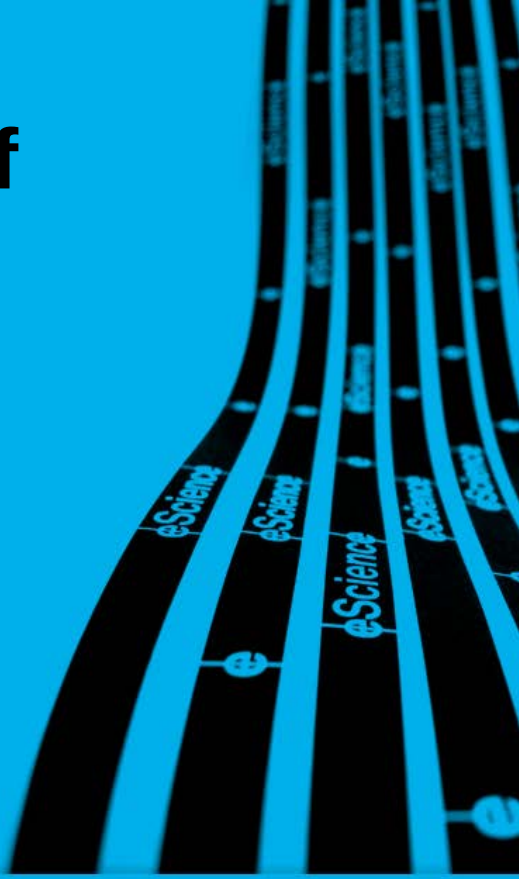
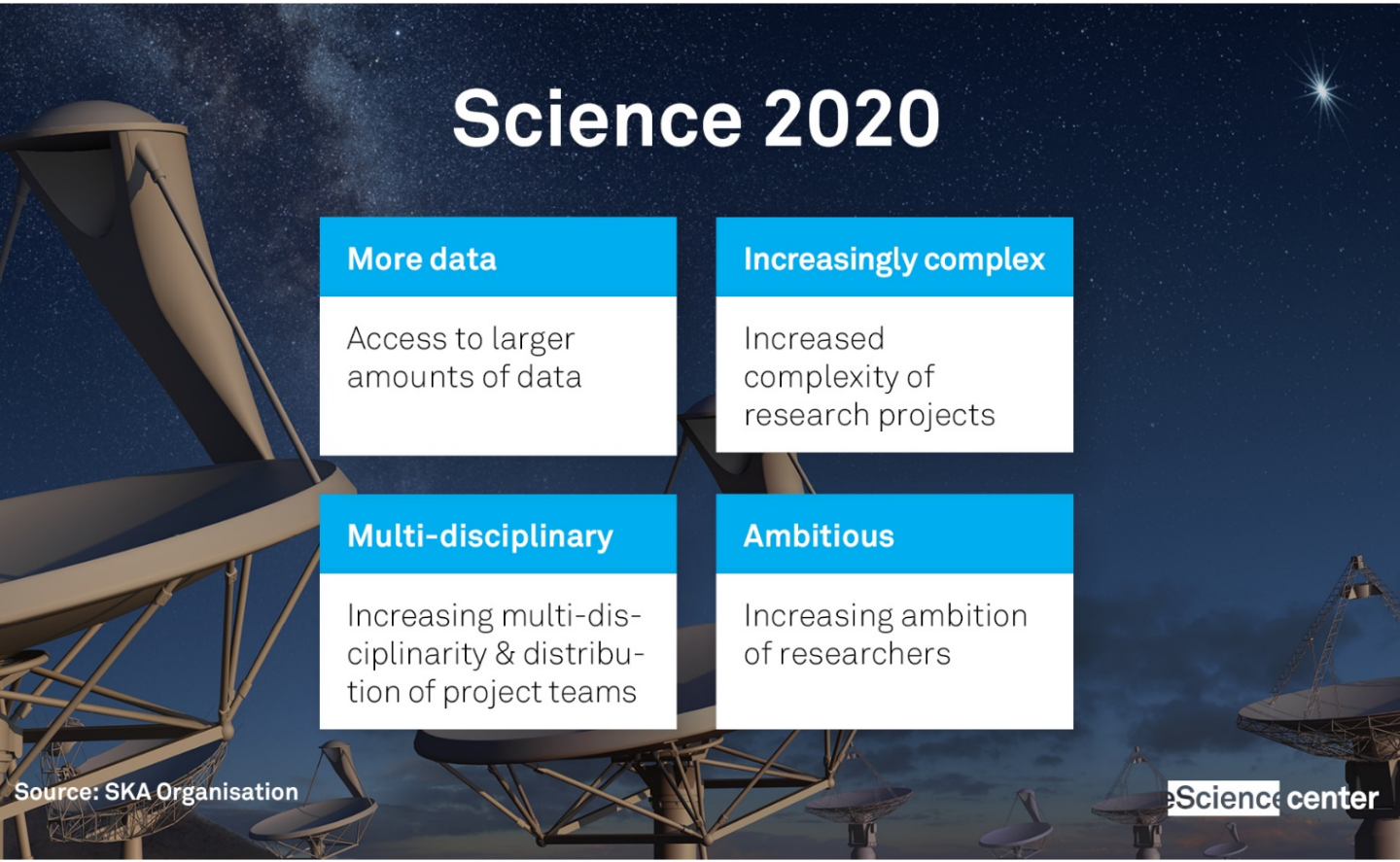
Getting science out of eScience

