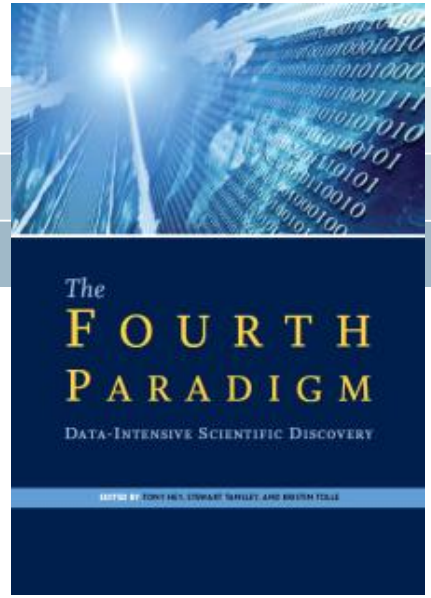
Gerben J de Boer
Sunday June 15th
Kijkduin, the Netherlands

JERICO summer school 2014
From Data to Decisions



4th paradigm
eScience

4th paradigm: 21st century = web-based community science

4 phases in science (Jim Gray †, Microsoft research)

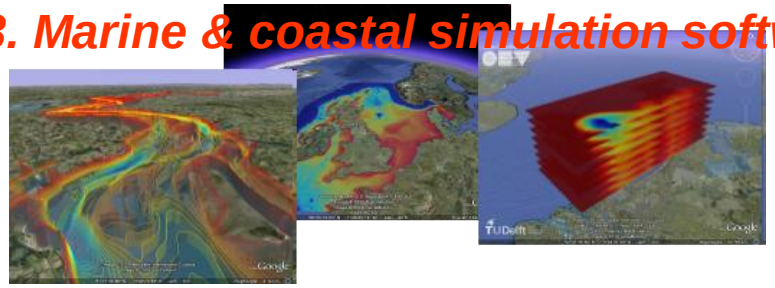
1. **Empirical** (last 1000s yr)
   - Archimedes …
2. **Theoretical** (last 100s yr)
   - Newton, Kelvin …
3. **Computational** (last 10s yr)
   - Mellor & Blumberg …
4. **Digital Data Deluge**: BIG data + merging of
   1. Experiments,
   2. Theory,
   3. Massive simulations
   4. Mass-data gathering: Lidar, iphones, PIV

1. Marine & coastal labs

2. Marine & coastal math. models

$$\frac{\partial k}{\partial t}+\frac{u}{\sqrt{G_{\xi\xi}}}\frac{\partial k}{\partial \xi}+\frac{v}{\sqrt{G_{\eta\eta}}}\frac{\partial k}{\partial \eta}+\frac{\omega}{d+\zeta}\frac{\partial k}{\partial \sigma}=$$
$$+\frac{1}{(d+\zeta)^2}\frac{\partial}{\partial \sigma}\left(D_k\frac{\partial k}{\partial \sigma}\right)+P_k+P_{kw}+B_k-\varepsilon.$$

