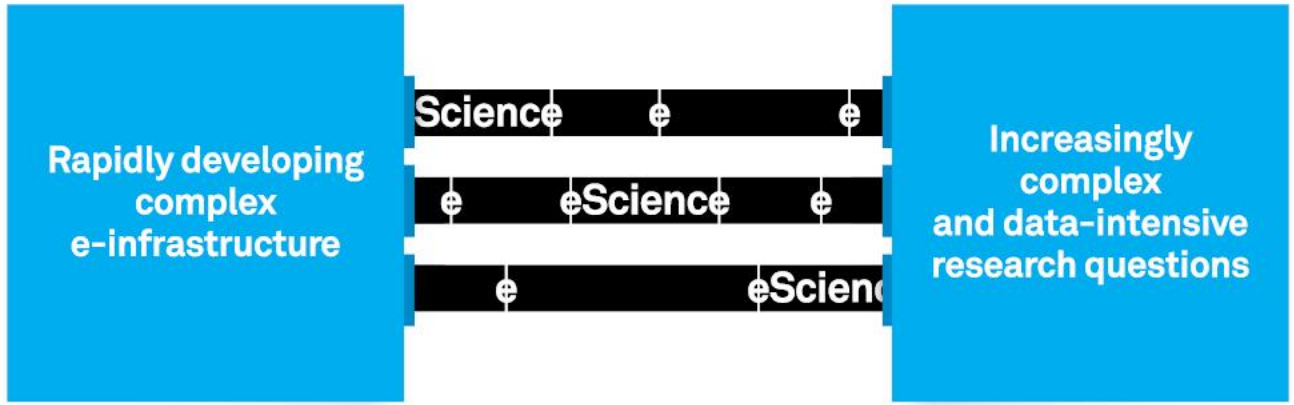
netherlands eScience center

Enabling digitally enhanced research

by SURF & NWO

Why eScience



eScience center

Priority domains

Humanities & Social Sciences

incl. SMART cities, text analysis, creative technologies

Physics & Beyond

incl. astronomy, high-energy physics, advanced materials

Sustainability & Environment

incl. climate, ecology, energy, logistics, water management

Life Sciences & eHealth

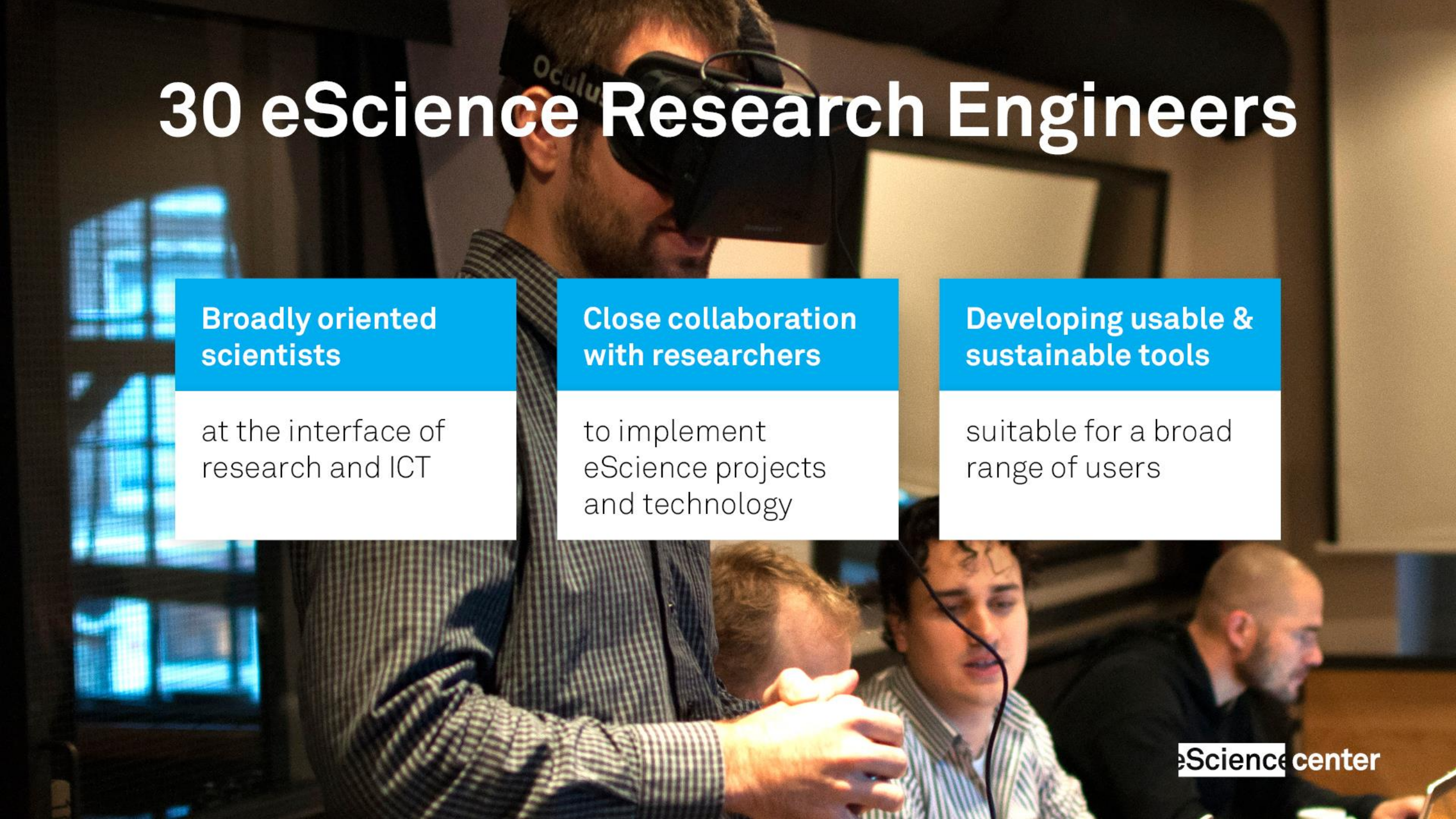
incl. bio-imaging, next generation sequencing, molecules

