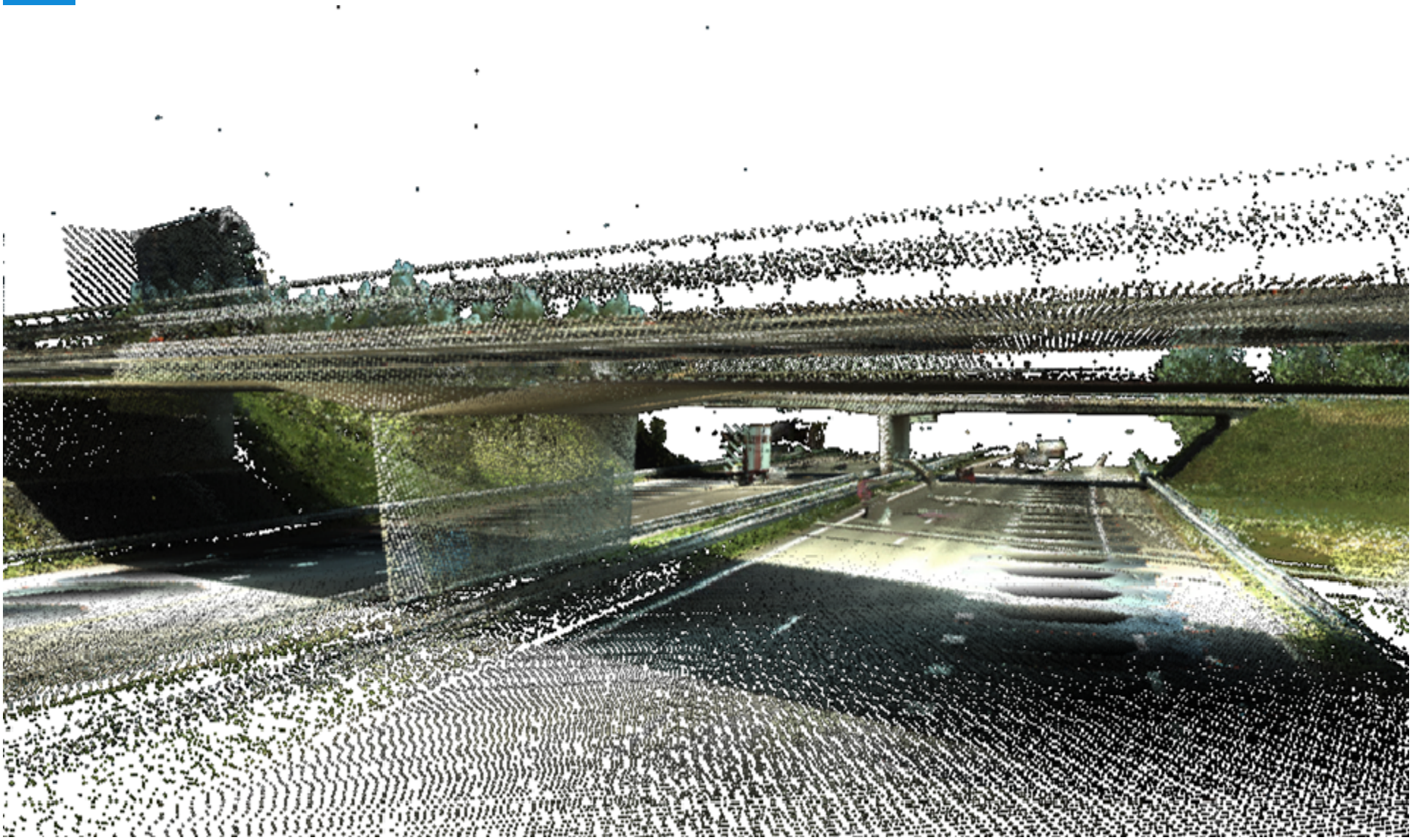
Managing large point clouds - an eScience project proposal

17-1-2013

Peter van Oosterom, GIS technology, OTB, TUD

GISt Lunch seminar, Thursday 17 January 2013, 12.45-13.30 hours,
Pieter Groetelaerszaal, OTB, TU Delft