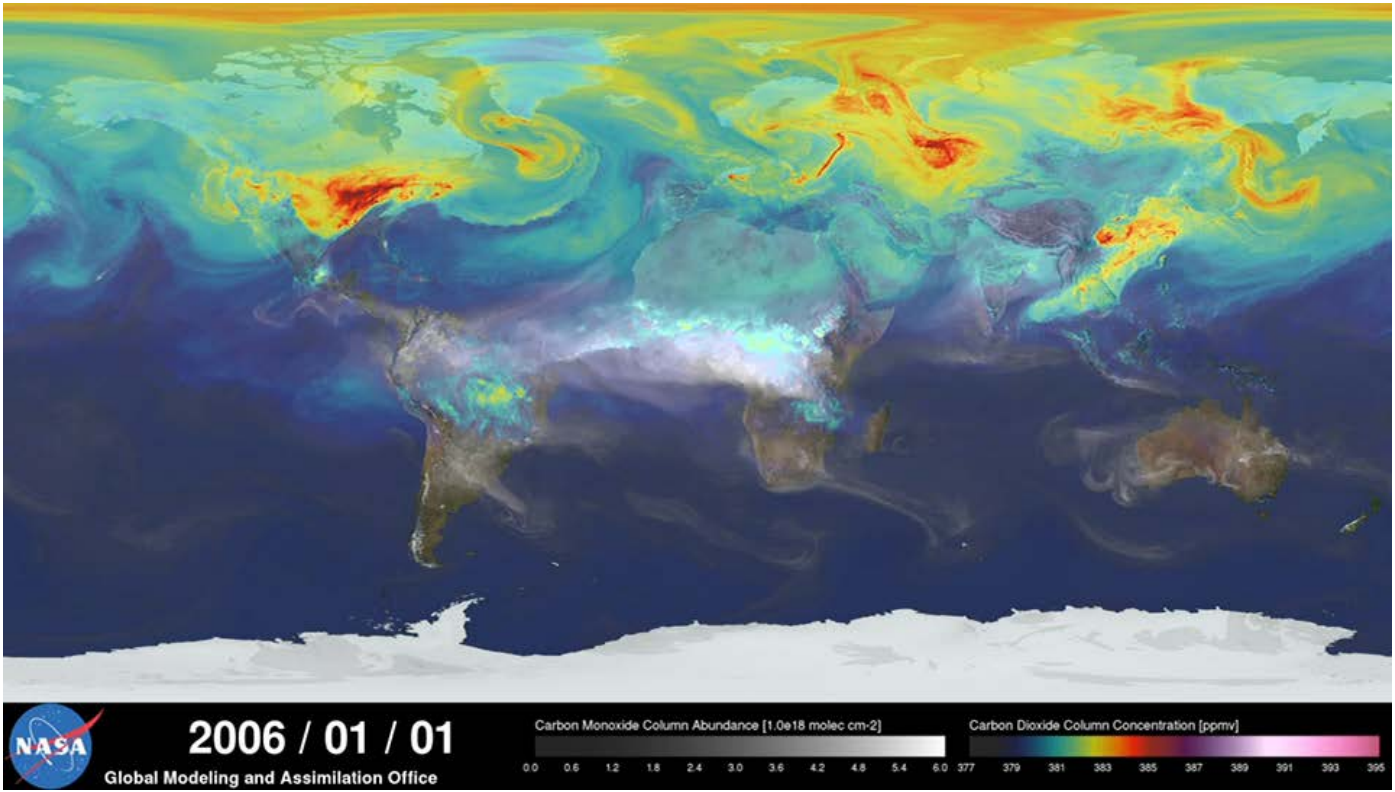


Wilco Hazeleger

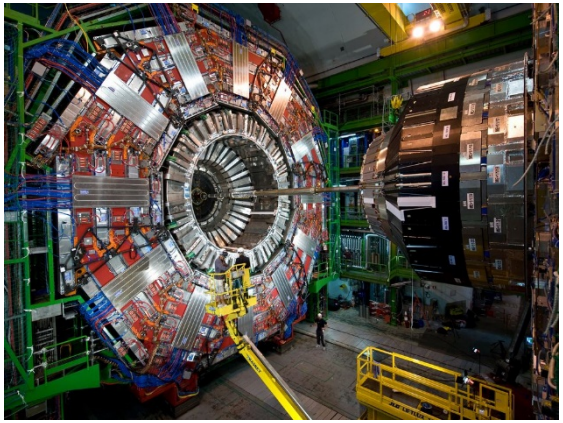
netherlands eScience center

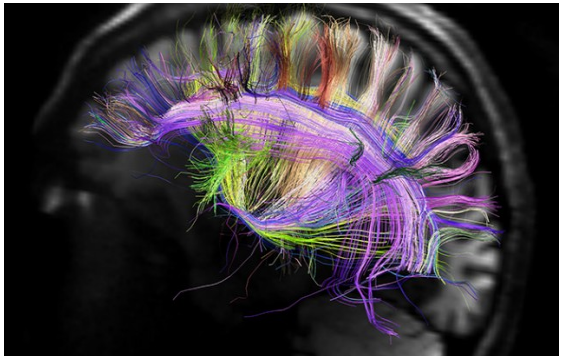
by SURF & NWO


Science 2020
More data
Access to larger amounts of data
Increasingly complex
Increased complexity of research projects
Multi-disciplinary
Increasing multi-disciplinarity & distribution of project teams
Ambitious
Increasing ambition of researchers
Source: SKA Organisation
eScience center

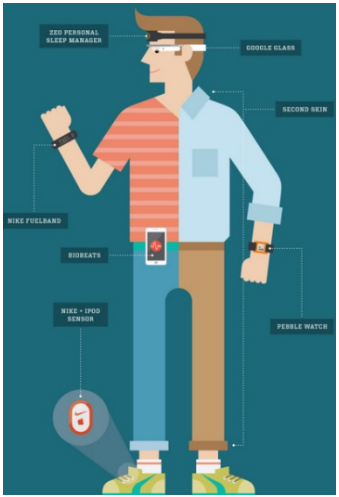

NASA
2006 / 01 / 01
Global Modeling and Assimilation Office
Carbon Monoxide Column Abundance [1.0e18 molec cm-2]
0.0 0.6 1.2 1.8 2.4 3.0 3.6 4.2 4.8 5.4 6.0
Carbon Dioxide Column Concentration [ppmv]
377 379 381 383 385 387 389 391 393 395

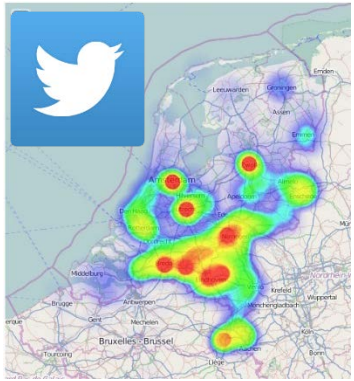
The ‘Big Science’ era





The other ‘Big Data’



