laptop
- Storage resources: easily accessible storage on disk or tape
- Data transfer: securely send very large files to other users
- Visualisation: advanced solutions and support to create visualisations
- Translation and integration support: dedicated support by experienced scientists
- Network infrastructure: fast end-to-end connections tailored to your research needs
- Collaboration infrastructure: single sign-on access to many commercial and non-commercial servic

Contact us for support

Choosing and using the right resources may be challenging. For that purpose partners within SURF initiated the SURF Outreach and Support Programme (SOS). The SURF SOS team can aid in deciding which infrastructure or solution to choose and can help to set up a research infrastructure that best fits your research

Courtesy of Jan Bot, SURFsara

Commitment, or co-funding?

‘We want to share your resources’

- need a sustainable model to funding shared infrastructure

Dutch strategy is likely to be mixed

- as long as the individual end-user will *not* get the bill
- we never turn scientists into customers – or they’ll go shopping (wasting time & resources)

Depends on community and size

- for individuals and small groups: central (small) funds
- for very large communities, expect some contribution ***to the shared infra***

but flexible: co-investment, bearing of operational cost – what is best depends on the community, and both work! (for Nikhef: LHCRoadmap)

David Groep
Nikhef
Amsterdam
PDP & Grid

All models exist, even within NL, and systems should allow all models to contribute in their own way to … but don’t turn communities into customers

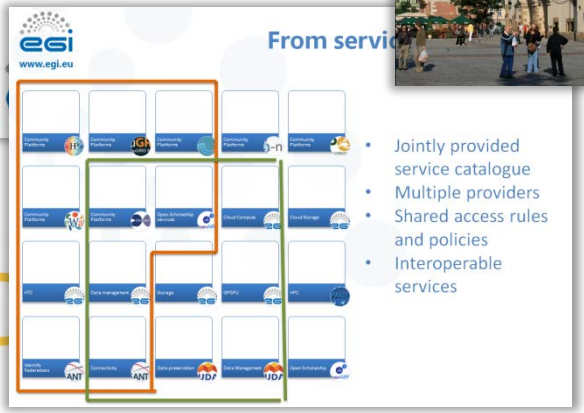
Meanwhile in Europe ...

Foto: NAS

Convergence and divergence …

Position Paper: European Open Science Cloud for Research

EUDAT LIBER OpenAIRE



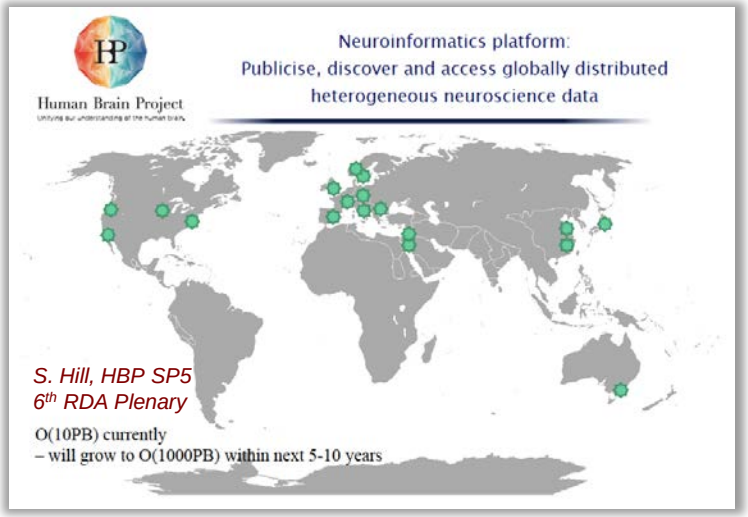
ENLIGHTEN YOUR RESEARCH GLOBAL 2015

David Groep
Nikhef
Amsterdam
PDP & Grid

EUDAT egi GÉANT … PRACE

Application driven infrastructure

Research is truly global



… and opportunistic

Larger communities

- can go ‘shopping’ for the best service,
- can pick the easiest place to ‘it done’
  – like AAs & PII at CERN, since it’s an extra-territorial treaty org ☺
- just set up their own if that’s expedient