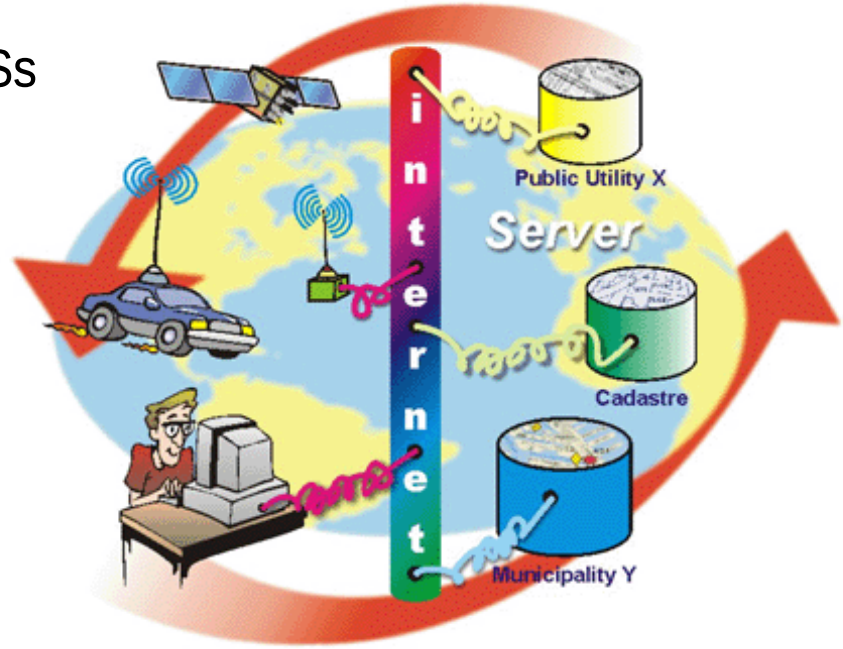


Contents

- Introduction
- Project proposal
- Point clouds
- Approach
- Innovation
- Conclusion

Towards the Geoweb (a.k.a the SII): Spatial Information Infrastructure

- Heading towards an information society with varying impact on sectors of our economy
- Impact on the geo-information sector is huge: geo-info and processing can be transported via (wireless) networks and can be combined with positioning
- Engines of the SII nodes: geo-DBMSs (Data Base Management Systems) filled with geo-data
- Fair pricing of data and services, no more endless copying of data (and loss of quality/up-to-dateness)



Context: Geoweb innovaties
Maps4Science NWO GOF voorstel

- geo-semantisch web ondersteuning
- goede en complete 3D/4D/5D ondersteuning
- expliciete en bruikbare links tussen de originele metingen en de resulterende kaart (interpretatie)
- omgaan met extreem grote datasets
- teruglevering door gebruikers (VGI)
- sensor netwerken, streaming life geo-data
- zowel data als processing (analyse, simulaties) van server
- standaarden, daar waar onvoldoende (basis GI of domein)
- GII organisatie (rechten, licenties, privacy, open data, evaluatie)

Zie ook: Craglia, M., et al, 2012.
Digital Earth 2020: towards the vision for the next decade.
International Journal of Digital Earth, 5 (1), 4-21.

Maps4Science sequels

1. terugblik uitslag NWO GOF roadmap:
    - 4 expert reviews: very good-excellent (on 11 aspects): 4.4
    - invited for interview with NWO GOF commission
    - besides partners, strong support network: GI industr+gov
    - not selected for roadmap

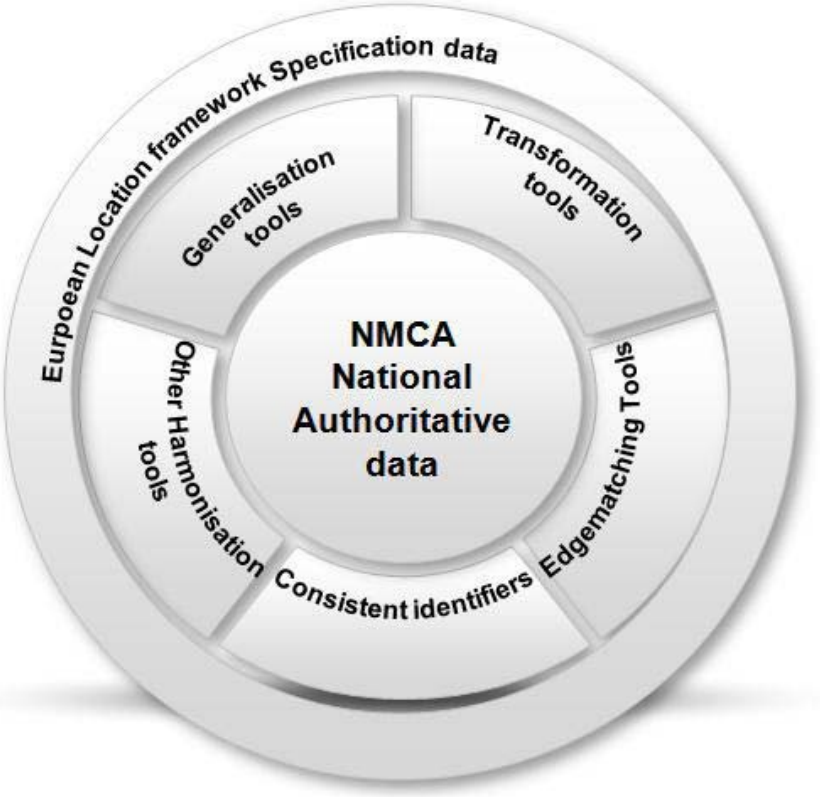
2. opties toekomst M4S
    - FP7/ ICP, EuroGeographics voorstel ELFI
    - STW Partnershipprogramma
    - overig, incl. GOF 2014/16
    - NWO/ Netherlands e-Science Center call

