



Mimir Consultancy



De mogelijkheden van DGGS en cloud distributed computing voor schaalbare analyses! Special guest: ongestructureerde geodata

Wie ben ik?



- **Studie Neuropsychologie**
- **Traineeship data science**
- **MBA big data & advanced analytics**
- **6 jaar (lead) data scientist @ Gemeente Rotterdam**
- **Freelance data scientist, o.a. gewerkt voor UWV, ministerie IenW, RWS. Currently working for Gemeente Rotterdam and Alliander**
- **Specialisatie in GIS, pointclouds, image data & digital twins**

Waar houd ik mij mee bezig?

Hoe interacteren mensen met de wereld om zich heen, en hoe kunnen we dat verbeteren?



De buitenruimte en onze assets daarin staan steeds meer onder druk en de uitdagingen zijn groot en complex. Hoe kan ik daarin helpen met slimme tools?

Wie zijn jullie?

Waarom deze workshop?

Steeds meer ongestructureerde data binnen GIS, waar AI/machine learning “betekenis” aan kan geven.

Maar modellen & analyses kosten vaak veel GPU/CPU kracht.
Dus: distribueer het over een cluster!

Waarom deze workshop?

Steeds meer cloud platformen ontwikkelen GIS functionaliteiten & we hebben steeds meer cloud-optimized geo formats

Waarom deze workshop?

We passen al veel standaarden voor data-uitwisseling toe, maar een Nederlands gestandaardiseerd grid vindt nog weinig adoptie, terwijl dit veel waarde kan toevoegen!

Hoofdpunten

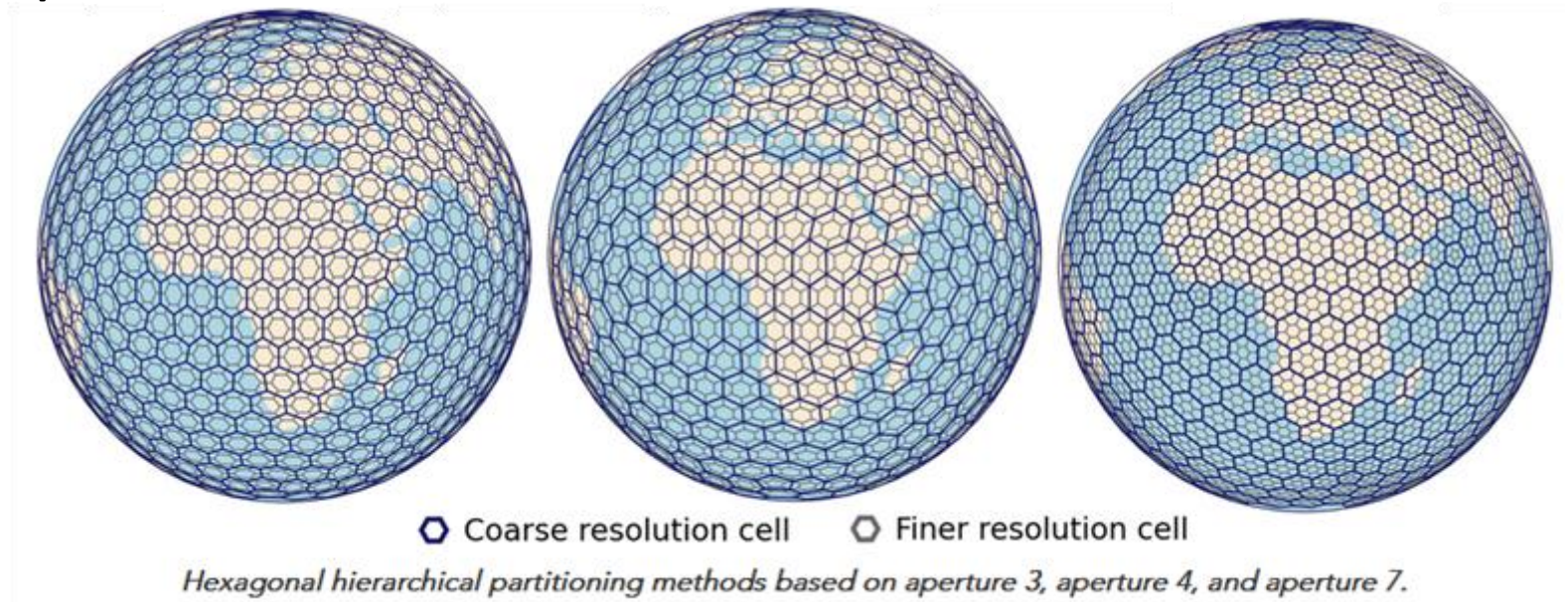
- Wat is het DGSS? En wat kan ik ermee?
- Wat is cloud/distributed computing?
- With our powers combined...
- Workshop
- Afronding

Wat is het DGSS en wat kan ik ermee?

Wat is het DGSS en wat kan ik ermee?

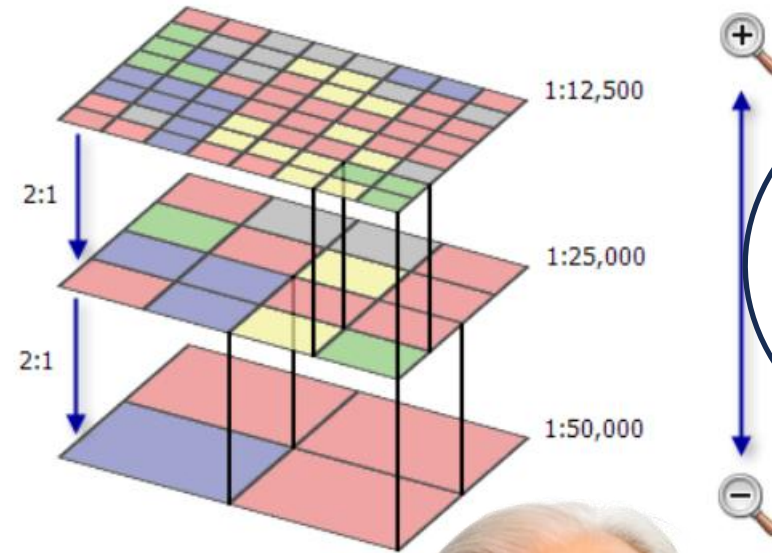
A “spreadsheet for Earth”

- DGSS= Discrete global grid system
- We delen de wereld op in cellen i.p.v. coördinaten



Wat is het DGSS en wat kan ik ermee?

- Cellen staan vast op de wereld
- Cellen kunnen kleinere subcellen bevatten
 - Hiërarchische opbouw
 - verschillende resoluties

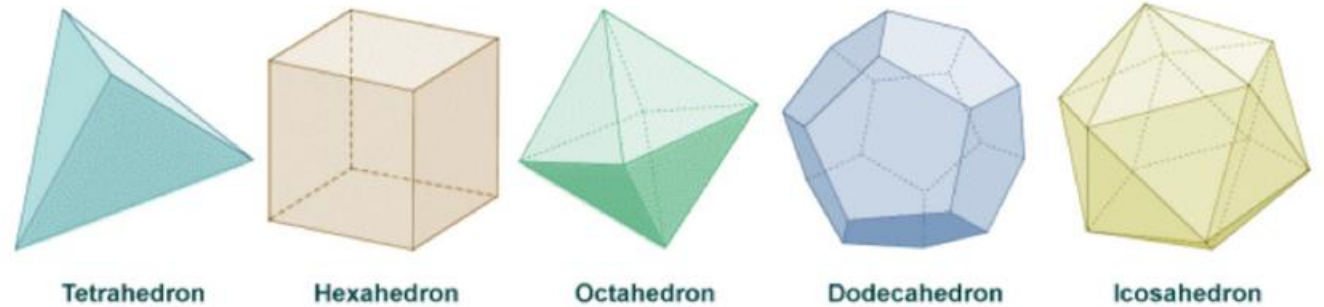


In mijn tijd
noemden we
dit een raster
pyramid!

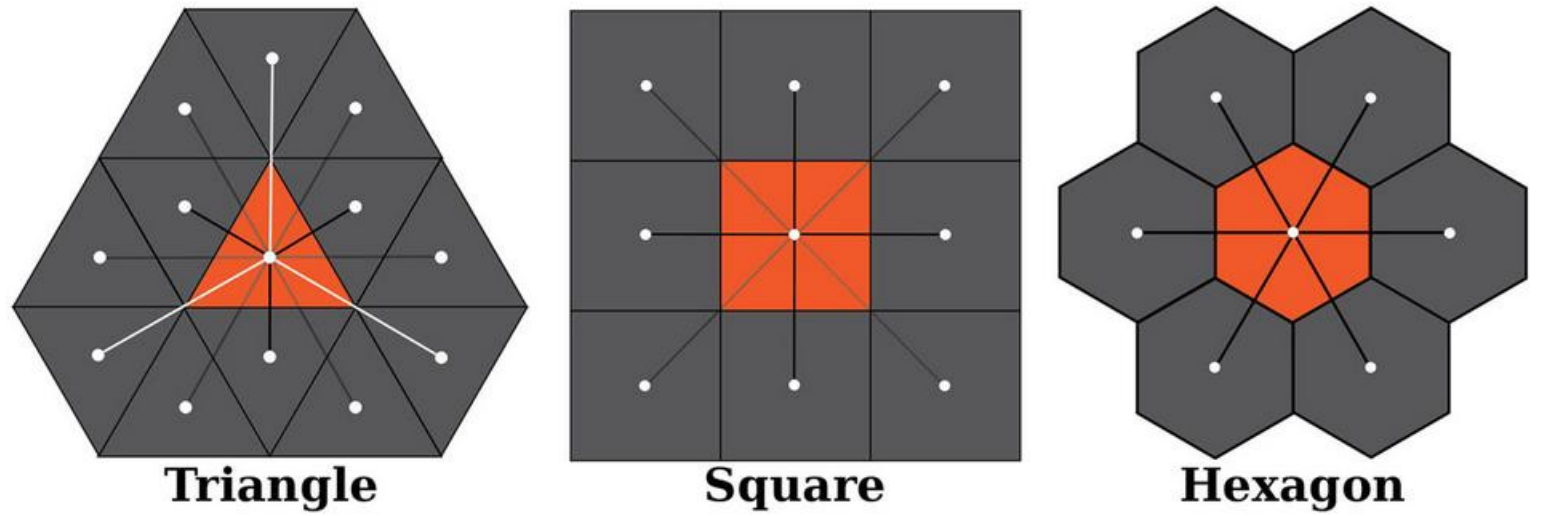


Wat is het DGSS en wat kan ik ermee?

