

OpenEarthTools =

Open source management of

Data, Models and Tools for

marine & coastal

science & technology



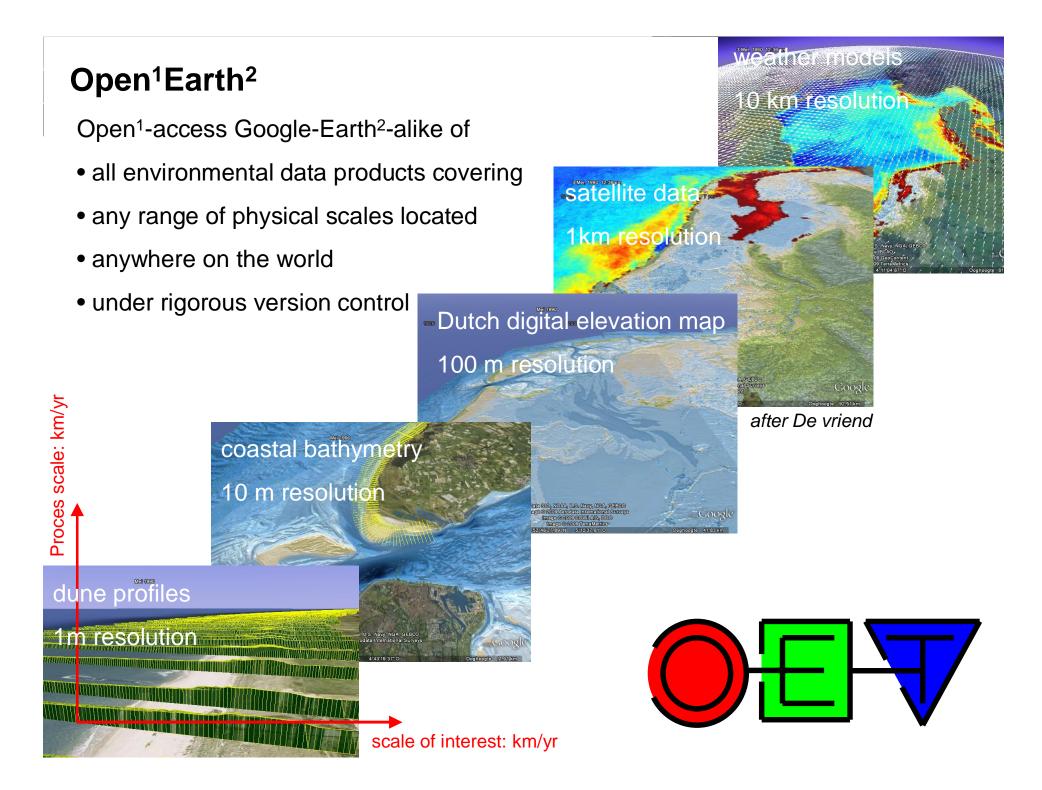




.. and what about



wwww.OpenEarth.eu



Contents

O. The power of collaboration

1. Google Earth: web service for images

2. OPeNDAP/netCDF: web service for data

3. Subversion: web service for version control



Version control and backup

tools data & scripts



Web access to datasets

from pluriform to standard format!



3 Straightforward visualisation

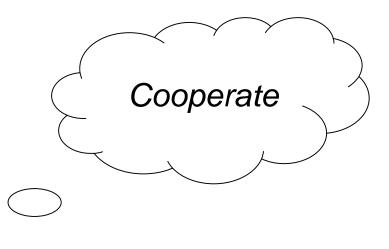
open source software





What is OpenEarth: philosophy

- 1. philosophy
- 2. community
- 3. repository
- 4. delivery





When you work on complex matters you soon discover that some things are too big to handle on your own. That you constantly reinvent the wheel. That others seem to reinvent your wheels. That some important wheels never get invented because they are too big to handle for a single project. The solution is to collaborate across the artificial boundaries of projects.

What is OpenEarth: community

- Prisoners dilemma
- Tragedy of the commons
- OpenEarth accepts some asymmetry in sharing: it is the player that keeps the sharing going

2 * 2 = 4
3 * 3 = 9
1 * 1 = 1
11*11 = 121
2 * 2 = 4
1 * 1 = 1
=== +
140

Berekening van
de score van rood

Journal of Personality and Social Psychology 2009, Vol. 96, No. 1, 83-103

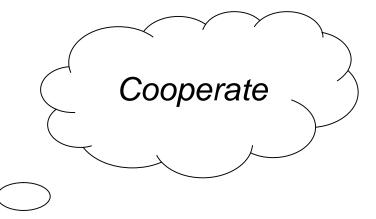
Promoting Cooperation and Trust in "Noisy" Situations:
The Power of Generosity

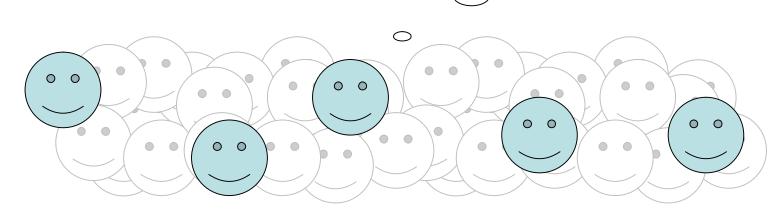
Anthon Klapwijk VU University Amsterdam Paul A. M. Van Lange VU University Amsterdam and Leiden University

The authors present an interdependence theoretical framework and advance the argument that generosity serves the important purpose of communicating trust, which is assumed to be of utmost importance to coping with incidents of negative noise (i.e., when the other every now and then behaves less cooperatively than intended). Using a new social dilemma task (the parcel delivery paradigm), it was

What is OpenEarth: community

- 1. philosophy
- 2. community
- 3. repository
- 4. delivery

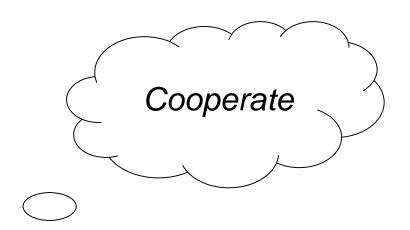


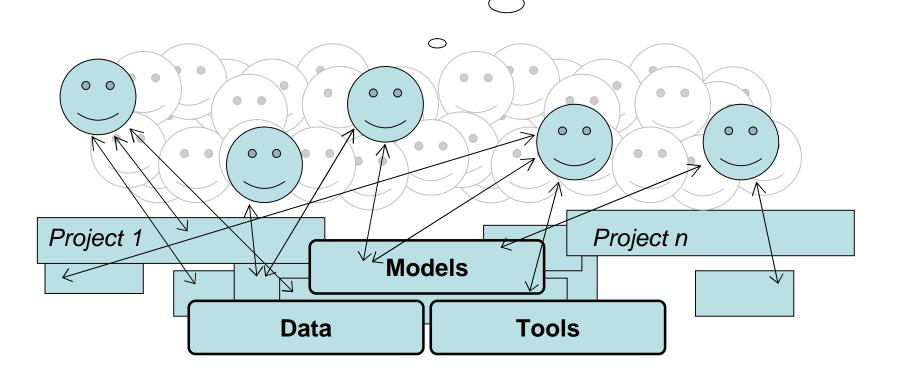


Fortunately there are usually more people than you think who are prepared to team up in related topics. So share the wheels you invented, never invent a wheel again, and embark a mission to invent those very big wheels we all need. You can split the required investment over different projects that afford the money otherwise needed for reinvention of wheels. Like this, OpenEarthTools made an advanced Google plot toolbox, a coordinate conversion toolbox, etc.

What is OpenEarth: community needs structure

- all collaborating is not enough
- chaos
- coordination needed
- but no overall boss
- like wikipedia



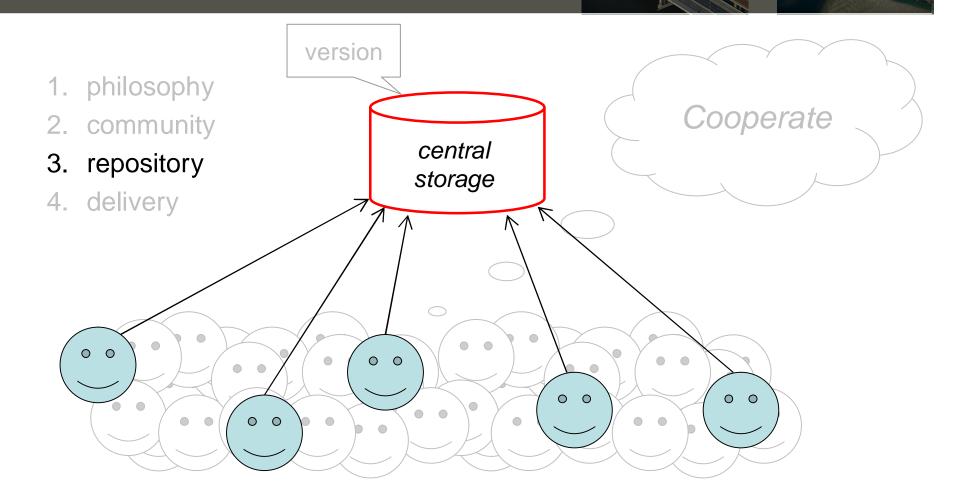


What is OpenEarth: community needs structure

- oh oh, ik heb een oude versie gebruikt!
- euh, wat/waar is de meest recente versie eigenlijk?
- o nee, maken we weer dezelfde fout!
- hmm, wat is hier eigenlijk gemeten?
- tja, waar is dit eigenlijk gemeten?
- oops, welke tijdzone is dit gemeten?
- aarghh, wie weet nog wat hier precies is gedaan?
- #\$*!, is die data weggegooid?!?
- ja hoor, weer een nieuw data formaat!!!
- zucht, dit probleem moet iemand al eens eerder hebben opgelost!
- QA/QC = SOMETHING¹ ADHERES TO² SOME CRITERIA³
 - 1. version control
 - no version control = no quality
 - 2. a community using it and improving it on the spot
 - no users = no quality
 - 3. the cumulative opinion and filosophy of all users
 - no opinion in quality = no quality