EU FP7 ELF voorstel

- EU KP7 CIP-ICT PSP Call 6 (Competitiveness and Innovation framework Programme – ICT Policy Support Programme)
- Deel van de door M4S beoogde internationale inbedding

- European Location Framework (ELF)
- Consortium 30 partijen (15 NMCAs) o.l.v. EuroGeographics
- NL: Kadaster/Geonovum en TUD
- 13.5 Meuro, 50% eigen bijdrage

- Score: relevantie 5, impact 4.5, en implementatie 4.5
- Start maart 2013, duur 3 jaar
- Tekenen: GA (met EC) en CA



STW Partnershipprogramma

M4S is netwerk (voor onderzoeks geo-informatie infrastructuur)
Van Maps4Science naar Maps4Society (22 → 3 meuro)

Praktijkpartners dus mogelijk ook netwerken:

1. Publieke Dienstverlening Op de Kaart (PDOK) met partners: ministeries I&M en ELI, RWS, Kadaster en Geonovum
2. Nationaal Modellen- en Data Centrum (NMDC) met partners: RIVM, PBL, TNO, Deltares, Alterra en KNMI

M4S faciliteit als een soort “innovatiemotor” voor praktijk
Financiering 50% STW en 50% praktijkpartners
Partners: RWS (groot), Kadaster en NSO (klein)

29 jan’13 laatste concept (en geld) bespreken, 22 feb’13 STW besluit

Contents

- Introduction
- Project proposal
- Point clouds
- Approach
- Innovation
- Conclusion

Link to Netherlands eScience Center 'Building the e-Infrastructure'

1. René van Schaik (CTO, Netherlands eScience Center)
2. discussion to support Maps4Science needs:

'traditional' datasets of an estimated volume growth of 200 TB/year and 'non-traditional' datasets (such as sensor network data, VGI, LBS and other sources) of between 500 TB and 1000 TB/year resulting in a total storage (and processing) capability of 7 PB. Some important connections must be based on optical pathways to ensure fast and seamless access.

3. Netherlands eScience Center/ SURF Foundation/ SARA:
   - generic grid computing infrastructure (BiGGrid),
   - fast network connections and
   - substantial available data storage

NWO/ Netherlands e-Science Center call Converging Sciences

Change scientific practice by enabling large-scale “Big Data” analysis across multiple disciplines by:

- optimizing scientific discovery in the era of “big-data”;
- stimulating new science only possible by advanced computing;
- challenging traditional research by creative & innovative use of eScience & other computer-based methods;
- pioneering new techniques needed to explore & connect relevant (massive) datasets.

Funded max €500K (Dutch university or research institute).