- Cellen kunnen verschillende vormen hebben
- Vorm verschilt per use-case
- ...maar hexagons are the bestagons (voor analyse)



Representation of the five Platonic solids

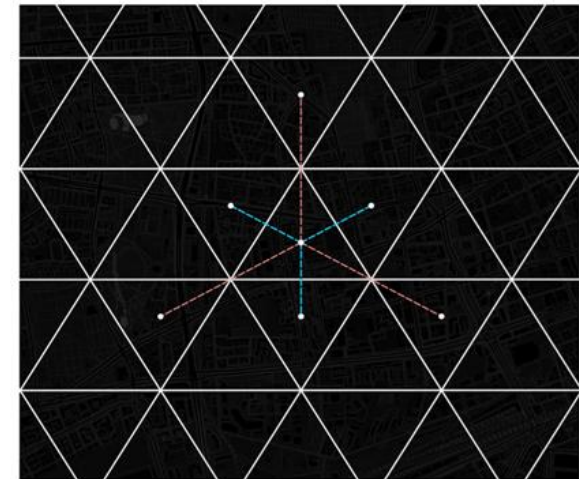
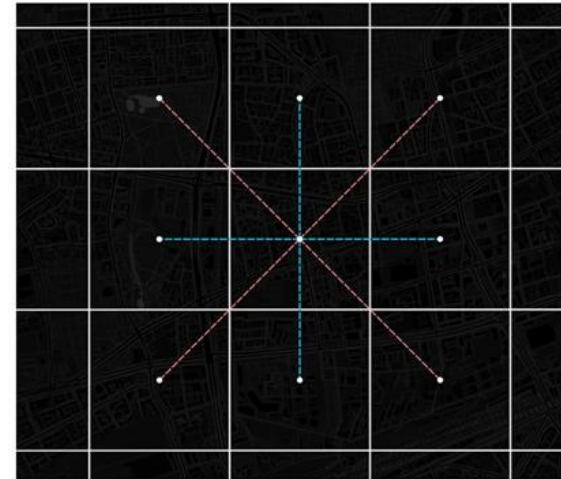
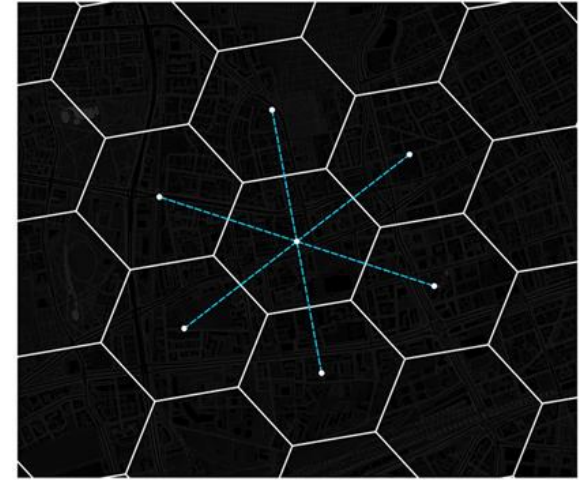


Geometric properties and neighbourhood relationships for triangle, square, and hexagon.

Wat is het DGSS en wat kan ik ermee?

**...maar hexagons are the bestagons,
omdat:**

- Ze liggen allemaal even ver van elkaar af
- Vorm dichtbij een cirkel terwijl ze volledig sluitend zijn
- Minder sampling bias door edge effects (edge to area ratio)

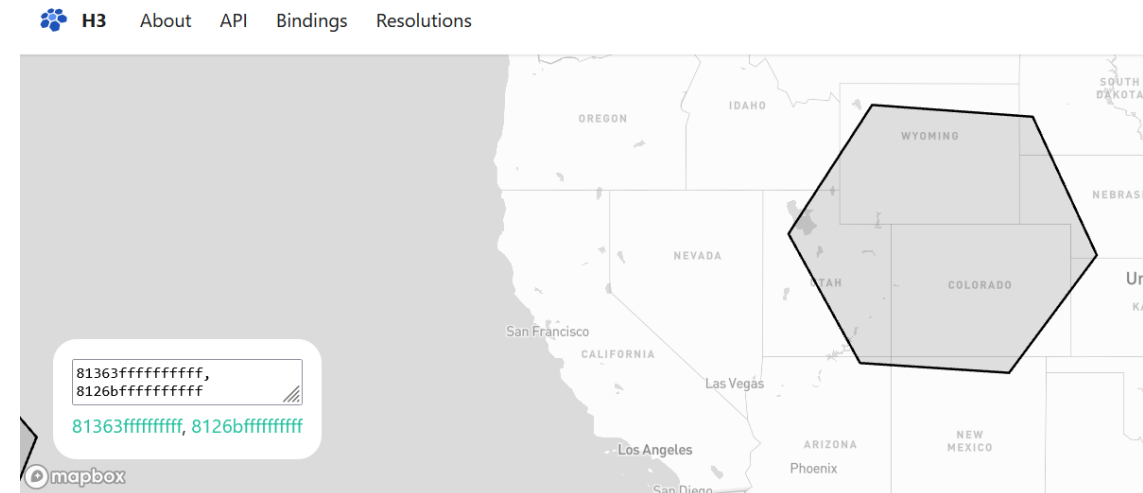


Bekende DGGS systemen & resources

- H3 (Uber)
- OGC spec (<https://ogcapi.ogc.org/dggs/>)
- DGGAL (<https://dggal.org/>)
- DGGRID (<https://github.com/sahrk/DGGRID>)
- DGGRIDR (<https://github.com/r-barnes/dggridR>)

OGC API - DISCRETE GLOBAL GRID SYSTEMS

OGC API - DGGS specifies an API for accessing data organised according to Discrete Global Grid Reference Systems (DGGRS).

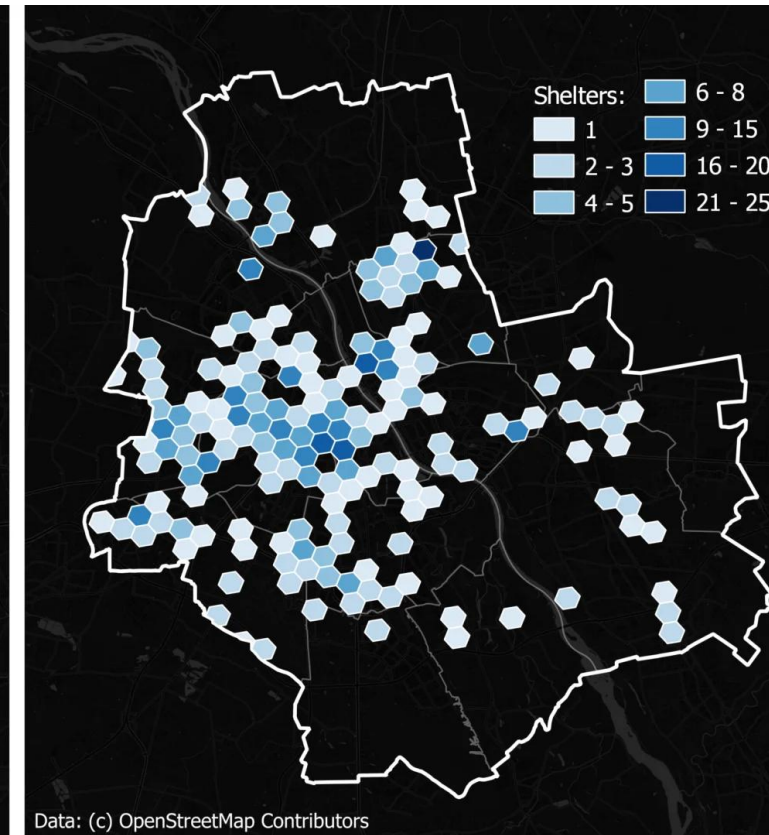
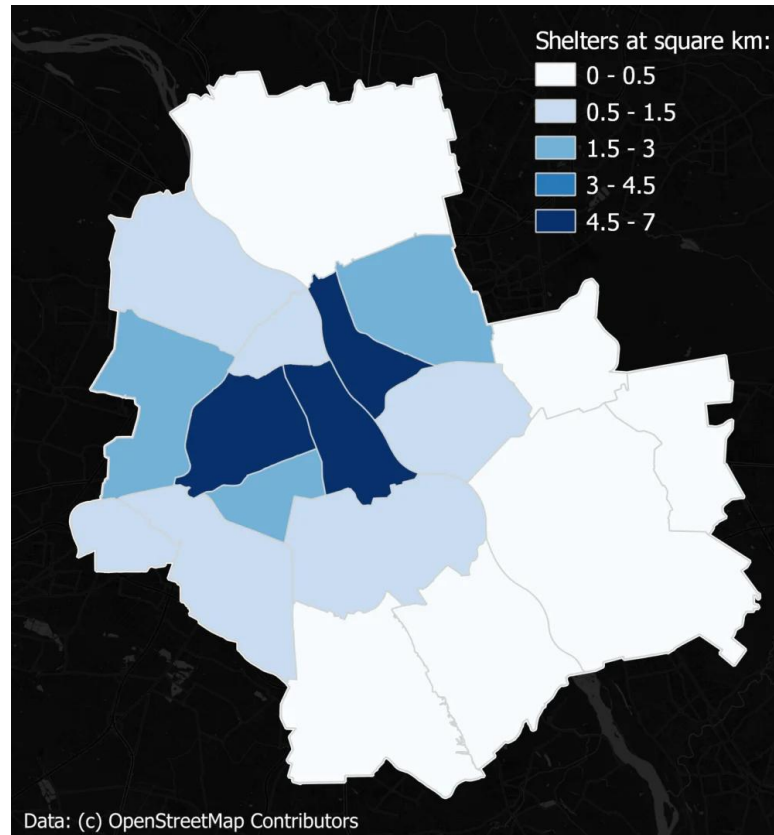


H3 indexes points and shapes into a hexagonal grid.

Dus...wat kan ik ermee?

Wat is het DGSS en wat kan ik ermee?