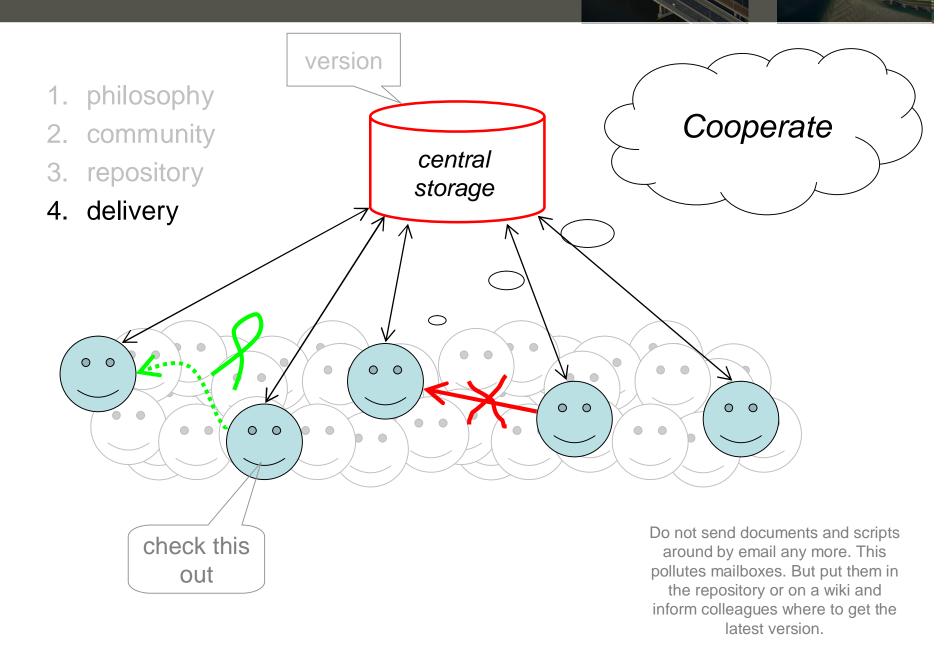


What is OpenEarth: repository



When working with a team, it is important to keep good track of the latest version of a tool/document. When working in pairs of 2 it is easy: either you work on it or the other person. In groups of 3 and bigger a formal system is required. Especially when more then 1 person is working on it at the same time. A repository allows more persons to work on it simultaneously, and it takes care of the essential merging process.

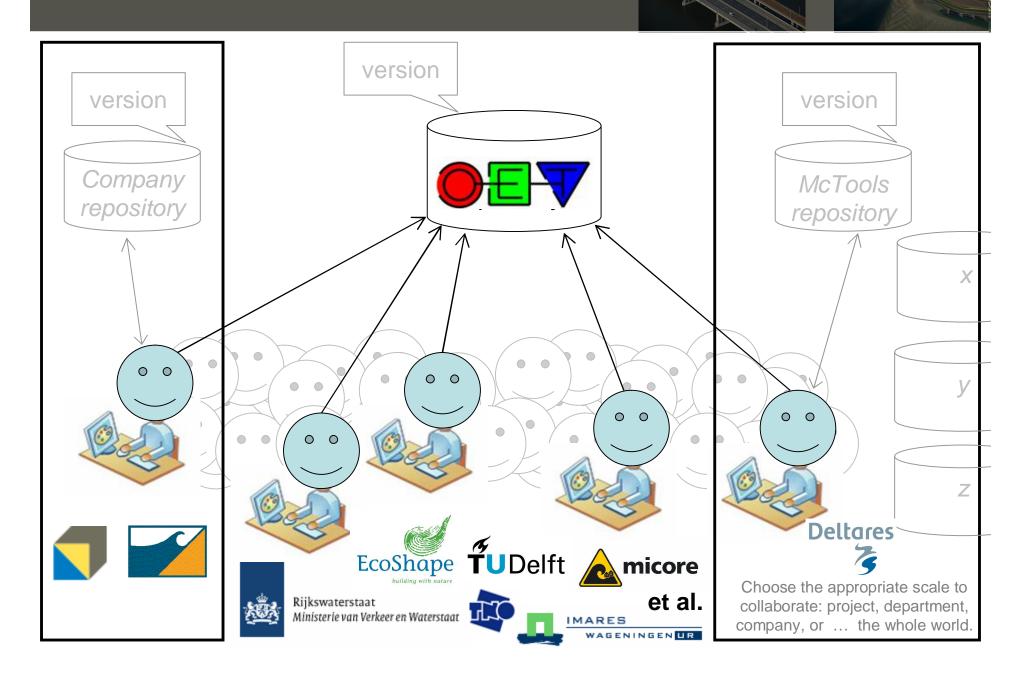
What is OpenEarth: web services



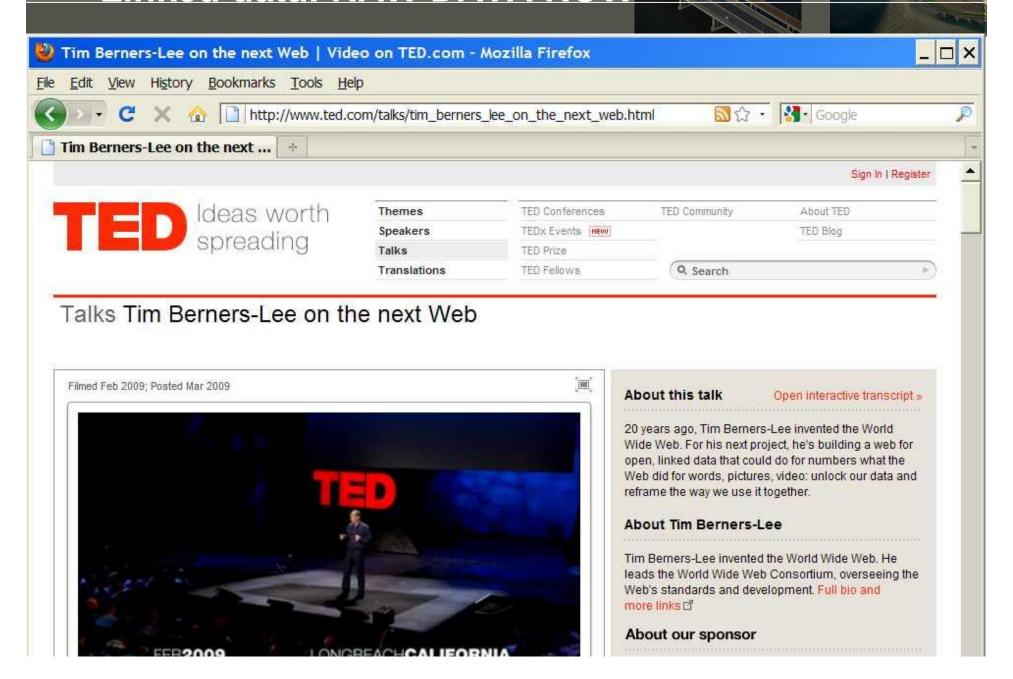
What is OpenEarth: example

- 1. General Mapping Toolbox: only *.eps on linux
- 2. arcGIS: \$
- 3. matlab mapping toolbox: \$
- 4. PCTRANS marine: powerful, but much ado with external files (grids)
- 5. PROJ 4: powerfull but requires it skill to link it
- 6. matlab CTRANSDV: Rijksdriehoek, UTM, LonLat
 - Zitman > Verploeg > Bonekamp > Elias > de Boer > ?
- 7. OpenEarth SuperTrans convert coordinates
 - Maarten van Ormondt (Deltares) & Thijs Damsma
 - GUI & matlab command line
 - ~ 3000 EPSG transformations
 - free if you join OpenEarthTools

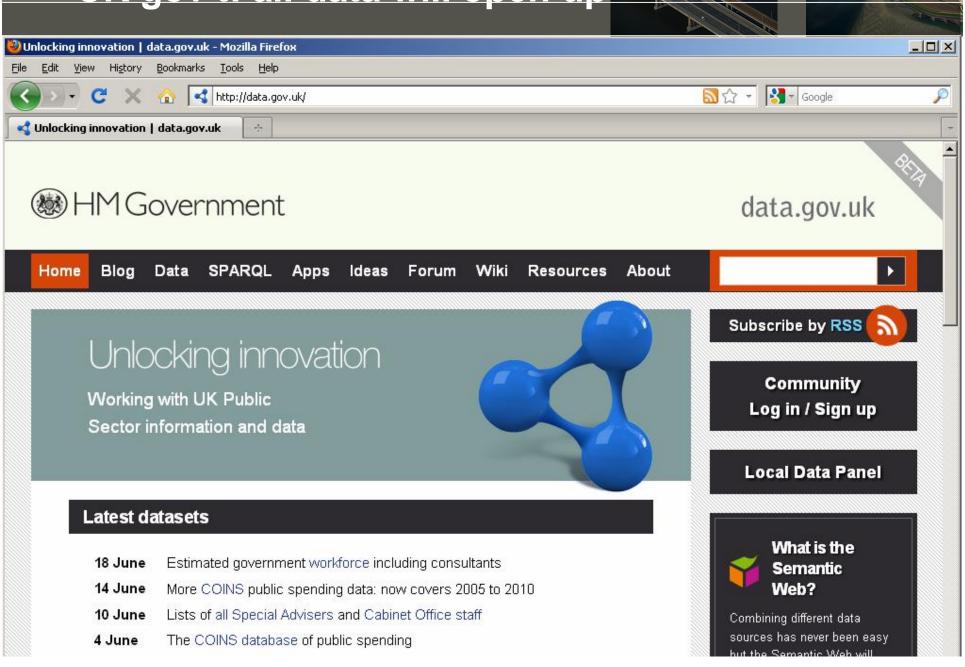
What is OpenEarth: uniform open and closed access



Linked data: RAW DATA NOW



UK gov't: all data will open up



What does OpenEarth

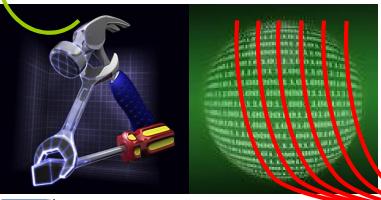
- 1. Stimulate collaboration
 - spread the message
- 2. Organise collaboration
 - design open source components IT infrastructure:
 - operational: proof-of-concept: ready for 'test-drive'
 - clones to spread approach (pilot > operational):
 - McTools, TU Delft library, Van Oord, Alkyon, MATROOS
 - on linux server: pilot takes ICT half a day
- 3. Training and minimizing divergence
 - <u>http:// openEarth.eu</u>

How does OpenEarth do that: WORKFLOW



Version control and backup

tools data & scripts



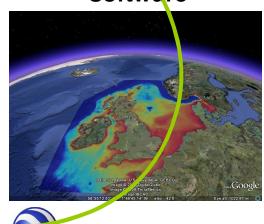
Web access to datasets

from pluriform to standard format.



