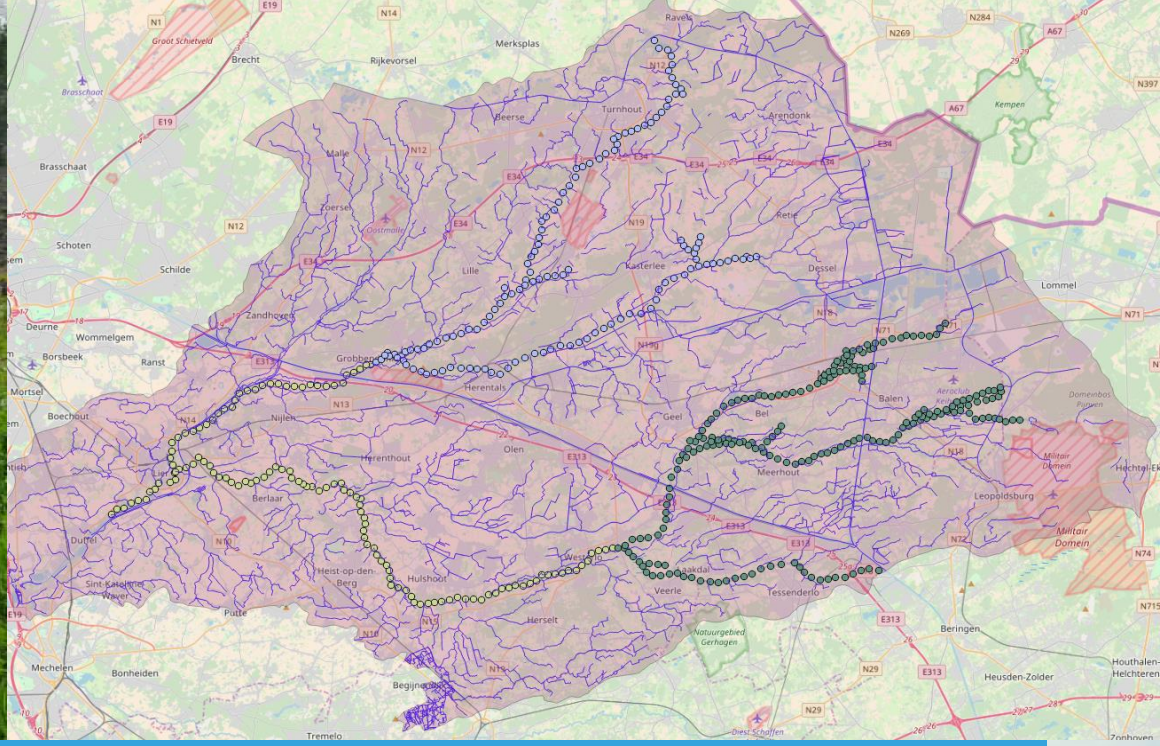


WORKING WITH POSTGIS

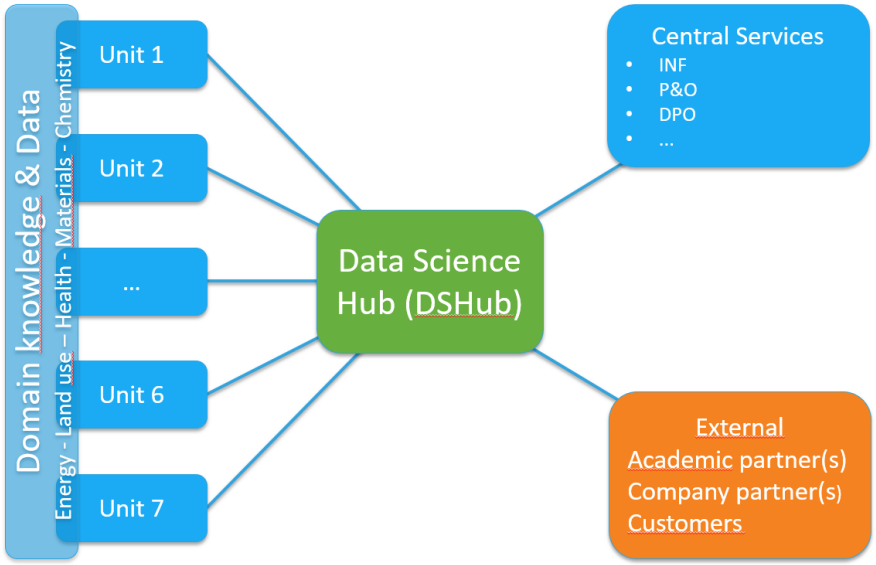
CASE: THE FLEMISH RIVER NETWORK

Rik Hendrix, VITO Data Science Hub
FOSS4G Belgium
Brussels, 24 Oct 2019

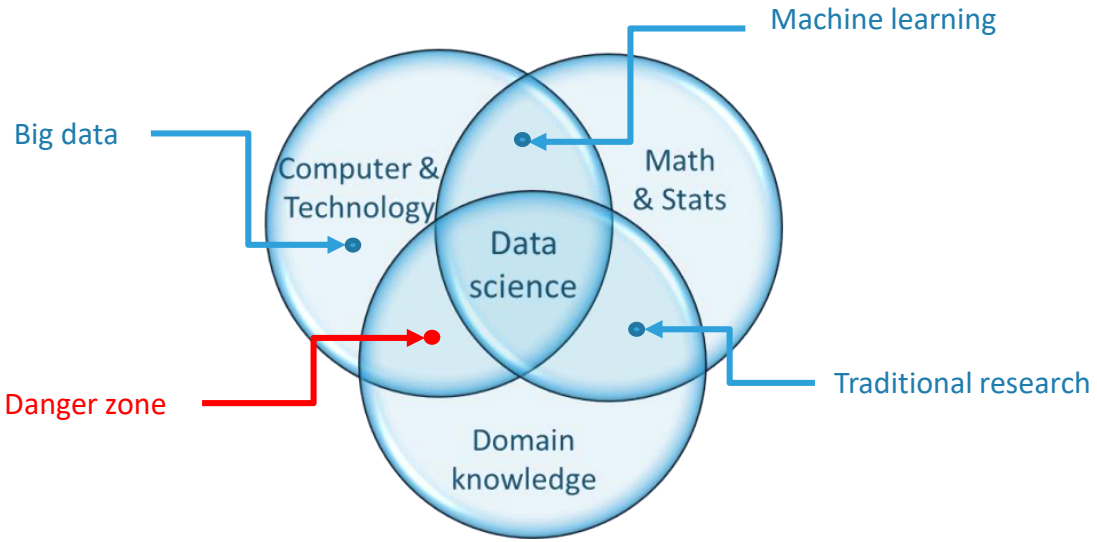
vito

VITO DATA SCIENCE HUB

- ENERGY
- CHEMISTRY
- MATERIALS
- HEALTH
- LAND USE



DATA SCIENCE?