3. Marine & coastal simulation software

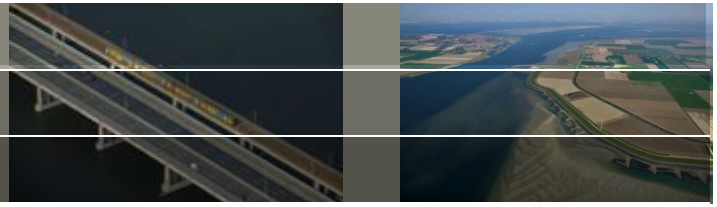


4. Marine & coastal open communities

1980 Floppy


AT&T
IBM FORMATTED
2HD
HD
AT&T
WOLF
CCCP
11-26-IV-59
КИЕВ
ПОЧТАМТ
ВТОРОЙ ВСЕСОЮЗНЫЙ КИНОФЕСТИВАЛЬ
123RF

1990 FTP



fedorbaart — ftp — 97×30 — ⌘1

bash | bash | bash | ftp

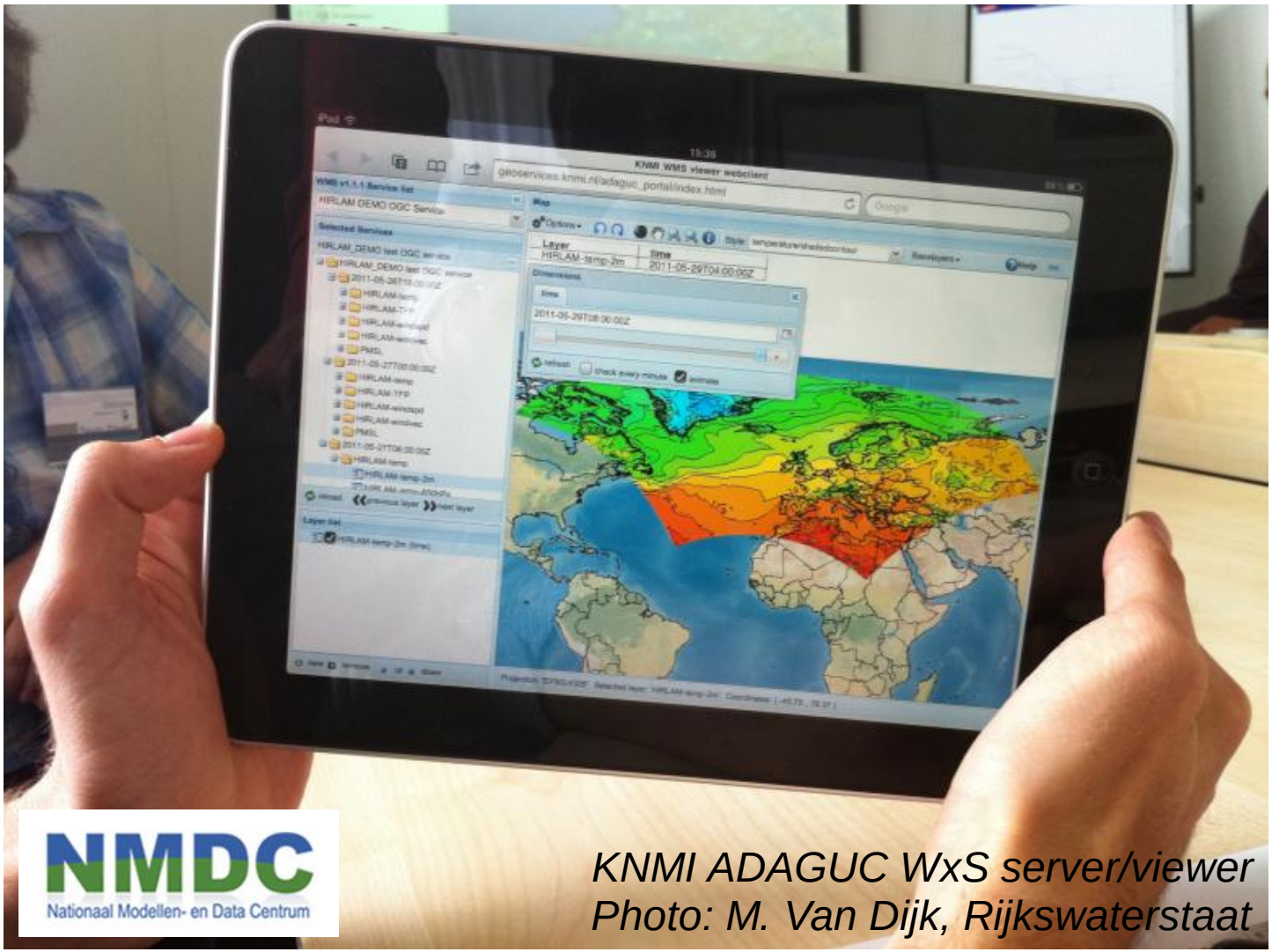
```
Name (nssdcftp.gsfc.nasa.gov:fedorbaart): Anonymous
331 Please specify the password.
Password:
230 Login successful.
Remote system type is UNIX.
Using binary mode to transfer files.
ftp> ls
229 Entering Extended Passive Mode (|||34366|)
150 Here comes the directory listing.
-rwxr-xr-x    1 0        0            1868 Apr 28  2006 00README.TXT
-rw-r--r--    1 0        0            4769 May 27  2005 Access_Warning_Priv.html
drwxr-xr-x    5 0        7017         4096 Dec 03  2012 PDS
drwxrwxr-x    3 0        0            4096 Jun 29  2006 admin
drwxr-x--x    3 0        5000         4096 Mar 02  2007 css
drwxr-xr-x    4 3562     462          4096 Sep 27  2012 documentation
drwxr-xr-x    2 2000     0            4096 Jun 26 13:45 icons
drwxr-xr--    2 0        7004         4096 Nov 30  2006 image
drwxr-xr-x    3 0        5000         4096 Apr 18 14:52 images
-rwxr-xr-x    1 0        5000         5993 Mar 20  2011 index.html
drwxr-xr-x   19 0        5000         4096 Sep 14  2009 miscellaneous
drwxr-xr-x   10 0        7005         4096 Jun 29  2006 models
drwxr-xr-x    5 0        7011         4096 Apr 14  2009 photo_gallery
drwxr-xr-x   12 0        5000         4096 Jun 29  2006 selected_software
drwxr-xr-x   56 0        462          4096 Jun 03 12:47 spacecraft_data
drwxrwxrwx   14 0        0            4096 Jun 28 15:43 staging
drwxr-xr-x    6 0        7010         4096 Feb 27  2008 standards
226 Directory send OK.
ftp> cd staging
250 Directory successfully changed.
ftp>
```