3 Straightforward visualisation

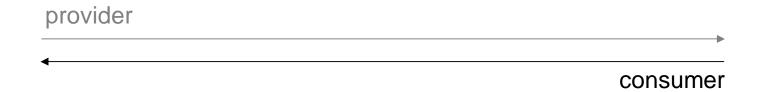
open source software





OpenEarth infrastructure





repos.deltares.nl



tools



SubVersion Server



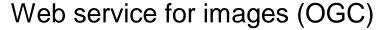
netCDF data

(2) opendap.deltares.nl (3) kml.deltares.nl (soon)

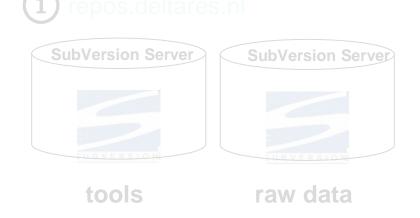


Google Earth data

OpenEarth infrastructure



- Visualizations of geo-data: 3D
- Interface to other visualizations
- Interface to data on OPeNDAP server

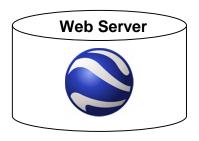




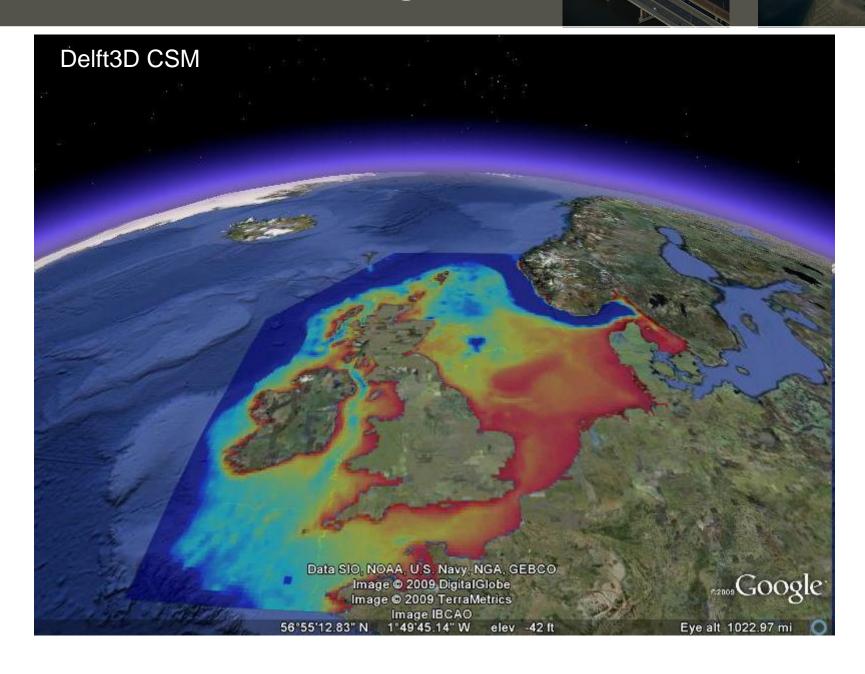


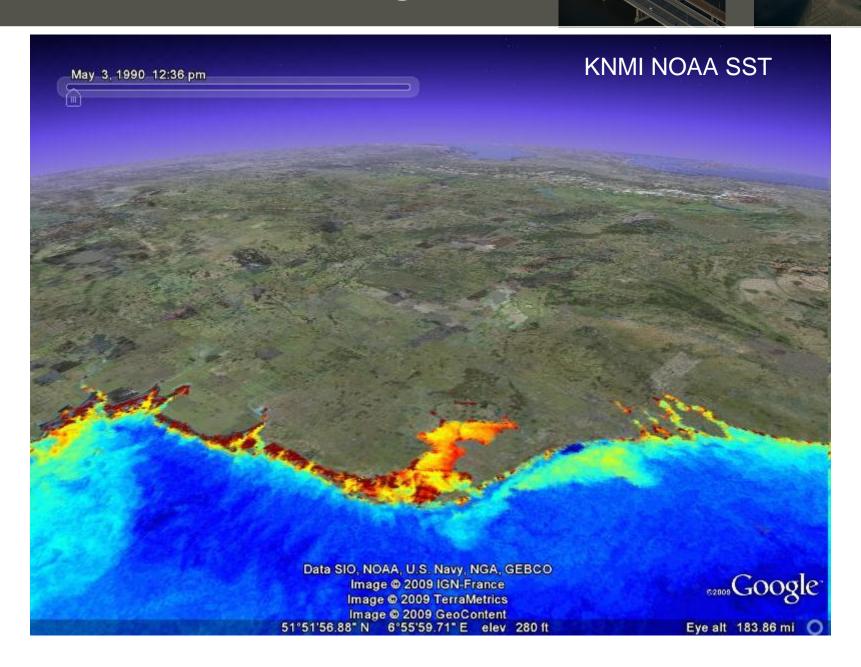
netCDF data

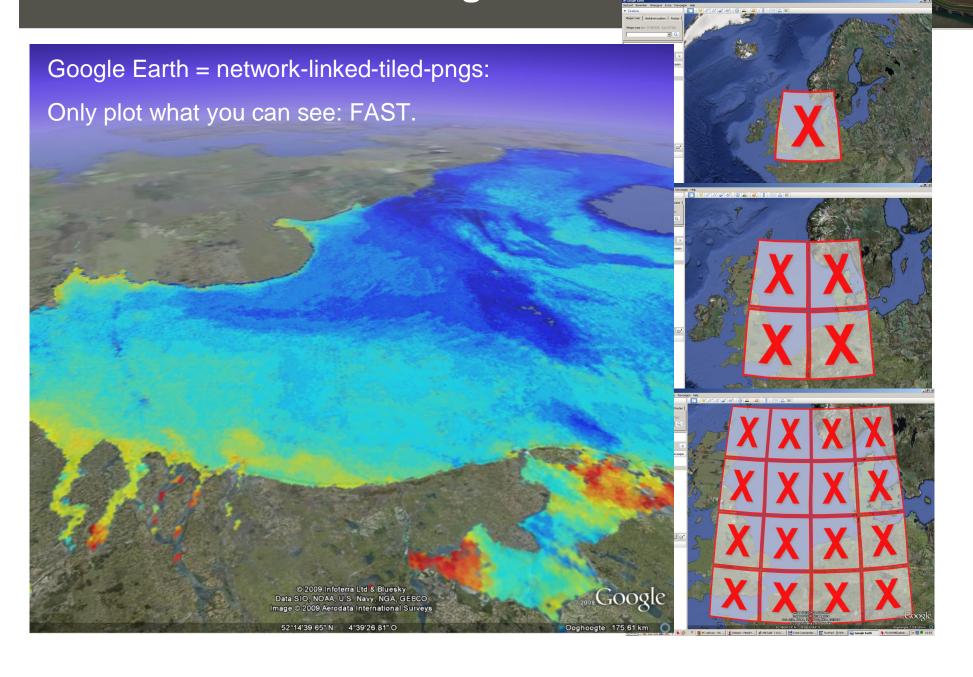




Google Earth

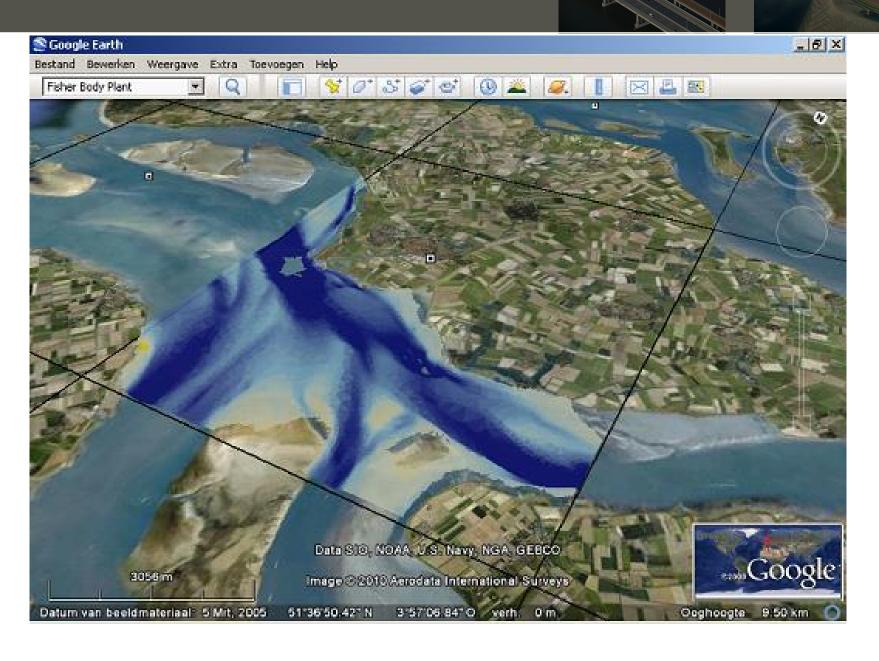








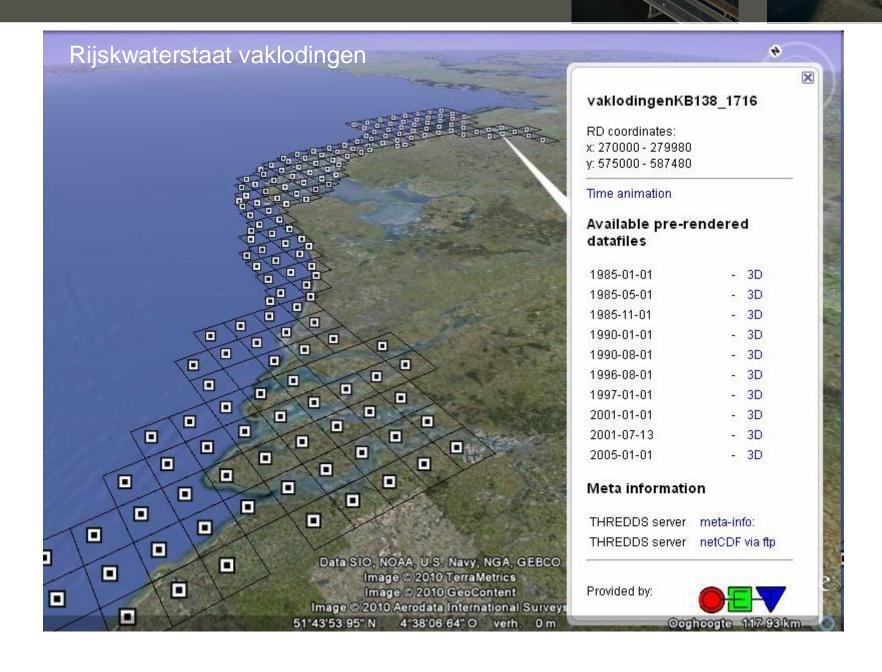
Google Earth can deeplink to other visualizations