data science is NOT a standalone discipline

data science is a *team* effort

Why PostGIS is awesome

It’s free
It’s fast
It’s fun

Why PostGIS is awesome

It's free
It's fast
It's fun

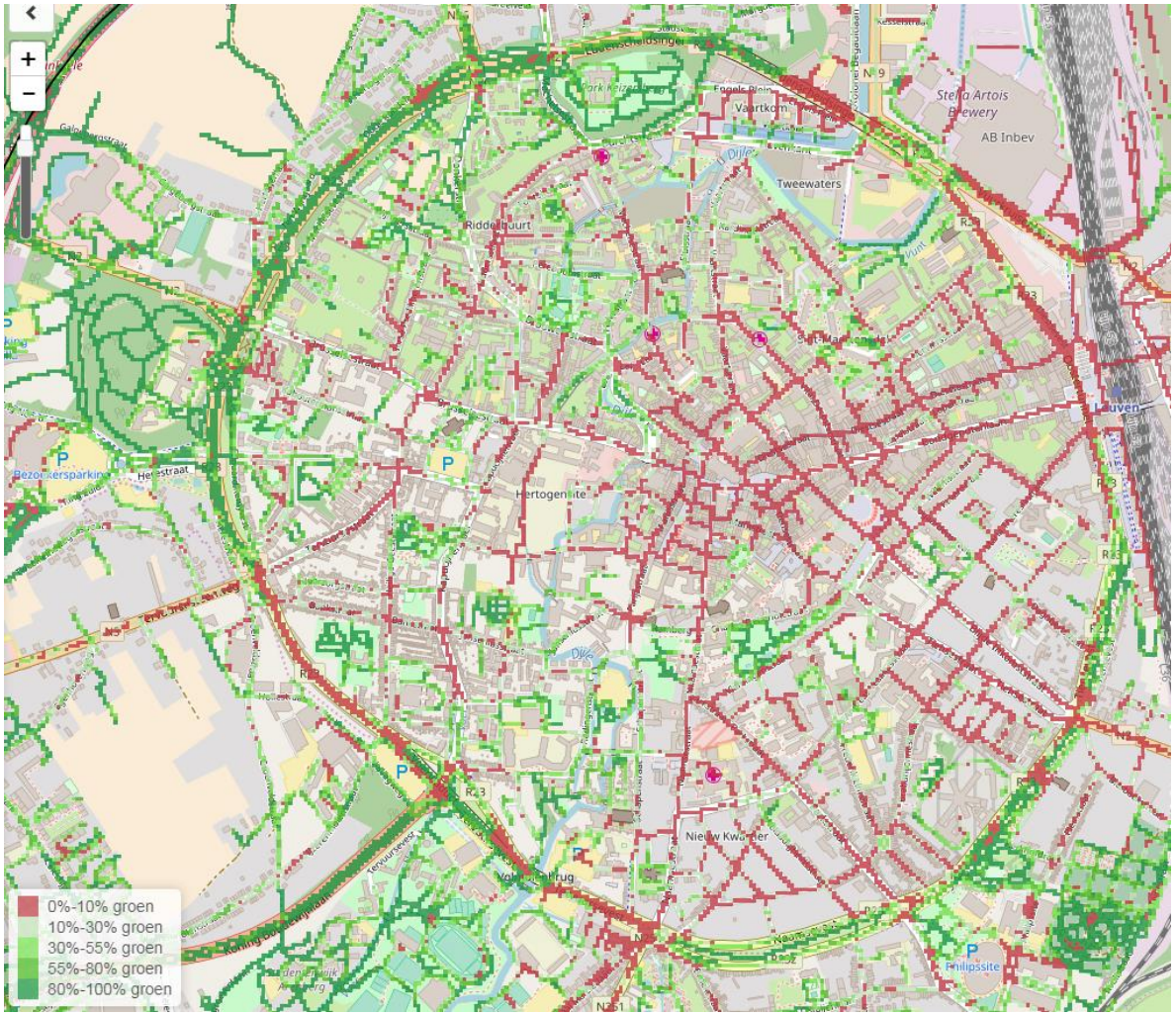
It's multi-platform

It combines the best of 2 worlds
- Spatial accuracy of desktop GIS
- Processing of large volumes