- 23 april 2012 call open
- 31 may 2012 deadline preprosal
- 19 sept 2012 full proposal
- 20 dec 2012 NWO decision
- in 2013 start selected projects

Partners project (proposal)

- TU Delft:
  1. GIS technology
  2. TU Delft, Library, contact with research & education users, dissemination & disclosure of point cloud data
  3. 3TU.Datacentrum, Long-term provision of ICT-infra (hardware/expertise)
  4. TU Delft Shared Service Center ICT, storage facilities
- NL eScience Center, designing and building ICT infrastructure

- Oracle spatial, New England Development Centre (USA), improving existing software
- Rijkswaterstaat, data owner (and in-house applications)
- Fugro, point cloud data producer

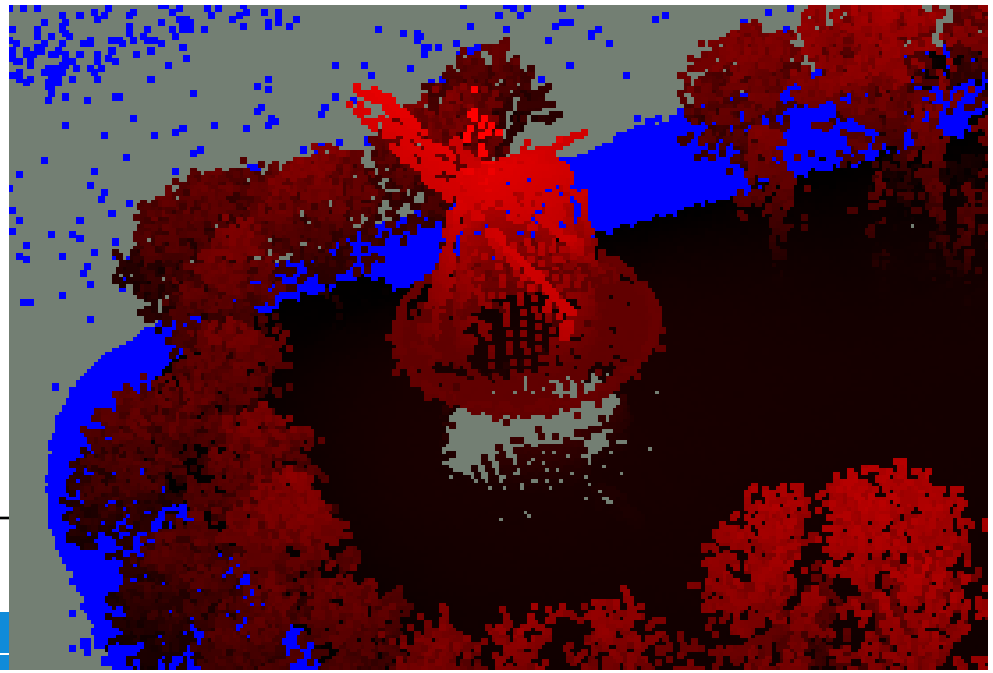
NWO/ Netherlands e-Science Center call Converging Sciences

- Massive point clouds for eSciences
  (project van 2 jaar, concreet AHN-2 effectief ontsluiten)

- Point cloud voorstel door pre-selectie (totaal 55 voorstellen)
  op 19 sept'12 definitief voorstel ingediend

- Selectie voorstellen door experts en daarna NWO
  (totaal 5 gaan door)
  3 reviews ontvangen:
  1. A (high quality)
  2. B (good)
  3. A+ (highest quality)

  (gem. 4.0 op 5 puntschaal)



Expert reviews on proposal

- R1: *"huge significance in eScience research and enables cross disciplinary research"*
- R2: *"project proposes a conventional paradigm in which data is still downloaded and moved around"*
- R3: *"in particular the importance of the proposed core support for point cloud data in spatial database management systems cannot be overstated, since this will allow to integrate point cloud data with existing types"* and
  "the project will be carried out by a strong research team with the broad expertise required for the project"

- Weerwoord 15 nov'12, vooral bedankt en aangegeven misverstand R2 omdat *"in proposal the database server is enriched with functionality making it possible to perform the operations as close as possible to the data"* (to make the cloud happen) en verwezen naar blz. in oorsponkelijke voorstel

Contents

- Introduction GIS tech
- Project proposal
- Point clouds
- Approach
- Innovation
- Conclusion