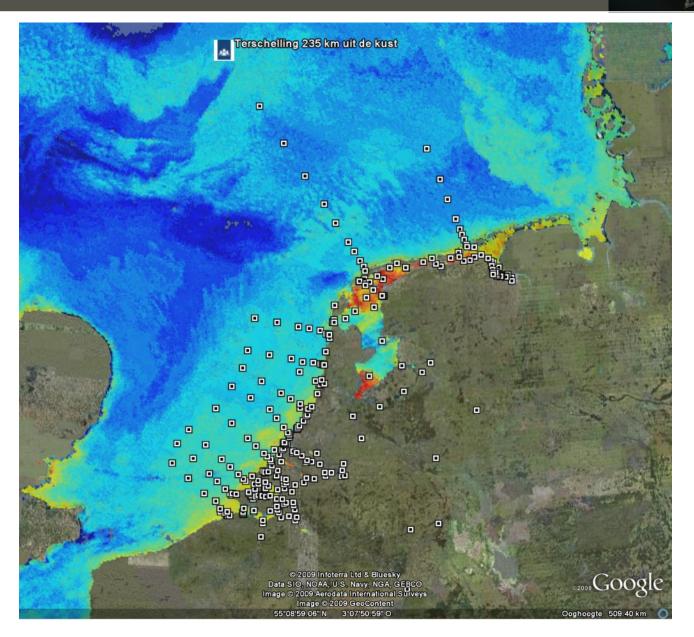
KML and other OGC web services

- OGC WxS: upcoming standard for web services for visualizations
 - WMS: grid images
 - WFS: polygons (shapefiles)
 - WCS: grid matrices
 - > no free viewer yet, so we just wait
 - > no simple server yet, so we just wait
 - > limited creativity: experts needed to set-up portal
 - > 2D (why is it that GIS people keep denying Galileo ...)
- OGC KML (Keyhole Markup Language, bought by Google)
 - > free viewer already ... with all aerial pictures as bonus
 - > simple file server, hardly expert needed
 - > ample creativity: simple syntax & description pop-ups: html+
 - > 3D (Google Earth only deals a round world, in globally accepted wgs84 roundishness)

Google Earth can deeplink to data



Google Earth can deeplink to visualizations & data

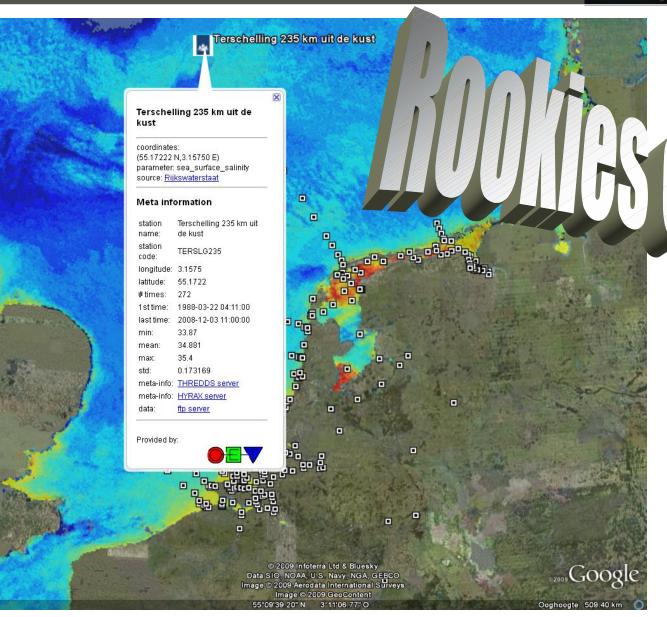


Instead of adapting viewers to data, adapt data to viewers.

World's best viewer is free, and has an open interface definition (kml) on how to put your own data in.

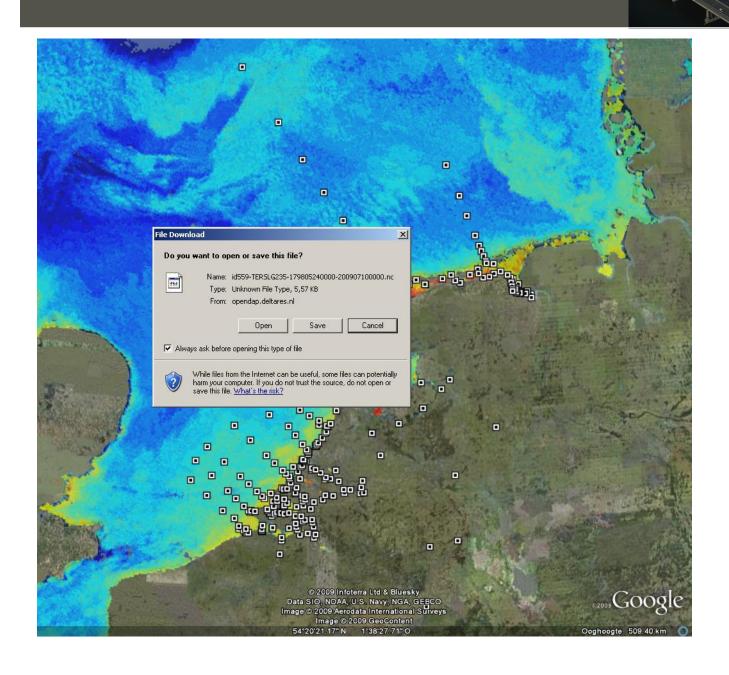
Put yout effort in making dedicated kml files, instead of building ever more expensive viewers.

Google Earth can deeplink to visualizations & data





Google Earth can deeplink to visualizations & data



Google

The

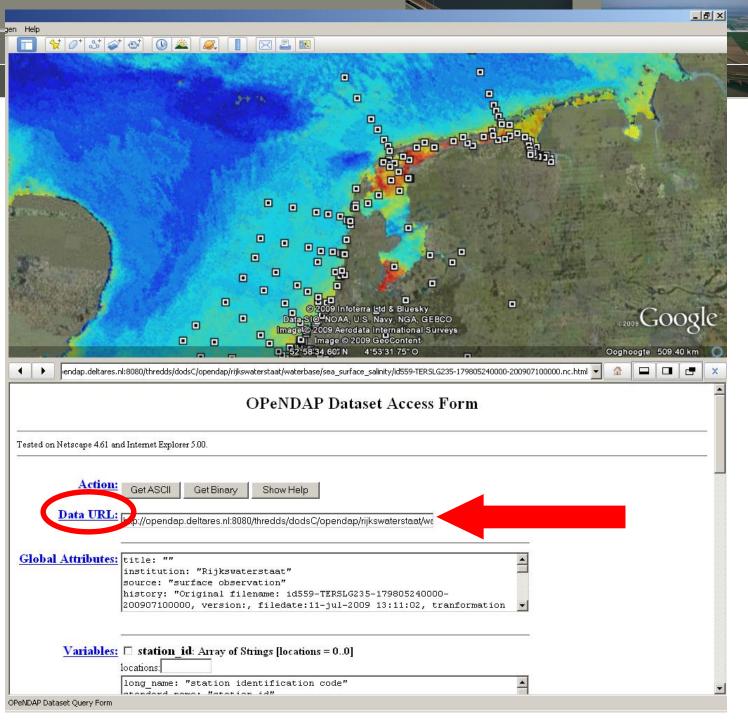
Data URL

can be

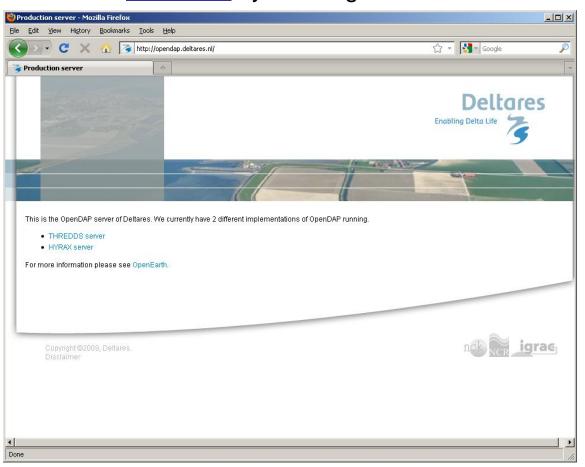
used

in various

clients

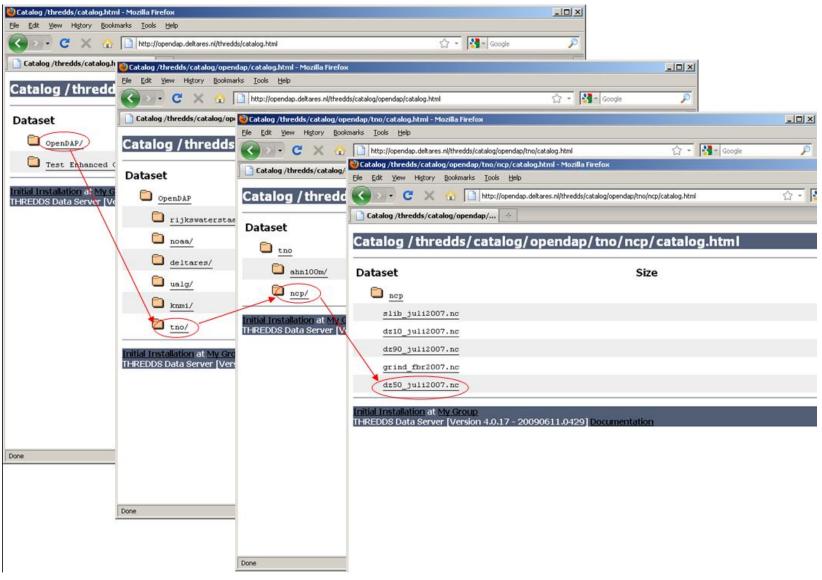


You can also obtain the this **Data URL** by scrolling down an OPeNDAP server.



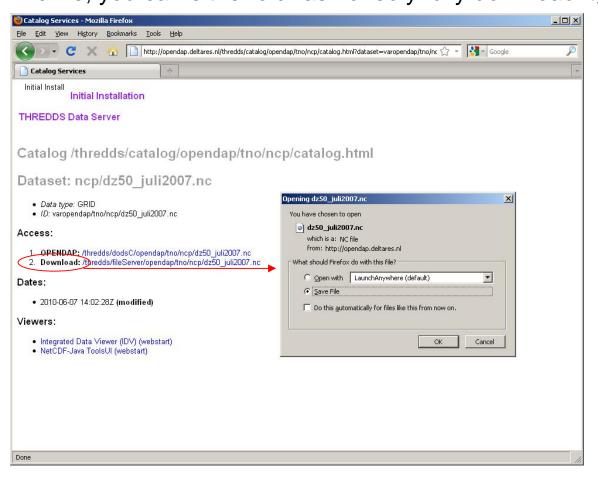
e.g. opendap.deltares.nl

The OPeNDAP server hosts an underlying directory structure, just click.