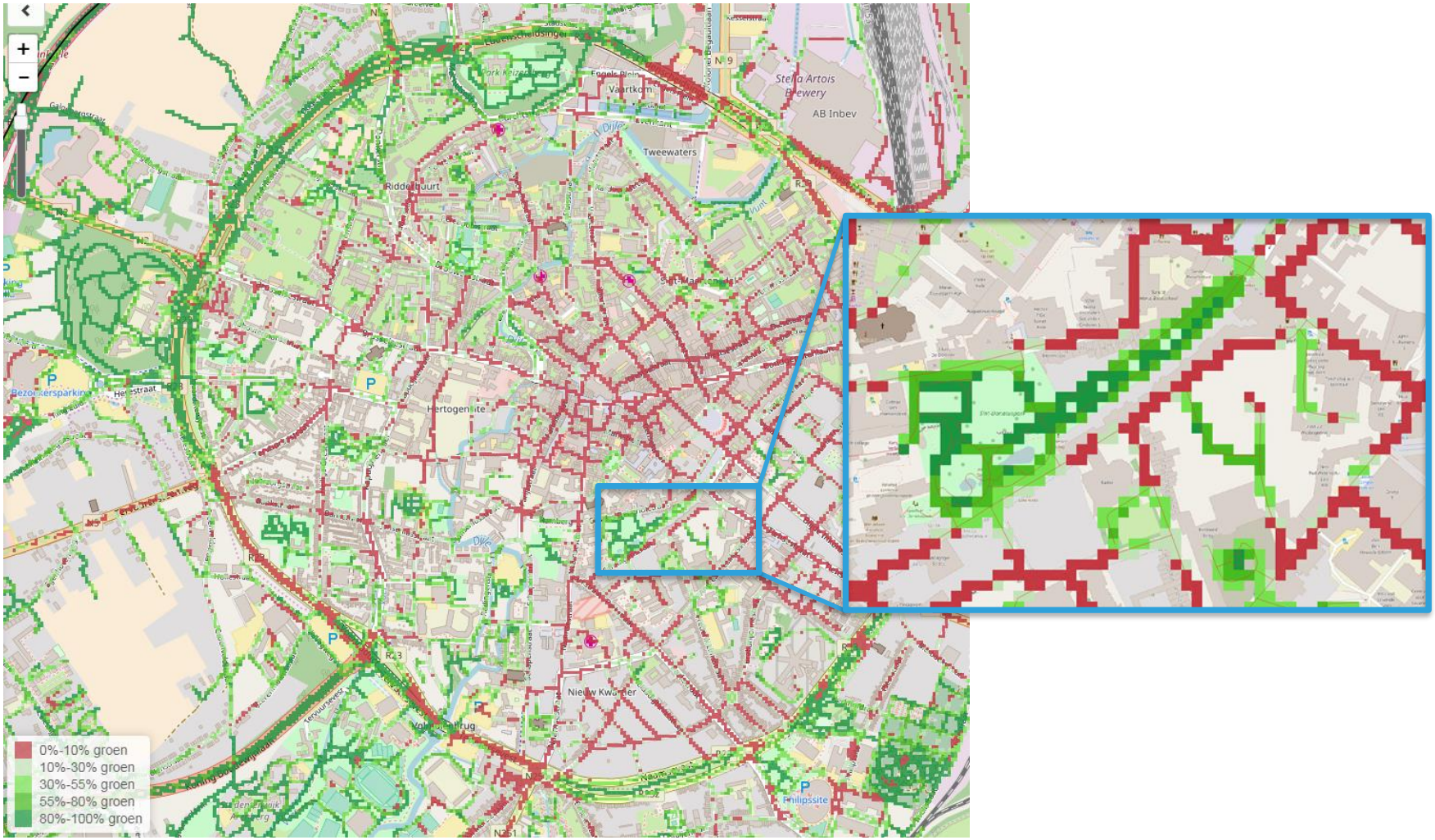
Query language SQL close to human language

Note: all maps have been made with QGIS

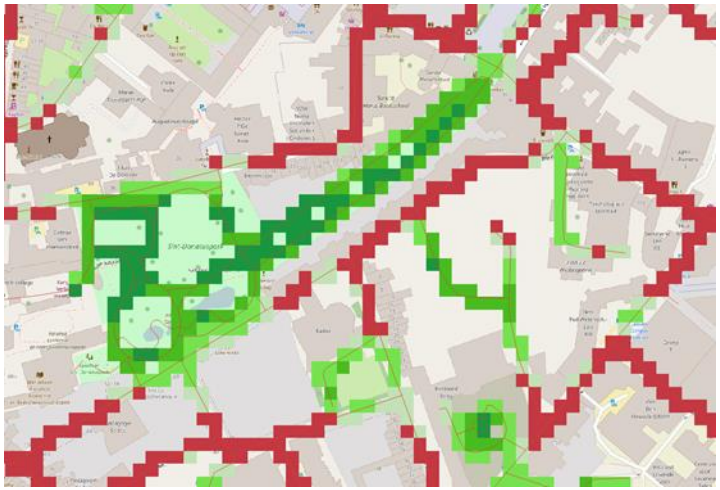
RASTER VS VECTOR: STREET GREENERY MAP (“STRAATGROENKAART”)



RASTER VS VECTOR: STREET GREENERY MAP (“STRAATGROENKAART”)



RASTER VS VECTOR: STREET GREENERY MAP (“STRAATGROENKAART”)



Vector map:

- Nicer result
- More accurate
- More suitable for analytics (e.g. routing)
- Less computing power required

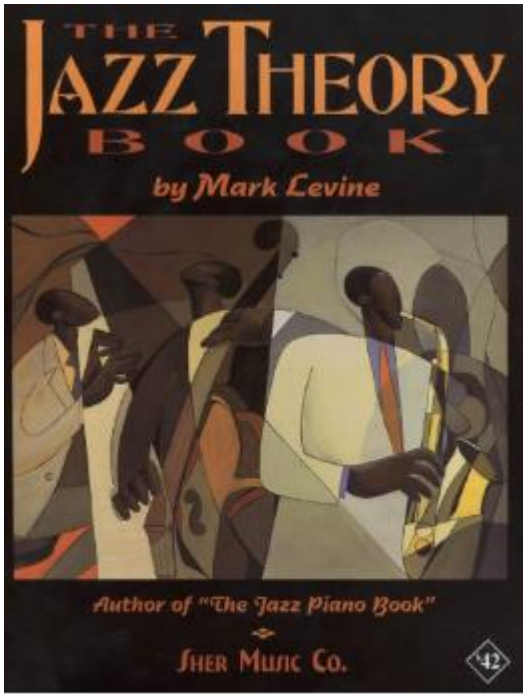
vito

WORKING WITH POSTGIS: 99% “READ THE MANUAL”, 1% EXPERIENCE

A great jazz solo consists of:

- 1% magic
- 99% stuff that is
  - explainable,
  - analyzable,
  - categorizeable,
  - doable.