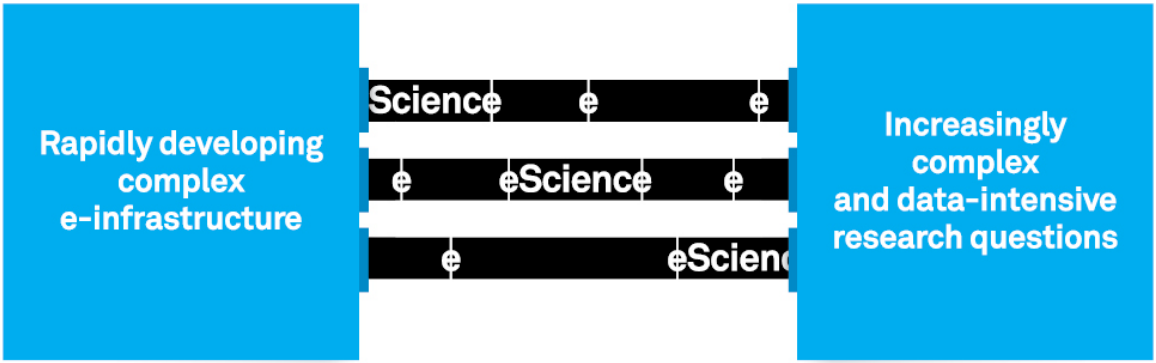


Grand challenges

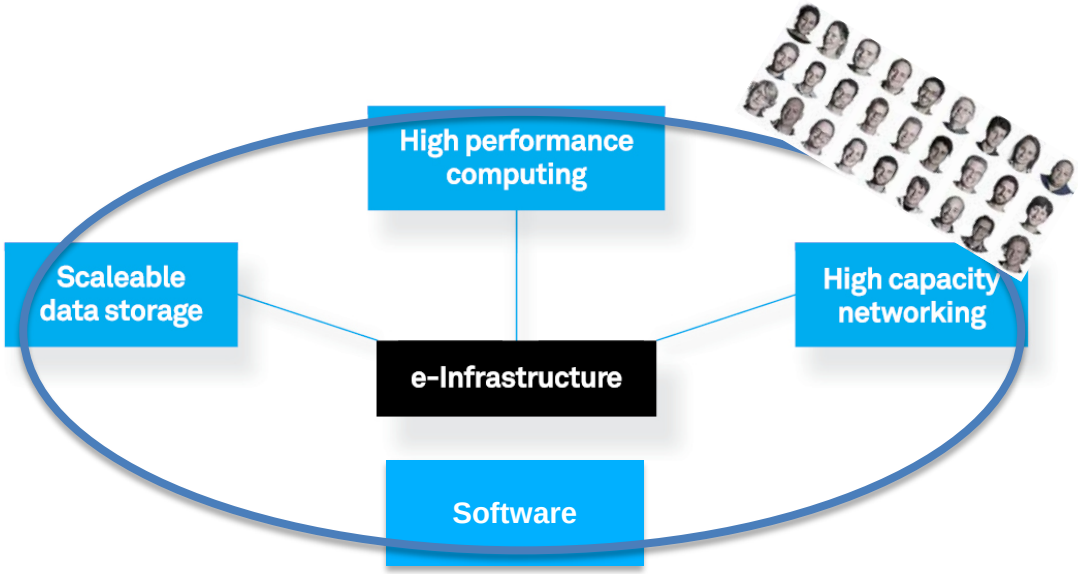
- **‘Big science’ driven by ‘big societal’ challenges**
  - **Multi- inter-, transdisciplinary research (*health & well being, sustainability & energy, society & economy*)**
  - **Nearly always a large ICT and data component**

- **Science itself ‘big’ and complex**
  - **Big data (volume, complexity, real time sensor data), computing, analytics**
  - **Interactions between systems and scales**