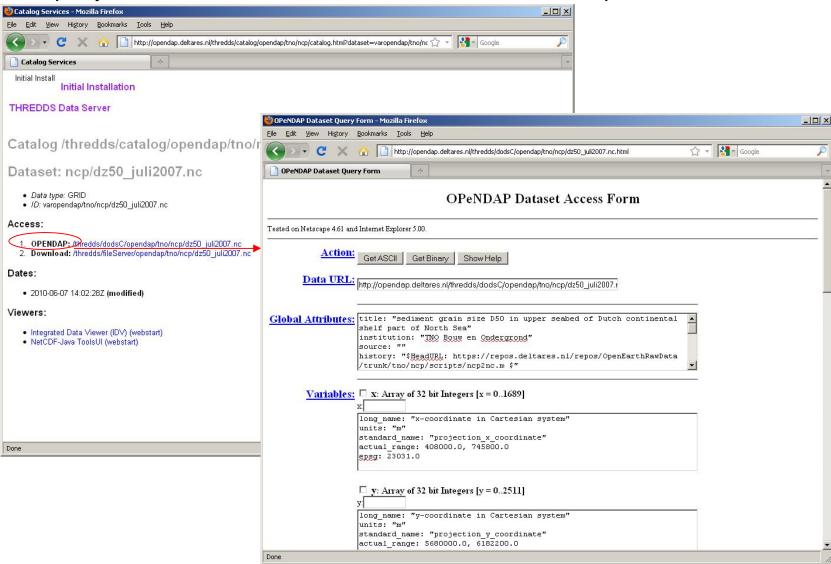


NCP example: download

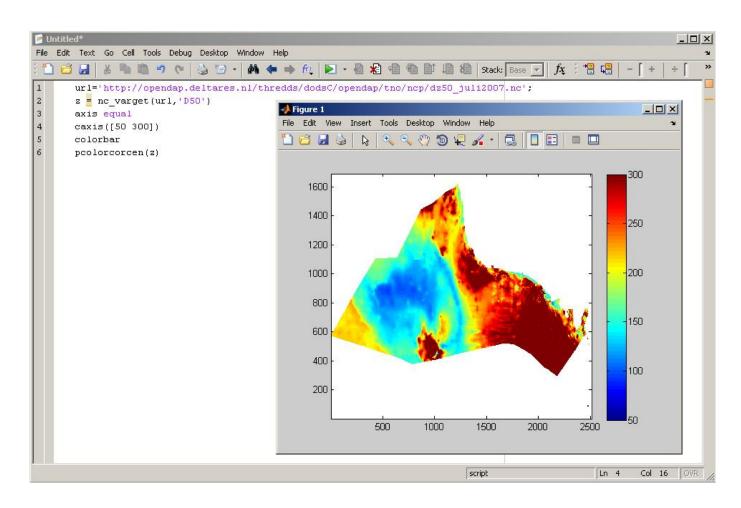
Once you hit an *.nc file, you can either old-fashionedly fully download it, or ...



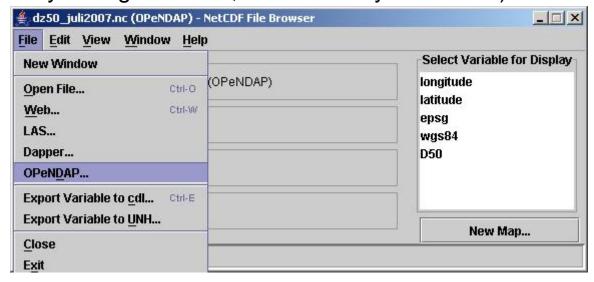
.. or you can query it's meta-info, or subset it, with the OPeNDAP protocol and...



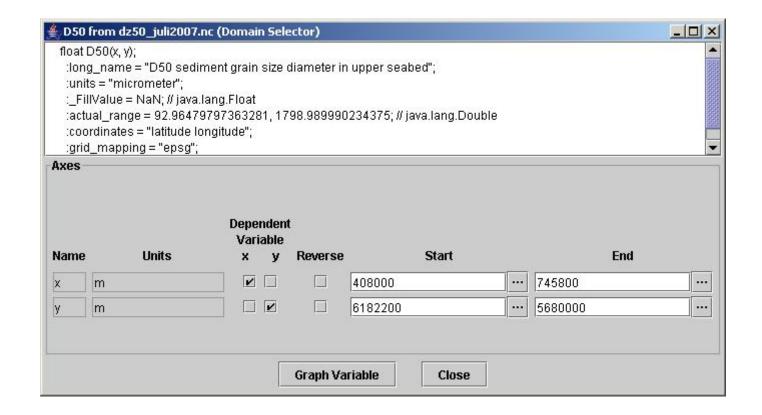
.. use the <u>Data URL</u> in OPeNDAP-enabled netCDF tools (recompilation is only change needed, software stays identical.).



.. use the <u>Data URL</u> in OPeNDAP-enabled netCDF tools (recompilation is only change needed, software stays identical.).







Alas, data is too biog in ncBrowse, better use Matlab, python, R, ...

OpenEarth infrastructure

Web service for data (huge)





tools



raw data



netCDF data



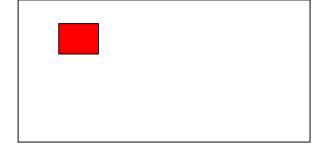


Google Earth

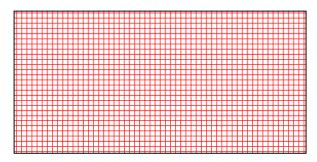
Unprecedented: all data at your fingertips

- Google Earth
 - web with all the worlds aerial images on a just-what-you-need basis
- OPeNDAP
 - web with all the worlds data on a just-what-you-need basis

> sub-region

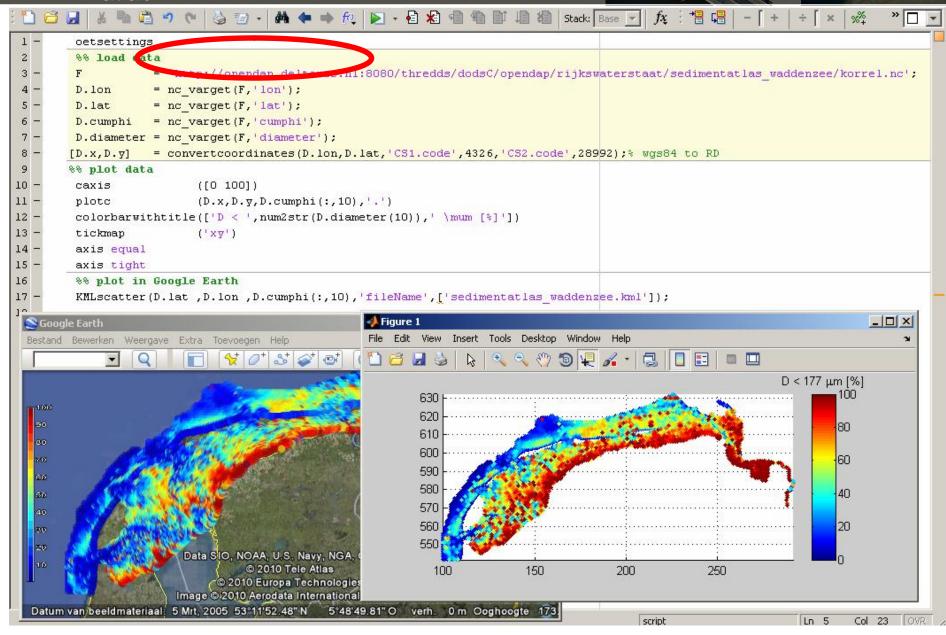


> sub-resolution

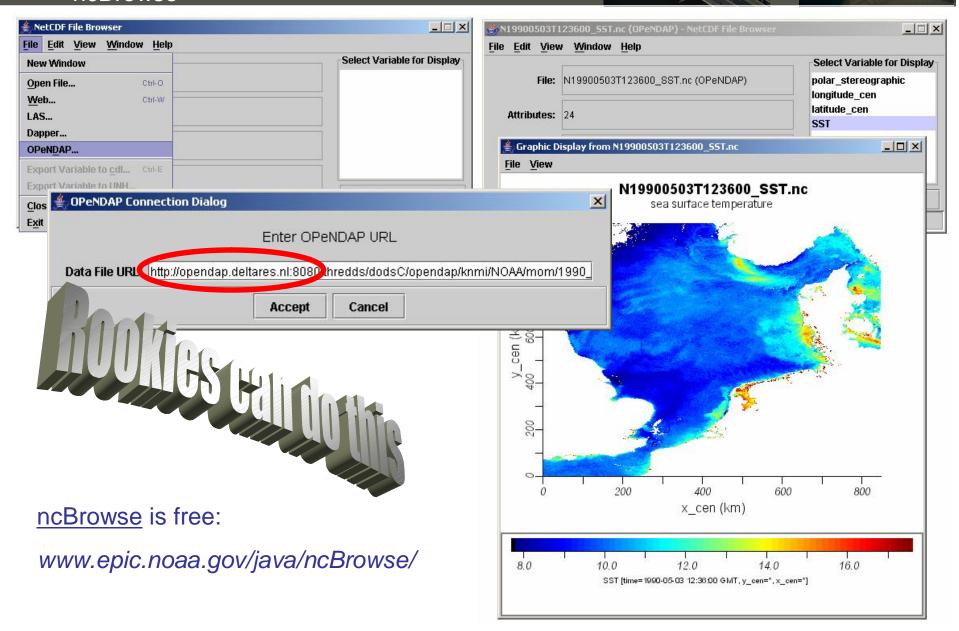


Delft3D-Quickplot Delft3D-QUICKPLOT _ | X **Delft3D Quickplot** File Macro Window Help Open File... Open URL... Diff Files... Delft3D-QUICKPLOT File Macro Window Help Figure No. 1 Eile Edit 9 90 http://opendap.deltares.nl:8080/ /id1-H0EKVHLD-17980524 🔻 point 1 Domain station identification code station identification code 3 station name station longitude station latitude time sea surface height sea surface height (m) → 19 M range and N range K range FAI 1 □ All F All Add to Plot Quick View Define Var. -3 L 1900 1920 1960 1980 1940 2000 time → ▶ V