eScience center

We develop domain-overarching expertise & software

Optimized Data Handling	Big Data Analytics	Efficient Computing
e.g. Sensor Networks	e.g. Natural Language Processing	e.g. Distributed & Heterogeneous Computing

eScience center



30 eScience Research Engineers

Broadly oriented scientists

at the interface of research and ICT

Close collaboration with researchers

to implement eScience projects and technology

Developing usable & sustainable tools

suitable for a broad range of users

eScience center

eScience Technology Platform

Management & Dissemination

of software & work-flows employed in project portfolio

Promoting the exchange & reuse

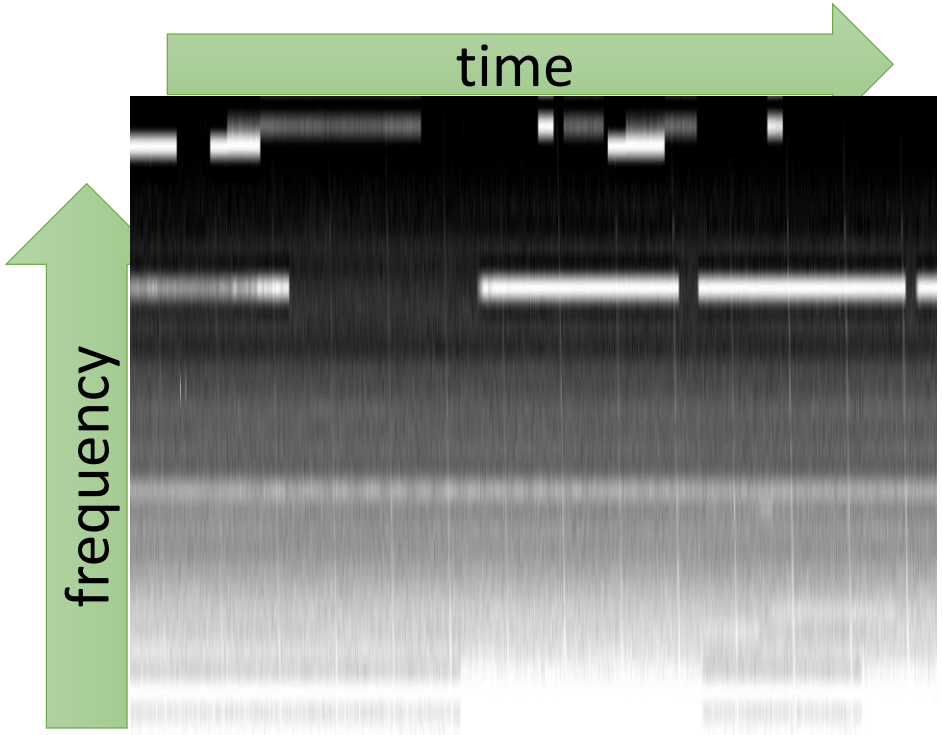
of best practices in software development

Open source & open access

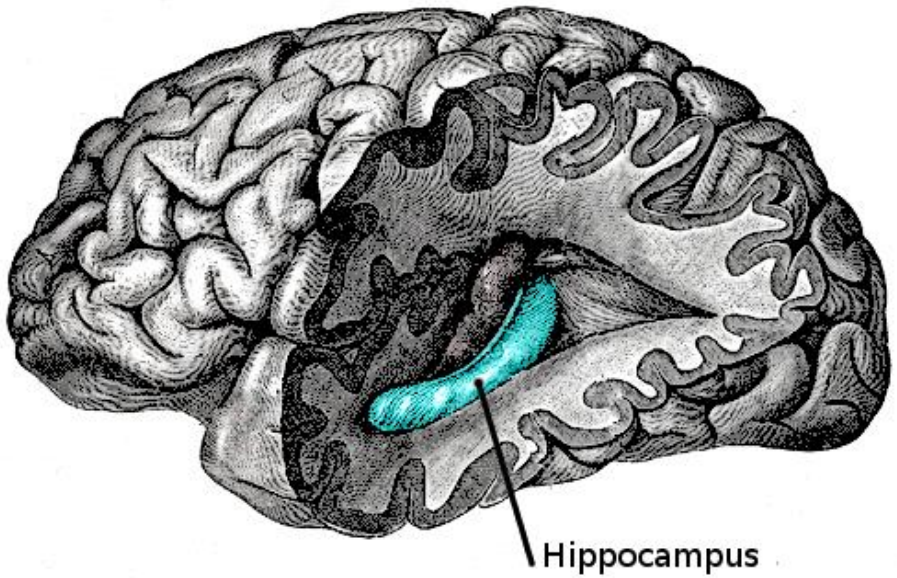
utilizing professional coding standards, unit & integration testing & documentation