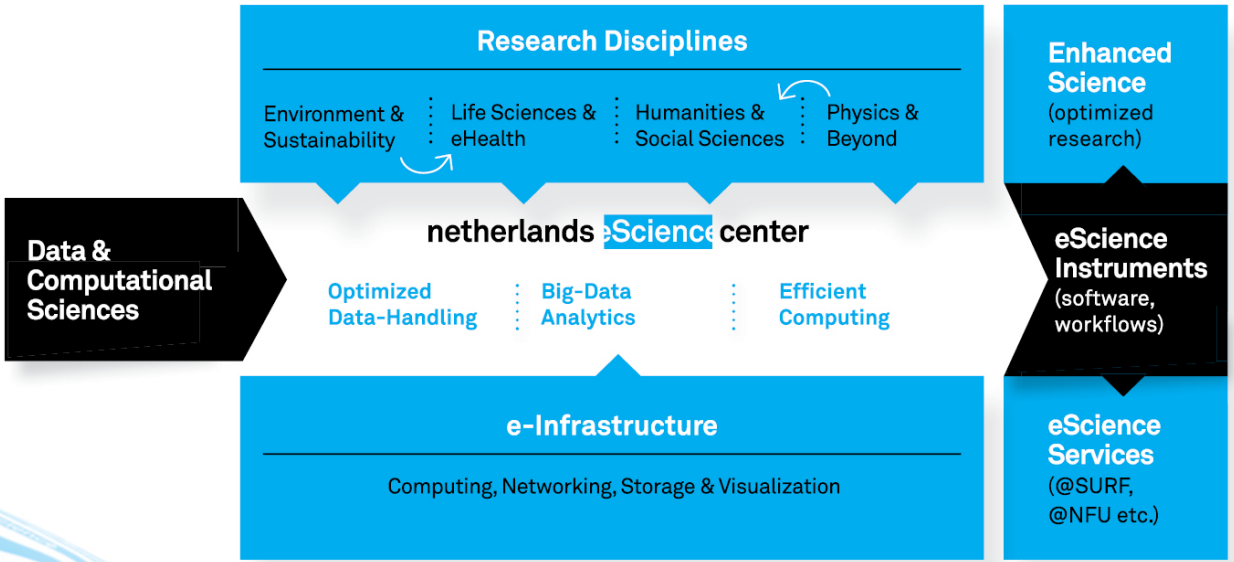
Why eScience



Making the power of the e-infrastructure accessible



How we work



eScience center

How we work

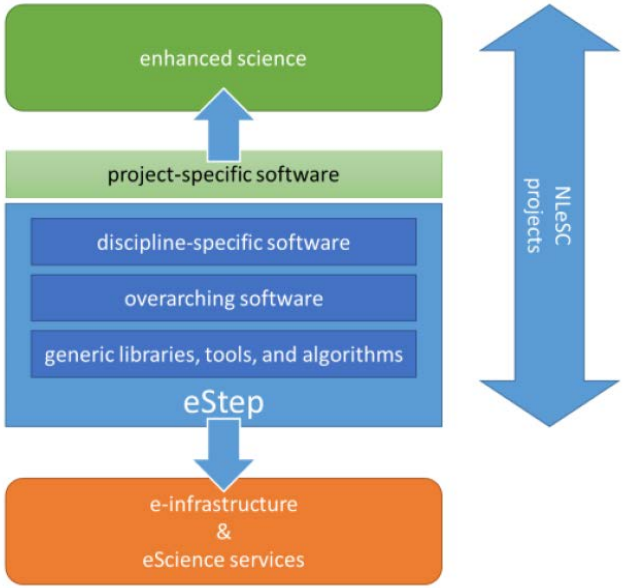
- **eScience Research Engineers**
- **Partnership with domain scientist**
- **We fund projects (‘in cash’ and ‘in kind’)**
- **Public private partnerships**
- **National and international coordination (PLAN-E for Europe)**