Mark Levine, The Jazz Theory Book

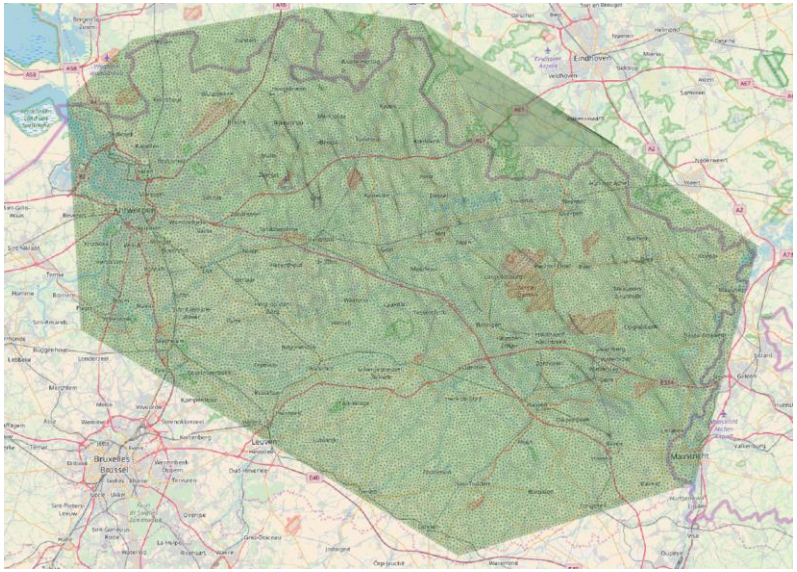


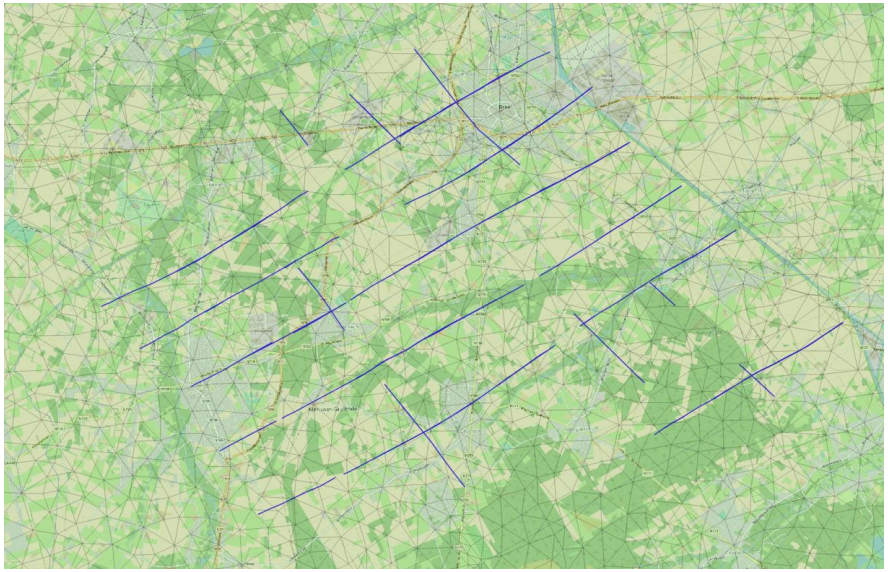
vito

WORKING WITH POSTGIS: 99% “READ THE MANUAL”, 1% EXPERIENCE

Calculation in 3D of 50 000 intersections of 1 plane, defined by 300 000 points, and 50 000 lines

