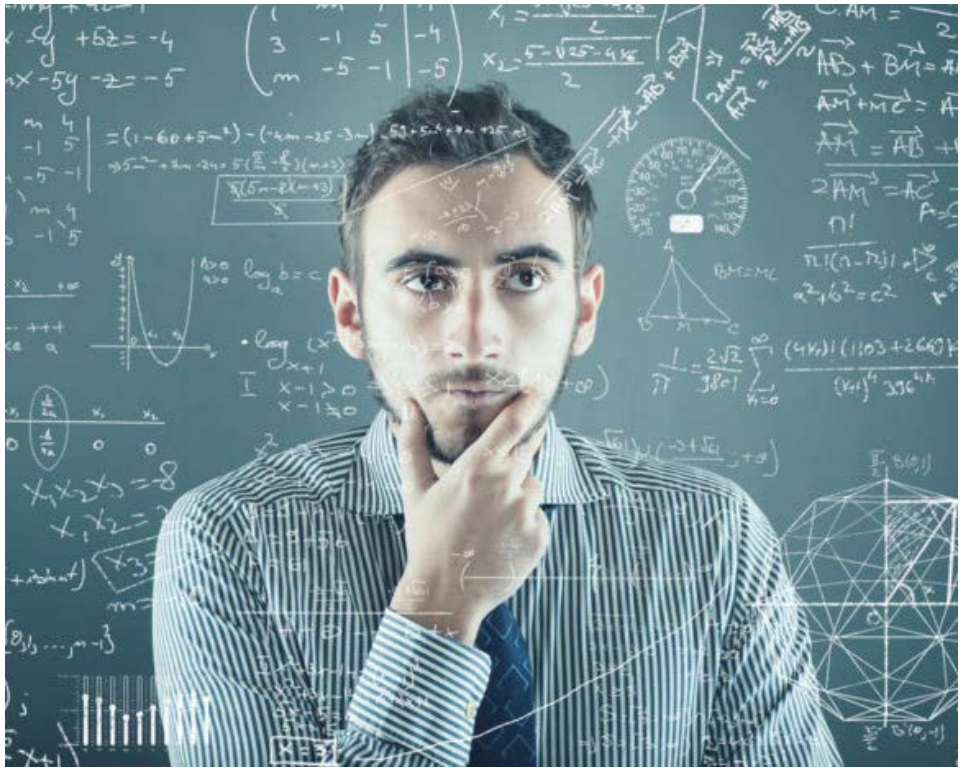
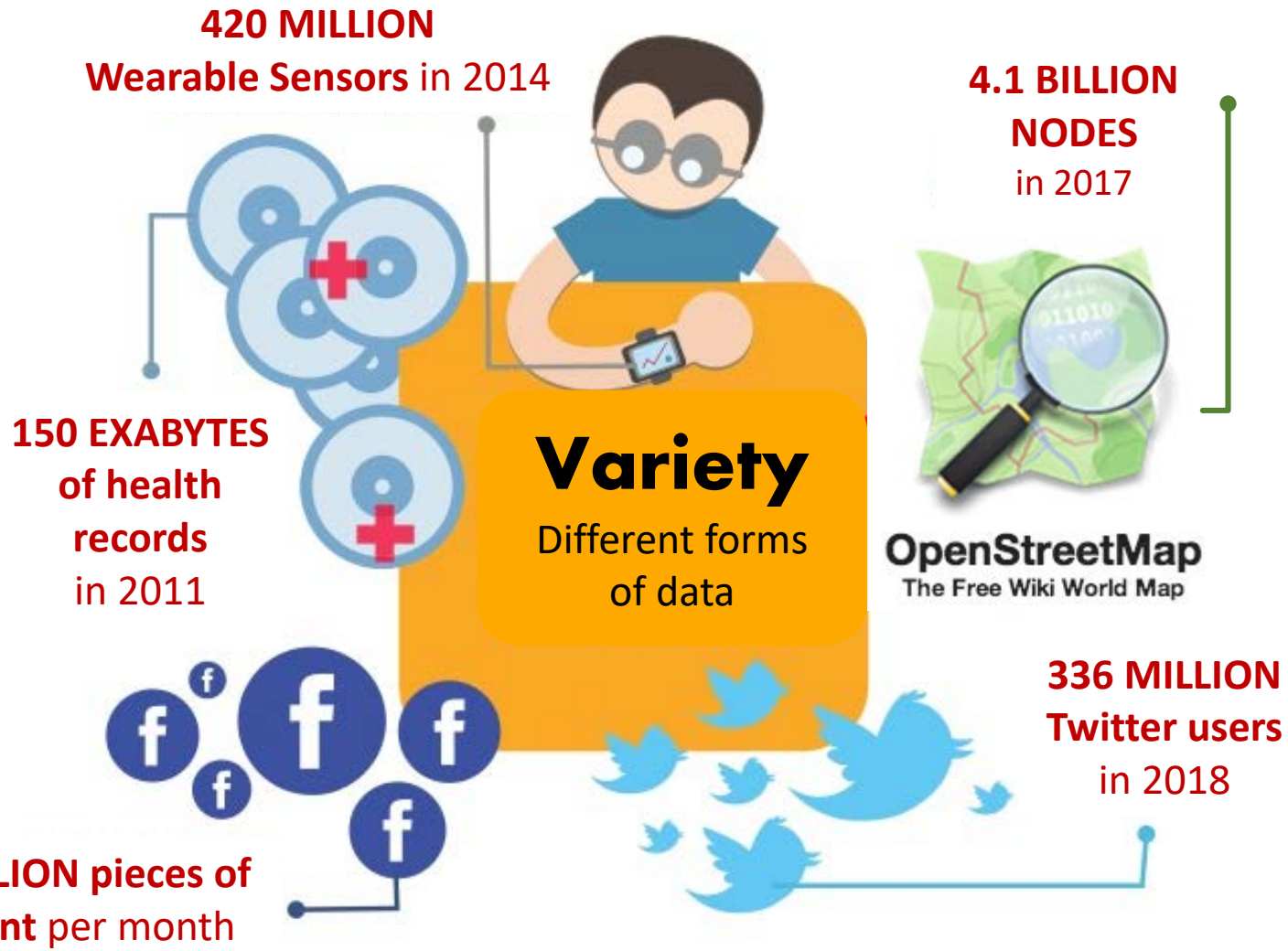
Question-answering with linked spatial data: the role of spatial core concepts