David Groep
Nikhef
Amsterdam
PDP & Grid

Collaboration vs. branding

- Continuing tension between collaboration (which in Dutch actually carries the meaning of ‘working with the enemy’) and giving a ‘corporate marketing’ face to a (national) e-Infrastructure
  - BiG Grid built on the idea that the best marketing is marketing that **others do on your behalf** – so promote (research) communities to say how great you are, don’t do it yourself
  - It was a hardware investment proposal, but only 8 out of the 40 pages, 20%, actually talked about hardware or investment money, rest was about research domains
  - we *on purpose* relinquished the brand later – to promote SURF as a sustained entity
- User communities must feel ‘in control’ to support you in your quest for sustainability
  - that’s more than just an advisory group – it’s a *steering group*
  - don‘t build for them, but also don’t let them build: target ‘joint ventures’
  - some uneasy tension will typically remain

David Groep
Nikhef
Amsterdam
PDP & Grid

Bottom-up continues to work

- We did and still do EUGridPMA, IGTF, TACAR, SCI, SirTFi
- Multi-year run-up to AARC is a demonstration of success
- WISE ng

- So maybe “Joint EYR Global” should not be too hard …

Image credits: Alessandra Scicchitano,

David Groep
Nikhef
Amsterdam
PDP & Grid

Collaboration need not wait for funding!

- *At times, working* on a joint challenge most important part:
  collaboration need vision more than money!

- Remaining ‘outward facing’ is important
  - A focus on changes within the org does not help – sorry!
  - SURF disengaged for ~2 years during the merger, and that ‘did not help’ in creating a national e-Infra collaborative spirit
  - Merging cultural differences is detrimental, since the most rigidly organised part tends to win early on (because it’s more pushy)
  - But it later then becomes clear that the most outward facing org offers better prospects for sustainability since it makes external folk (its users) promote it on its behalf:
    “having someone else tout your excellence is far better than banging yourself on the chest”

David Groep
Nikhef
Amsterdam
PDP & Grid

http://pericles.ipaustralia.gov.au/ols/auspat/applicationDetails.do?applicationNo=2001100012

Re-inventing the wheel …

(12) INNOVATION PATENT (11) Application No. **AU 2001100012 A4**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Circular transportation facilitation device

(51) International Patent Classification(s)
B60B 001/00

(21) Application No: **2001100012** (22) Date of Filing: **2001.05.24**

(45) Publication Date: **2001.08.02**
(45) Publication Journal Date: **2001.08.02**
(45) Granted Journal Date: **2001.08.02**

(71) Applicant(s)
John Keogh

(72) Inventor(s)
Keogh, John Michael

(74) Agent / Attorney
Sandercock & Cowie 69 Robinson Street Dandenong Victoria AU