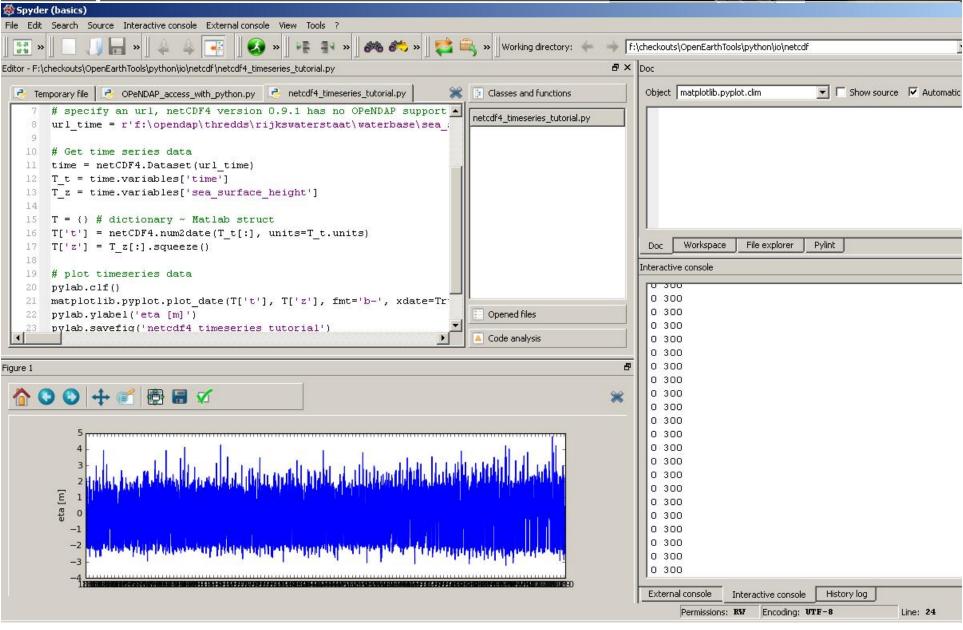
Matlab



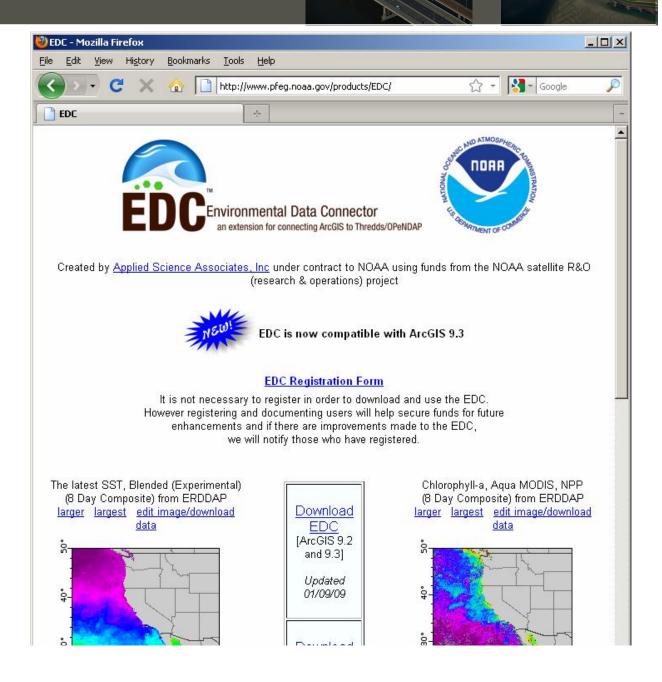
ncBrowse



PythonXY

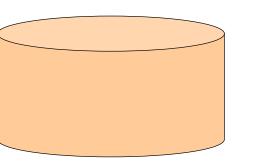


EDC for arcGIS 9.2+



OPeNDAP vs RDBMS: no fight, use both

Relational Database
 RDBMS / SQL



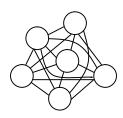
- lots of relations
- not web-based
- central filling
- central access point

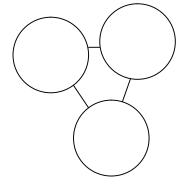




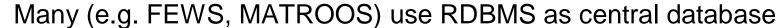


few relations
web-based
decentralized filling
distributed access





OPeNDAP vs RDBMS: RDBMS is basis



- Clean-up strategy keeps RDBMS limited in size
- Increasing demand for keeping full archive for calibration/validation
- Limitations of RDBMS with huge data sizes (grids) (TBs)
- Performance conflicts with completeness

RDBMS

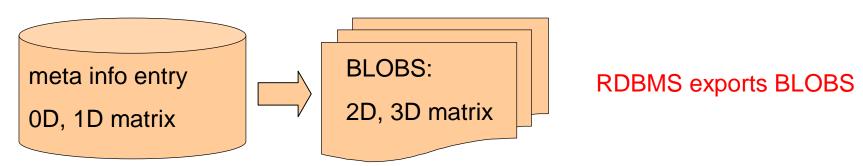
SQL etc

RDBMS contains **BLOBS**

OPENDAP vs RDBMS: export BLOBS



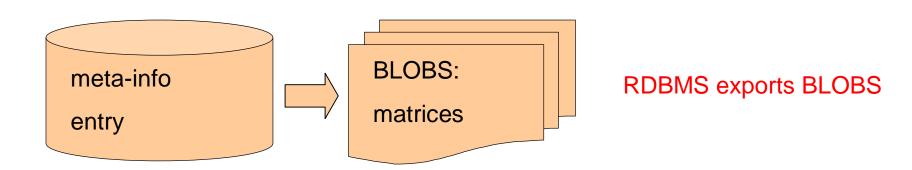
- export BLOBS from RDBMS to file server
- BLOB = Binary Large OBject, e.g. 2D of 3D matrix
- keep entry in RDBMS for fast query
- BLOB server for all separate BLOBS
- light-weight RDBMS remains
- Basic architecture of RWS-MATROOS
- Possibilities explored for FEWS (talked with NOAA-NWS)



OPENDAP vs RDBMS: export BLOBS

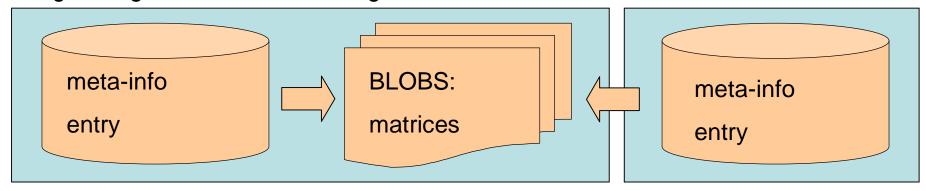
OPeNDAP BLOB server: difference with e.g. ftp BLOB server

- standardized BLOBS: netCDF (NASA standard)
- <u>aggregate</u> BLOBS to one virtual big BLOB (slices > sausage)
 - make from 2D (x,y) grid time series a 3D (x,y,t) grid
- <u>subset</u>
 <u>BLOBS</u> and only requets what you need
 - a la Google Earth: only get what you need (FEWS in Africa...)
 - a slice or stick from a 3D model grid
 - a patch from a big 2D grid (NL from European weather model)

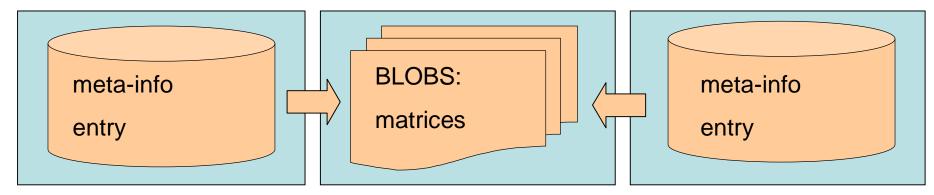


OPeNDAP vs RDBMS: fast vs. complete

a light-weight installation co-using another ones BLOB archive server.

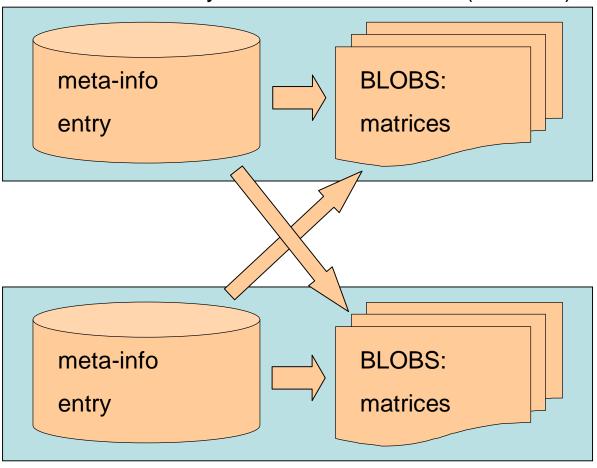


a central BLOB archive server hosted for for multiple light-weight installations



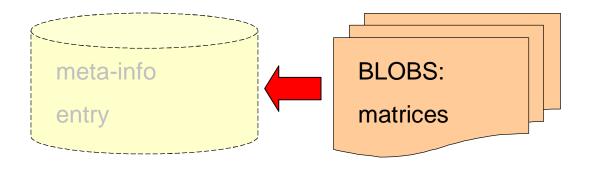
OPeNDAP vs RDBMS: fast vs. complete

mirror use of heavy BLOB archive servers (> 10 TB): redundancy



OPENDAP vs RDBMS: index BLOBS into SQL

- RDBMS: difficult when data is very pluriform as in:
- www.OpenEarth.nl
 - RWS historische tijdseries waterbase
 - KNMI historische tijdseries wind en meteo
 - RWS historische grids lodingen
 - KNMI historische satelliet beelden
 - AHN100
 - TNO NCP data
 - •



RDBMS imports meta-info: cache results of harvesting scripts into RDBMS

OpenEarth infrastructure

- 1. Web service for tools, raw data and model schematizations:
 - provides paramount QA/QC for:
 - write access for everyone with moderators
- 2. webservice for data: netCDF
 - editorial board only
- 3. webservice for graphics: Google Earth
 - editorial board only





tools















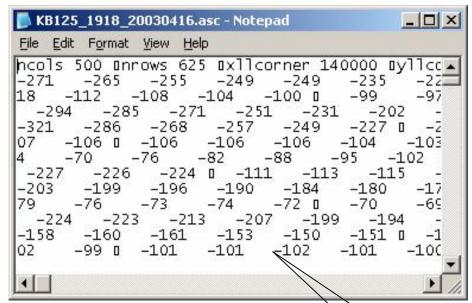
Google Earth

Every peanut butter jar has: date + factory batch no.