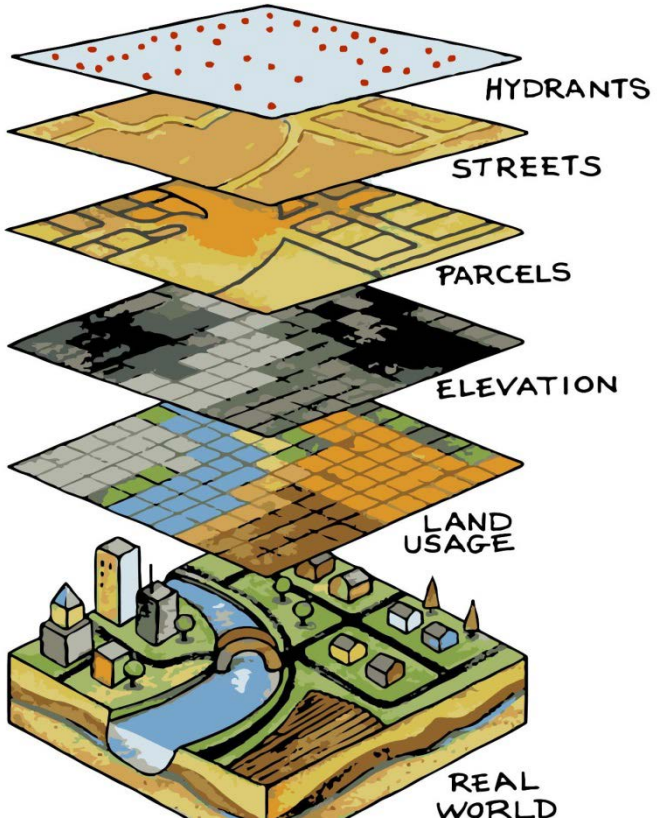
Universiteit Utrecht

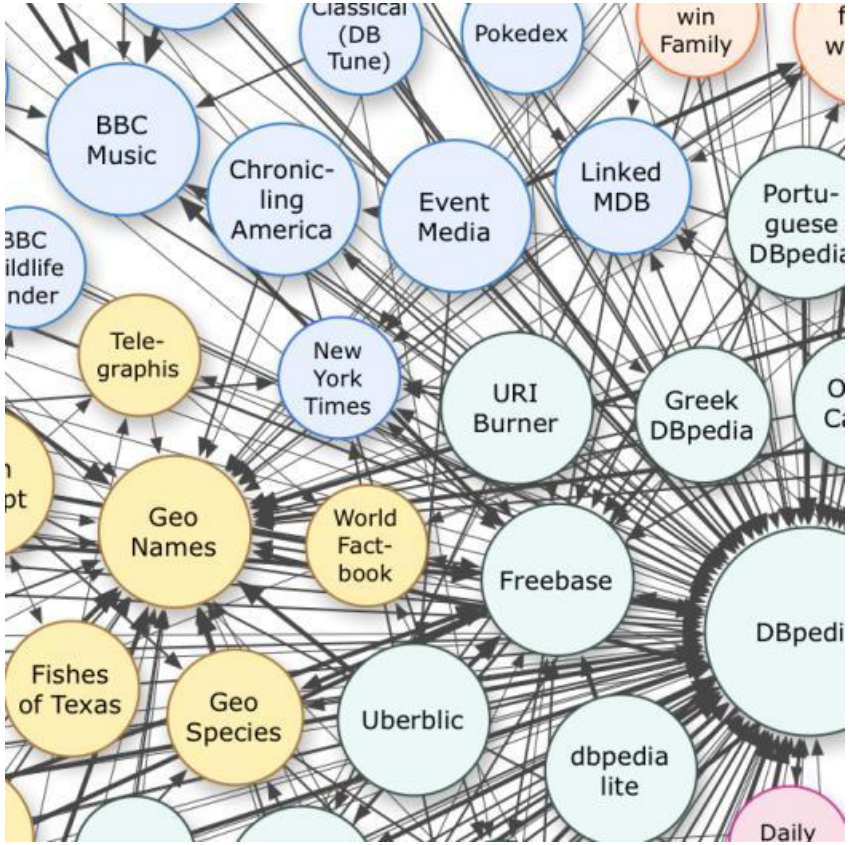
Simon Scheider, Dep. Human Geography and Planning,
SDI.Next: Linked Spatial Data in Europe, 12-3-2019