2000 Download forms

	(•) Map	( ) Time serie	
ZONE	Global		
SATELLITE	(•) Reference*(1) ( ) *Topex/Poseidon*	( ) *Jason*-1 ( ) *Jason*-2	( ) Envisat ( ) Multi-mission(2)
CORRECTIONS	**Applied corrections** (•) Inverted barometer; *Radiometer* wet troposphere* ( ) Inverted barometer; Model wet troposphere ( ) Inverted barometer correction not applied; *Radiometer* wet troposphere		
	Download the image	Download the data (NetCDF) Download the data (ASCII)	

2010 Web services voor thin clients

Sped-up by fact that first time in history that personal computers become smaller: Thin clients as browser, tablet or phone

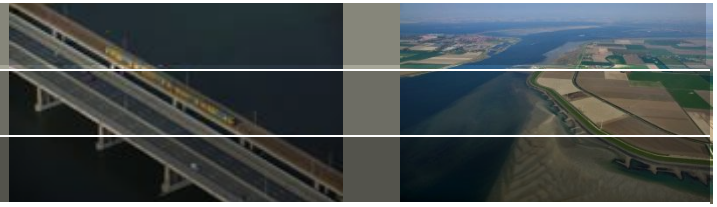


KNMI ADAGUC WxS server/viewer
Photo: M. Van Dijk, Rijkswaterstaat

DataTube

NMDC
Nationaal Modellen- en Data Centrum

Sentinel



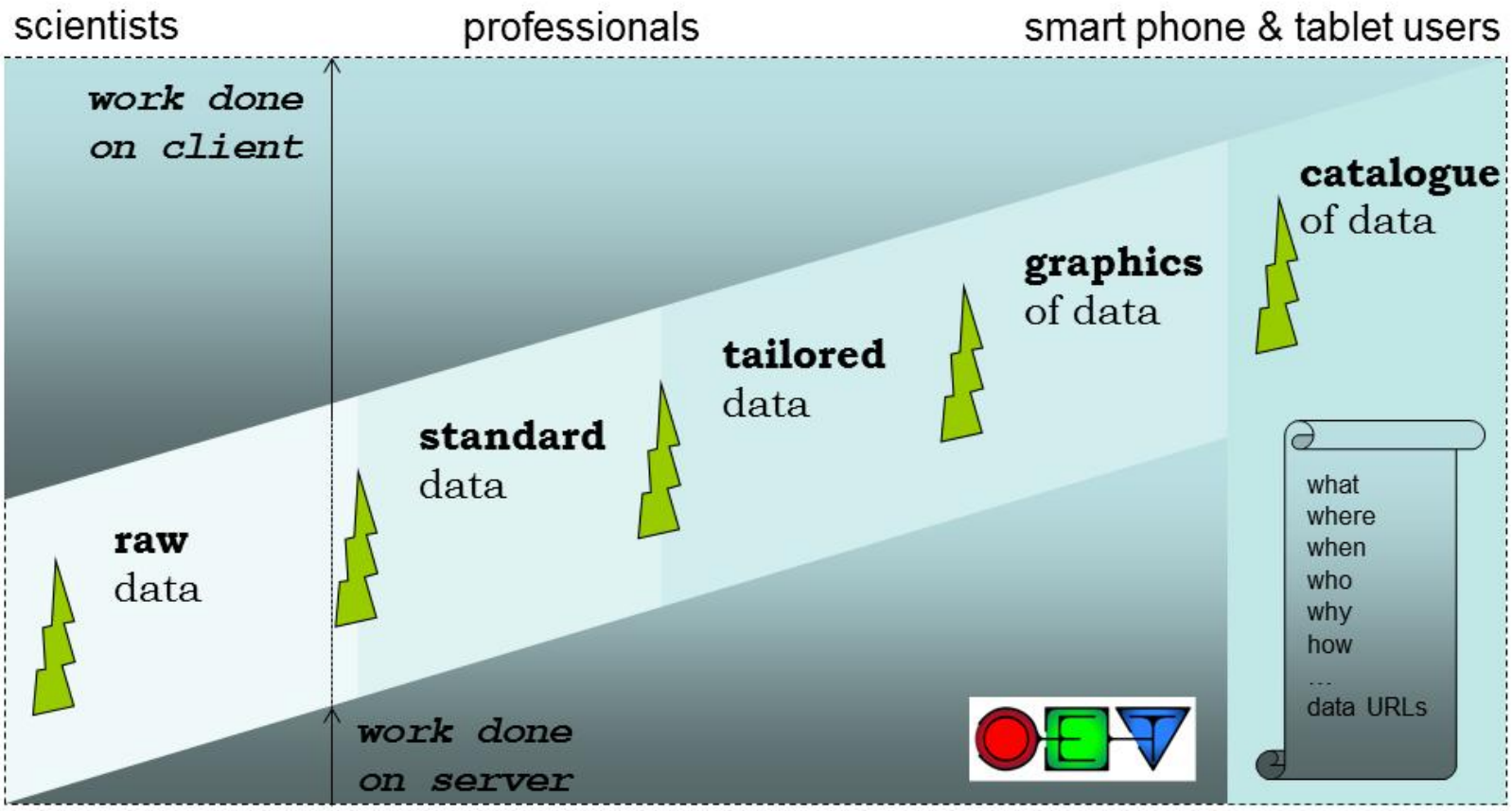
- New series of ESA satellites
- Generate order 1TB/day
- Not a single researcher can have a copy any more
- Even institutes have to make proper considerations
  - (1TB = 100 EUR as external hard drive)
  - 1TB = 1000 EUR for professional hosting in datacenter
  - 1 year: 365 TB
  - 400,000 EUR to host a single copy of 1 year of data

Move data to analysis, or move analysis to data

- Old: data moves to scientific analysis [once]
  - 1980: Floppy, Cd, or DVD by mail
  - 1990: Ftp
  - 2000: Download forms
- Now: data moves to scientific analysis [live]
  - 2010: OGC web services “DataTube”
- Now: scientific analysis moves to data [once]: catalog of functions
  - 20xx: Web Processing Service: EMODnet Chemistry case friday
- Future: scientific analysis moves to data [live]: log-in to data
  - 20xx: Back to mainframe: NASA NeX, ESA GPOD