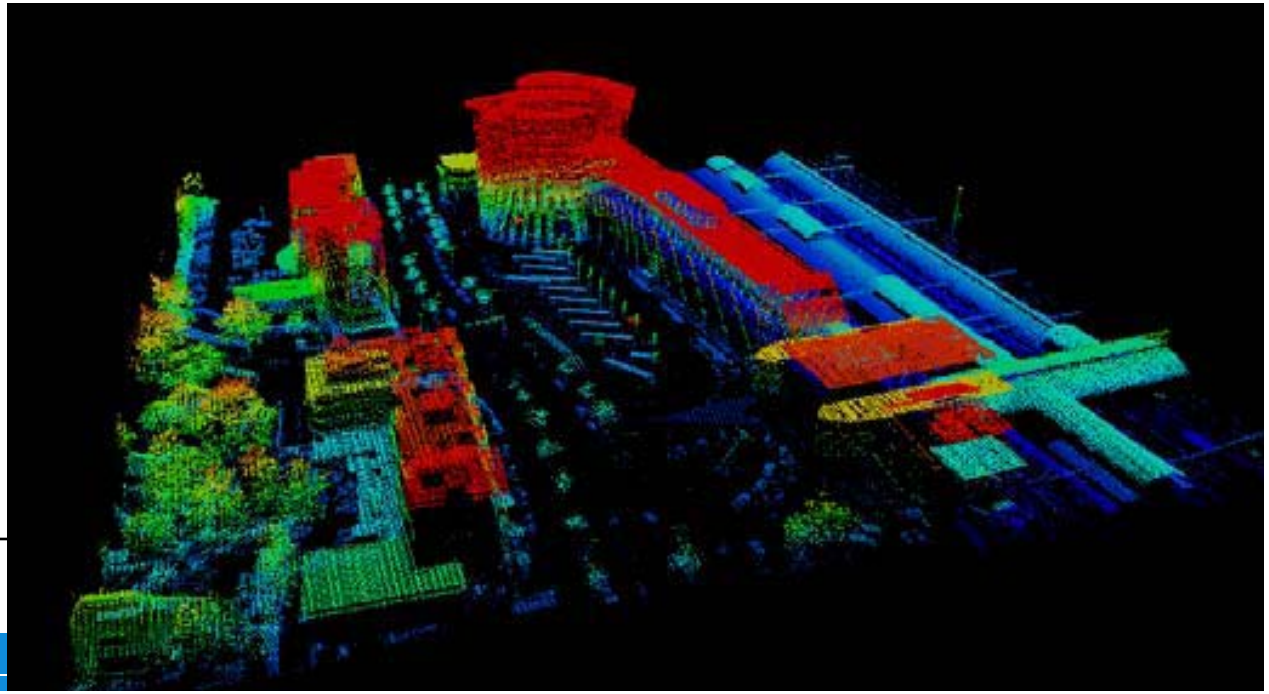
Point cloud data

- Not new, but growing rapidly (with increasing data management problems)
- Some producing technologies:
  - Laser scanning (terrestrial, airborne),
  - Multi-beam echo sounding
  - Stereo photogrammetry (point matching)