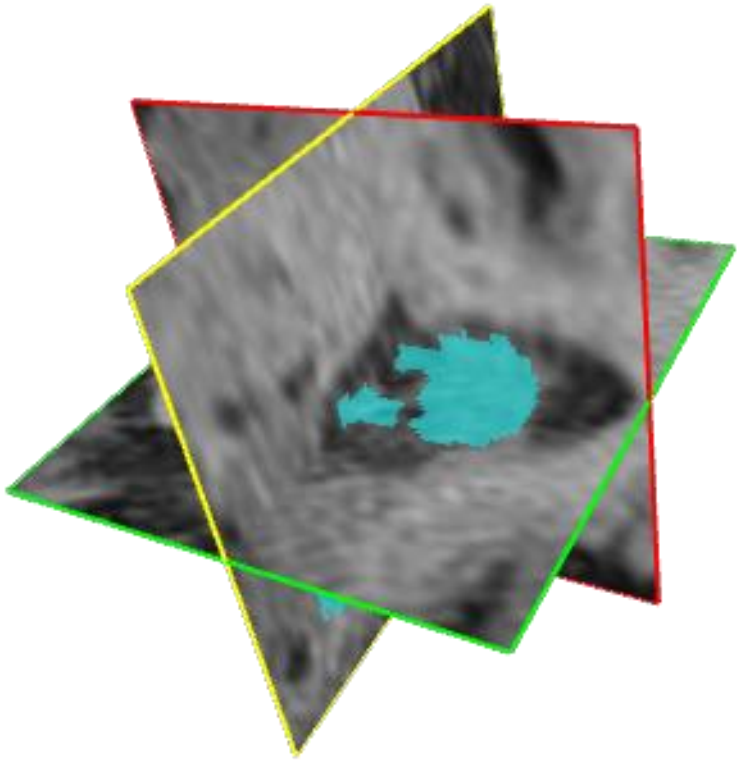
eScience center

Big data for the Big Bang

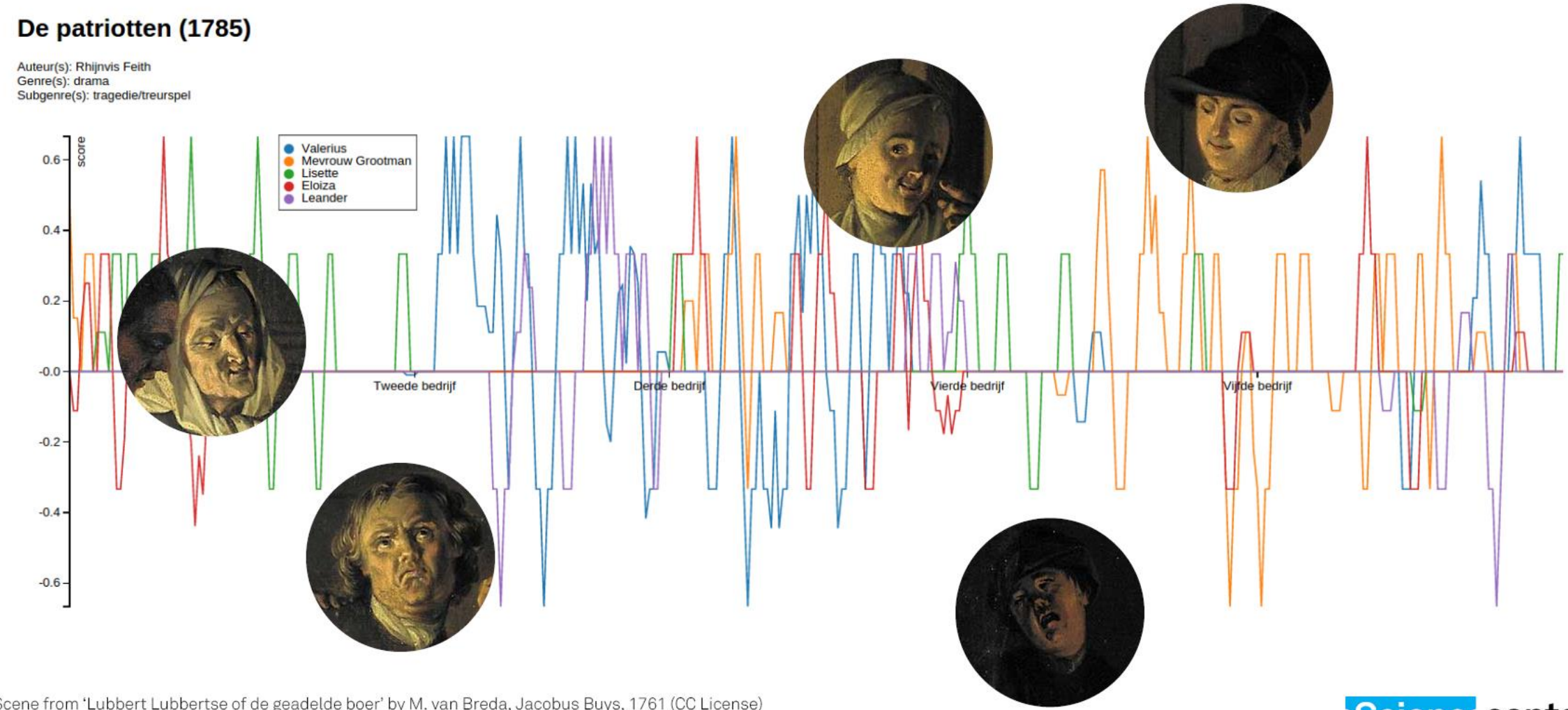


New markers for human health





Emotional styles on the Dutch stage



Scene from 'Lubbert Lubbertse of de geadelde boer' by M. van Breda, Jacobus Buys, 1761 (CC License)