- **Generic *eScience Technology Platform, eSTeP* (software, open access, knowledge basis, training in e-skills)**

Some characteristics of eScience technology

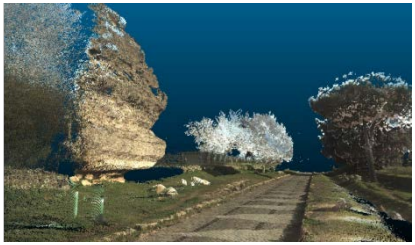
- **Versatile, overarching applicability**
- **Code quality and best practices**
- **Integration of software**
- **Analytics (incl visualization)**
- **Scaling of software**

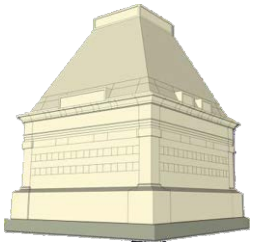
→ Tailored to domains



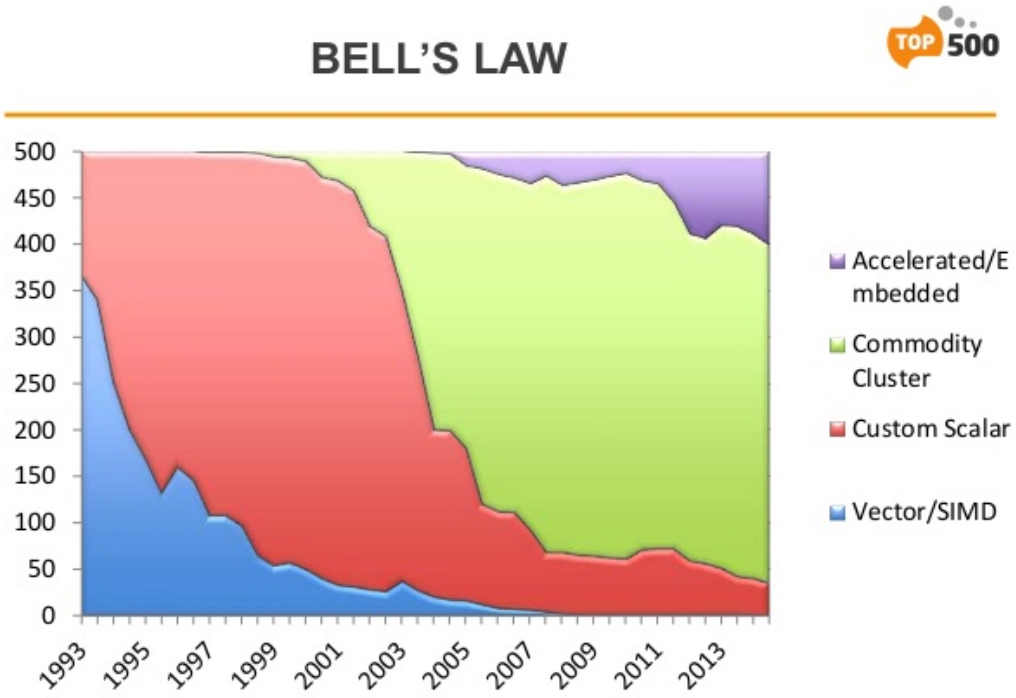
eScience as enabler (4D GIS in archeology)