- Results is huge data sets, very detailed (precise), information rich

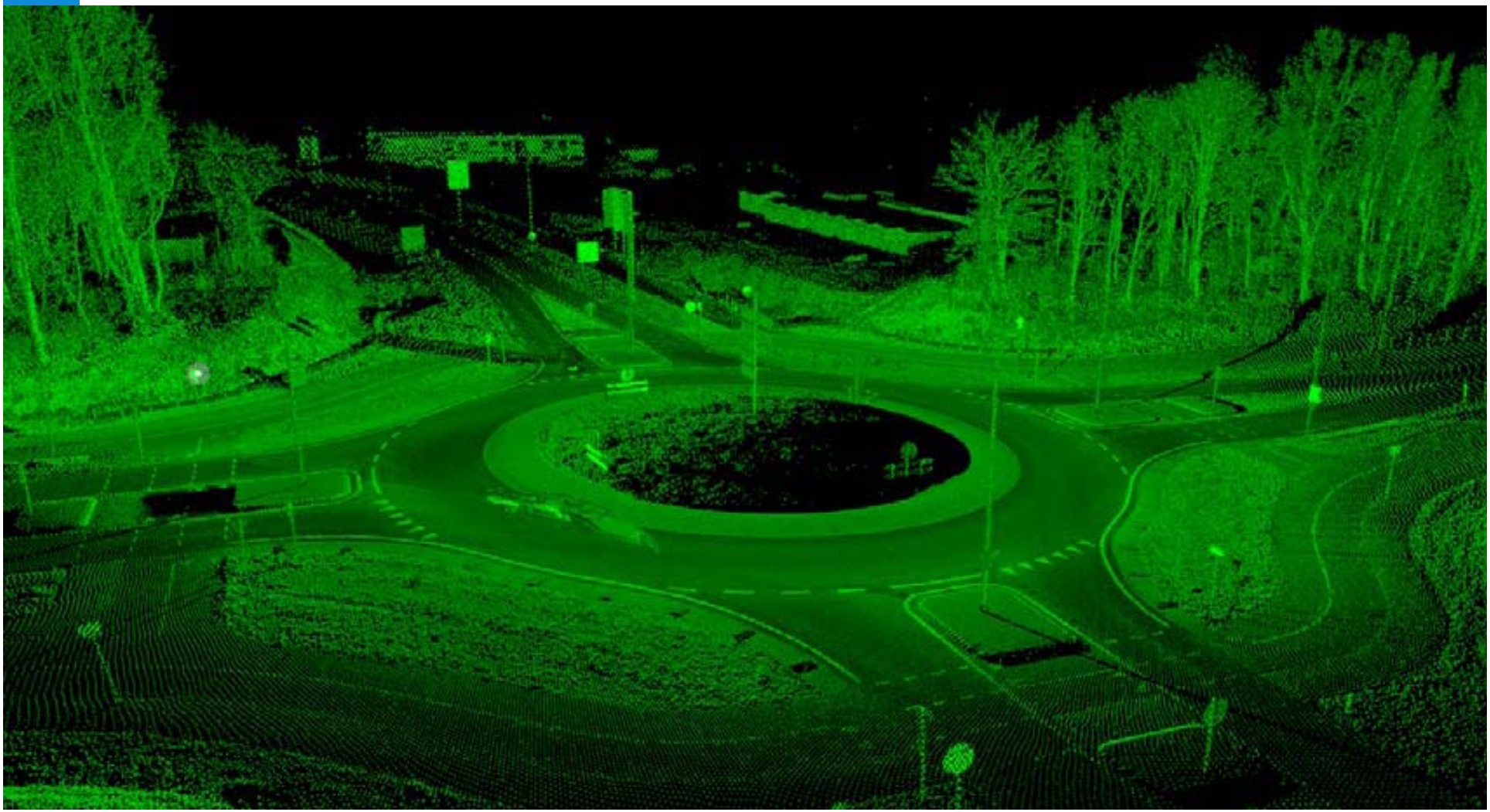
- AHN:
  1 point per 16 m^2
  →AHN2:
  10 points per 1 m^2

  Many Terabytes



Case AHN2

- TU Delft Library: distribution point for (geo-)data, including AHN2
- Users include:
  - Architecture (Urbanism, Landscape architecture),
  - Civil Engineering and Geosciences (Water Management, Geo-engineering),
  - Aerospace Engineering (Mathematical Geodesy & Positioning),
  - Electrical Engineering, Mathematics and Computer Science (Computer Graphics & Visualisation)
  - 'Outside' the TU Delft (but on campus): Deltares
  - More and more students are using this data.
- Challenge: how to get the data to the users?



Applications, often related to the environment

- Examples:
  - flood modeling,
  - dike monitoring,
  - forest mapping,
  - generation of 3D city models, etc.
- It is expected that AHN3 will feature an even higher point density (as already in use at some today; e.g. Rotterdam)
- Because of a lack of (processing) tools, most of these datasets are not being used to their full potential (e.g. first convert 0.5m grid or 5m grid, the data is losing potentially significant detail)
  → Sitting on a gold mine, but not exploiting it!

Contents

- Introduction GIS tech
- Project proposal
- Point clouds
- Approach
- Innovation
- Conclusion

Approach

- Develop an infrastructure for the storage, the management, the analysis and processing, the dissemination, the visualisation and the manipulation of massive point clouds
- Scalable solution: if data sets becomes 100 times larger and/or if we get 1000 times more users (queries), it should be possible to configure based on same architecture
- Generic, i.e. also support other (geo-)data and standards based, if non-existent, then propose new standard to ISO (TC211/OGC): Web Point Cloud Service (WPCS)