Variety of big geodata: A blessing and a curse



GIS and linked spatial data





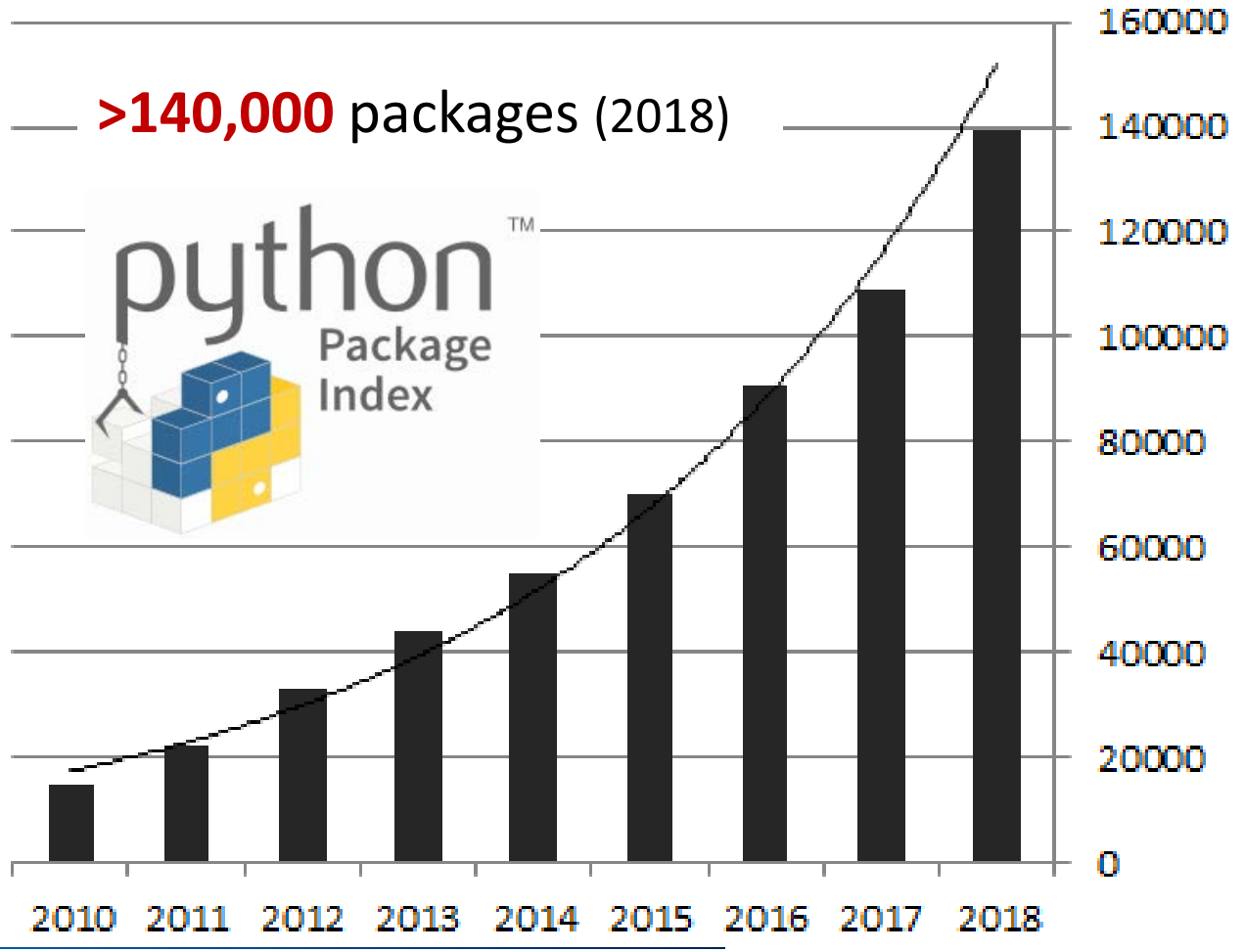
Variety of GIS tools:
> 8000 tools in main software programs

40 most well known software programs together contain **> 8,000** tools (Ballatore et al. 2018)