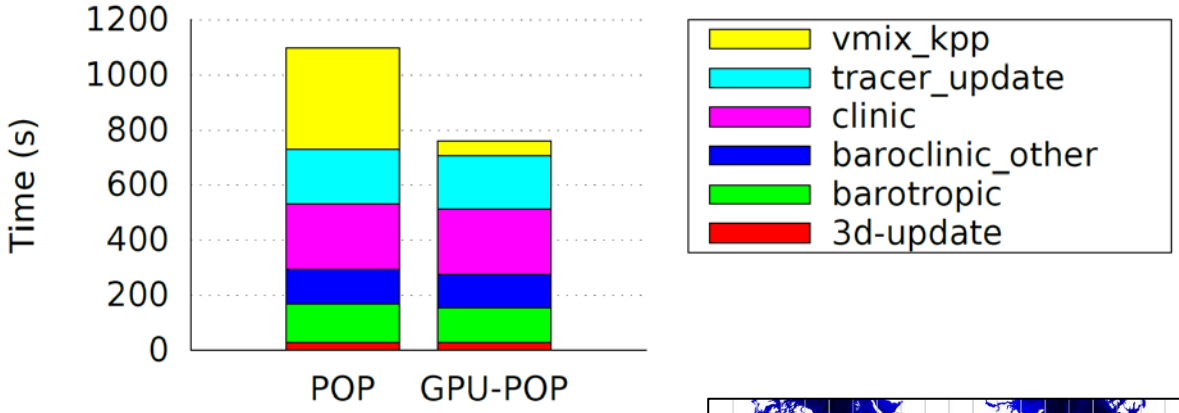


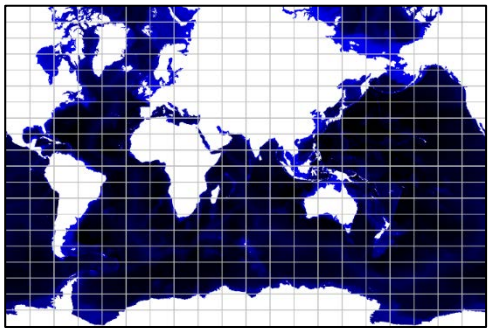




eScience as accelarator

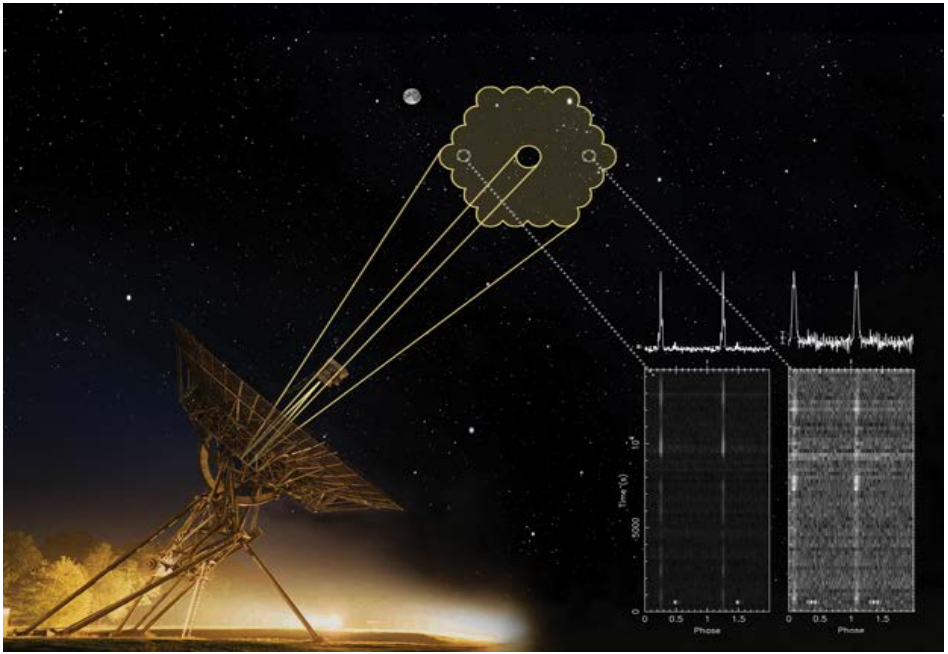


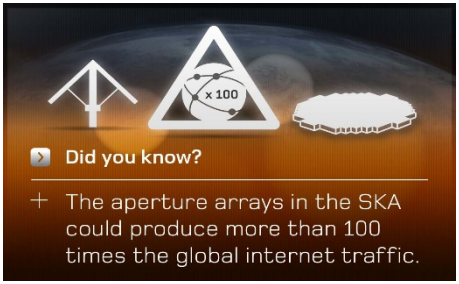
eScience as accelarator





eScience as enabler and accelerator





Paralellization of pulsar search pipeline

- **Folding instead of FFT for period search**

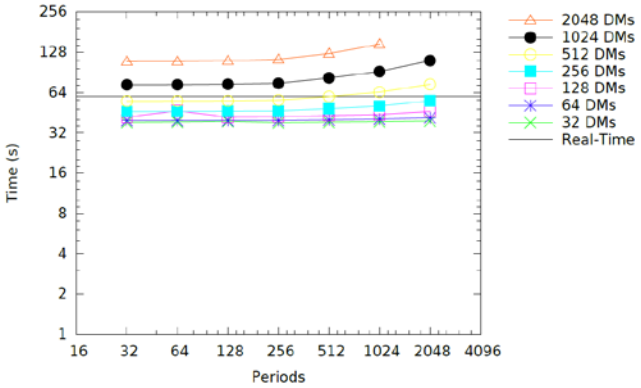


Fig. 5. Pipeline performance for the AMD HD7970, SKA1 scenario.