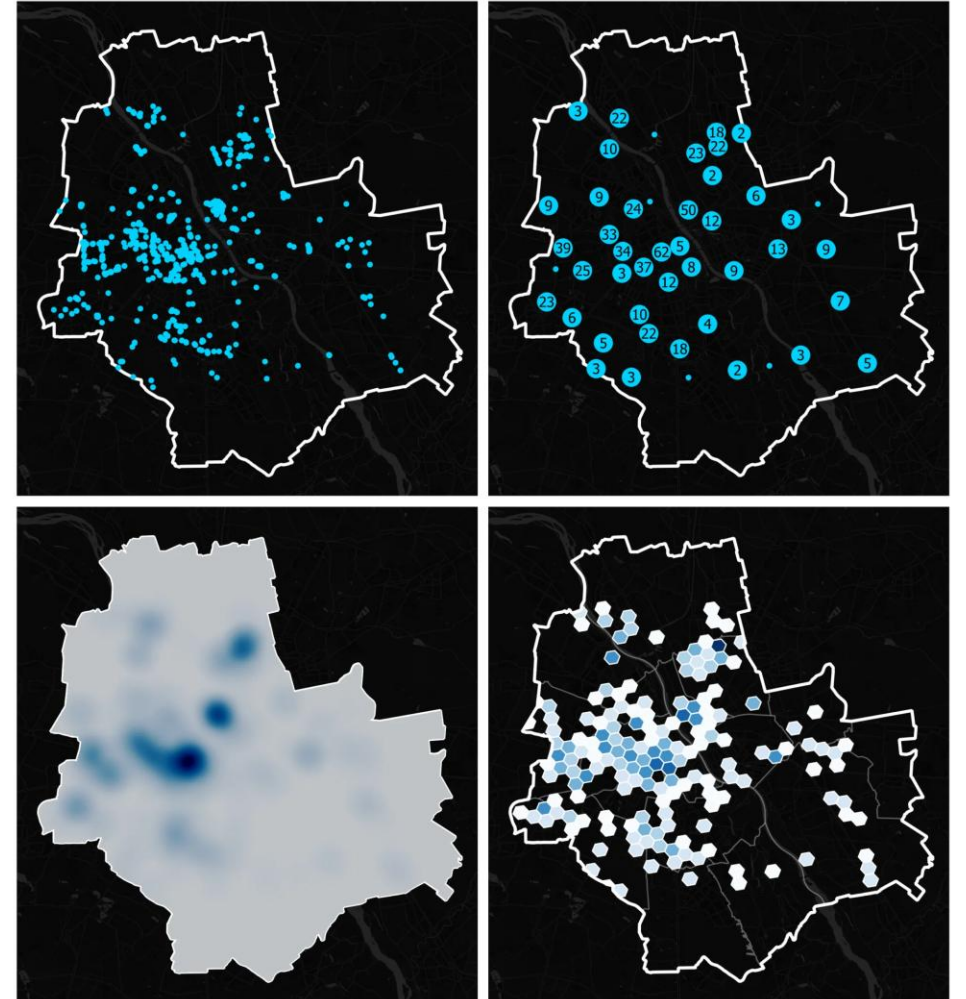
- Standaardisatie en integratie van verschillende typen data
 - Raster
 - Vector
 - Metadata
- Visualisatie: Veel analyses worden voor gebieden gedaan (wijken, steden, landen), maar klopt dat wel?



Wat is het DGSS en wat kan ik ermee?

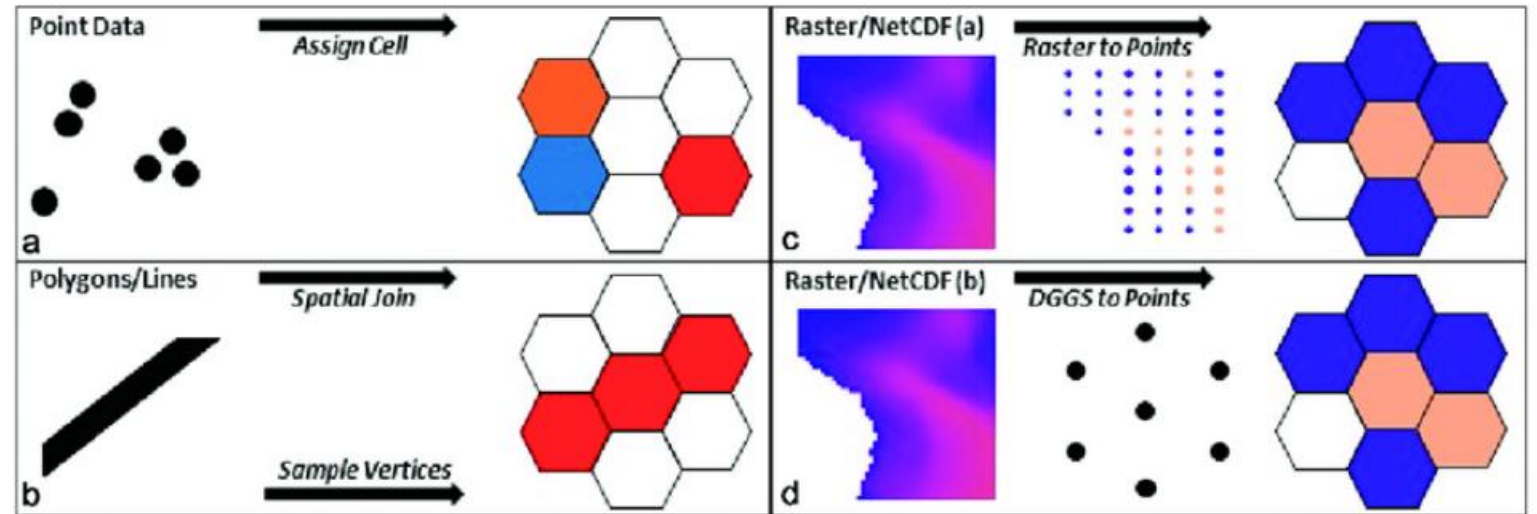
- Data anonimiseren
- Spatiele “aggregatie”

It can be tricky to visualize a large set of spatial data in a way that will provide relevant information. With a point map, for instance, you can attempt to display each data point with a tiny dot; yet these will inevitably overlap due to the size of the dataset. This approach can be problematic as the actual visual representation may depend on the points' rendering order – or how the data is stored.



Wat is het DGSS en wat kan ik ermee?

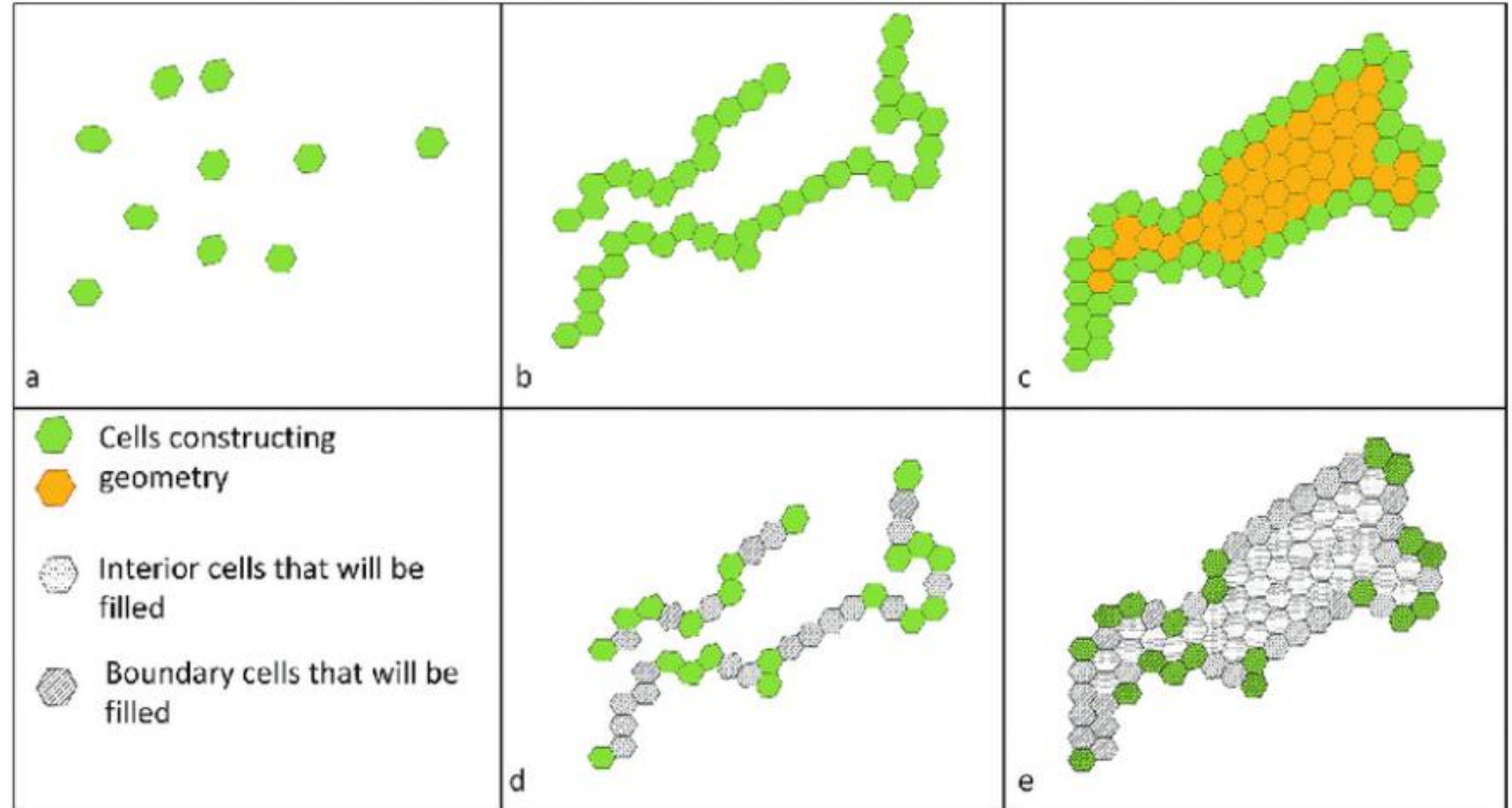
- Data op een gestandaardiseerde manier deelbaar maken met elkaar



Methods of assigning DGGS cells to geospatial data.

Wat is het DGSS en wat kan ik ermee?

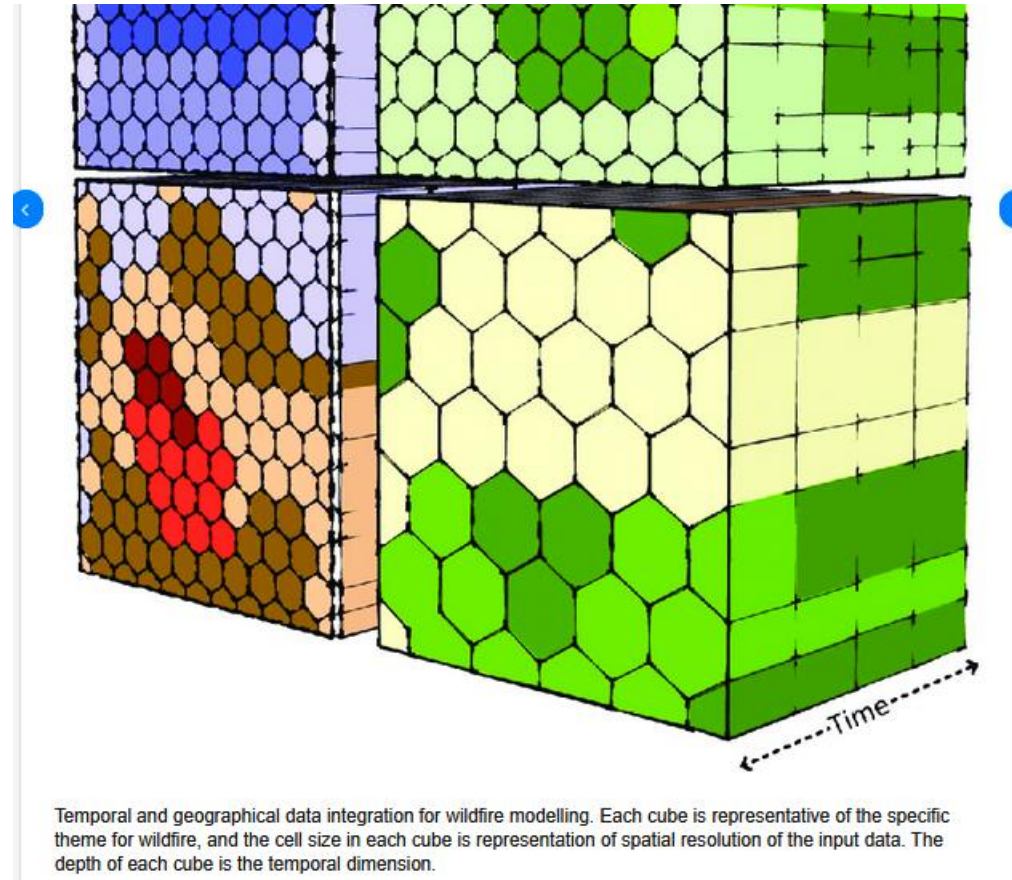
- Meerdere gangbare manieren



Two methods of geographic object representation in the DGGS model. (a-c) Cell subset method and (d-e) Vertex cell subset. (a) point objects, (b) and (d) line objects, and (c) and (e) polygon objects.