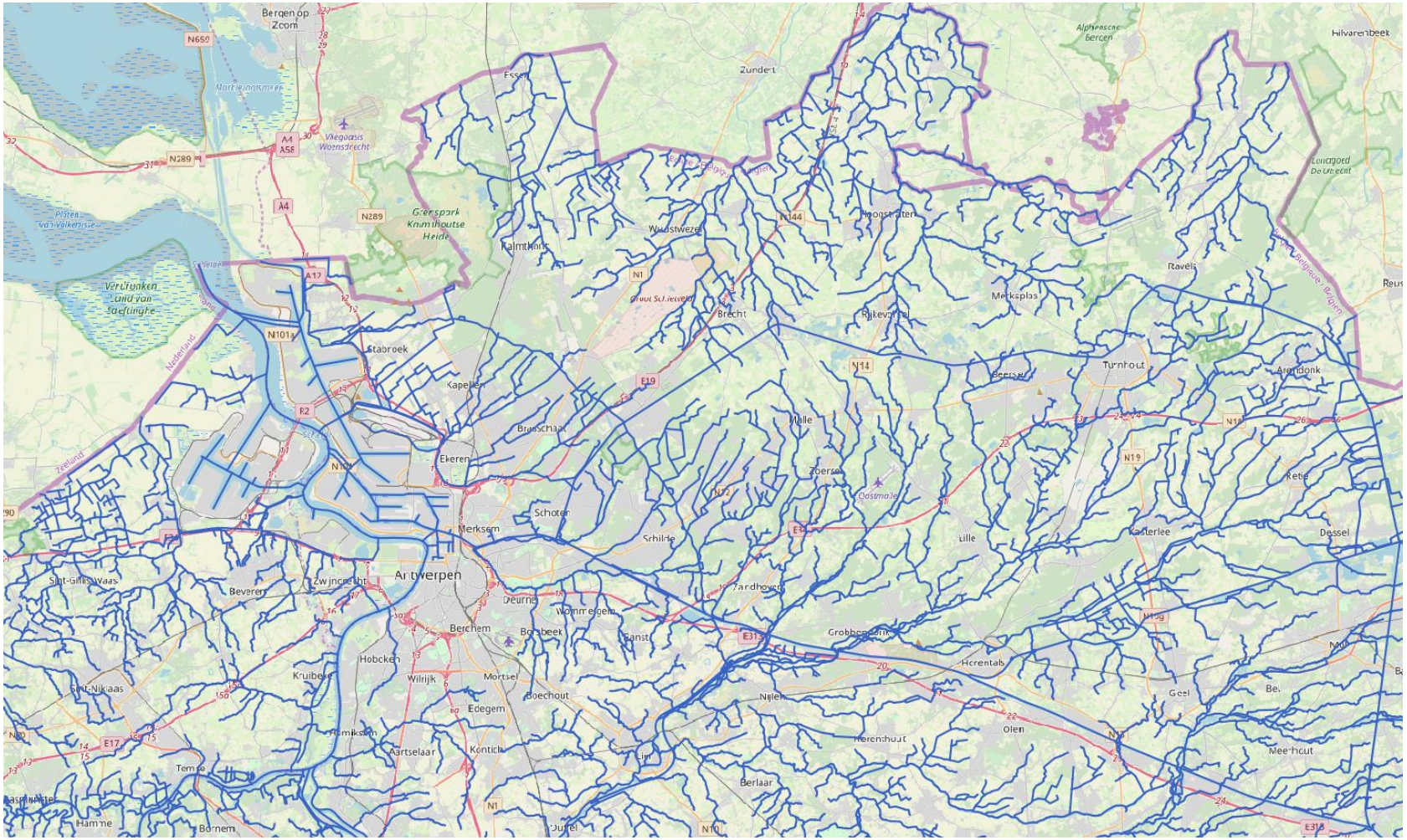
Original code: > 1 week
After changes in data model: 1 ½ hour





vito

LINKING THE RIVER SEGMENTS



LINKING THE RIVER SEGMENTS

Table river_segments
Attributes: id, name, geom, …

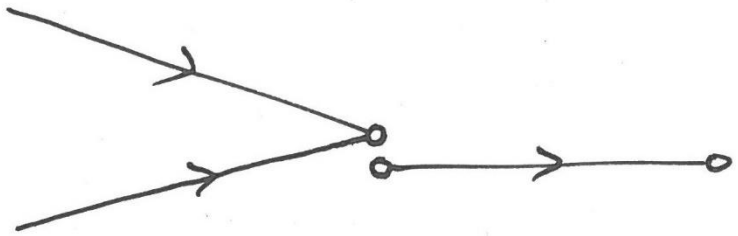
```
alter table river_segments
 add column id_downstream integer;

update river_segments a
   set id_downstream = b.id
  from river_segments b
 where ST_StartPoint(b.geom)= ST_EndPoint(a.geom)
```

HIERARCHICAL DATABASE OF RIVER SEGMENTS

LINKING RIVER SEGMENTS: ISSUES

ISSUE 1: Start and end points do not match exactly



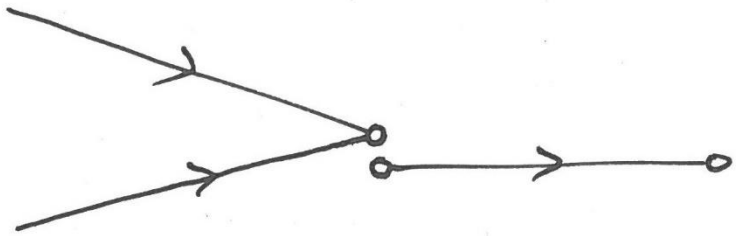
SOLUTION

```
update river_segments a
   set id_downstream = b.id
  from river_segments b
 where ST_StartPoint (b.geom) = ST_EndPoint (a.geom)
       ST_Distance (ST_StartPoint (b.geom),
                    ST_EndPoint (a.geom)) < 4
```


vito

LINKING RIVER SEGMENTS: ISSUES

ISSUE 1: Start and end points do not match exactly



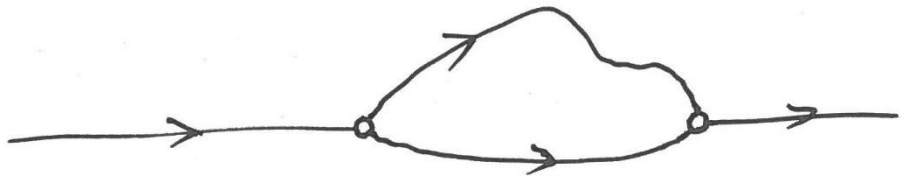
SOLUTION

```
update river_segments a
   set id_downstream = b.id
  from river_segments b
 where ~~ST_StartPoint (b.geom) = ST_EndPoint (a.geom)~~
       ST_Distance (ST_StartPoint (b.geom),
                    ST_EndPoint (a.geom)) < 4
   and not exists (SELECT 1 from river_segments c
                   WHERE  (ST_Distance (ST_StartPoint (c.geom),
                                        ST_EndPoint (a.geom))
                           < ST_Distance (ST_StartPoint (b.geom),
                                          ST_EndPoint (a.geom)))
   and b.id <> a.id
```


vito

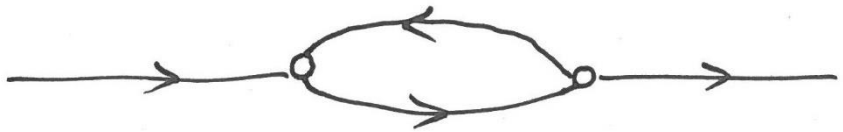
LINKING RIVER SEGMENTS: ISSUES

ISSUE 2: Branching



Only local impact. Upstream and downstream are OK.