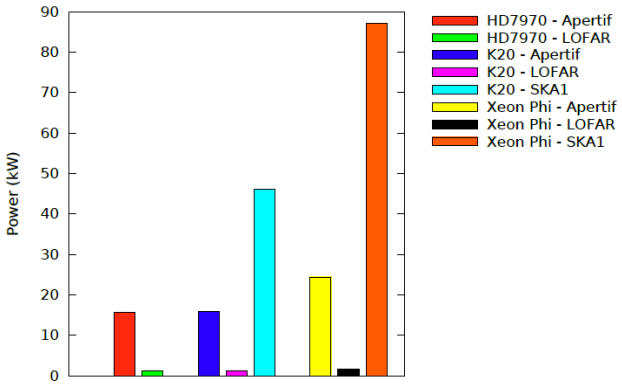


Fig. 12. Total consumed power, $2,048 \times 2,048$ case.

Sclocco et al 2015, IEEE eScience

eScience as enabler and accelerator: ‘Embodied’ emoties

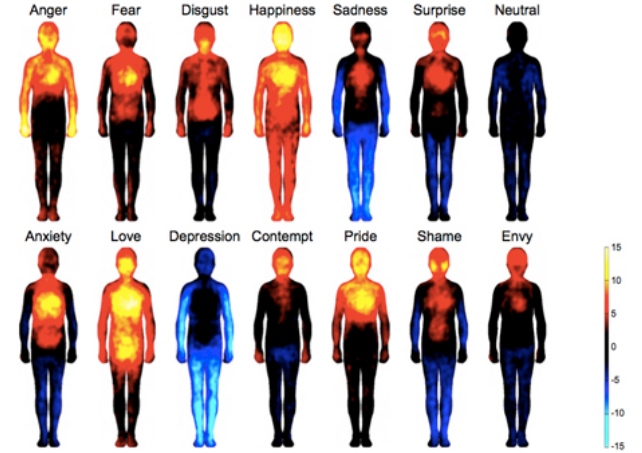
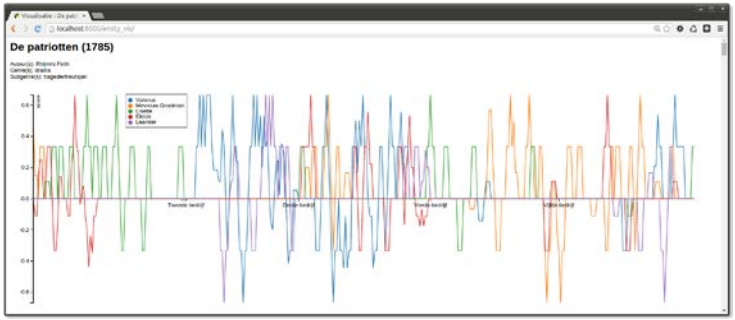
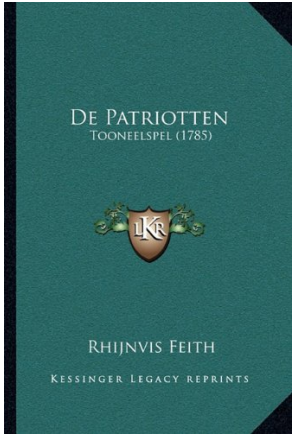


Fig. 2. Bodily topography of basic (*Upper*) and nonbasic (*Lower*) emotions associated with words. The body maps show regions whose activation increased (warm colors) or decreased (cool colors) when feeling each emotion. ($P < 0.05$ FDR corrected; $t > 1.94$). The colorbar indicates the *t*-statistic range.