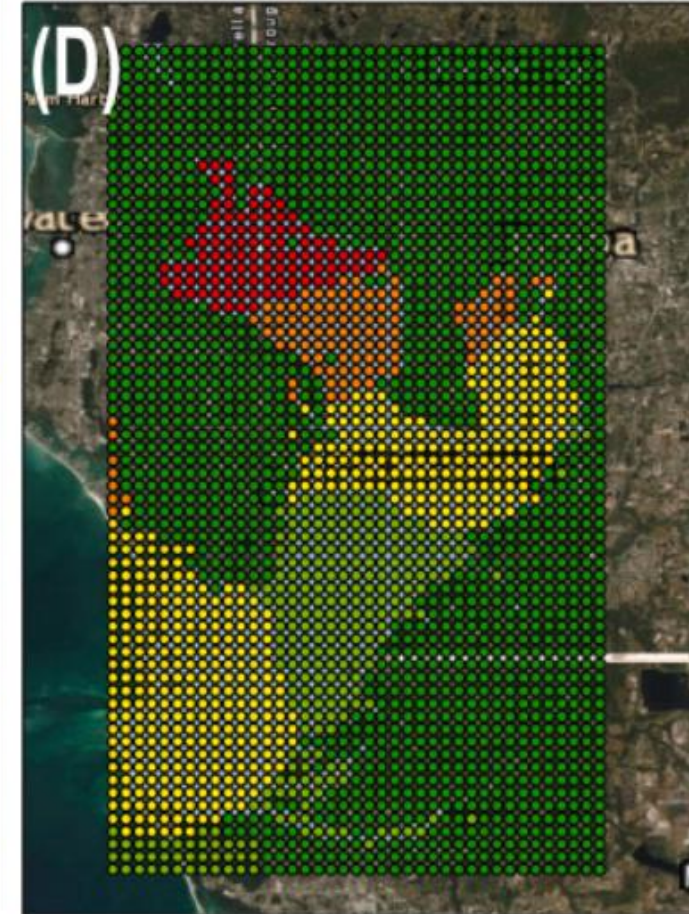
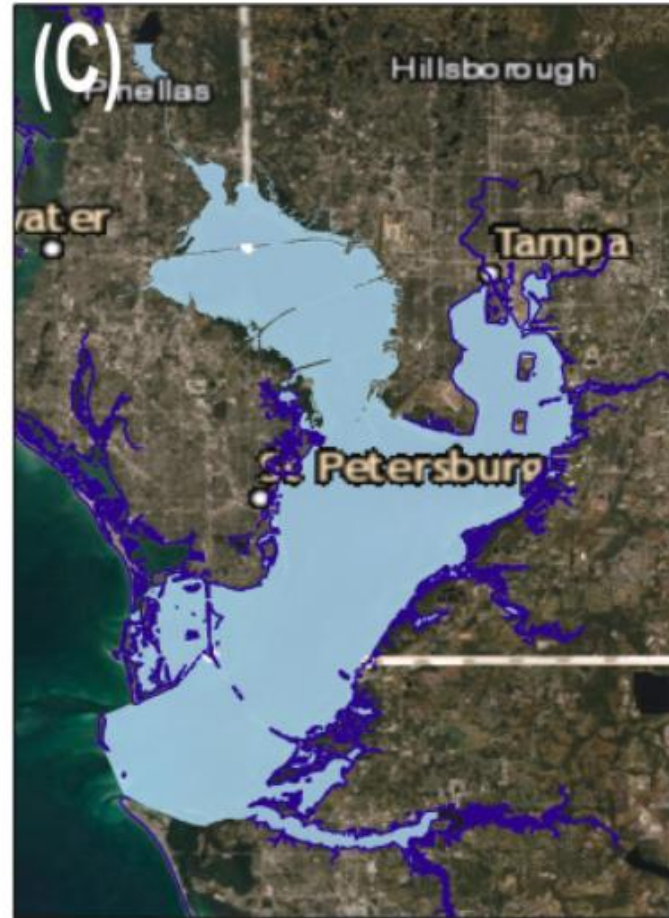
Wat is het DGSS en wat kan ik ermee?

- Ook temporele data natuurlijk!



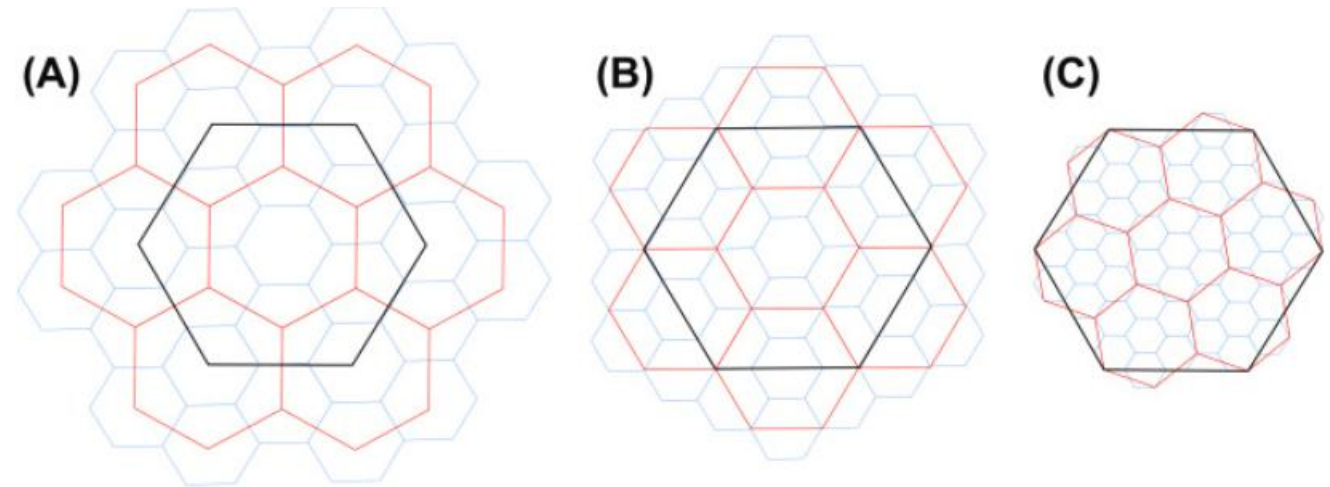
Wat is het DGSS en wat kan ik ermee?

- Complexe analyses opschalen!
- Computers zijn goed in uniforme grid analyses
- Maar zie je het probleem?



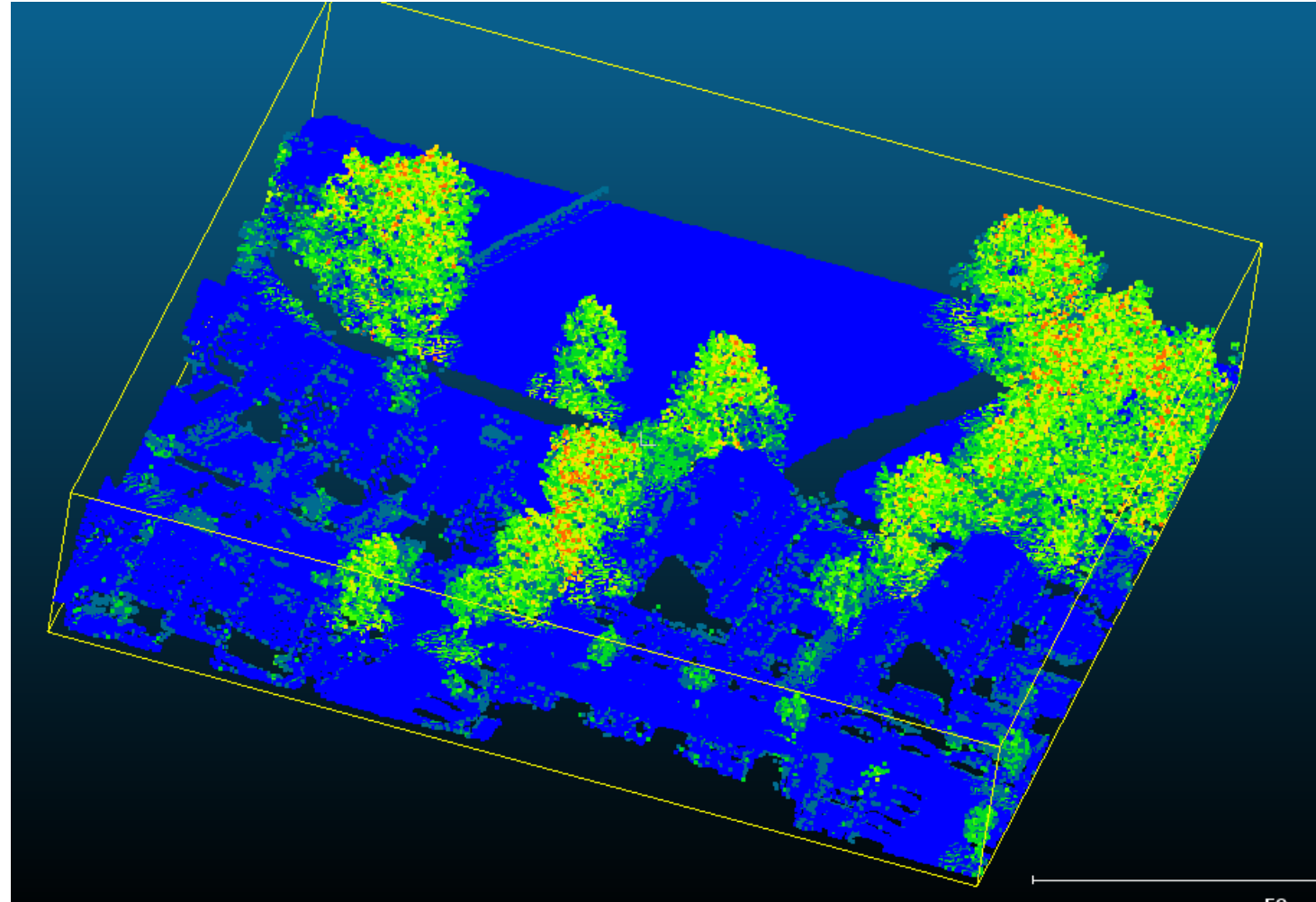
Uitdaging 1

- Nadeel hexagons: ze passen niet mooi in elkaar met verschillende grid levels
- Oplossing: overlappende cellen



Uitdaging 2

- Wat doe je met data tussen grid cells, of die in meerdere cells vallen?
- Nog grotere uitdaging met ongestructureerde data (pointcloud, luchtfoto's, panoramafoto's)



Uitdaging 3

- Wat doe je met projectie vervorming?