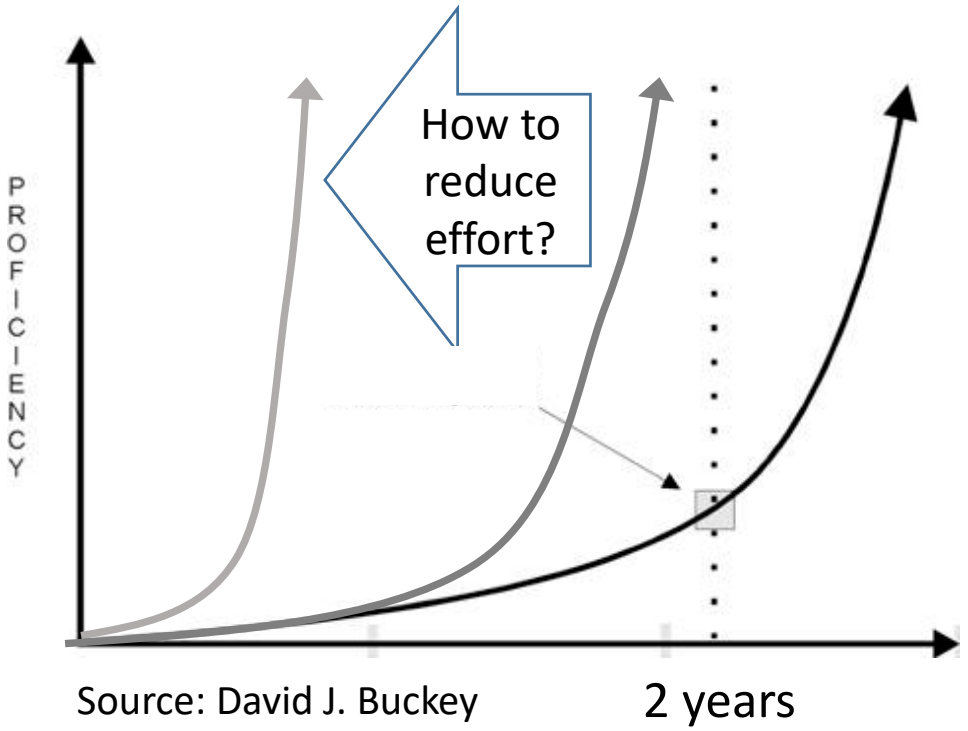
Variety of (GIS) tools: Exponential growth of packages

>140,000 packages (2018)

python™ Package Index



Learning a GIS tool simply takes too much time



Source: David J. Buckey

Learning …

- a new toolset : ~ 2 months
- to become proficient in GIS: ~ 2 years

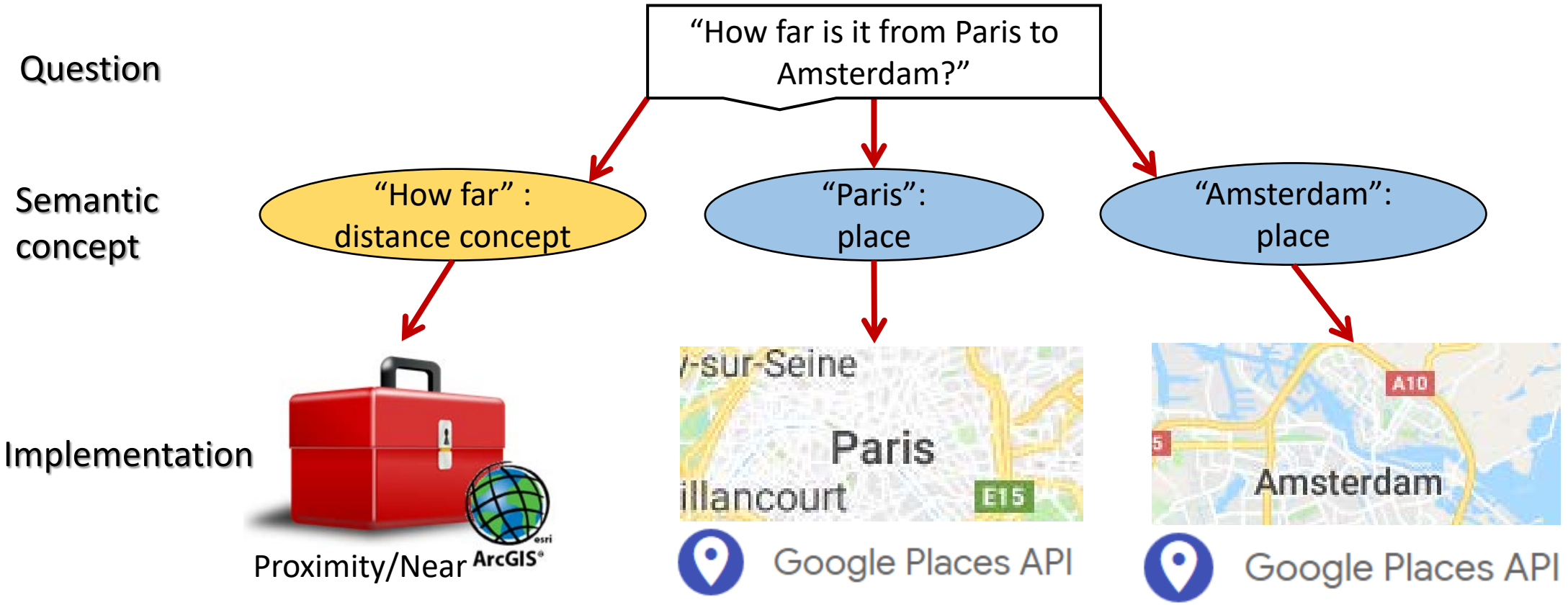
Ask a spatial question - and find the right tools and data in an instant