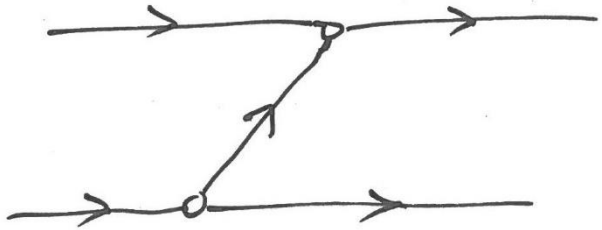
ISSUE 3: Circular streams



Solution: dependent on requirements of the application.

Id_downstream_1 =
Id_downstream_2 =

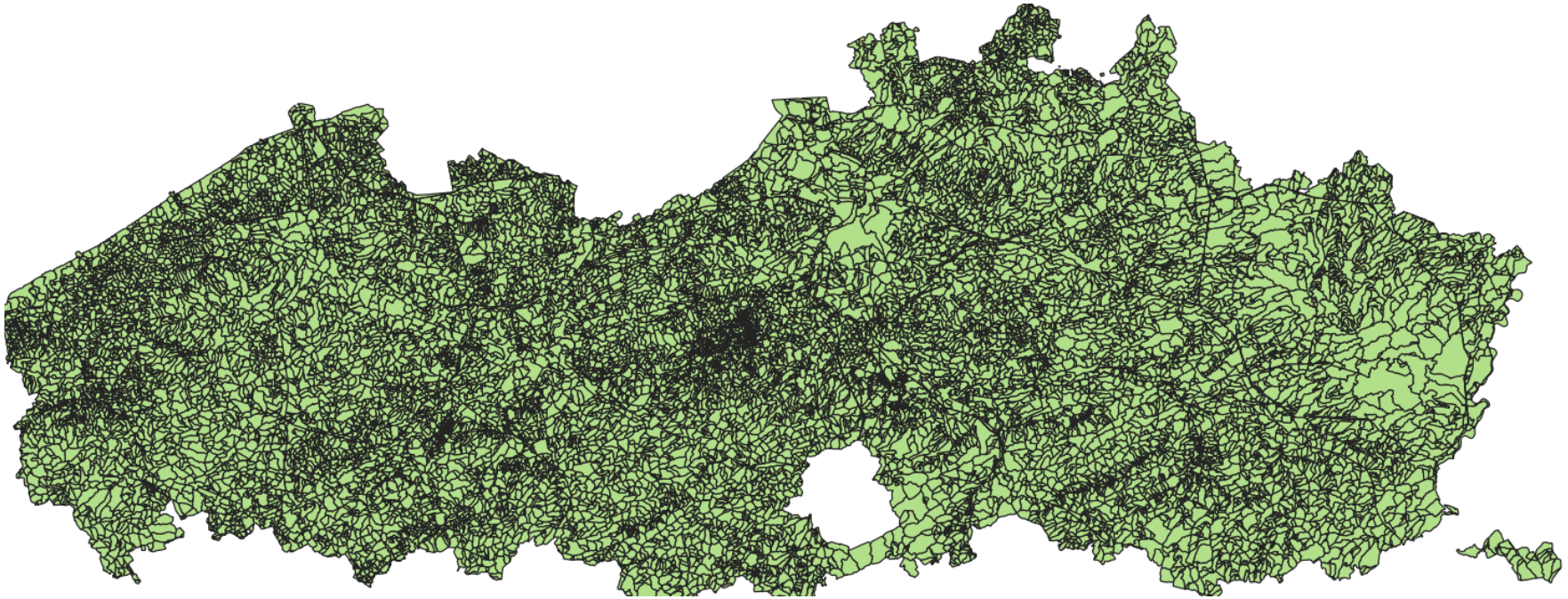
ISSUE 3: Connections to other river segments



Side effect : you might lose the power of a relational database for some types of queries

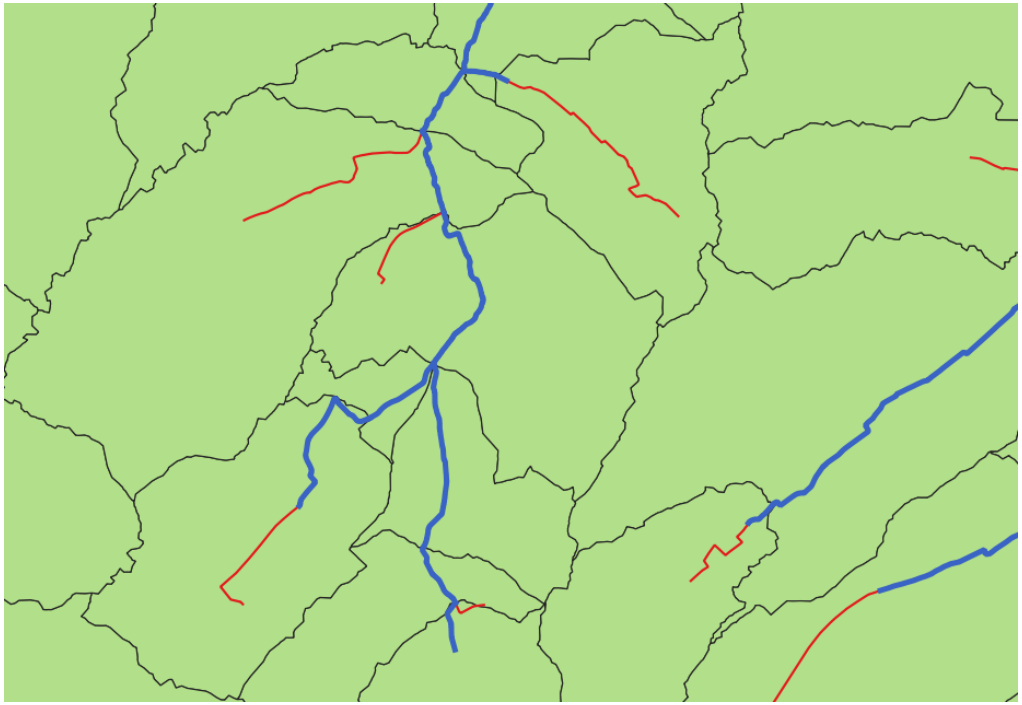
vito

NEXT STEP: LINKING RUN-OFF AREAS (“AFSTROOMGEBIEDEN”)



vito

NEXT STEP: LINKING RUN-OFF AREAS



Link every run-off area with 1 river segment (drop river segments that do not have their own run-off area).