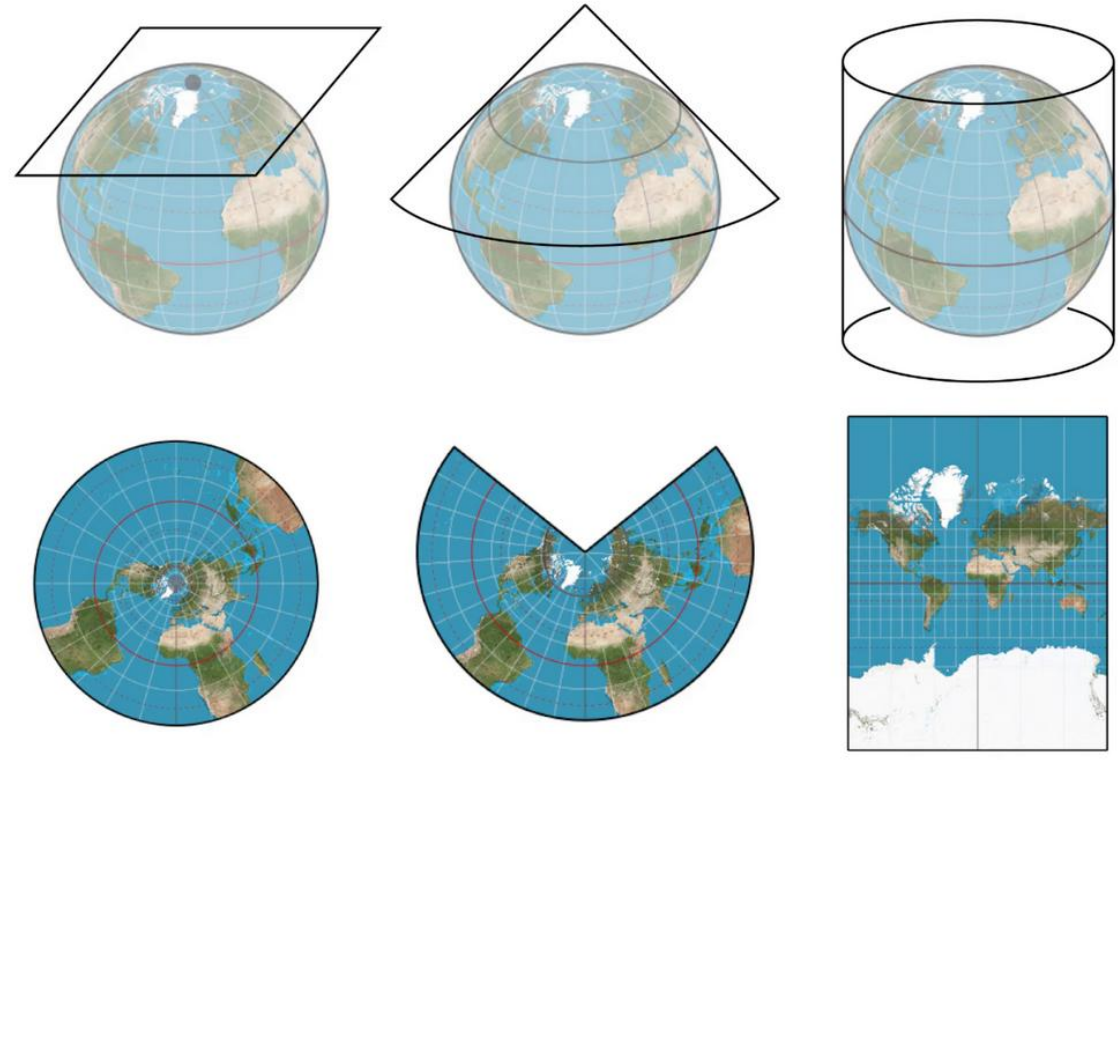


Figure 5. We chose to use gnomonic projections centered on icosahedron faces (left) for H3's map projection, projecting Earth as a spherical icosahedron (right).

Uitdaging 4

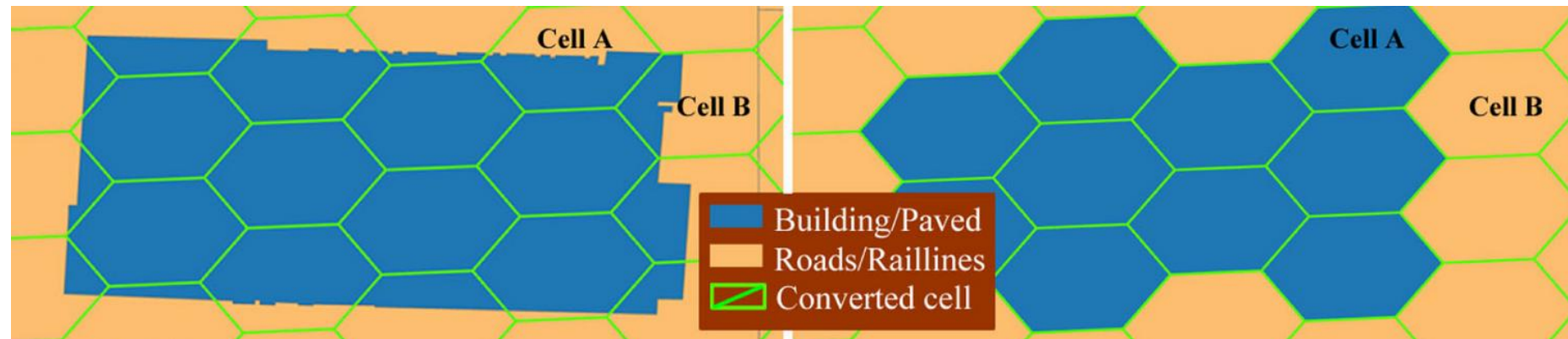
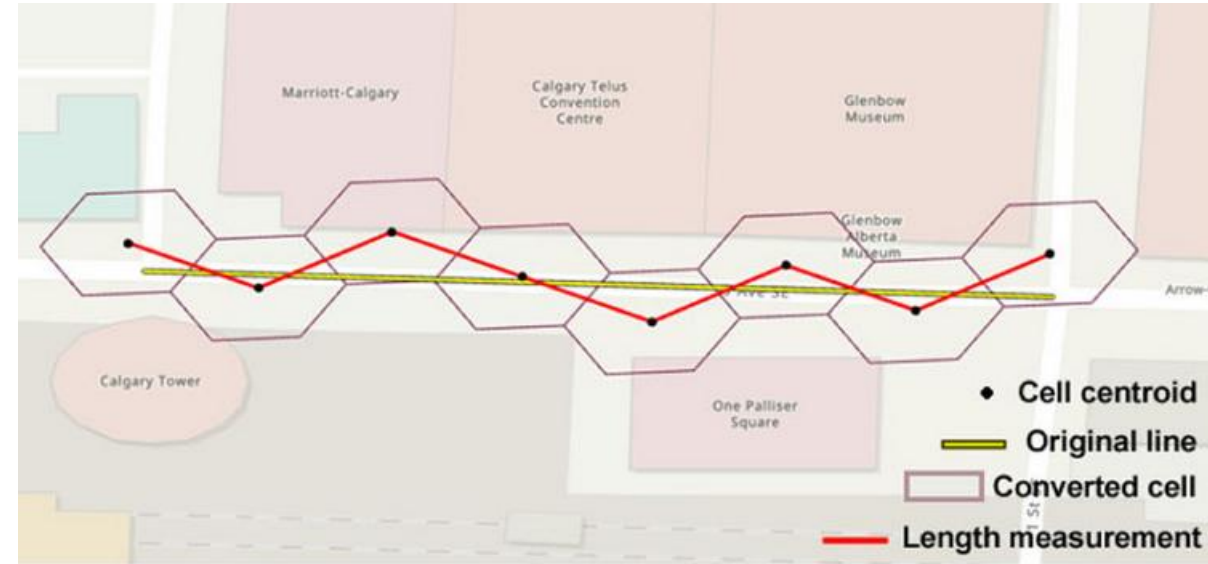
- Je accuracy verlaagt



Uitdaging 5

Het is geen “magische” oplossing voor alles. Complexe polygonen in een grid proppen kan:

- veel rekenkracht kosten
- Imperfecties en kapotte geometrieën opleveren
- Veel tijd kosten om werkbaar te maken



Wat is het DGSS en wat kan ik ermee?

Onderzoek

- Vergeleken dggridR en H3
- dggridR simpeler en meer flexibel
- H3 sneller en efficiënter

Discrete Global Grid Systems as scalable geospatial frameworks for characterizing coastal environments