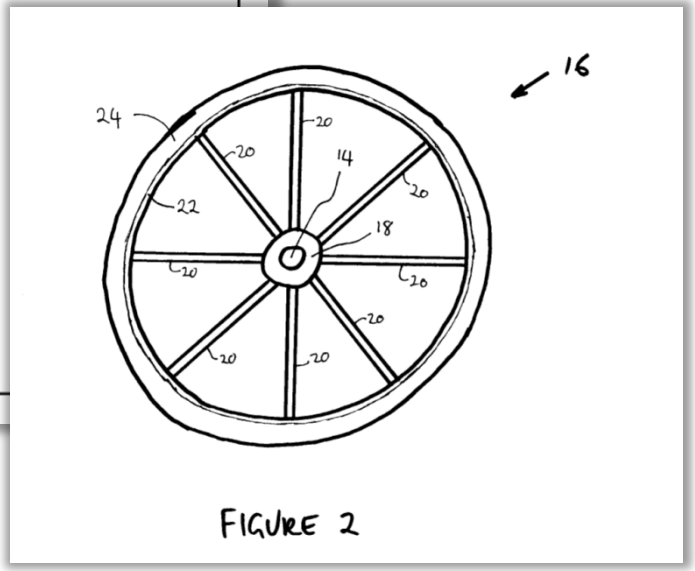


FIGURE 2

David Groep
Nikhef
Amsterdam
PDP & Grid

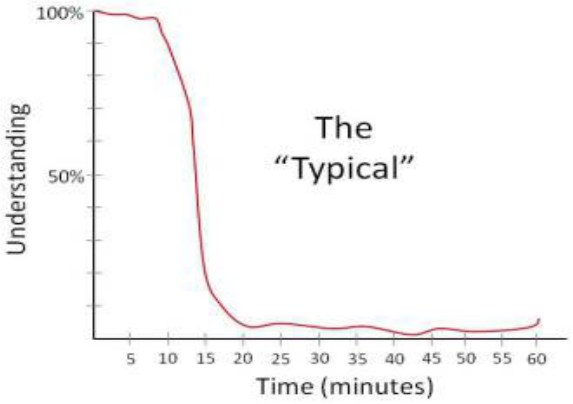
Jointness may appear shaky at times –

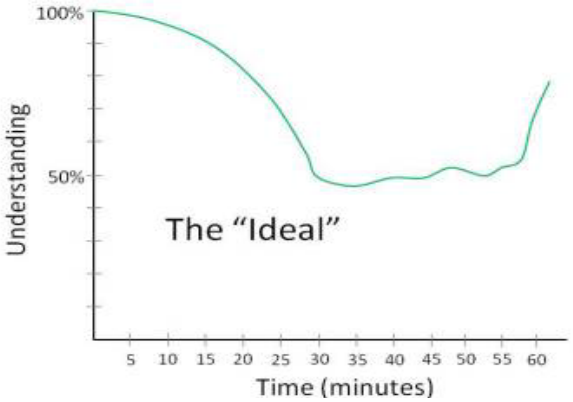
but just continue to build!

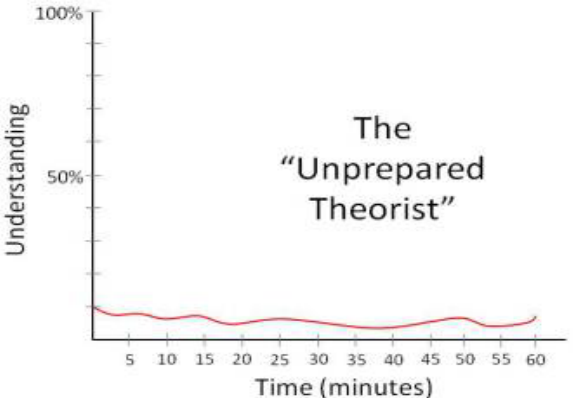
https://www.youtube.com/watch?v=LVJOedTBogU

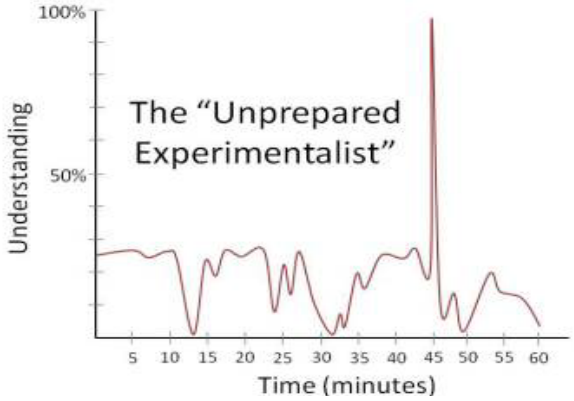
www.hornbach.de http://www.globalflagproject.com/de