5 data levels in client-server rationale

inspired by 5-stardata of Tim-Berners Lee



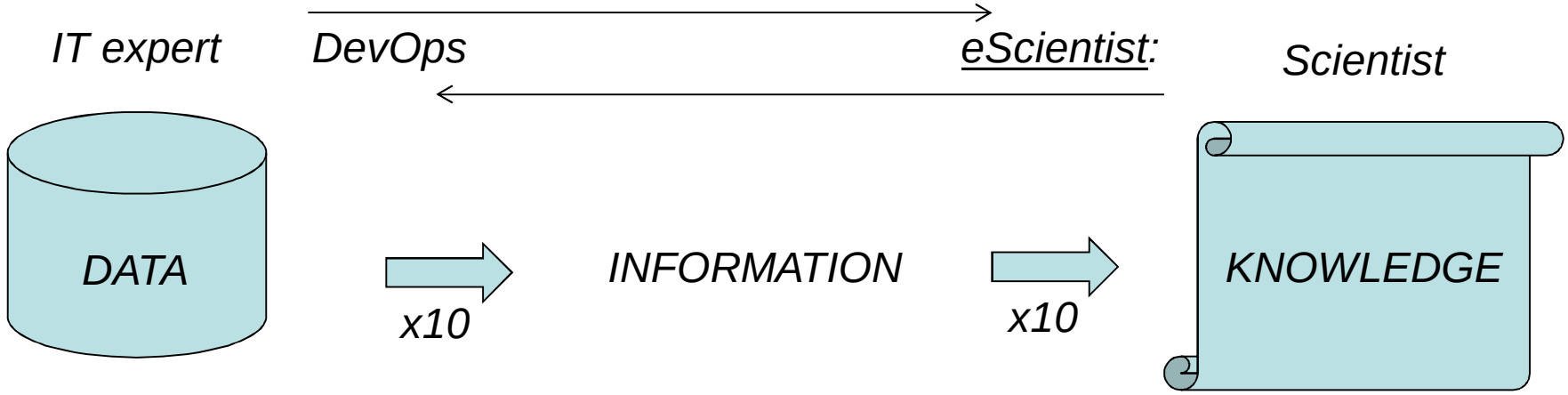
IT and science: eScience + DevOps

Top Sector: involve companies in science: Dutch companies requested our government to invest in better use of existing IT rather than developing more IT

R&D *in* IT: data gathering (hardware): Sentinel satellites: 1 TB /day
R&D *with* IT: data processing (software): How to process all this?

IT and science need to team up

- IT cannot handle science any more
- Science cannot handle IT any more

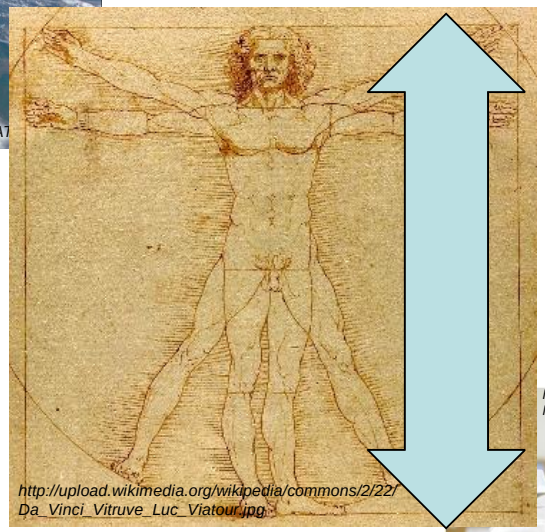


Especially for marine & coastal science/engineering


http://nupedailynews.com/wp-content/uploads/2013/01/telescope..jpg


http://oeatech.net/wp-content/uploads/2011/03/RADARSAT

6400 km:
North Atlantic Oscillation (NAO)


http://upload.wikimedia.org/wikipedia/commons/2/22/Da_Vinci_Vitruve_Luc_Viatour.jpg