Data scientists always start (and end) with an analytic question

- … to capture what they actually want to know
- … to share and discuss their results
- … independently from a tool or data set (Vahedi et al 2016)

Building a question-based GIS

GIS currently do not understand questions! The challenge lies in *semantically translating* questions into tools and data (Gao 2013):

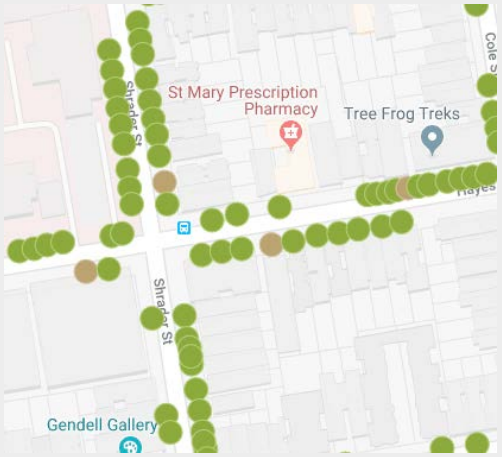


Example: Translation of a question into GIS

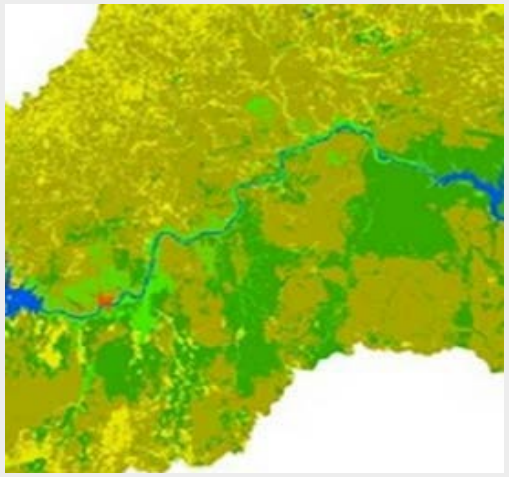
"How much is Tom exposed to green space while running through the city?"

Data and tools that might be of relevance …

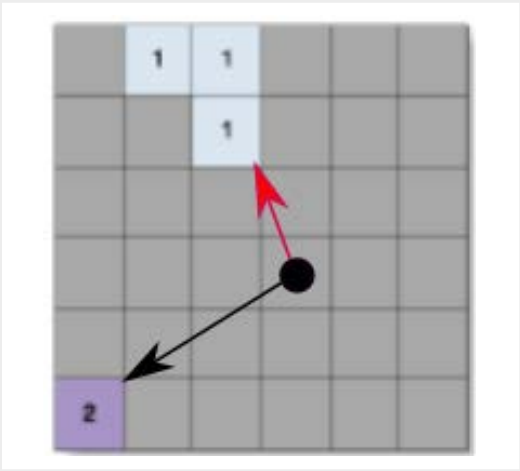
Map of urban trees



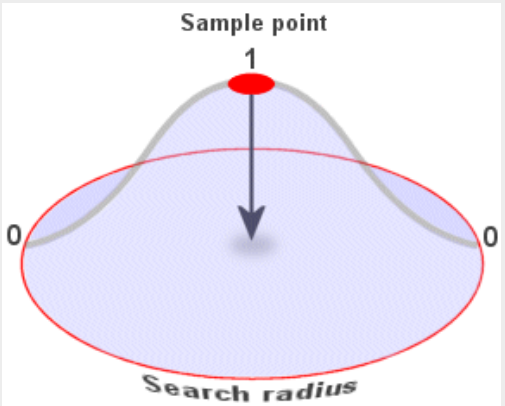
Landuse polygons



ArcGIS: Euclidean Distance



ArcGIS: Kernel Density



…

Which ones of these resources answer the question?

State-of-the-art: KB QA

Knowledge based Question-Answering (KB QA):

Match questions with *facts* stored in a fact DB (Diefenbach et al 2017)

1. Build a semantic representation of the question as query
2. Run this query on a facts DB:
   - Geospatial DB
   - Ontologies (DBPedia, WordNet, Yago)
   - Scientific databases

Many steps are based on *machine learning* (QA learning)

State-of-the-art: KB QA

For example, IBM Watson's winning reply in Jeopardy!:

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE PRINCIPALITIES OF
WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL