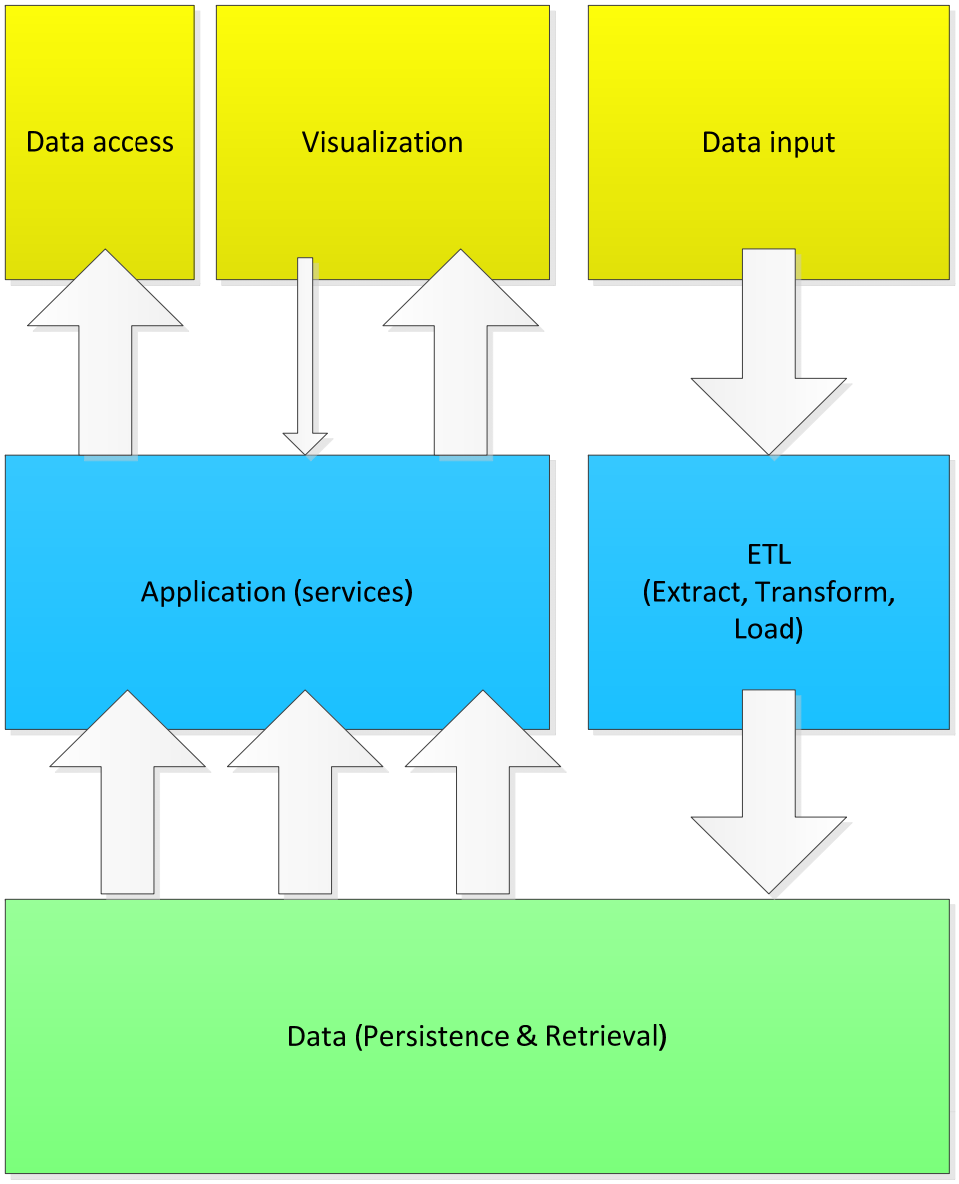
The 9 WPs of the project, duration 2 years

1. Compile specifications/ requirements & Create Benchmarks (data, queries, etc.)
2. Design web-based interface, protocol between server and clients (WPCS supporting streaming data)
3. Create facility, server side (hardware: storage, processing, communication)
4. DBMS extension (with point cloud storage and analysis functionality)
5. Design database schema, load data, tune (spatial clustering/indexing)
6. Develop client side (pose queries, visualize results)
7. Execute Benchmark (incl. scaled up versions of data set)
8. Investigate operational model for facility after the project
9. Project management