gap 10^{-6} > 10^{+6}

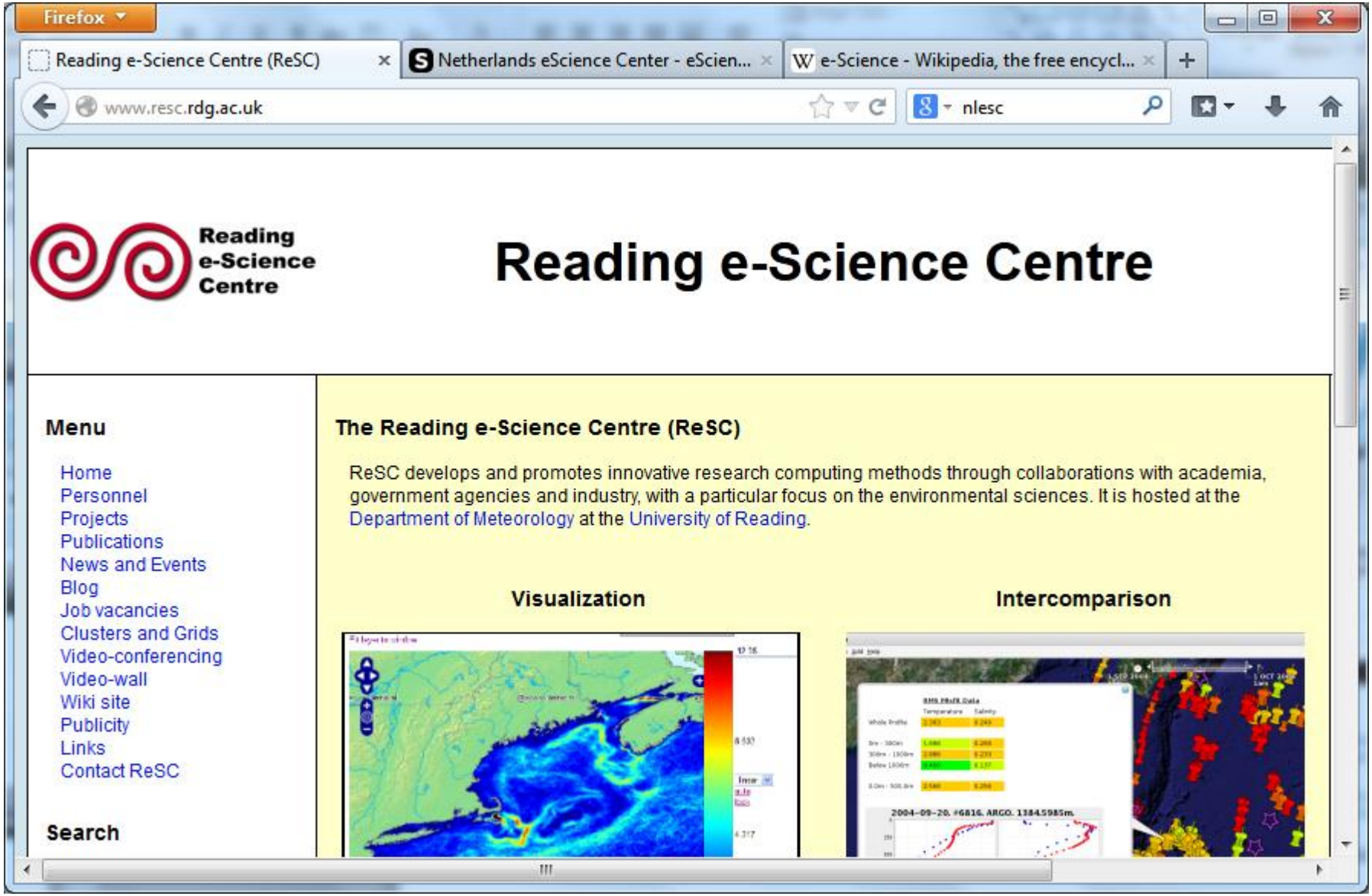
Water scarcity
Flood protection
Water pollution