eScience center

Get in touch

Netherlands eScience Center
Science Park 140
1098 XG Amsterdam
The Netherlands

+31 (0)20 4604770
info@eScienceCenter.nl

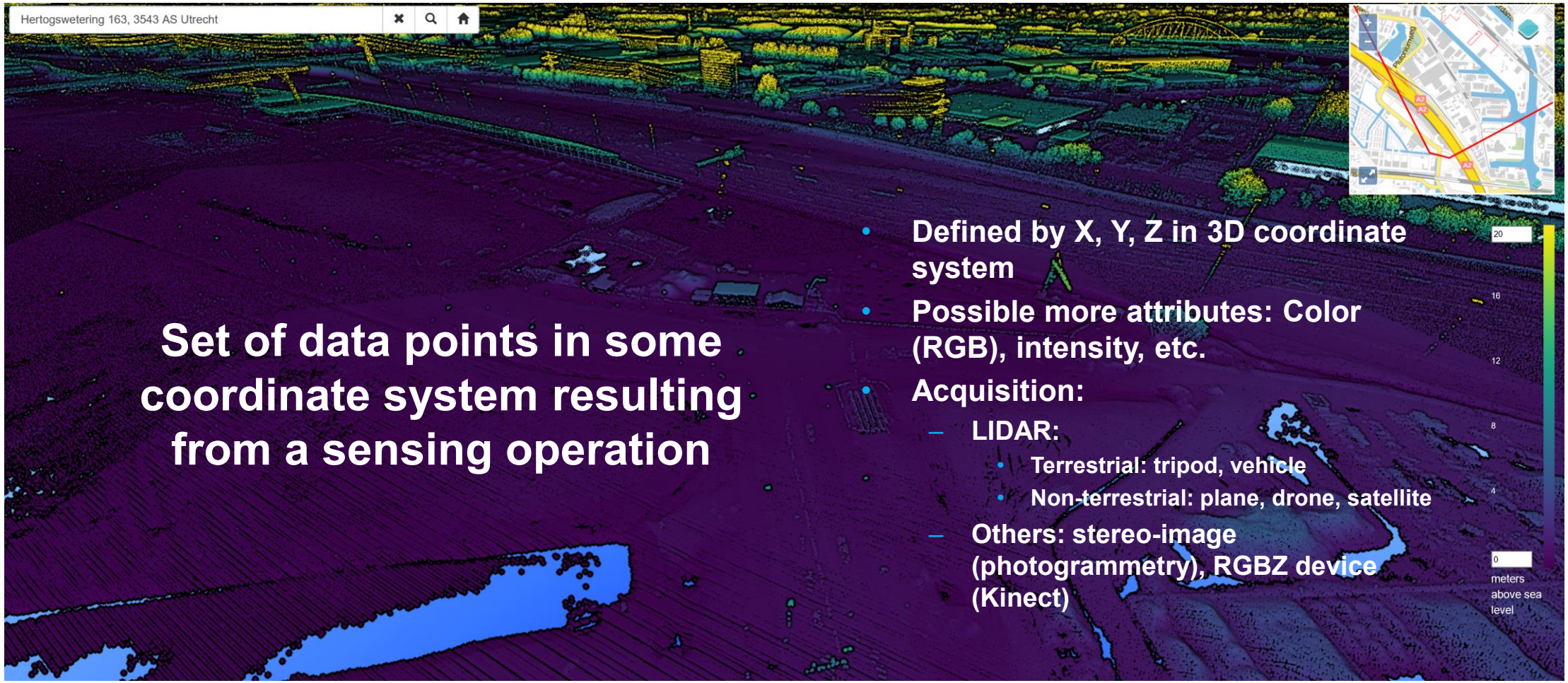
www.eScienceCenter.nl

linkd.in/1j2uS8S

vimeo.com/eScienceCenter

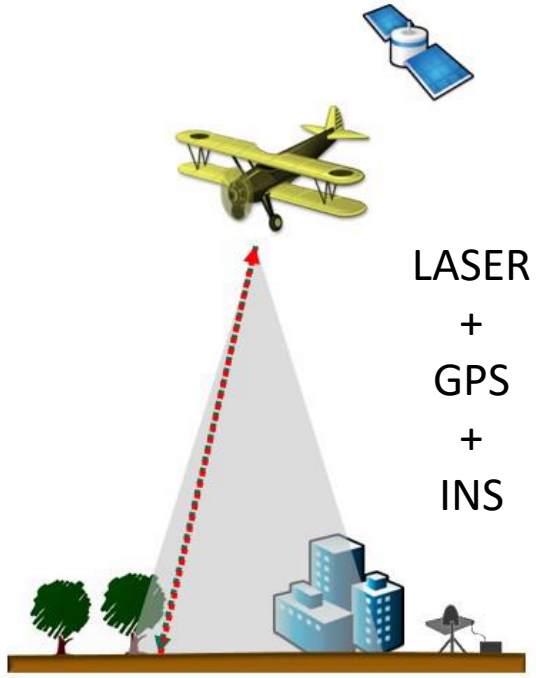
twitter.com/eScienceCenter

Point Clouds



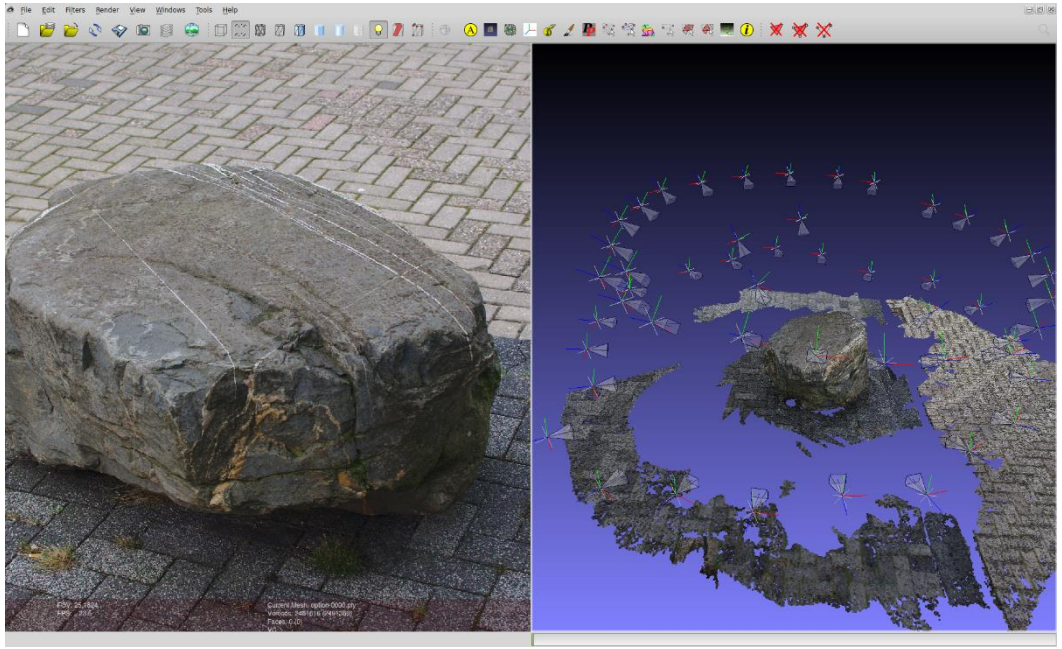
