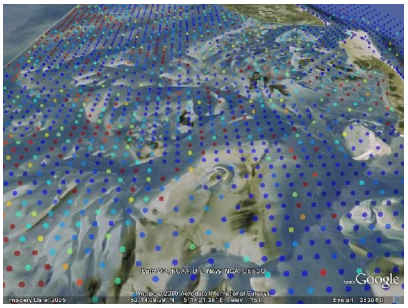
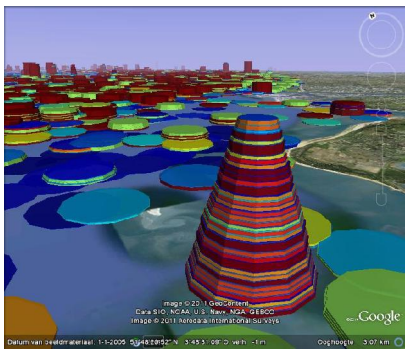
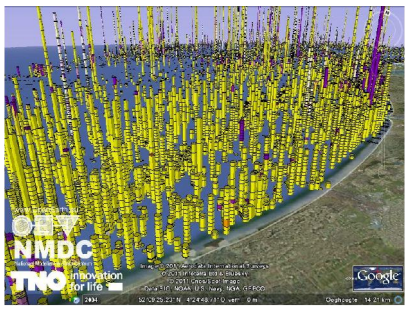
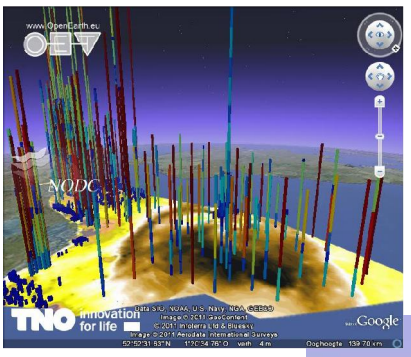
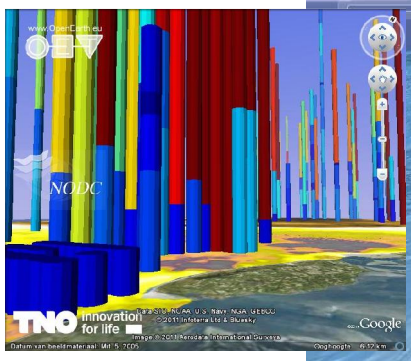
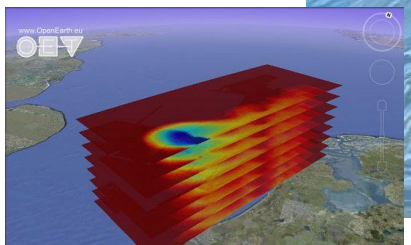












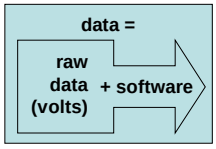
www.openearth.eu

OpenEarth is an open source workflow for management of data, model and tools. OpenEarth strives for 3 paradigm shifts in coastal + marine science & technology:

1. We think that all **raw data** and scripts should be kept under **version control**, an aspect known as provenance. OpenEarth advocates using the open source **SubVersion** system to store not only raw data but **also processing tools** under version control. We adhere to the **Wikipedia approach** to quality control: we advocate **crowd sourcing** (web 2.0) by allowing anyone to help us improve data by adapting the processing.

How **NASA** does it: Processing raw data (volts, counts) into data products is cumbersome, cutting-edge science where we are prone to make some errors. Fortunately we can fix these errors over time, resulting in different versions of the data though. To avoid a version hell where scientists work on outdated versions, data should have rigorous version numbers. This requires strict distinction between raw data, processing and data products. Raw data (volts, counts) will never change (history doesn't change after all). But the processing software or scripts do change (new insights & coefficients). Data versioning is therefore in fact versioning of our scripts.

NASA adopted this strict distinction early on. They label raw data L0 and keep that forever, frozen. At regular intervals they re-apply the latest version of open source software to L0 to produce data product levels (L1 to L4). These products have a version number, so previous versions can be deleted destroyed when new are available, but regenerated if needed with older script-version.



2. Data should be open and accessible via live **web services**. The concept of versions of data implies that local copies of datasets should be avoided. They should either be considered as cache: temporary working copies to be deleted asap or subject to automatic an updating system. But the best solution is not to have local copies any more at all. OpenEarth adopted the open source **netCDF-CF / OPeNDAP** system where users can access datasets live via the web. We introduce this as **DataTube** because it's like YouTube: use data just by streaming them live via the web, directly from within Matlab, Python or R.

3. Laymen should be able to view all our data and model results. We think that free **Google Earth** is the most advanced viewer that in fact **everyone can use**. We made an open source community toolbox in Matlab OpenEarthTools that allows scientists and engineers to plot all their data easily in **KML format**. KML is the open standard that underpins Google Earth. We regularly organize hands-on sprint sessions to explain this toolbox.