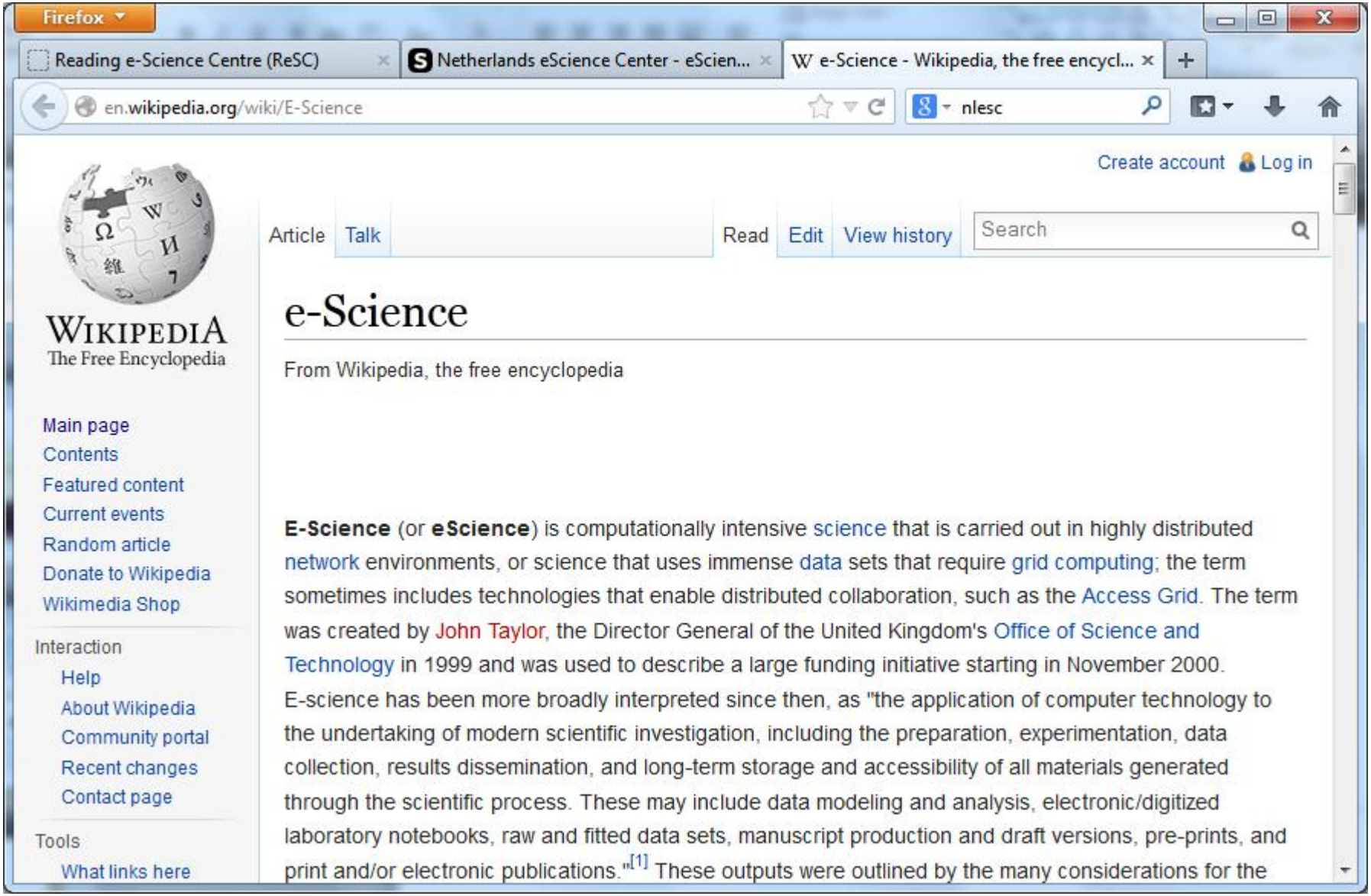

http://chromblog.thermoscientific.com/Portals/49739/images/lims for biobanking1.png


http://www.eiroforum.org