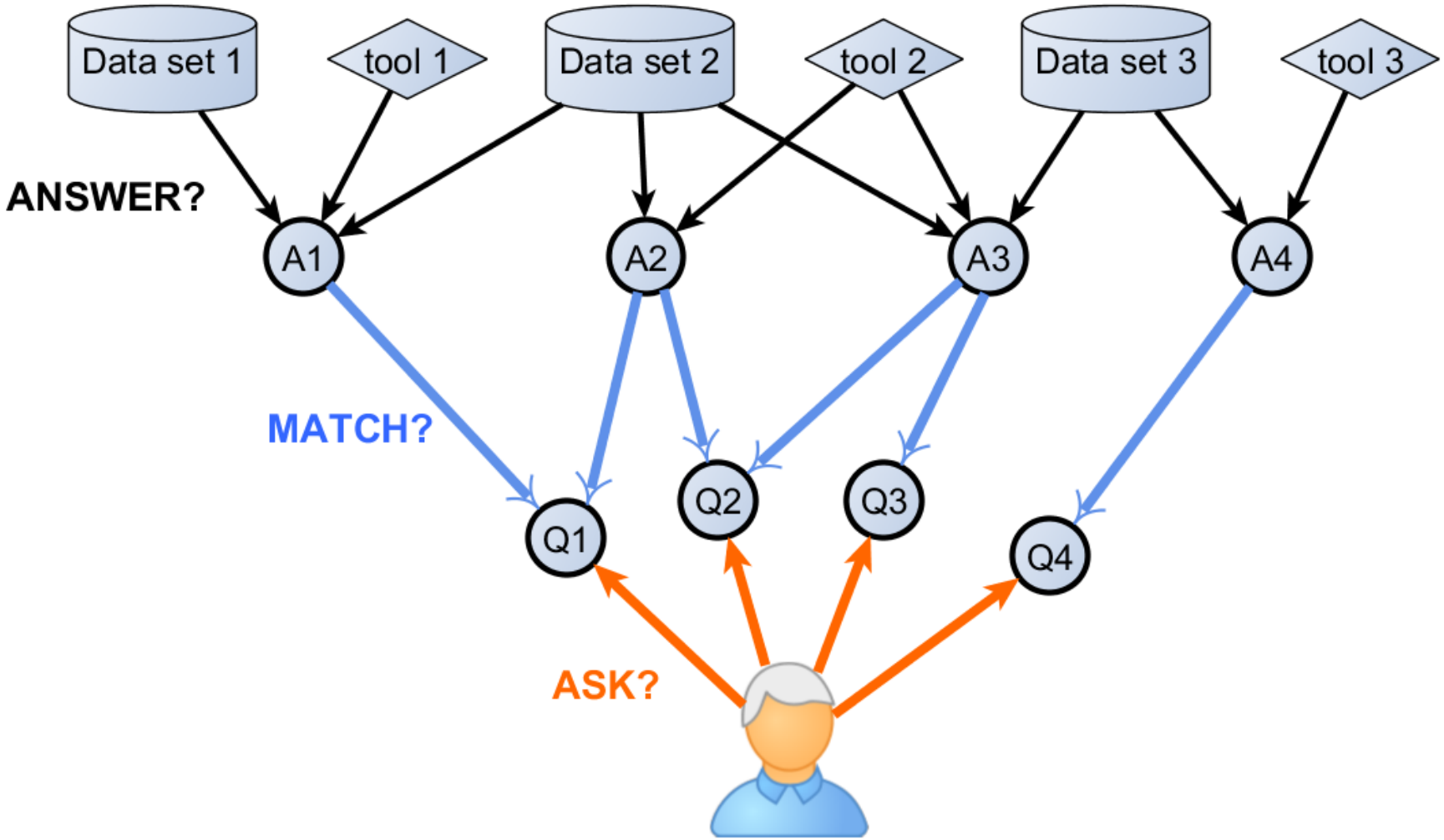
→ Bram Stoker

Or Apple's Siri:

The challenge I: Creativity in asking/answering questions

… makes it
a non-trivial
learning task …

- Problem for QA learning: training samples lacking
- Problem for QA learning: many possible answers