ood-info.net

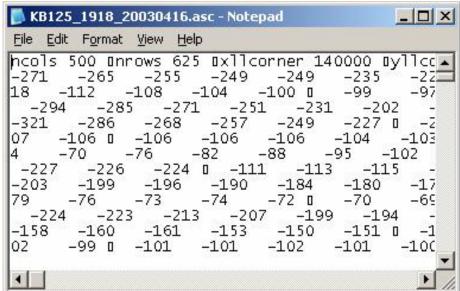
- date x
- batch no. x
- ingredients x
- calories x
- $\bullet \ e \ \mathsf{gram} \ \boldsymbol{\mathsf{x}}$



- date x
- svn no. x
- parameters x
- coordinatesx
- dimensions √

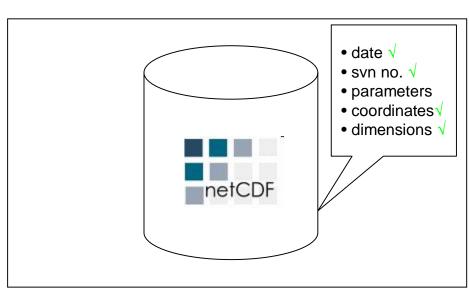
Every peanut butter jar has: date + factory batch no.





food-info.net





lloapp.gibodriehoek.be

1st principle: data = raw data + processing

NASA satellite data with open source SeaDas processing toolkit (in IDL)

- L0: dump of recorded voltages, only averaged over 16 pixels
- L1: voltages + satellite track
- L2 ~ physical quantities
- L3 ~ binned in space (1 grid instead of zillions of warped photos)
- L4 ~ binned in time (climatology)

version number

RAW DATA (volts)

History will never change!

one parameter at one place at one time

PROCESSING

Interpretation does change!

e.g. instrument deterioriation, recalibration

1st principle: data = raw data + processing

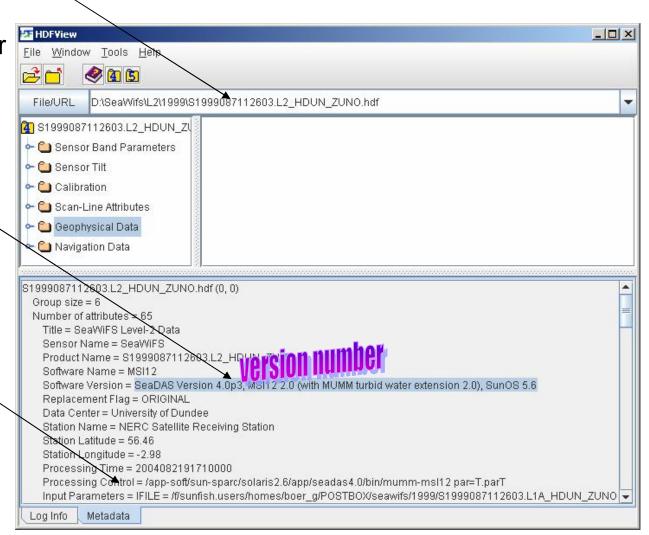
DATA: SeaWiFS L2 ocean color data product

=

TOOL: version number of open SeaDAS IDL toolbox

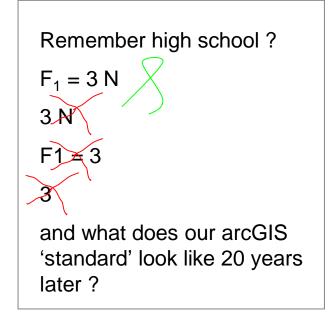
+

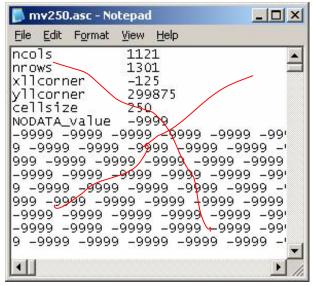
VOLTS: name of raw L1 input datafile



OpenEarth principles: data with meta-data

- ASCII file means nothing
- we need meta-information
- ... standardized
- at very least
 - quantity
 - units
- et voila, netCDF
 - allows for unlimited meta-information
 - offers machine-readable standards
 - vocabulary of quantity and units
 - can be loaded in any language/program
 - is open
 - designed for huge datasets
 - robust yet simple (NASA standard)



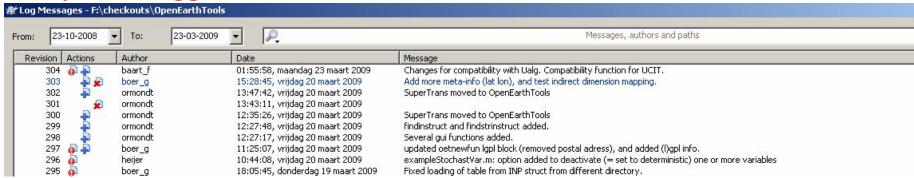


Repository username

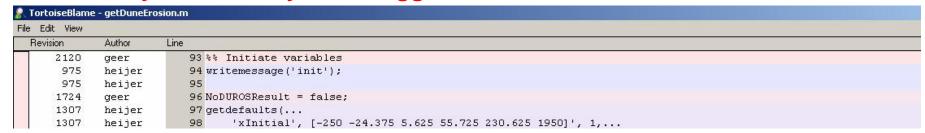
SUBVERSION

- Get <u>username and password.</u>
- Why, OpenEarth is open, right? Yes, but not anonymous.

Every file is logged ...



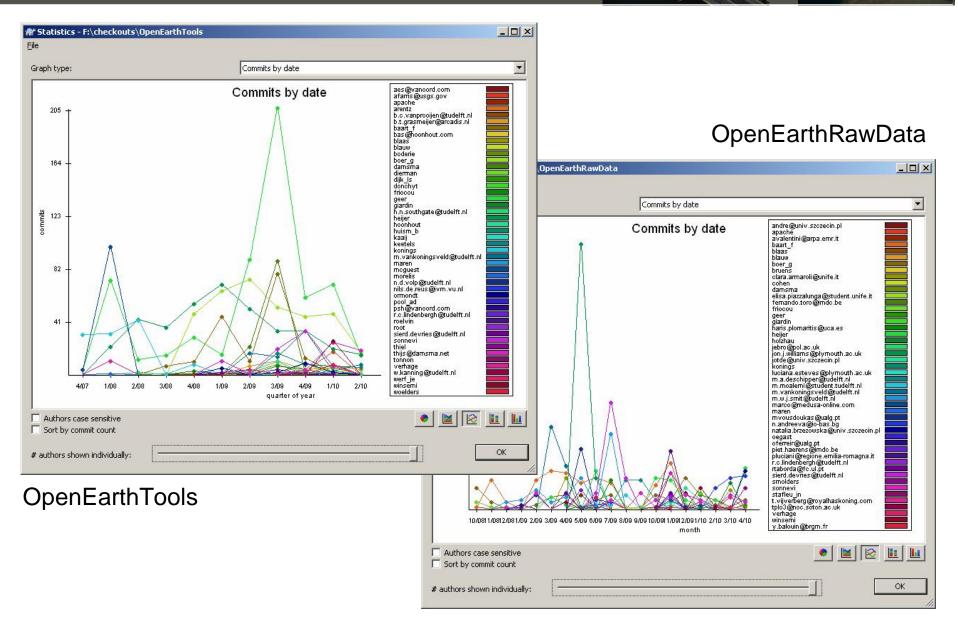
... and every line in every file is logged.



- Everyone has write acces to facilitate lots of micro-improvements.
- This allows anyone can be allowed to join.

