Linked Open (Geo) Data for Agriculture

Obstacles in Standards & Spatial Thinking

September 2017, Rob Knapen

WAGENINGEN
UNIVERSITY & RESEARCH

Lessons Learned from a few projects:

AGINFRA+

E-ROSA

#geo4web

SemaGrow

LIAISE

AgMIP

Computers can use *linked data* to figure out interesting things and find related information.

Sources: LIAISE, eROSA

However this will not be ‘perfect’, and users will have high expectations (they are used to Google).

It helps to keep humans in the loop and to make things less ‘magical’.

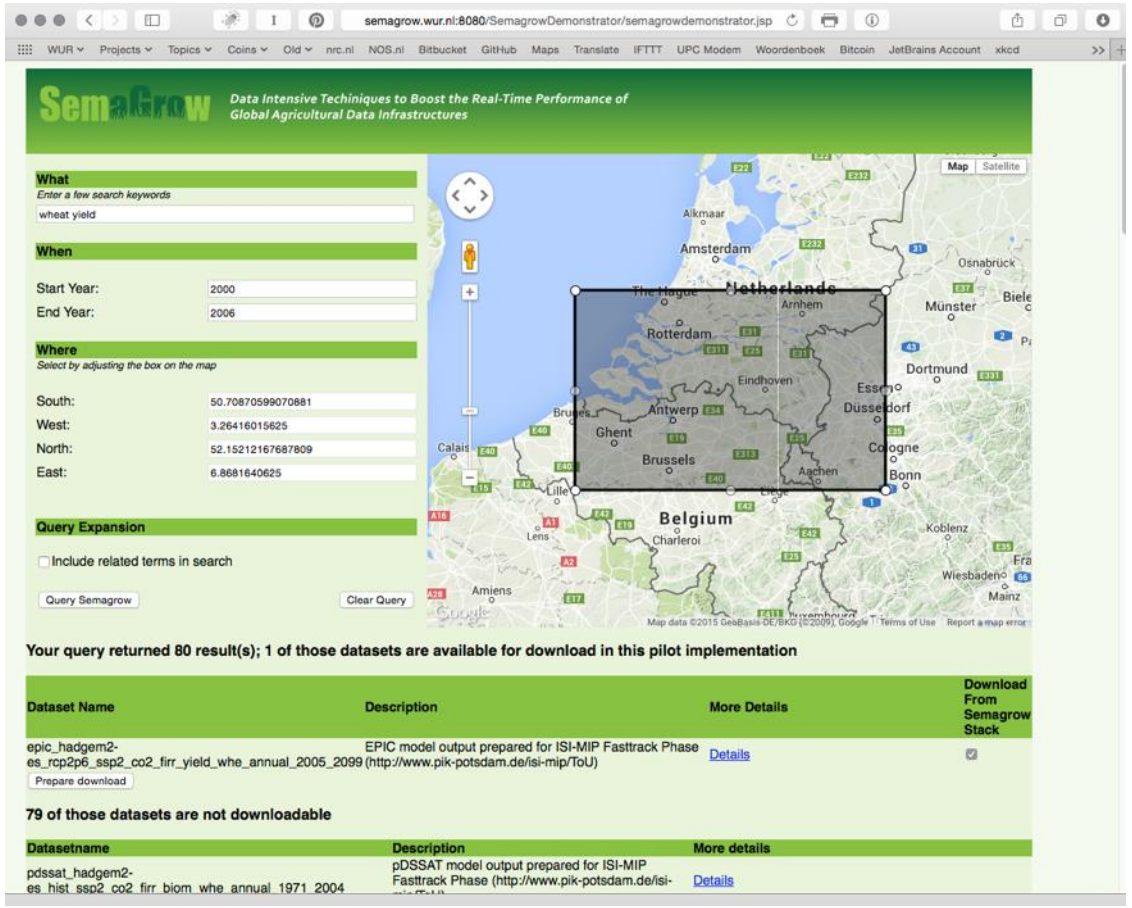
Sources: LIAISE, eROSA

Considerations

- Unfamiliarity with handling incomplete and low quality (Big) Linked (Open) Data.
- Closed World vs Open World Assumptions, how to use them in software applications?
- Invest in NLP, domain specific corpus, language models, machine learning.
- How to sell it to the users?

Sources: LIAISE, eROSA

The LD technology is interesting but research projects might lack the infrastructure to use its full potential.



WAGENINGEN
UNIVERSITY & RESEARCH

Sources: SemaGrow, LIAISE

Having to scale back will cause questions and remarks, and loss of trust.

We have to find a practical mix of traditional and new technologies (e.g. GIS and Big Data).

Sources: SemaGrow, LIAISE

Considerations

- Geo-Information and linked data are not (yet) a good match.
  - Triplifying gridded data is practically not possible.
    - e.g. 150 MB netCDF expands to 1 Gtriples
    - Too large to load into 4Store
    - Slow to transport over the Internet
  - Geo-Information datasets often are not precise and still map-oriented.
- Helps (at the moment) to combine spatial database and triple stores to solve e.g. performance issues.

Sources: SemaGrow, LIAISE

More Considerations

- Data might not be prepared for assigning global identifiers.
  - Geo-Information processing often is based on implicit spatial relations (e.g. overlays, buffers) or vector - raster operations (e.g. zonal statistics). Data is stored for these purposes and might not already have ‘things’ that can be given a unique URI.
- Geo datasets might have limited and sparse metadata (still have to ask the owner).
- LD competes with existing geo standards and Spatial Data Infrastructures.