Lidar (Light & Radar)

- **Remote sensing** technology
- Distance = *ƒ (*laser reflected light*)*
- High density **point clouds of surface**



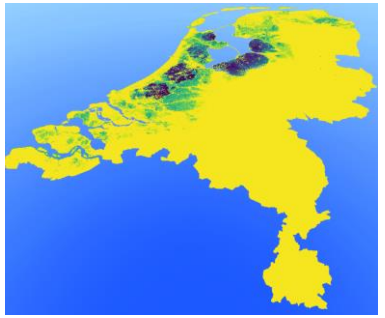
Photogrammetry

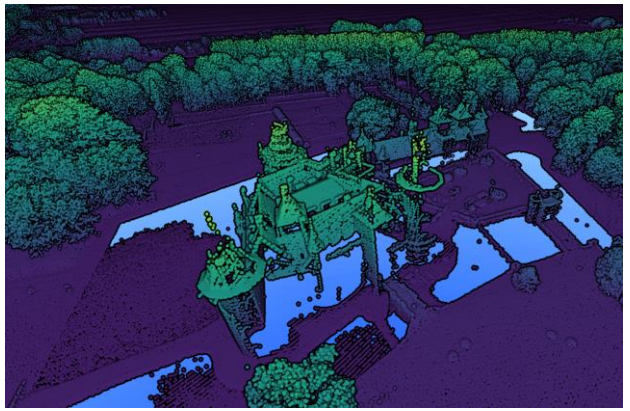
- **Remote sensing** technology
- **3D model from images**
- Very high density **point clouds of surface**