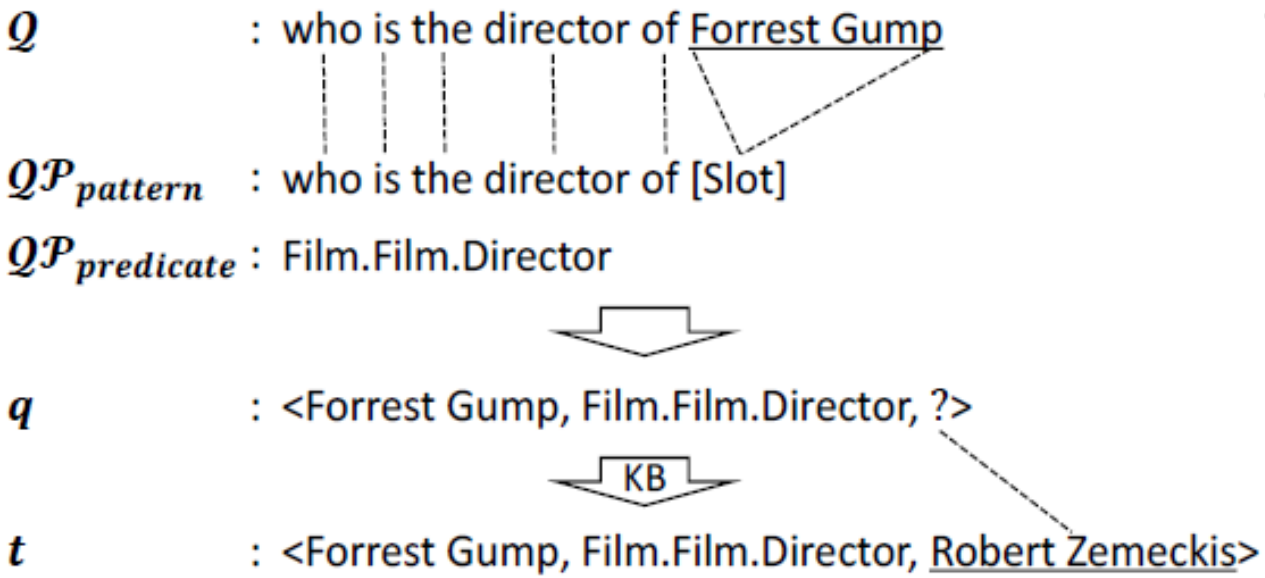
The challenge II: Unknown answers

Example: "*Who is the director of Forrest Gump*?" (Bao et al 2014)

Answer derived by matching *question patterns* to facts in DBpedia:

Yet,

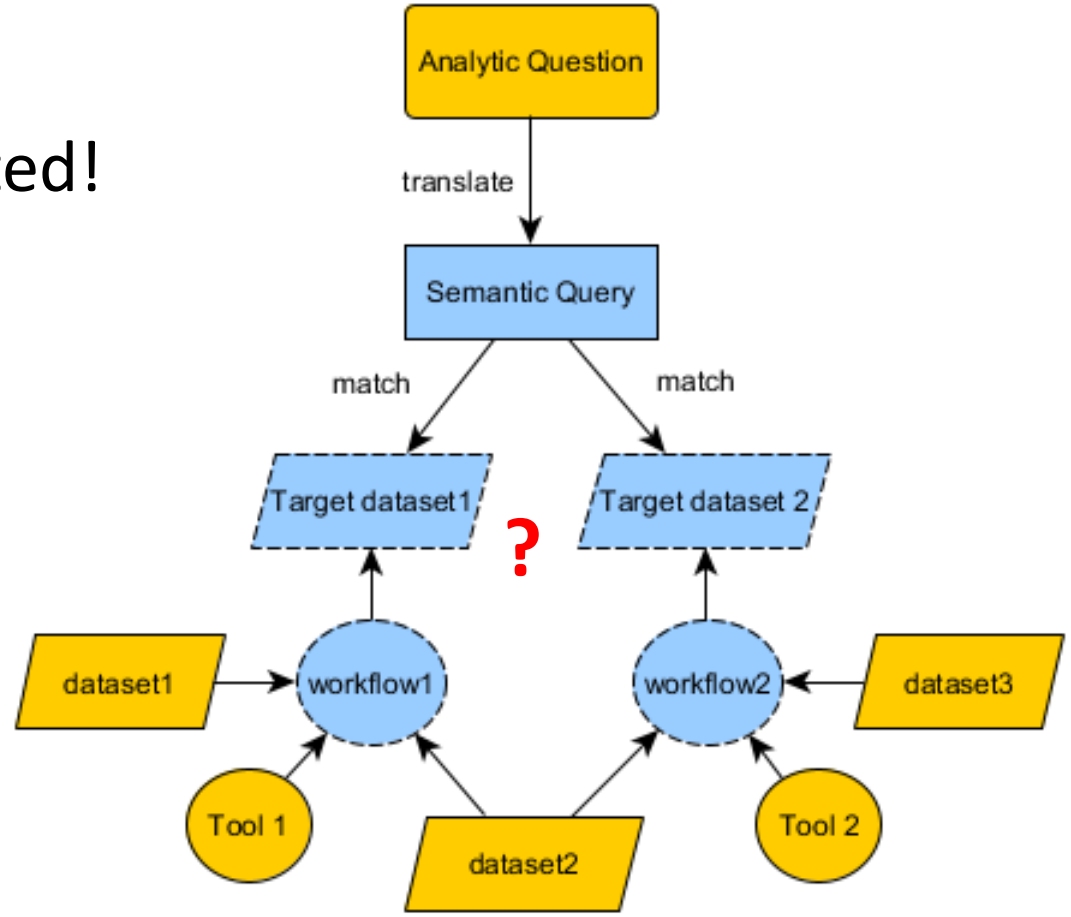
- The most interesting GIS questions are *analytical*, they have unknown answers!
- So: we cannot *learn* QA!



The challenge III: Analytic QA: *potential* of tools and data

Analytic QA:

- Matching questions to data which *may not exist* (yet), but may be generated!
- Finding a way to describe the *analytic potential* of
  - geodata
  - GIS tools
- … across datatypes (raster, vector)
- … across software implementations



QuAnGIS

1. ASK: Which spatial questions can be **asked**?
2. ANSWER: Which spatial questions can be **answered** by
    a) GIS tools?
    b) Web data sources?
3. MATCH: How to **match** geoanalytic resources and questions?

https://questionbasedanalysis.com/

@2019 -2024: Funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 803498).