Layer I

Layer II

Layer III



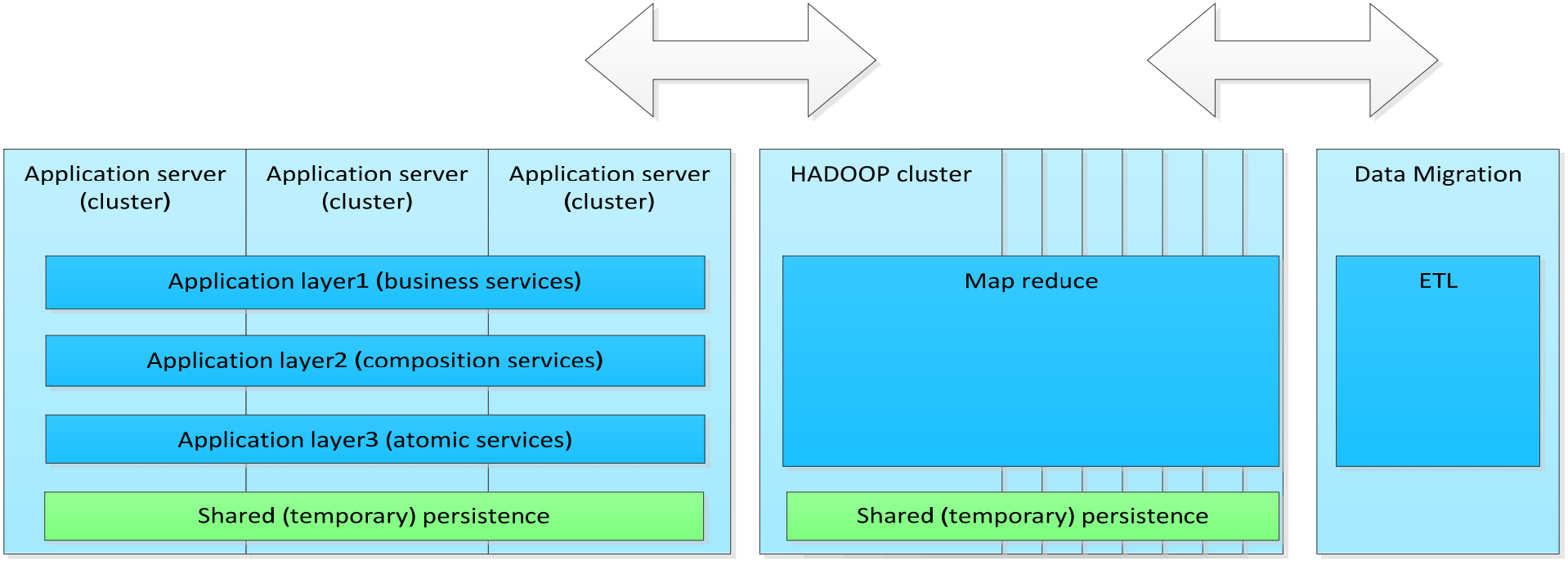
Business functions

Application (services)

Data

Application layer II can be implemented in various ways

- A cluster of web-services (small tasks, quick response)
- A HADOOP cluster with shared persistence, importing and manipulating large volumes of data (big jobs, one time only)
- Specialized ETL tools (big jobs, one time only)



Contents

- Introduction GIS tech
- Project proposal
- Point clouds
- Approach
- Innovation
- Conclusion

Innovations needed

1. parallel ICT architecture solutions with well-balanced configurations (including HW, OS, DBMS)
2. new core support for point cloud data types in the existing spatial DBMS-besides existing vector and raster data types
3. specific web-services point cloud protocol (supporting streaming, progressive transfer based on Level of Detail (LoD)/multi-resolution representation in form of a data pyramid)
4. coherent point cloud blocks/pages based on spatial clustering & indexing of the blocks and potentially within the blocks
5. point cloud compression techniques (storage & transfer; e.g. compact 'local block' coordinates, binary encoding)
6. point cloud caching strategy at client (supporting panning and zooming of selected data)
7. exploit the GPUs (at client side) given the spatial nature of most GIS related computations
8. integrate & fine tuning above parts within the overall system

Instrumentation

- Modern hardware, based on i86 processor technology
- Storage array (SAN or NAS based)
- Oracle Enterprise Linux (RedHat based) or Solaris; this is not a business requirement, but a project requirement to stay in pace with Oracle Product Development (patching)
- Oracle Database (Enterprise Edition 11gR2 or 12c Beta)
- Oracle Spatial
- Oracle Real Application Cluster & Oracle Automatic Storage Management; for system en storage virtualization
- Oracle Partitioning/Oracle Advanced Compression; for data partitioning and compression

Contents

- Introduction GIS tech
- Project proposal
- Point clouds
- Approach
- Innovation
- Conclusion

Conclusion

- Very innovative and risky project
- No solutions available today
  (big players active; e.g. Google with street view also collects point clouds, but has not be able to serve these data to users)

- Excellent team, world leading partners

- Wait for final letter NWO and start as soon as possible
  (this afternoon informal consortium meeting → first steps)