Applications

- **Geography** / Geology (digital elevation models)
- Urbanism / **Architecture** (3D city models)
- **Archaeology** / Tourism (virtual sightseeing)
- Autonomous vehicles / Robotics
- Biology and conservation (forestry)
- Military / Warfare
- Artistic
- …





Challenges

- **Acquisition** methods (**indoor** point clouds, **photogrammetry**, etc.)
- **Management** (trillions of points)
- **Processing** (Nearest neighbors, etc.)
- Object **recognition**
- **Visualization** (desktop, web, streaming, etc.)
- **Registration** (aligning un-references points)
- …

Usage

- **Traditionally converted** to rasters, vector data, polygonal models, etc.
- Lately growing **direct usage**:
  - Most **precise** information / accuracy
  - Most **reliable** representation of reality
  - **Visualization**
  - Technologies for their handling “becoming available”

Point Clouds @ NLeSC

- **Massive Point Clouds for eSciences** (http://pointclouds.nl)
  - Dealing with massive point clouds (billions/trillions): Improving management and visualization (Oracle is partner)

- **Mapping the Via Appia in 3D** (http://mappingtheviaappia.nl/)
  - 3D GIS for exploration of complex sites based on point clouds

- **Big Data Analytics in the Geo-Spatial Domain**
  - Improving point cloud and voxel support in the column-store MonetDB
- **3D geospatial data exploration for modern risk management systems**
  - Modernizing generation and manipulation of geospatial datasets by using a Geospatial Database Management System (DBMS)
- **Improving Open-Source Photogrammetric workflows for processing big datasets**
  - Making photogrammetric workflows work in consumer-grade computers

Which system to use?

- **Simple analysis** (minimal relation with other type of data) → If existing, **file-based** solutions offer (possibly) most efficient approach
- **Complex analysis** (complex queries and/or other-data integration) → **DBMS**
- **Visualization**
  - Small data sets → Both DBMS and file-based
  - Large data sets → Specific data management solutions (data structures)
    - More at the 8th December Studiedag “Management of massive point cloud data: wet and dry (2)”. Info on http://www.ncgeo.nl/ under agenda. Register sending email to info@ncgeo.nl

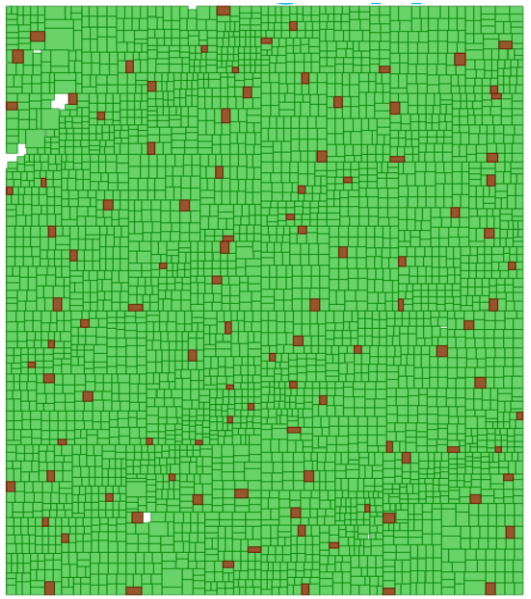
Point Cloud Data Management Systems (PCDMS) - DBMS

- Flat-Table storage model

X	Y	Z
85999.57	442667.63	-1.69
85999.82	446267.15	-1.69
85998.93	446267.38	-1.57
85999.42	446266.47	-1.65
85999.67	446265.99	-1.65
85999.92	446265.50	-1.64
85999.94	446251.73	-1.65
85999.10	446250.23	-1.69
85998.53	446250.04	-1.71
85970.83	446255.10	-1.71
⋮		