Sources: SemaGrow, AgInfra+, eROSA

Geo professionals and Linked Data professionals are both groups with specific knowledge.

```
// ----- 2. Leaflet with PDOK ---

var RD = new L.Proj.CRS.TMS(
    'EPSG:28992',
    '+proj=sterea +lat_0=52.15616055555555 +lon_0=5.38763888888889 +k=0.9999079
    [-285401.92,22598.08,595401.9199999999,903401.9199999999], {
    resolutions: [3440.640, 1720.320, 860.160, 430.080, 215.040, 107.520, 53.76
});
```

```
window.onload = function() {
    function onMapClick(e) {
        $.ajax({
          "accepts": {"json": "application/vnd.geo+json"},
          "data": {
            "graph": "http://lodlaundromat.org/data/c39e1092fd8387233e60222952f11a2a",
            "lng": e.latlng.lng,
            "lat": e.latlng.lat,
            "properties" : "yes",
            "page_size" : 30
          },
          "dataType": "json",
```


Sources: #geo4web, AgInfra+, eROSA

This knowledge is not yet easily accessible and understandable by outsiders.

We have to translate it into simple and easy usable APIs.

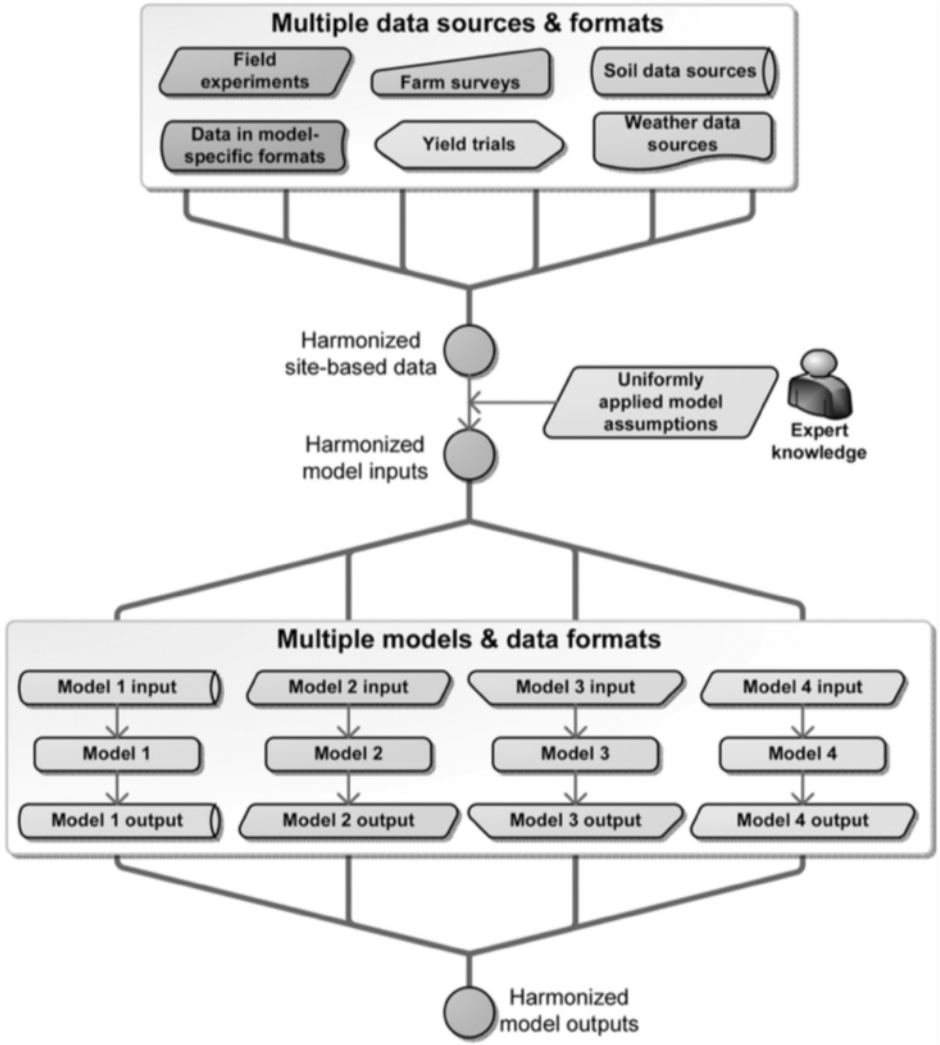
Sources: #geo4web, AgInfra+, eROSA

Considerations

- The Semantic Web is not (yet) suitable for spatial queries.
  - No support for Region Connection Calculus.
  - Have to use GeoSPARQL, which is a complex full scale geo standard. W3C currently only defines a spatial point (lat, long).
  - Linked geospatial data is being researched and standards are in the making but not widely used yet.
- LD works well with vector data ('things'), but big spatial data processing is best done on raster data.
- linked data vs Linked Data, e.g. json-ld, geojson.

Sources: #geo4web, AgInfra+, eROSA, LIAISE

It takes a lot of effort to come to shared definitions and semantics.



Source: AgMIP, LIAISE, eROSA

Usually ends up with a “lowest common denominator” solution, and practical things that can be implemented.

We have to promote reuse and standardization of ontologies.

Source: AgMIP, LIAISE, eROSA

Considerations

- Lack of shared semantics results in complex LD.
- Existing LD standards need to be further improved and promoted.
- Lack of standardized, adopted semantics and variable-types in agronomy (e.g. varieties, units).
  - AgroVOC and GACS thesauri
  - Ontology for Units of Measure
  - E.g. ontology for meteo data

Sources: AgMIP, LIAISE, eROSA

Thank You!

Rob Knapen

Wageningen
Environmental
Research

rob.knapen@wur.nl

WAGENINGEN
UNIVERSITY & RESEARCH