HIERARCHICAL DATABASE OF RUN-OFF AREAS

vito

LINKING RIVER SEGMENTS AND RUN-OFF AREAS: POSSIBLE APPLICATIONS

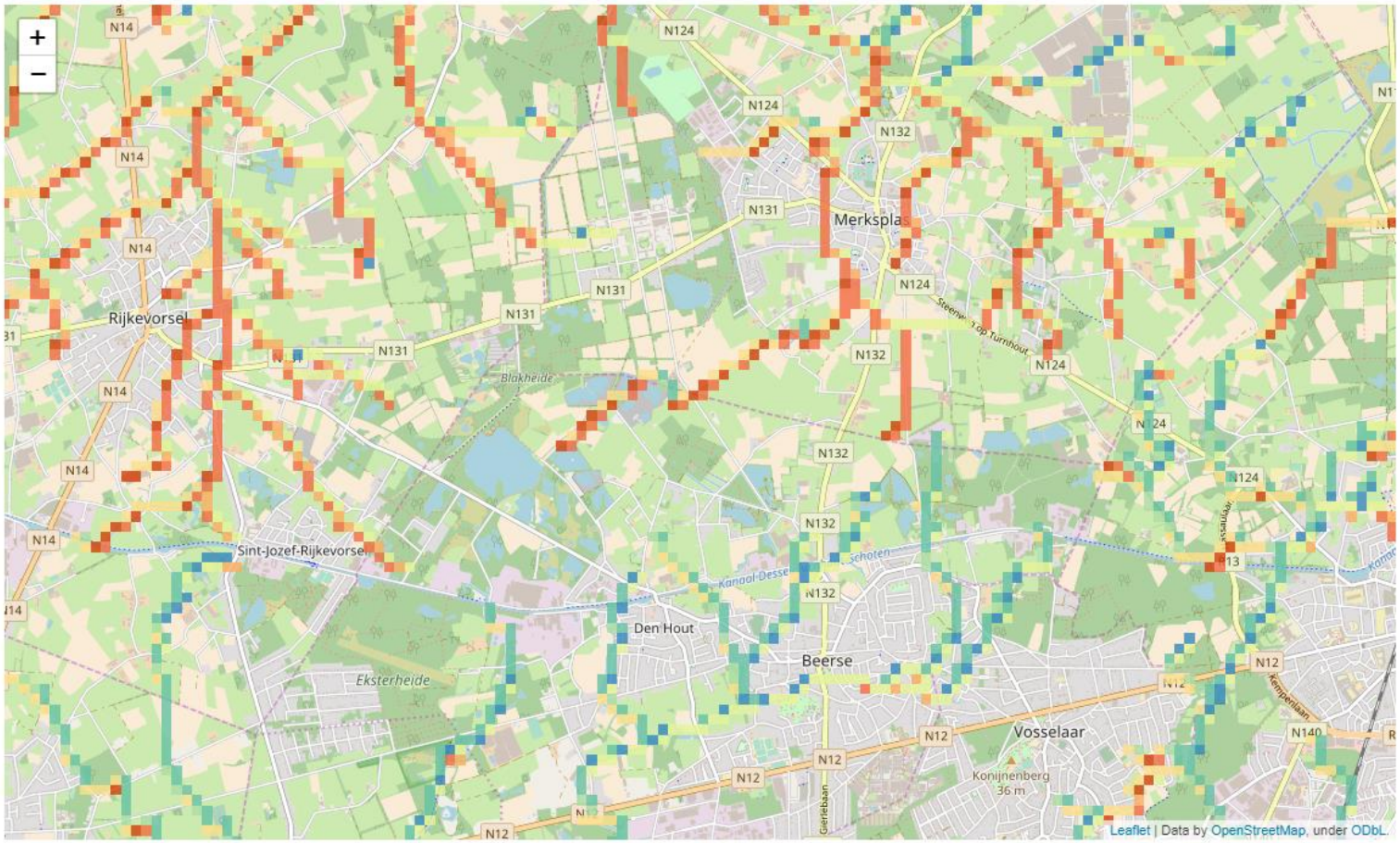
- From any given point, follow the path downstream for a given distance
- Local Drain Direction map
- Find the supply area for any given river segment
- …

This is only the beginning
This a starting point for analytics (supervised and unsupervised learning)