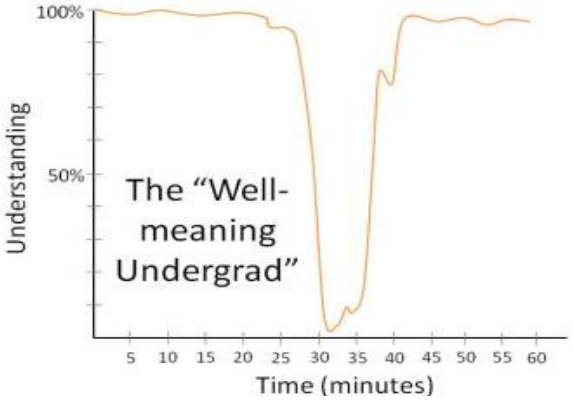
David Groep
Nikhef
Amsterdam
PDP & Grid

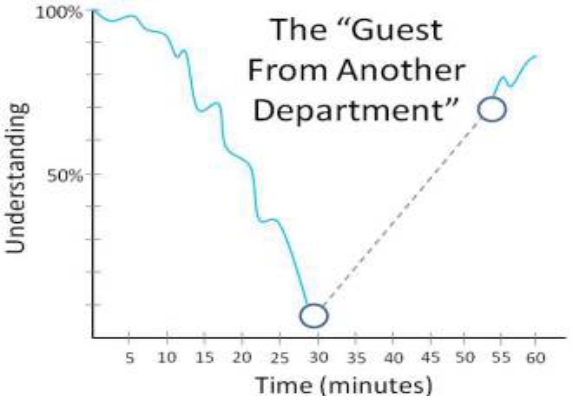


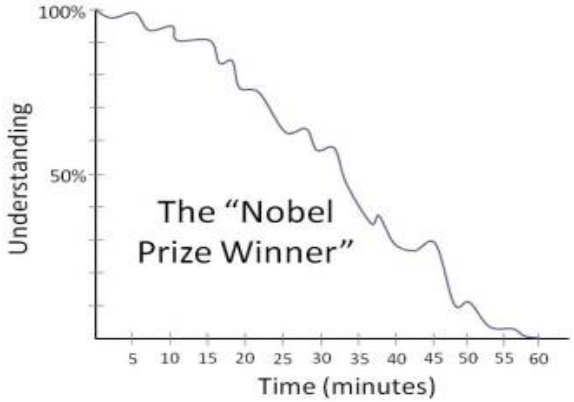


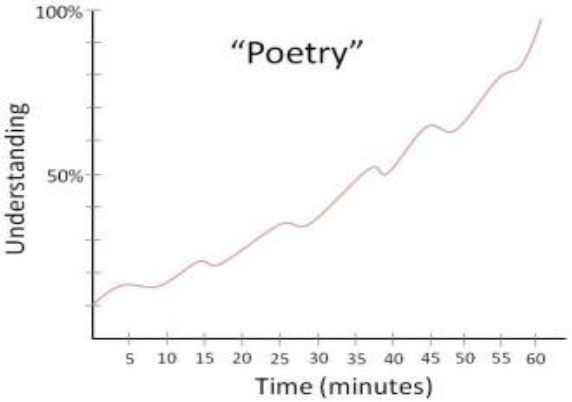


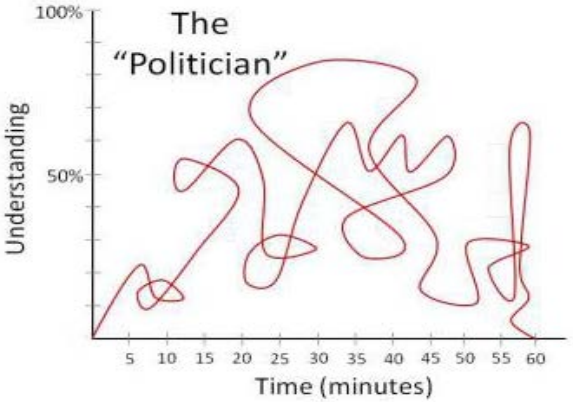












http://manyworldstheory.com/2013/10/03/the-9-kinds-of-physics-seminar/

Nikhef

WHAT SURF CAN DO