

Linked Open (Geo) Data for Agriculture

Obstacles in Standards & Spatial Thinking

September 2017, Rob Knapen



Lessons Learned from a few projects:

AGINFRA+

E-ROSA

#geo4web

SemaGrow

LIAISE

AgMIP

- 1 -

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



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'.

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?

- 2 -

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

semagrow.wur.nl:8080/SemagrowDemonstrator/semagrowdemonstrator.jsp

WUR Projects Topics Coins Old nrc.nl NOS.nl Bitbucket GitHub Maps Translate IFTTT UPC Modem Wordenboek Bitcoin JetBrains Account xkcd

SemaGrow


Data Intensive Techniques to Boost the Real-Time Performance of Global Agricultural Data Infrastructures

What
Enter a few search keywords
wheat yield

When
Start Year: 2000
End Year: 2006

Where
Select by adjusting the box on the map
South: 50.70870599070881
West: 3.26416015625
North: 52.15212167687809
East: 6.8681640625

Query Expansion
☐ Include related terms in search
Query Semagrow Clear Query



Map data ©2015 GeoBasis DE/BKG (©2009), Google T Terms of Use Report a map error

Your query returned 80 result(s); 1 of those datasets are available for download in this pilot implementation

| Dataset Name | Description | More Details | Download From Semagrow Stack |
|---|--|-------------------------|-------------------------------------|
| epic_hadgem2-es_rcp2p6_ssp2_co2_firr_yield_whe_annual_2005_2099 | EPIC model output prepared for ISI-MIP Fasttrack Phase (http://www.pik-potsdam.de/isi-mip/ToU) | Details | <input checked="" type="checkbox"/> |

79 of those datasets are not downloadable

| Datasetname | Description | More details |
|--|--|-------------------------|
| pdssat_hadgem2-es_hist_ssp2_co2_firr_biom_whe_annual_1971_2004 | pDSSAT model output prepared for ISI-MIP Fasttrack Phase (http://www.pik-potsdam.de/isi-mip/ToU) | Details |

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).

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.

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.

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

```
15 // ----- 2. Leaflet with PDOK ---
16
17 var RD = new L.Proj.CRS.TMS(
18     'EPSG:28992',
19     '+proj=sterea +lat_0=52.1561605555555 +lon_0=5.38763888888889 +k=0.9999079
20     [-285401.92,22598.08,595401.9199999999,903401.9199999999], {
21     resolutions: [3440.640, 1720.320, 860.160, 430.080, 215.040, 107.520, 53.760]
22 });
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71 window.onload = function() {
72     function onMapClick(e) {
73         $.ajax({
74             "accepts": {"json": "application/vnd.geo+json"},
75             "data": {
76                 "graph": "http://lodlaundromat.org/data/c39e1092fd8387233e60222952f11a2a",
77                 "lng": e.latlng.lng,
78                 "lat": e.latlng.lat,
79                 "properties": "yes",
80                 "page_size": 30
81             },
82             "dataType": "json",
```

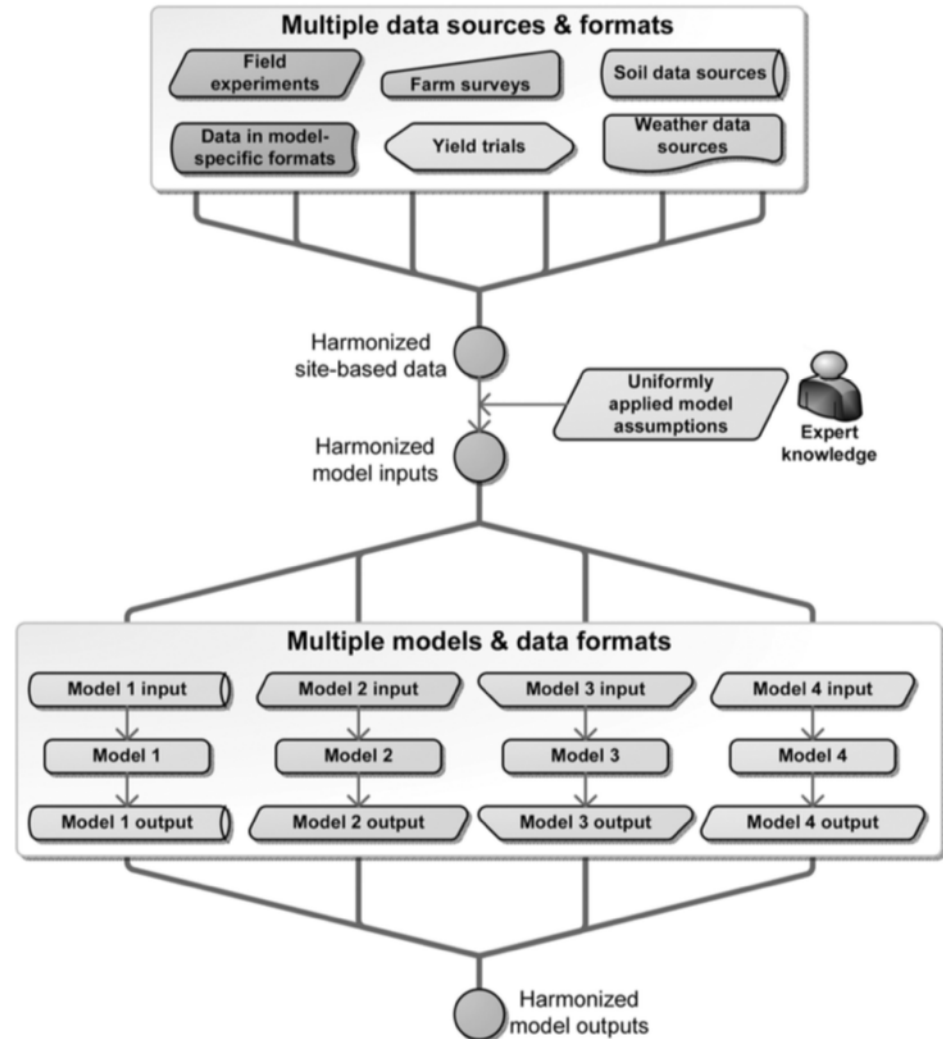
This knowledge is not yet easily accessible and understandable by outsiders.

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

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.

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



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.

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

Thank You!

Rob Knapen

Wageningen
Environmental
Research

rob.knapen@wur.nl