/media/photo_galleries/cern/cern-02l.jpg

suspended mud particle 64 µm

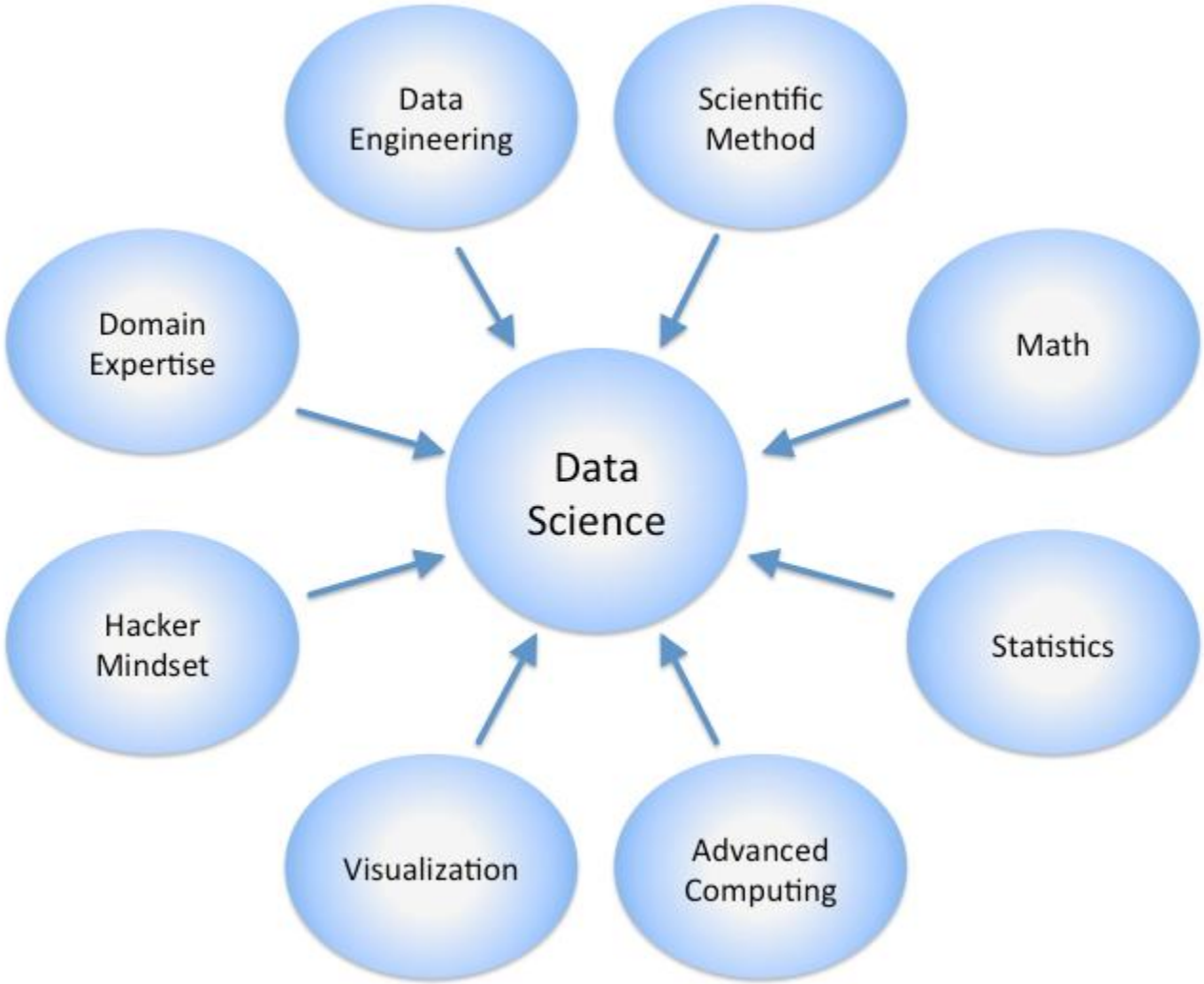
MyOcean uses ncWMS software from:




Firefox
Reading e-Science Centre (ReSC)
Netherlands eScience Center - eScien...
e-Science - Wikipedia, the free encycl...
en.wikipedia.org/wiki/E-Science
nlesc
Create account
Log in
Article
Talk
Read
Edit
View history
Search
WIKIPEDIA
The Free Encyclopedia
Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikimedia Shop
Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page
Tools
What links here
e-Science
From Wikipedia, the free encyclopedia
E-Science (or eScience) is computationally intensive science that is carried out in highly distributed network environments, or science that uses immense data sets that require grid computing; the term sometimes includes technologies that enable distributed collaboration, such as the Access Grid. The term was created by John Taylor, the Director General of the United Kingdom's Office of Science and Technology in 1999 and was used to describe a large funding initiative starting in November 2000. E-science has been more broadly interpreted since then, as "the application of computer technology to the undertaking of modern scientific investigation, including the preparation, experimentation, data collection, results dissemination, and long-term storage and accessibility of all materials generated through the scientific process. These may include data modeling and analysis, electronic/digitized laboratory notebooks, raw and fitted data sets, manuscript production and draft versions, pre-prints, and print and/or electronic publications."[1] These outputs were outlined by the many considerations for the

Asking for the moon (Dutch: *schaap met 5 poten*)

what a marine scientist needs to master in the 21st century



http://en.wikibooks.org/wiki/Data_Science:_An_Introduction/A_Mash-up_of_Disciplines

Job description example

what we aim to train you for this week

netherlands eScience center

Job description

eScience engineers are digital scientists able to work at the interface of their own scientific disciplines and enhanced ICT. eScience engineers are (mostly) PhD scientists with a history of developing and applying scientific approaches within their previous research domains and are able to utilize ICT to make scientific breakthroughs. They are also interested in sharing their knowledge and experience outside their historical domains.

As an eScience engineer you will operate in a scientific environment and be responsible for the translation of scientific questions into solutions that effectively apply advanced ICT technologies. You will ensure that methodologies and applications can be used easily by scientists within a coordinated eScience infrastructure. You will work closely with other scientists both within NLeSC and its partner organisations. eScience engineers work as a close team, creatively combining and sharing knowledge between projects. It is likely that you will work part time at NLeSC and part time at the universities participating in the collaborative projects.

NLeSC is recruiting eScience engineers at various levels able to work closely with scientists from various disciplines. Experience working in an academic or related environment is required. The tasks include acting as an interface between users and NLeSC, contributing to software development for scientific research and project management. Experience and specialist knowledge will determine the precise level of the engineer.