vito

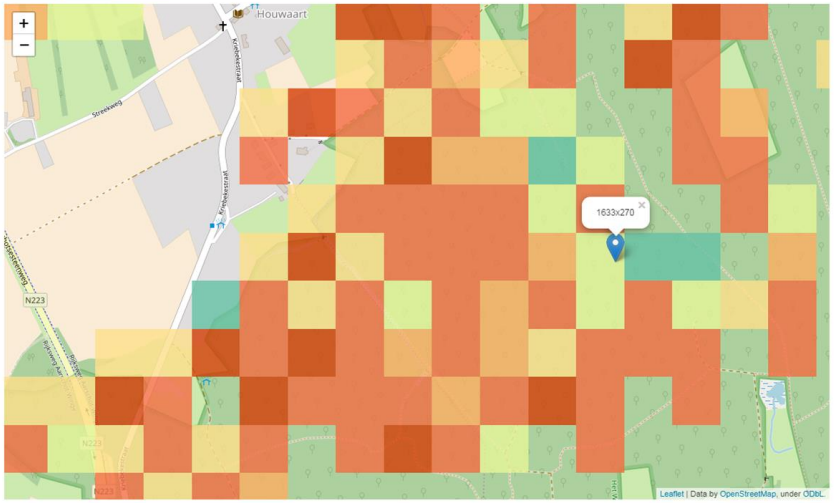
LOCAL DRAIN DIRECTION MAP

7	8	9
4	5	6
1	2	3

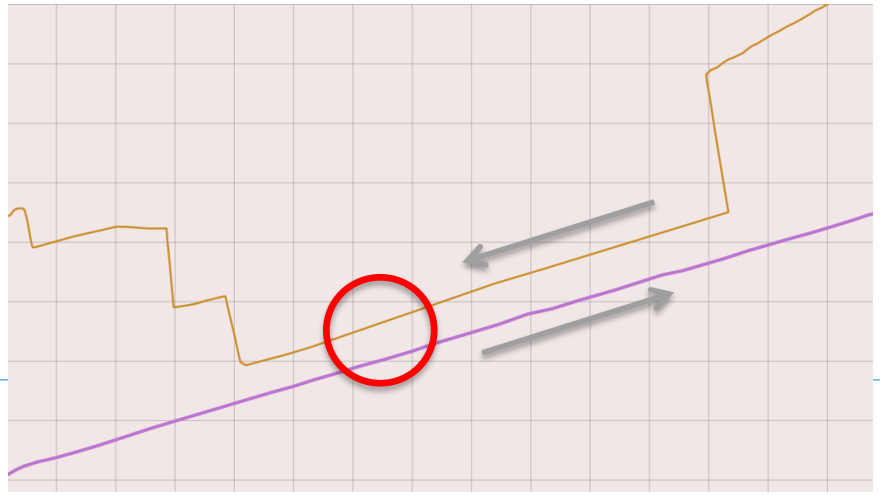


LOCAL DRAIN DIRECTION MAP: INFORMATION LOSS WITH RASTER MAPS

Problem 1: inaccurate for dense network



Problem 2: more than 1 value needed for 1 cell



vito

FIND THE SUPPLY AREA FOR ANY GIVEN RIVER SEGMENT (“TOELEVERGEBIED”)

Hierarchical query: find the run-off area id’s that make up the supply area
This takes < ½ sec, even for > 10 000 results

```
WITH RECURSIVE ids_supply_area AS (
    SELECT id, id_downstream
    FROM runoff_areas
    WHERE id = <given_id e.g. 6033854>
  UNION ALL
    SELECT run.id, run.id_downstream
    FROM runoff_areas run, ids_supply_area sup
    WHERE run.id_downstream   = sup.id
)
SELECT *
FROM ids_supply_area
```

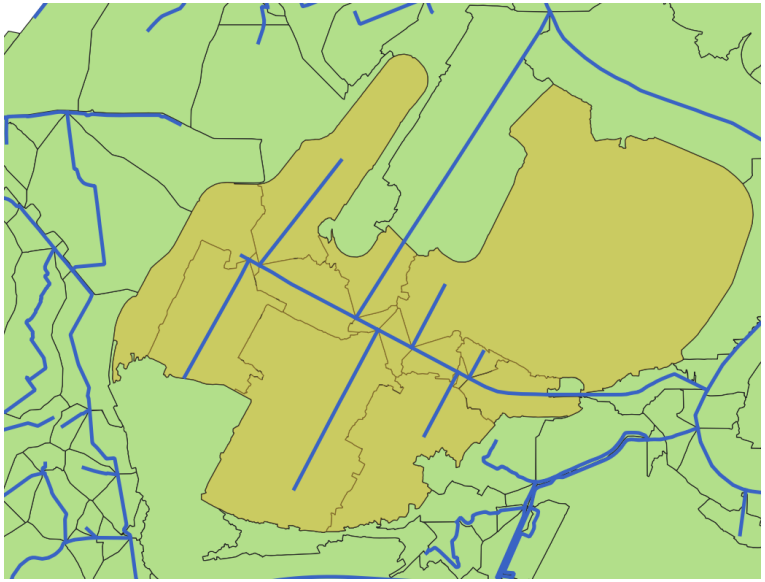
	numeric (10)	numeric (10)
1	6033854	6033853
2	6033842	6033854
3	6033833	6033854
4	6033841	6033842
5	6033837	6033842
6	6033840	6033841
7	6033832	6033841
8	6033843	6033840
9	6033838	6033840
10	6033851	6033843
11	6033839	6033843
12	6033855	6033851
13	6033844	6033851

vito

FIND THE SUPPLY AREA FOR ANY GIVEN RIVER SEGMENT

Next step: create supply areas from run-off area id's
This can take > 1 min in case of > 10 000 id's

```
SELECT ST_Union (geom) AS geom
  FROM runoff_areas
 WHERE <id in the list>
```



	numeric (10)	numeric (10)
1	6033854	6033853
2	6033842	6033854
3	6033833	6033854
4	6033841	6033842
5	6033837	6033842
6	6033840	6033841
7	6033832	6033841
8	6033843	6033840
9	6033838	6033840
10	6033851	6033843
11	6033839	6033843
12	6033855	6033851
13	6033844	6033851

You can precalculate this for every river segment and store everything in a table

vito

BONUS: MEANDERS IN RIVERS

For the whole of Flanders, it takes 40 seconds to identify all river segments with meanders.
Possible applications: ecology, water retention (space for rivers), …

vito

PostGIS is awesome ☺

Thank you for your attention

Questions?

rik.hendrix@vito.be
datascience@vito.be
https://vito.be/nl/over-vito/data-science