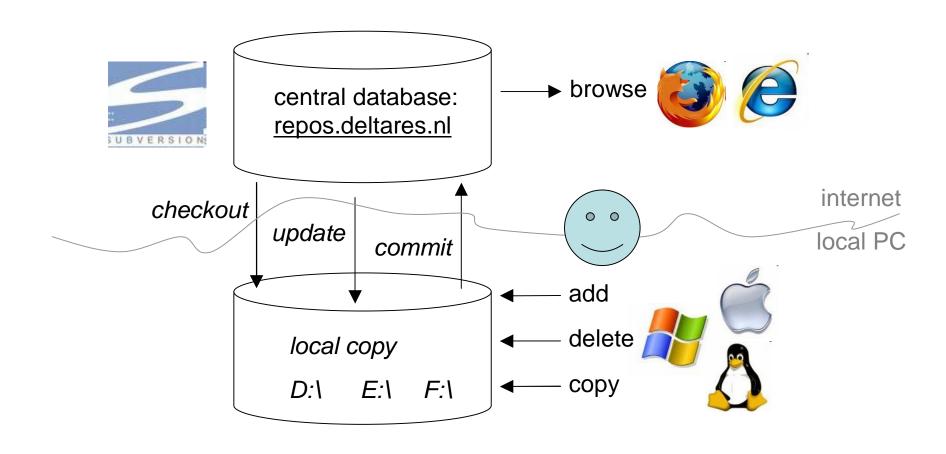
OpenEarth is already truly a community





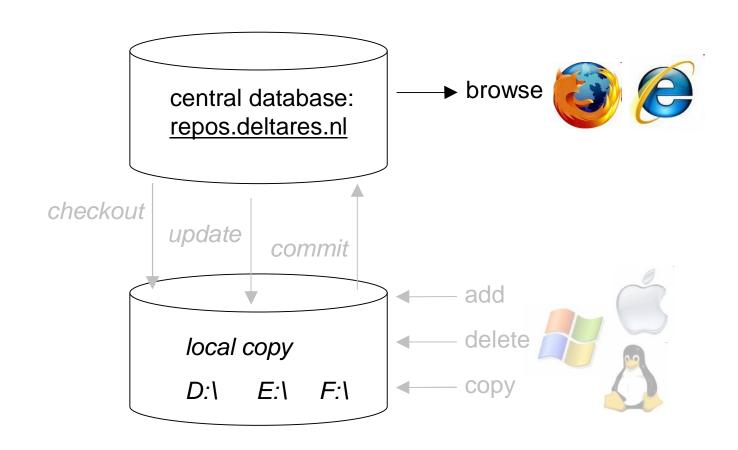
REPOSITORY basics





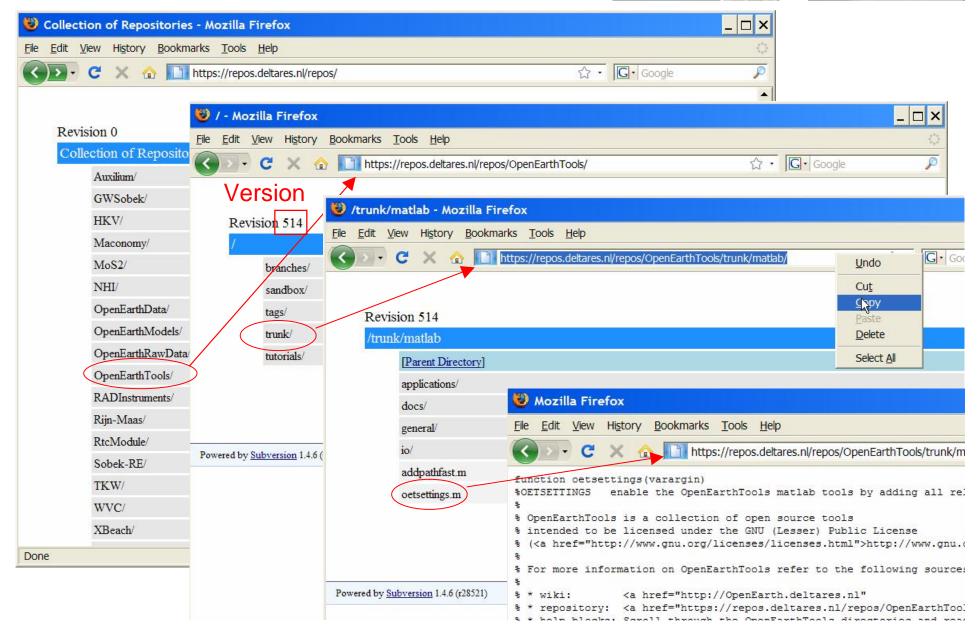
REPOSITORY browse





REPOSITORY browse

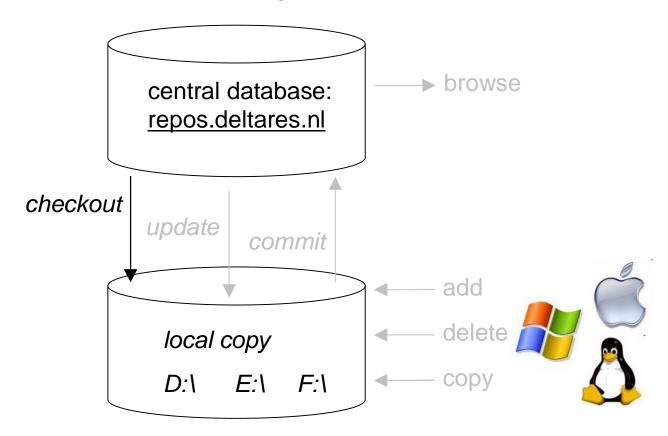




REPOSITORY & SVN Checkout...

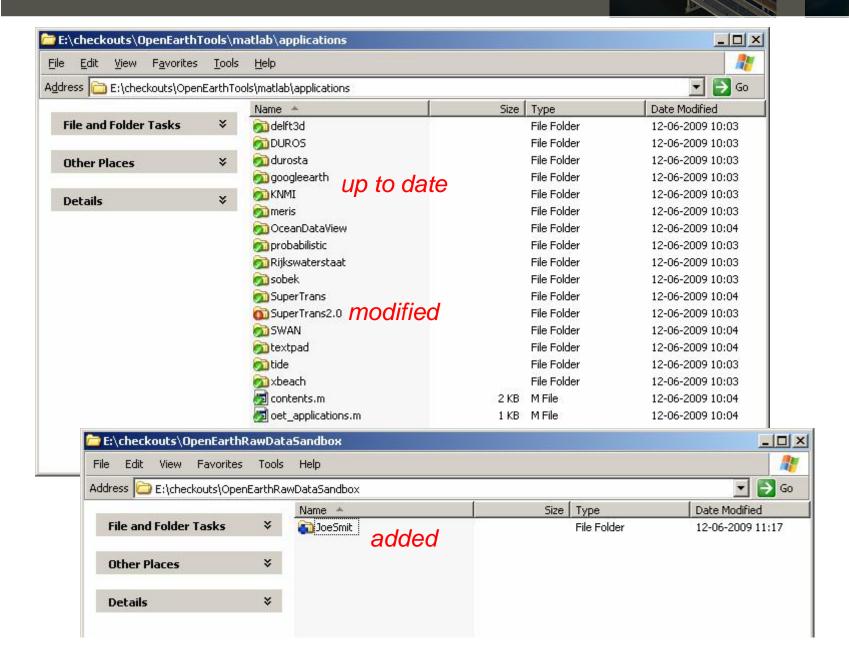


 It is not handy to get files one by one with a browser: Get them all at once with free Subversion client program.



Overlay icons show status of local copy



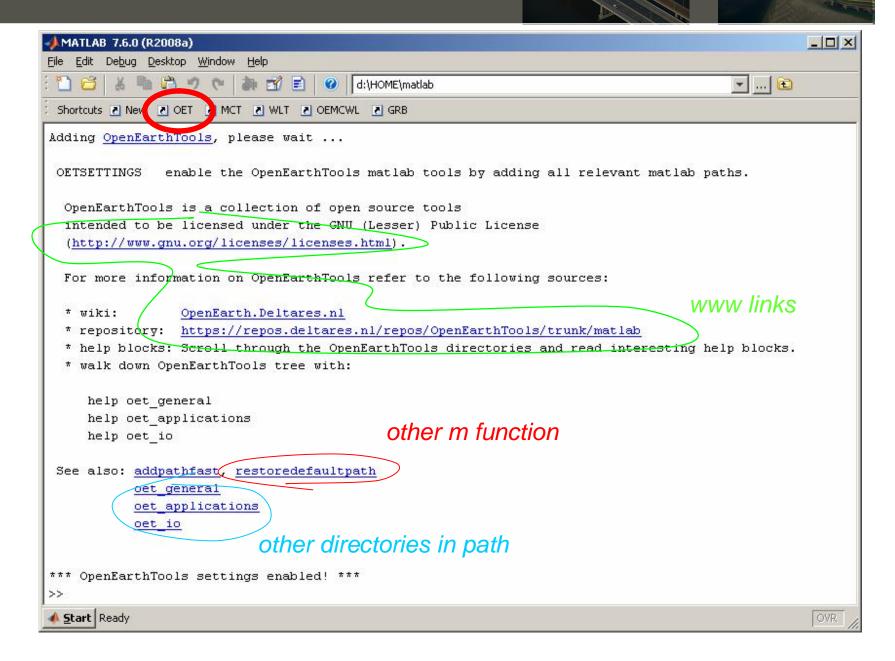


Tools protocol (Matlab)



- name space, precursor to Object Oriented
 - VS_*
 - swan_*
 - delft3d_*
 - nc_*
 - ucit *
- *_testfor unit tests
- *_tutorial
 for tutorials
- h1 line for table of contents
- example for cut-'n-pastable use
- alignment for colleagues
- setProperty for <keyword,value> pairs

Good Programming Practise: documentation



Good Programming Practise: documentation

>> help applications OpenEarth applications rijkswaterstaat - data: rijkswaterstaat data types (donar) ← contents.m delft3d - delftd file formats - read knmi wind and meteo timeseries knmi - swan toolbox swan >> help(rijkswaterstaat donar_read - read ASCII text file from www.waterbase.nl - get data from waterbase.nl getwaterbase getwaterbasestation - get stations names for parameter H1 line - get parameter from waterbase.nl getwaterbaseparameter - read meteo file from hmcz@ rws hmcz_read >> help donar_read donar read.m DONAR_READ) read ASCII text file from www.wate DONAR READ read ASCII text file from DAT = donar read(fname, <keyword, value>) % DAT = donar read(fname, < keyword, val)</pre> See also: getwaterbase, www.waterbase.nl % See also: getwaterbase, www.waterbas

recap

Contents

O. The power of collaboration

1. Google Earth: web service for images

2. OPeNDAP/netCDF: web service for data

3. Subversion: web service for version control



Version control and backup

tools data & scripts



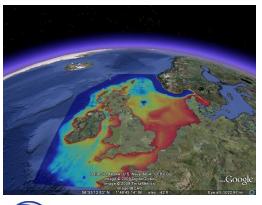
Web access to datasets

from pluriform to standard format!

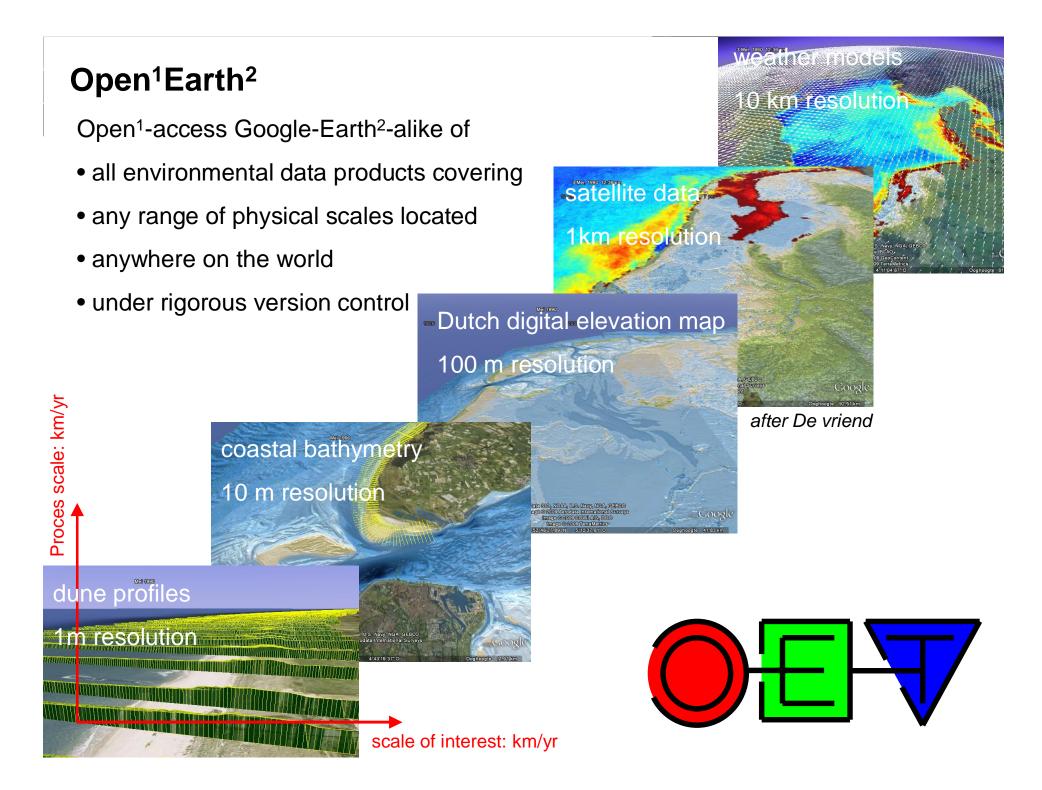


3 Straightforward visualisation

open source software















Community of practice: OpenEarth has a wide range of users (Building with Nature, EU FP7 MICORE, Delft Cluster etc.). Building with Nature facilitates training sessions.