PostgreSQL ORACLE monetdb

- Blocks storage model

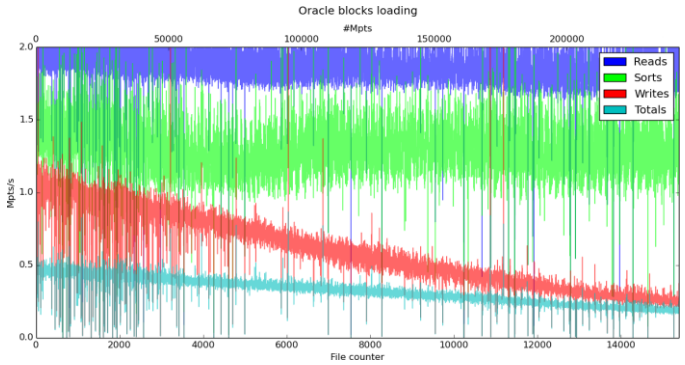


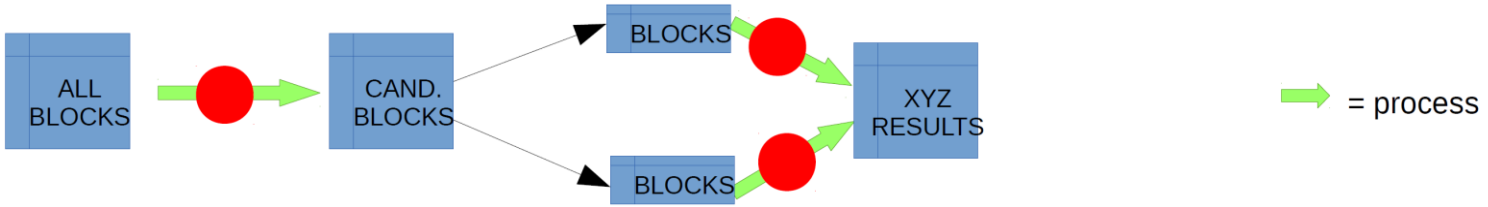
BlockId	Block
1	0101010101…
2	0100010010…
3	0000111010…
4	0010111001…
5	0100000100…
⋮	

PostgreSQL ORACLE

PCDMS – Oracle Spatial and Graph

- **SDO_PC package** – blocks storage model
  - **Loaders**:
    - Global: load all points from all files and create blocks globally
      - Too slow for large data sets (0.03 Mpts/s)
    - **Incremental**: “file by file”, blocks are created for each file
      - Faster (5 Mpts/s using multiple loading processes)
      - Decreasing performance issue for large data sets
  - Compression: BLOB compression too slow
  - **Queries**:
    - scalable (same performance for larger datasets) but slow for simple queries
    - **parallel algorithms** researched together and implemented in next release of SDO_PC



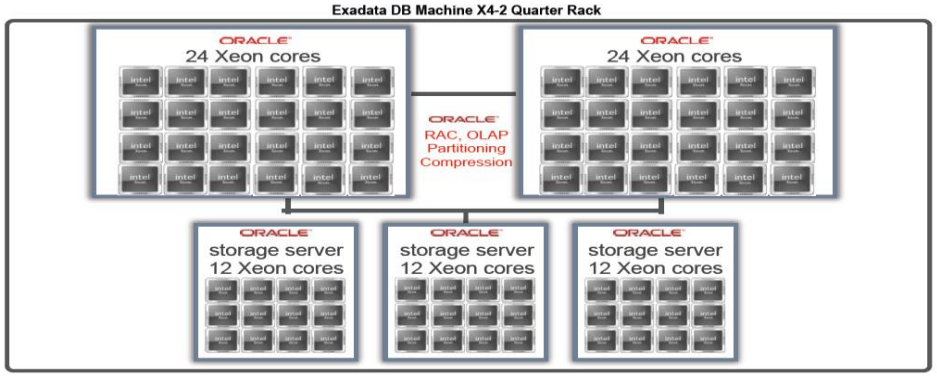


PCDMS – Oracle Spatial and Graph

- **Oracle-PDAL** – blocks storage model
  - Blocks created using third-party PDAL library
  - **Fast loading** (5 Mpts/s using multiple processes)
  - Laz-perf **compression** (10x more compressed)
  - **Not compatible with SDO_PC**
  - Queries:
    - Data retrieval through PDAL
    - Faster for simple queries

PCDMS – Oracle Spatial and Graph

- **Oracle Flat-Table**
  - Suitable for Oracle **Exadata**
    - Fast loading (40 Mpts/s)
    - Compression similar to LAZ with QUERY_HIGH mode
    - Fast queries
  - Non-scalable out of Exadata but used for research



“Massive Point Clouds for eSciences” project results

- **Mastered** several Point Cloud Data Management Systems (**PCDMS**).
- Defined, executed (and re-define) PCDMS **benchmarks:** *Define PCDMS’s, what is needed from them and actually test them*
- **Research**: **Scaling flat table** approach (with space-filling curves), **parallel** algorithms for **queries**, **parallel** and **smart loading** procedures, etc.
- Contributed (sometimes directly) to PCDMS’s **improvements:** *NOW PCDMS’s can deal with massive point clouds*
- Point cloud **web visualization**: *NOW it is possible to visualize massive point clouds over the web*
- **6 papers** including peer-reviewed (1), professional magazines (1), conference proceedings (3) and invited papers (1). Also many **benchmark reports**.
- **8 conference talks** including FOSS4G, IQmulus workshop, Data Science Symposium, NLeSC symposium, 3DGeoInfo (2014 and 2015), SPAR/ELMF (2014 and 2015) and several invited (minor) talks.
- **Point Cloud domain contribution**: **standardization** (PC-DWG at OGC), we have created **strong European consortium** with PC experts and submitted a H2020 proposal with them.