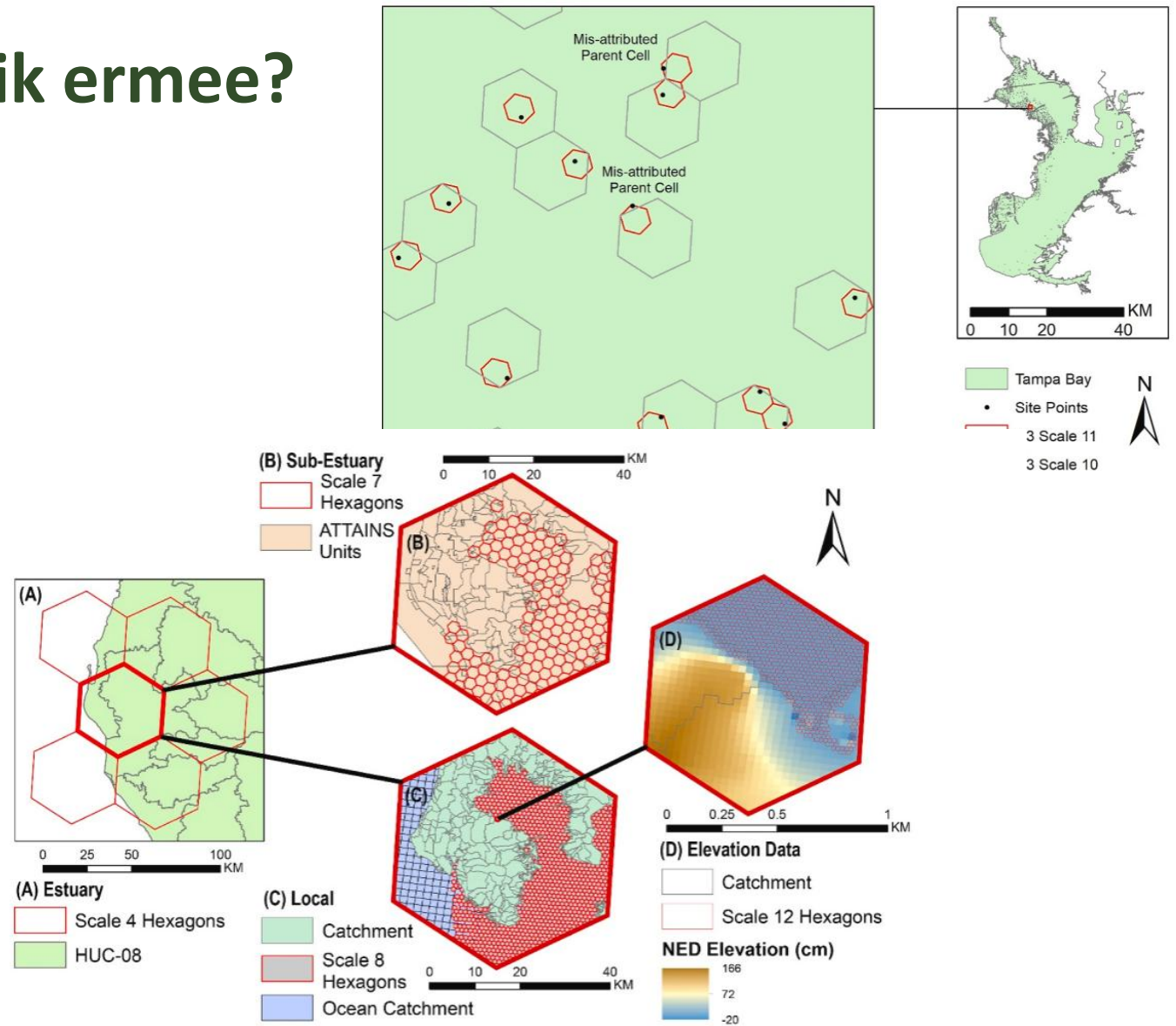


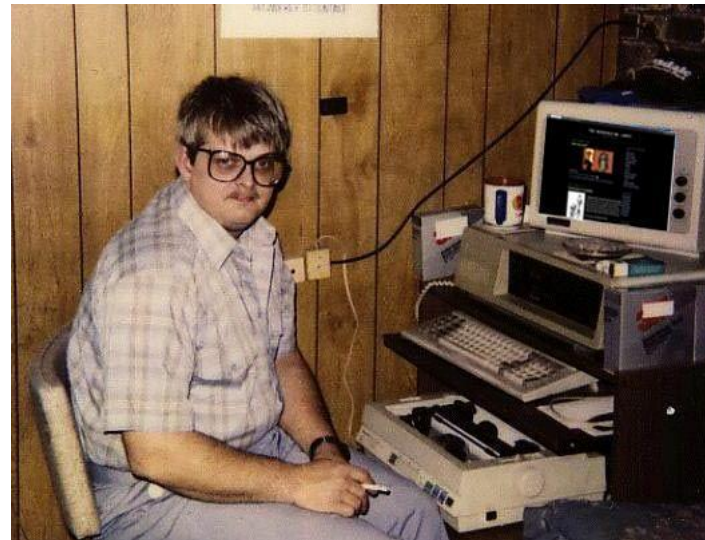
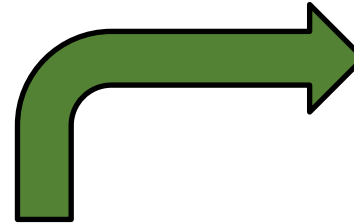
Fig. 3. Scaled H3 hexagon unit representation against other units in Tampa Bay. (A) Scale 4 H3 hexagons compared to HUC-08, (B) Scale 7 H3 hexagons compared to ATTAINS units, (C) Scale 8 hexagons compared to NHDPlus catchments and ocean catchments, (D) Scale 12 H3 hexagons compared to 30 m NED elevation raster.

Wat is cloud/distributed computing?

Een definitie

Het leveren van online computing services, zoals

- Rekenkracht
- Opslag
- Databases
- Netwerk
- Software



Voordelen

- Je hoeft zelf geen hardware te kopen
 - Je betaalt alleen voor wat je gebruikt
 - Je hoeft niet fysiek bij je hardware te zijn
 - Out-of-the-box technische security (disaster recovery, backups, etc.)
 - Betrouwbare hardware (maar ook niet altijd)
 - Gemakkelijk bij elkaar te klikken security en governance → wij regelen en garanderen het voor je!
-
- Don't be the biggest fish in the pond



Nadelen

- Afhankelijk van een externe partij (downtime, security)
- Vendor lock-in
- Complexe platforms met specifieke benodigde kennis
- Lange-termijn Kosten managen



Takeaways from the July 2024 Microsoft Azure Service Outage

Aaron Sherrill September 19, 2024 IT News

Breach at Turkey's Pegasus Airlines Exposes 6.5TB of Data

The Misconfigured AWS S3 Bucket That Led to the Breach Has Now

RANSOMWARE

Compromised AWS Keys Abused in Codefingerr Ransomware Attacks

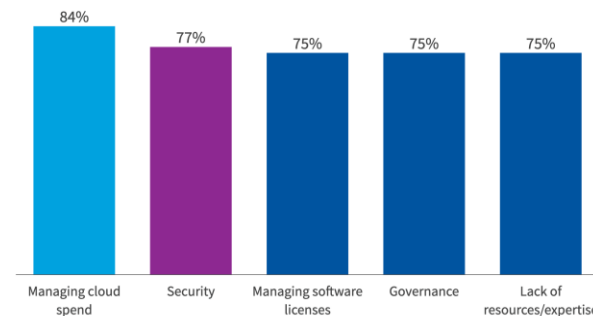
A ransomware group tracked as Codefingerr is using compromised AWS keys to encrypt S3 bucket data using SSE-C.



By Ionut Arghire | January 14, 2025 (7:12 AM ET)

A threat actor has been observed abusing compromised AWS keys to encrypt data in S3 buckets and demand a ransom payment in exchange for the encryption keys, cybersecurity firm Halcyon reports.

Top cloud challenges for all respondents



N=759
Source: Flexera 2025 State of the Cloud Report
flexera

De keuze is reuze!

- Veel cloud providers beschikbaar
- Kunnen grotendeels hetzelfde
- Echter, verschil zit in:
 - ❖ Configureerbaarheid
 - ❖ Kosten
 - ❖ Open/closed source
 - ❖ Integratie met andere systemen en providers



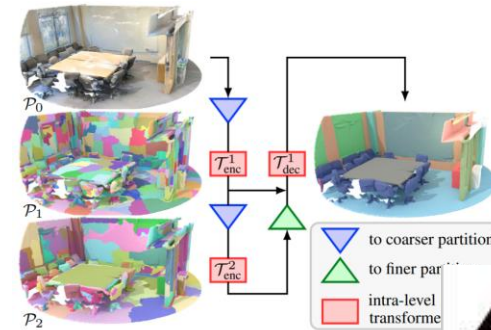
databricks



Wat is cloud/distributed computing?

Wat maakt dit uit voor GIS?

- De tijd van ‘monolieten’ lijkt ten einde, en toolsets worden steeds meer open source
- Steeds meer grote volumes ongestructureerde data. Niet meer lokaal te managen.
- Begrijpen, analyseren en bruikbaar maken kost rekenkracht
- Veel datapreparatie en analyse handwerk kan gescript, schaalbaar & herhaalbaar worden gemaakt
- 5 jaar geleden had data science hetzelfde probleem.
- Het antwoord toen: data engineering!
- De tools: Schaalbare cloud pipelines!



PostGIS



OPEN3D

Point clouds

3D Mesh

3D Model



COG

CLOUD OPTIMIZED
GEOTIFF

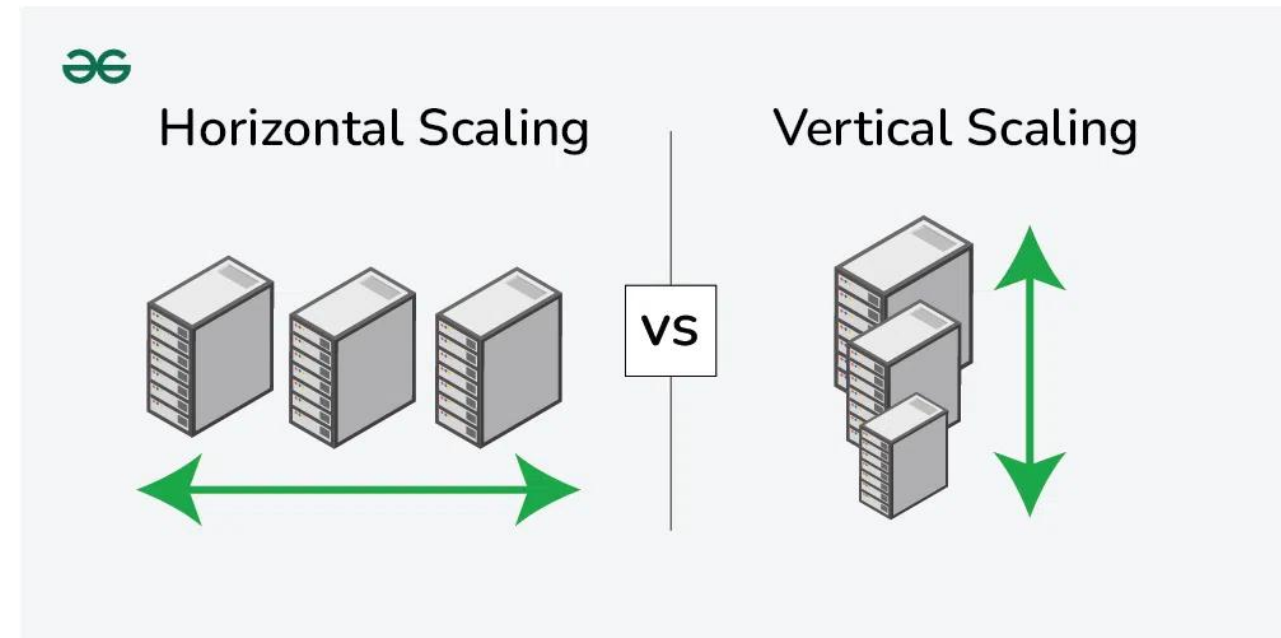
Hoe maak ik hier goed gebruik van?