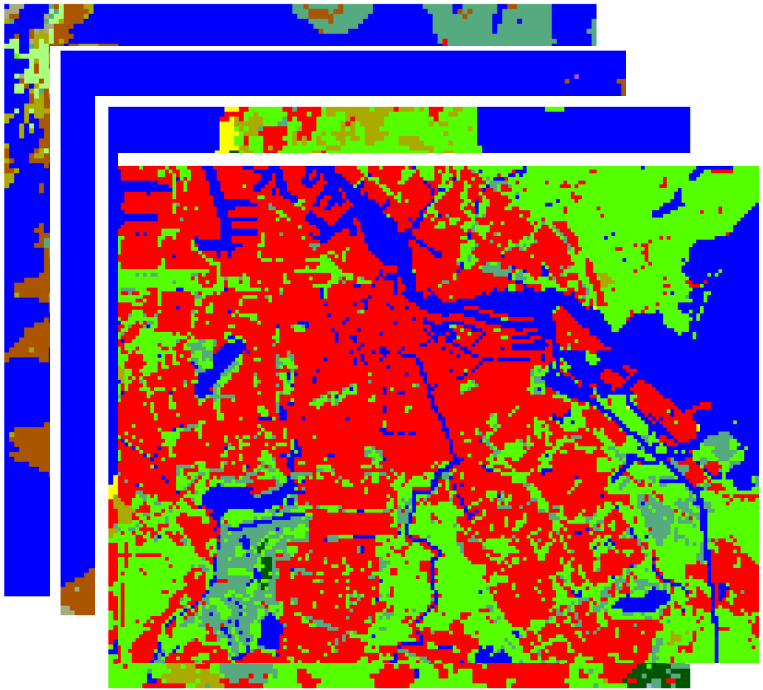




Computing and data challenge

https://www.esciencecenter.nl/project/summer-in-the-city

Computing and data challenge



Domain 1: 12.5km
default setup

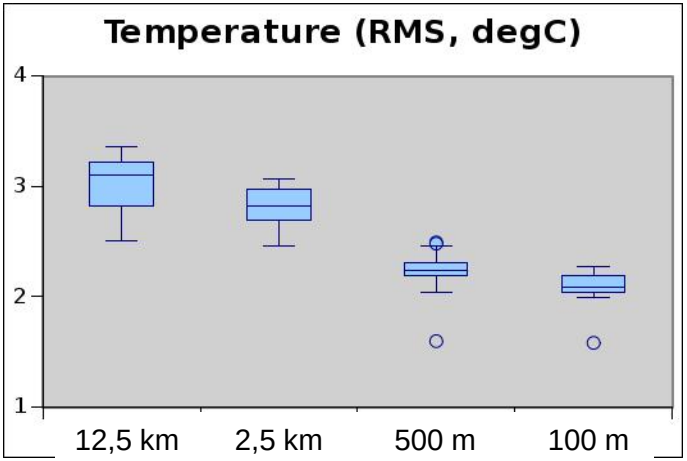
Domain 2: 2.5km
default setup

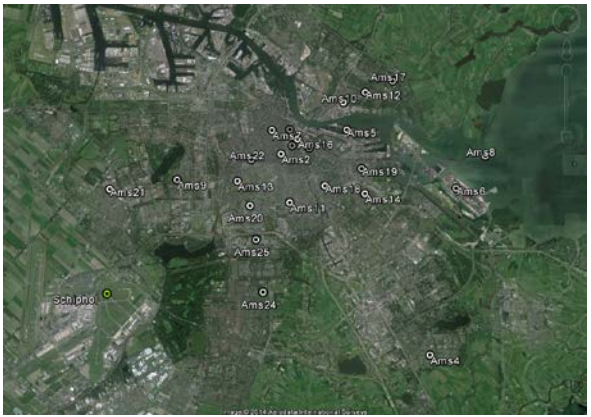
Domain 3: 500m
hi-res landuse, ec.
Rijkswaterstaat river
temperatures

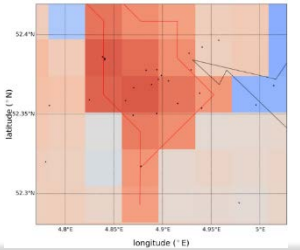
Domain 4: 100m
Rijkswaterstaat river
temperatures,
TOP10NL, satellite
imagery, AHN2
(height map), CBS
data

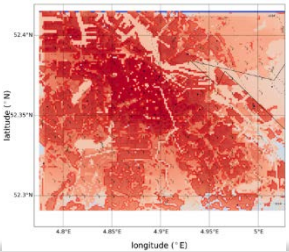
Daily forecasts
WRF3.5 + urban module (SLUCM)
48 hour runs, 24 hour spin-up

Impact of resolution and land use (summer 2014)









- **eScience**
  - **bridges e-Infrastructure and scientific research**
  - **crucially takes a domain science perspective**
  - **multidisciplinary scientists and engineers**
  - **Overarching technologies & cross-fertilization**

- **Common themes**
  - **Data stewardship**
  - **Software quality & sustainability**
  - **e-Skills (human capital)**

  → Open science

Up for discussion

- **e-Infrastructure for scientific research:**
  - **Data, data, data**
  - **Increase use commercial cloud services**
  - **Large infrastructures develop own specialized e-infra**

- **Tools closer to science problem:**
  - **Scalable software solutions**
  - **Data anfalytics close to the (distributed, real time, …) data**
  - **Workflow tools, scheduling etc etc.**

- **Need for strong human capital agenda (eScientists)**
- **Continuous need for innovation in e-Infrastructures**
- **Relation with large infrastructures (e.g. ESFRI)?**

www.eScienceCenter.nl

netherlands eScience center

by SURF & NWO