TU Delft Delft University of Technology

netherlands eScience center

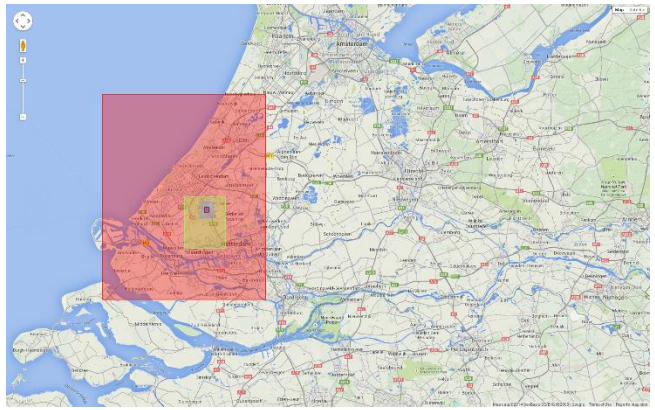
ORACLE SPATIAL

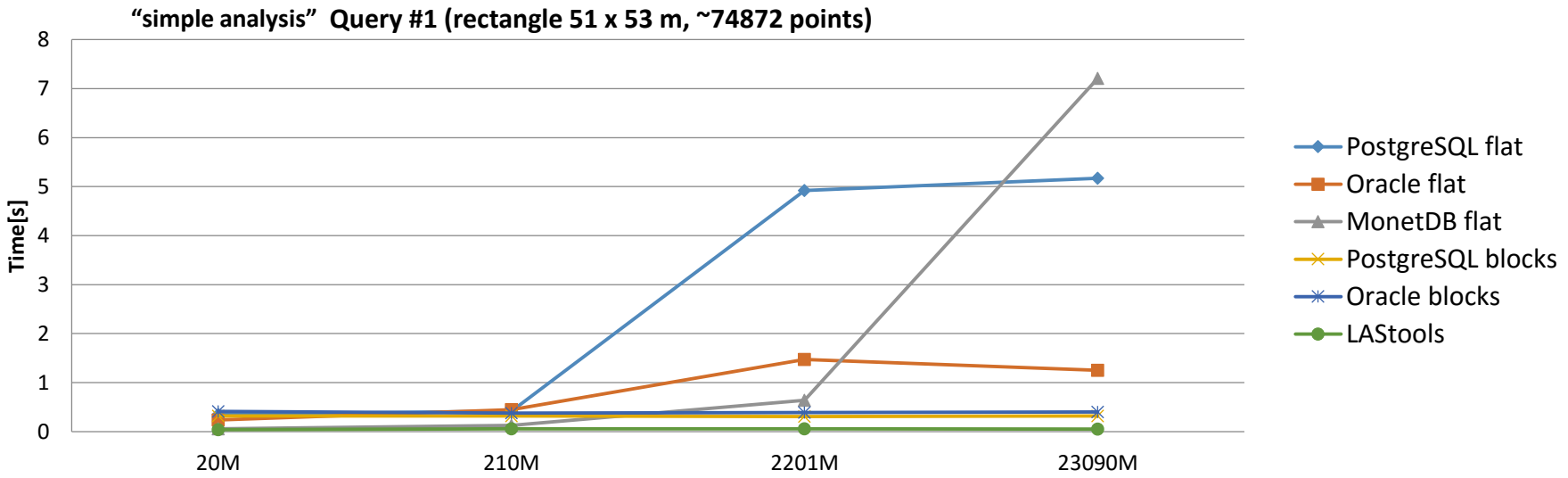
monetdb

FUGRO

Rijkswaterstaat
Ministerie van Infrastructuur en Milieu

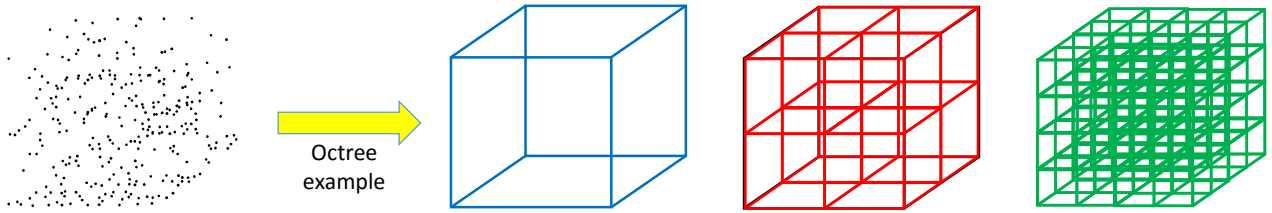
PCDMS benchmarks





Point clouds visualization

- Existing PCDMS's aim at analytic purposes
- Existing PCDMS's not efficient LoD support
- Specific data management solutions for visualization
  - Large point clouds → Multi-resolution data structures (quadtree, octree, kd-tree, etc.)



  - Many desktop-based solutions (and some really cool and efficient)
  - WebGL → new point cloud web renderers (FOSS: Potree and Plasio)
  - Massive point clouds → Bottleneck: expensive creation multi-resolution data structures

Divide and conquer approach → Divides creation of massive octree into multiple independent tasks (creation of small octrees) that can run in multiple systems and cores. Finally the small octrees are merged into massive octree

http://ahn2.pointclouds.nl

Get in touch

Netherlands eScience Center
Science Park 140
1098 XG Amsterdam
The Netherlands

+31 (0)20 4604770
info@eScienceCenter.nl

www.eScienceCenter.nl

linkd.in/1j2uS8S

vimeo.com/eScienceCenter

twitter.com/eScienceCenter