Herbruikbare, gestandaardiseerde en overdraagbare functionaliteiten bouwen

- Standaard OS, packages, etc.
- One-click-run van je pipelines
- Betrouwbare deployment
- Software development best-practices

Kosten en compute zoveel mogelijk optimaliseren

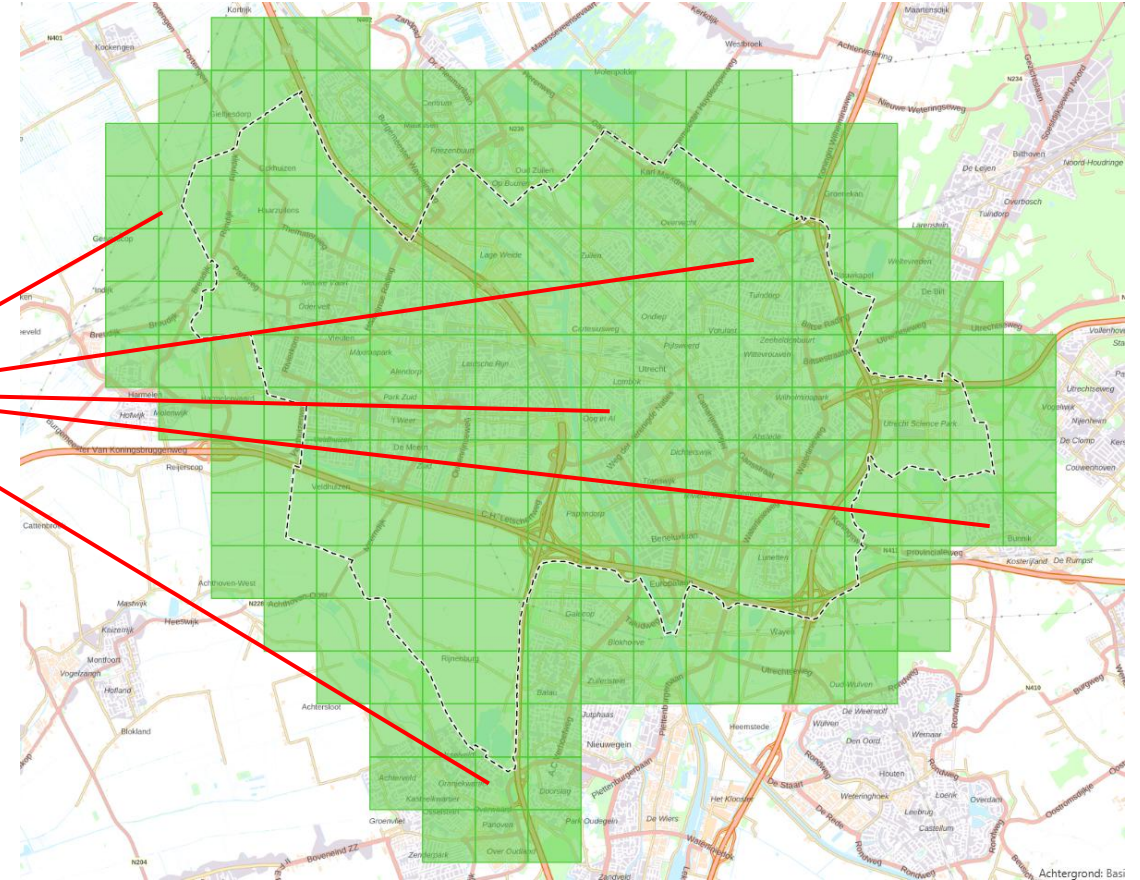
- Horizontaal schalen
- Verticaal schalen



With our powers combined...

With our powers combined...

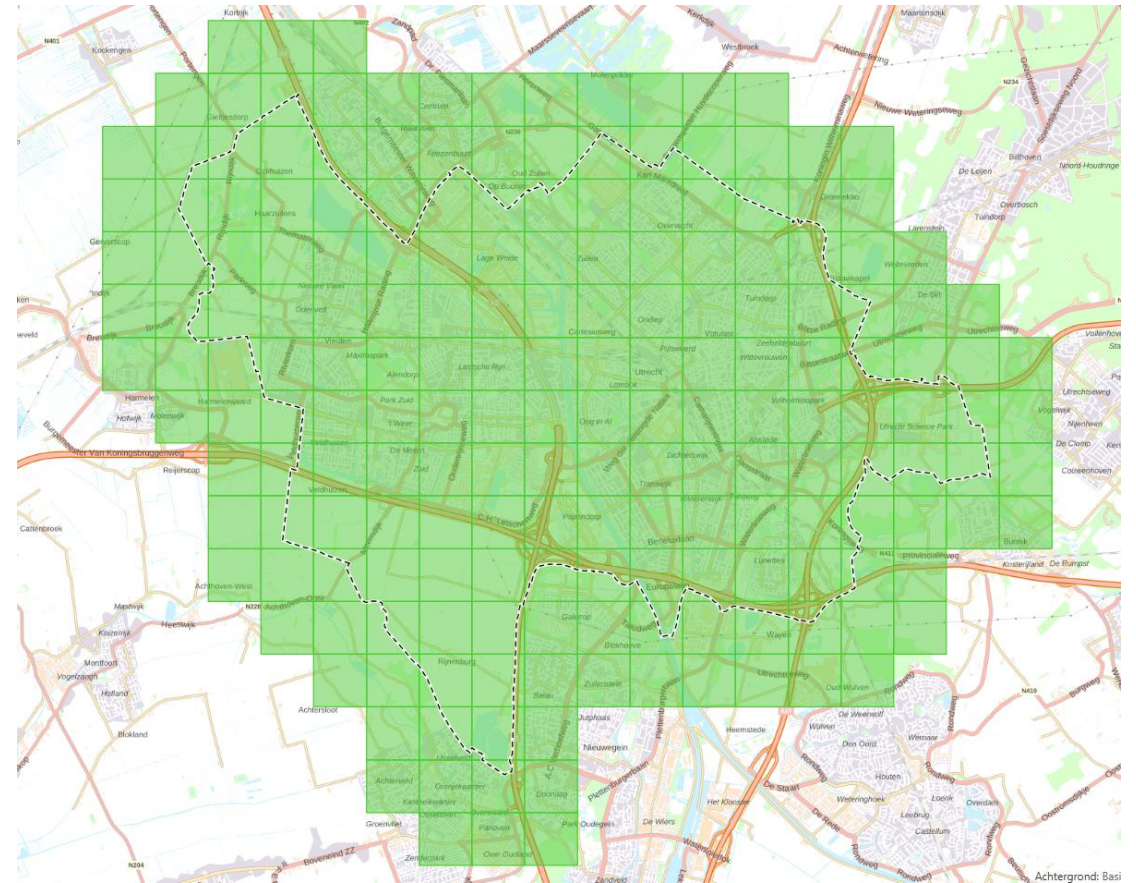
- We maken grote GIS analyses schaalbaar!
- Wat is groot?
 - Complex
 - Grote vd dataset
 - Resource-intensief



Workshop

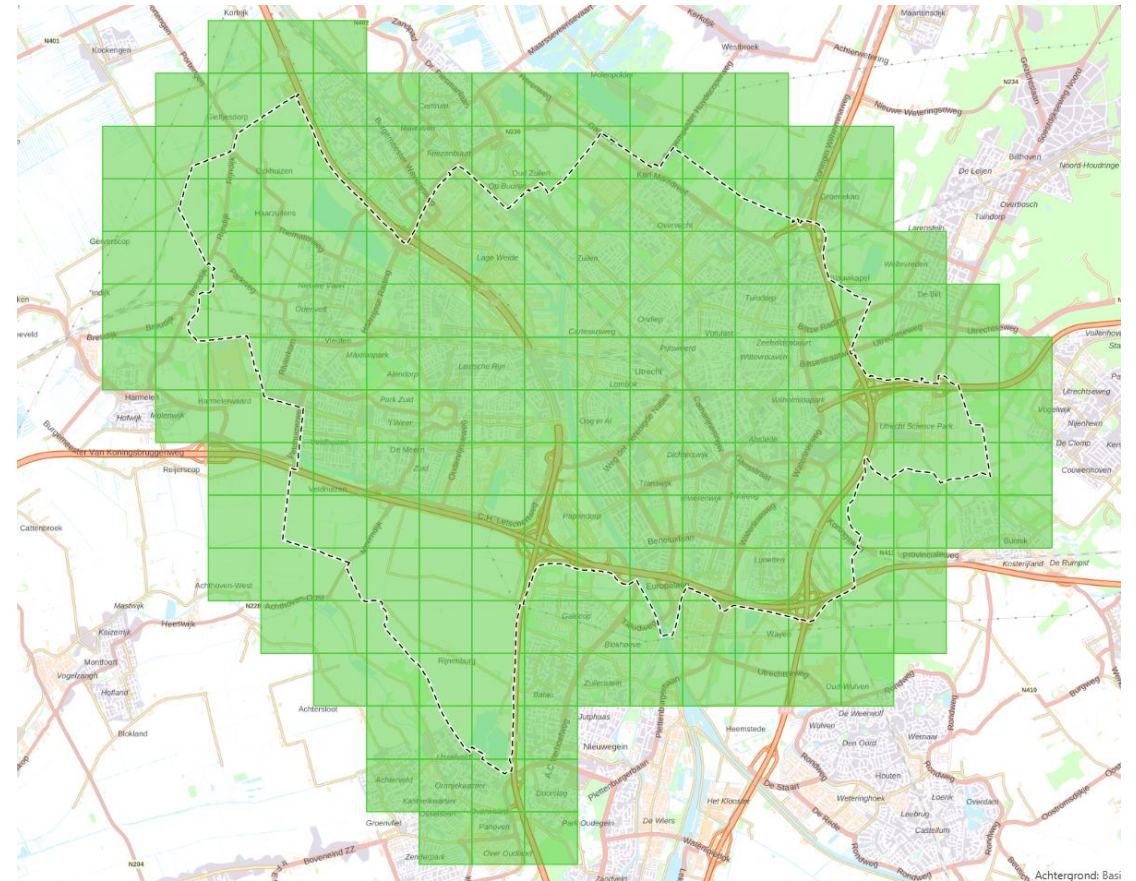
Workshop: datasets

- AHN Data van Utrecht, uit [Downloads Actueel Hoogtebestand Nederland en Beeldmateriaal](#)
- Ongeveer 32GB aan data
- Werken met een subset (1 tile) van 1km bij 1km, opgehakt in 10x10 tiles.
- Dit zie je ook terug in de naamgeving



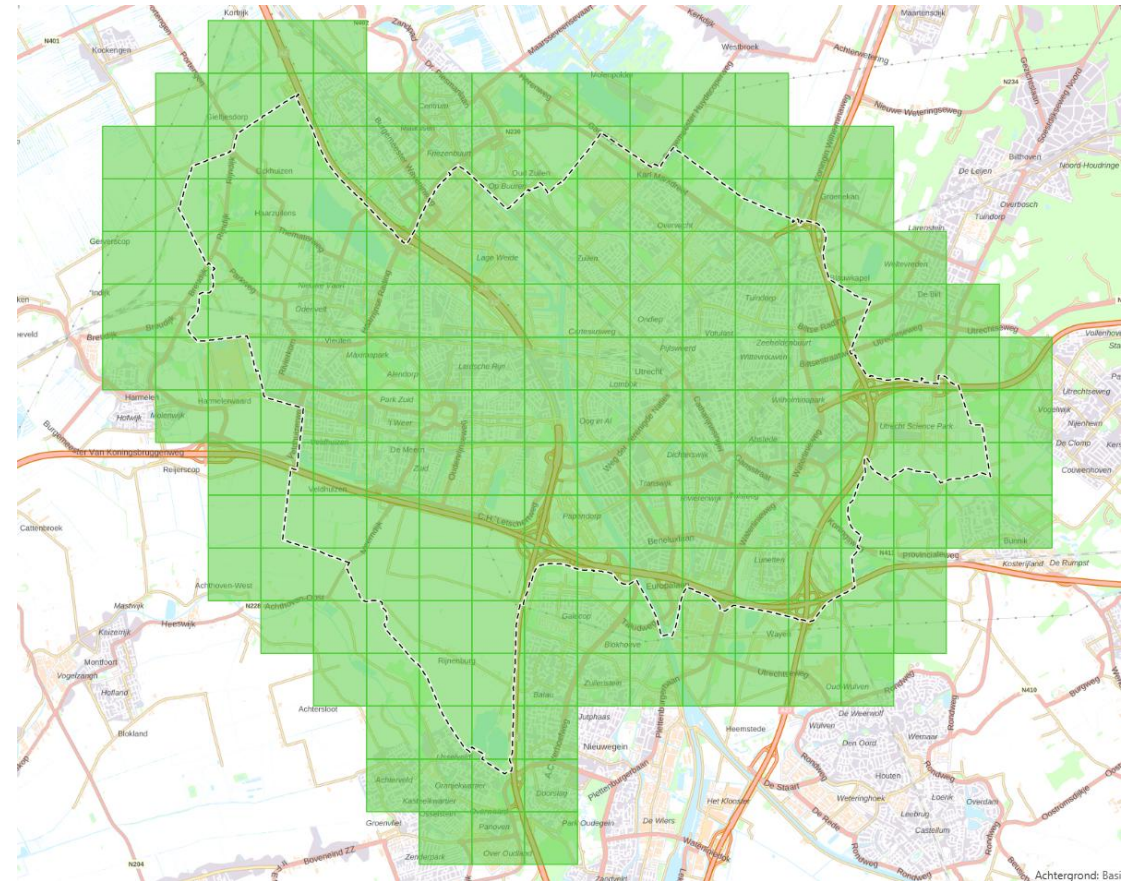
Workshop: proefening 1

- Bekend raken met de data
- Ophakken van de tile in 10 bij 10 subtiles (grid cells)
- “simpele” classificatie van grid cells



Workshop: proefening 2

- Voor 1 subtile analyse uitvoeren:
- BGT data van het gebied gebruiken voor segmentatie

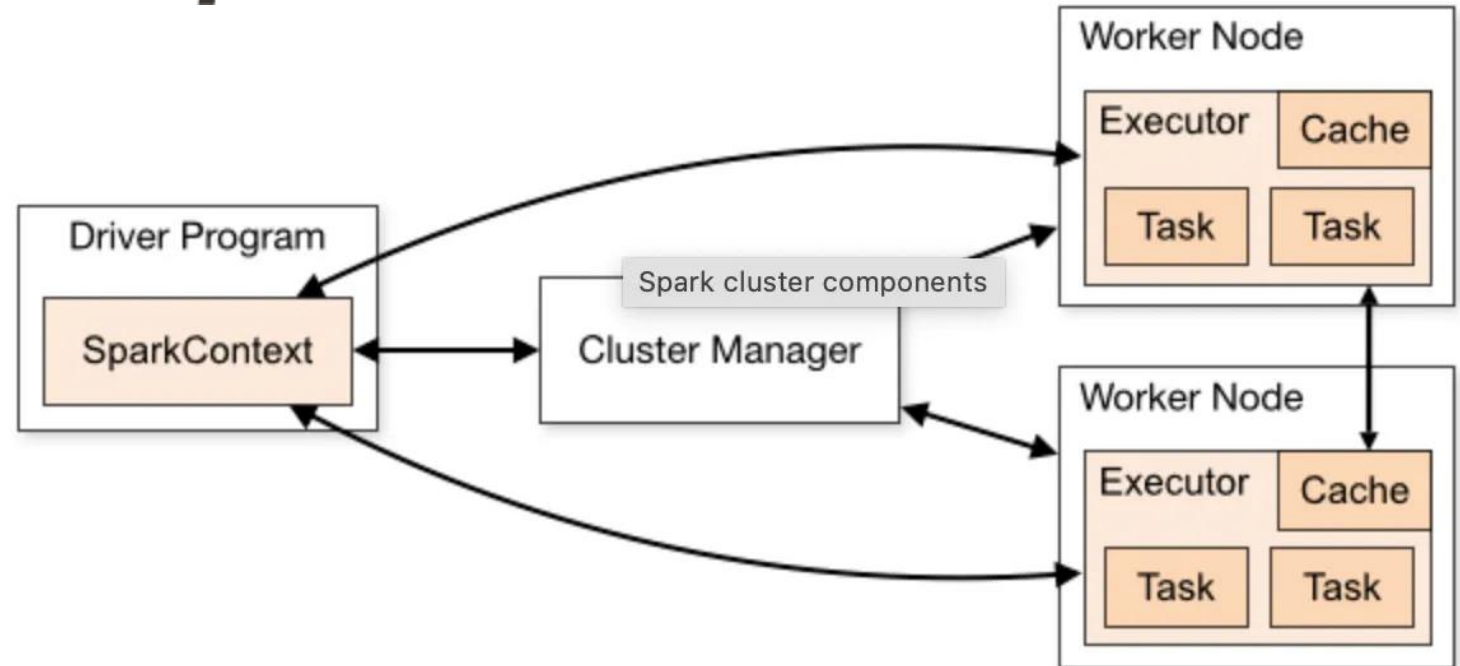


Afronding

Wat hebben we gezien?

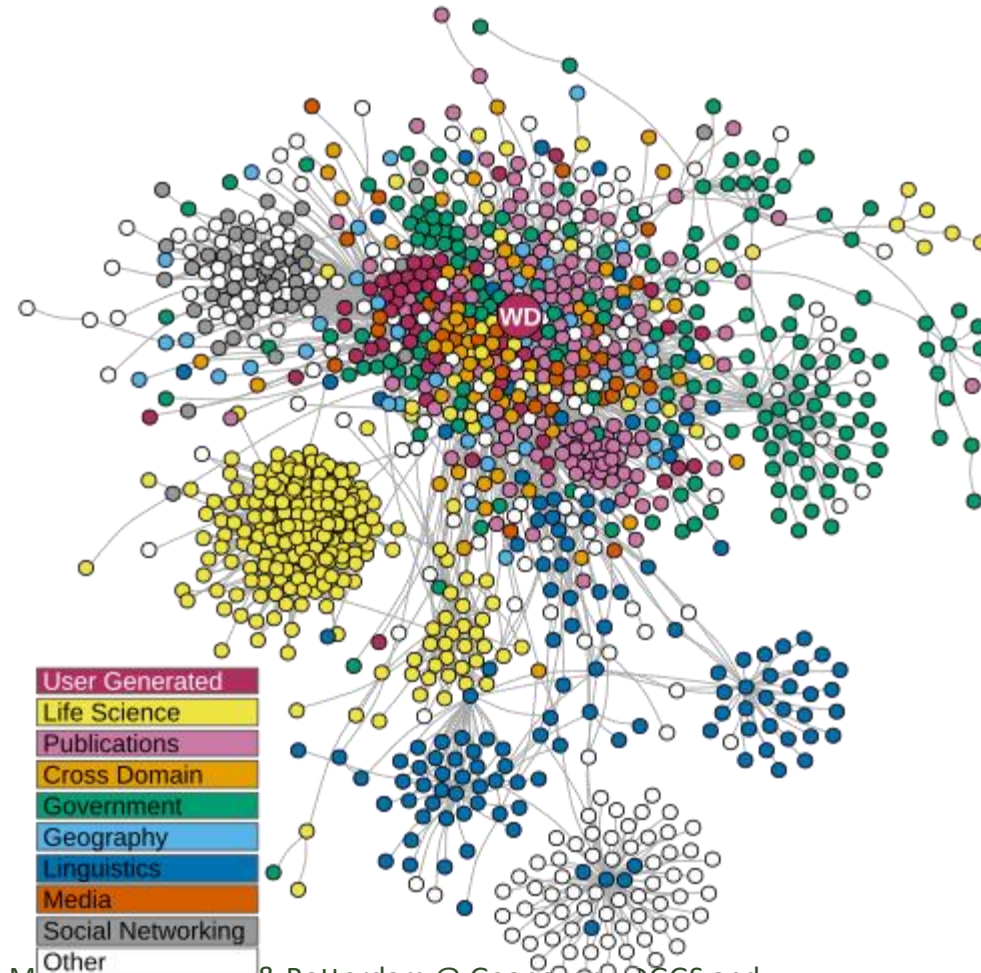
Overhead by distributed computing

- Extra spullenboel: Manager met eigen env, worker nodes, etc.
- Verdelen van werk naar nodes
- Netwerk latency
- serialization
- Managen van taken in parallel
- Verkeer tussen workers
- Alle workers hebben een eigen omgeving (X* je omgeving draaien)



Afronding

- Wat niet meer in de workshop paste: DGGs & linked data



Mimir consultancy & Rotterdam @ Geonovum: DGGs and cloud computing!

Afronding

- Nieuwe ontwikkelingen



Artificial Intelligence Discrete Global Grid System (DGGS) for Disaster Management: Call for Participation (CFP)

Version 1.1 - 17 April 2025

Table of Contents

Artificial Intelligence Discrete Global Grid System (DGGS) for Disaster Management: Call for Participation (CFP)

1. Introduction
2. Background
3. Eligibility Criteria

Afronding

Wat vonden jullie ervan!

Bedankt!

Email: s.braaksma@mimirconsultancy.org

Linkedin: Sjoerd Braaksma

Github: SjoerdBraaksma

Referenties

1. <https://www.sciencedirect.com/science/article/pii/S1364815221002528>
2. https://www.researchgate.net/publication/358095638_GIScience_research_challenges_for_realizing_discrete_global_grid_systems_as_a_Digital_Earth
3. <https://www.sciencedirect.com/science/article/pii/S1195103624000703>
4. https://www.researchgate.net/publication/354545167_Intelligent_Geospatial_Maritime_Risk_Analytics_Using_The_Discrete_Global_Grid_System
5. <https://pro.arcgis.com/en/pro-app/3.4/tool-reference/spatial-statistics/h-whyhexagons.html>
6. <https://www.kontur.io/blog/why-we-use-h3/>
1. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-whyhexagons.html>