Spatial core concepts (Kuhn 2012)

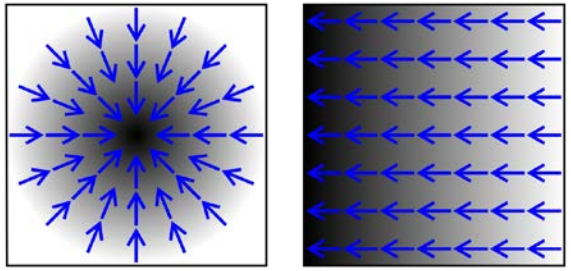
Objects of study
in Geographic
Information Science?
(… like “cell” in biology
or “value” in economics)

Constraints for…

1. Typing geodatasets and tools
2. Posing spatial questions
3. Finding answers with GIS workflows

Spatial core concepts ctd.

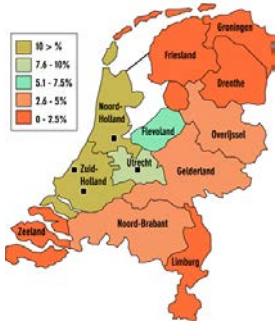
Field



- continuous phenomenon
- Space -> Quality (value field)
- boundaries are irrelevant

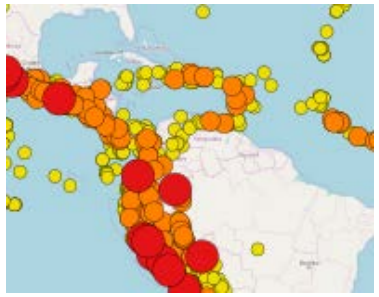
Object

Points: hotels



- discrete phenomenon with qualities
- Object -> Space (projected in space, not time)
- boundaries are relevant
- Object -> Quality (are not self-similar)

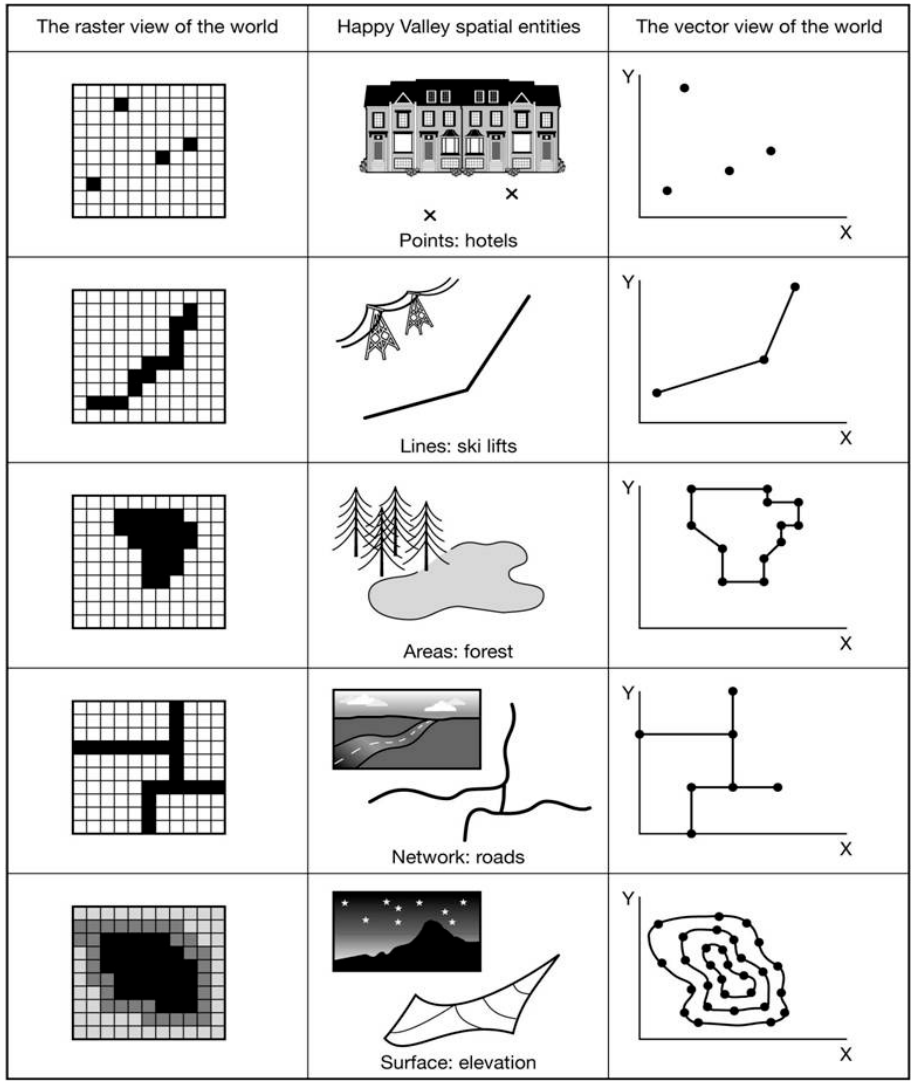
Event



- like objects, but…
- Event -> Time, Space (projected in time and space)

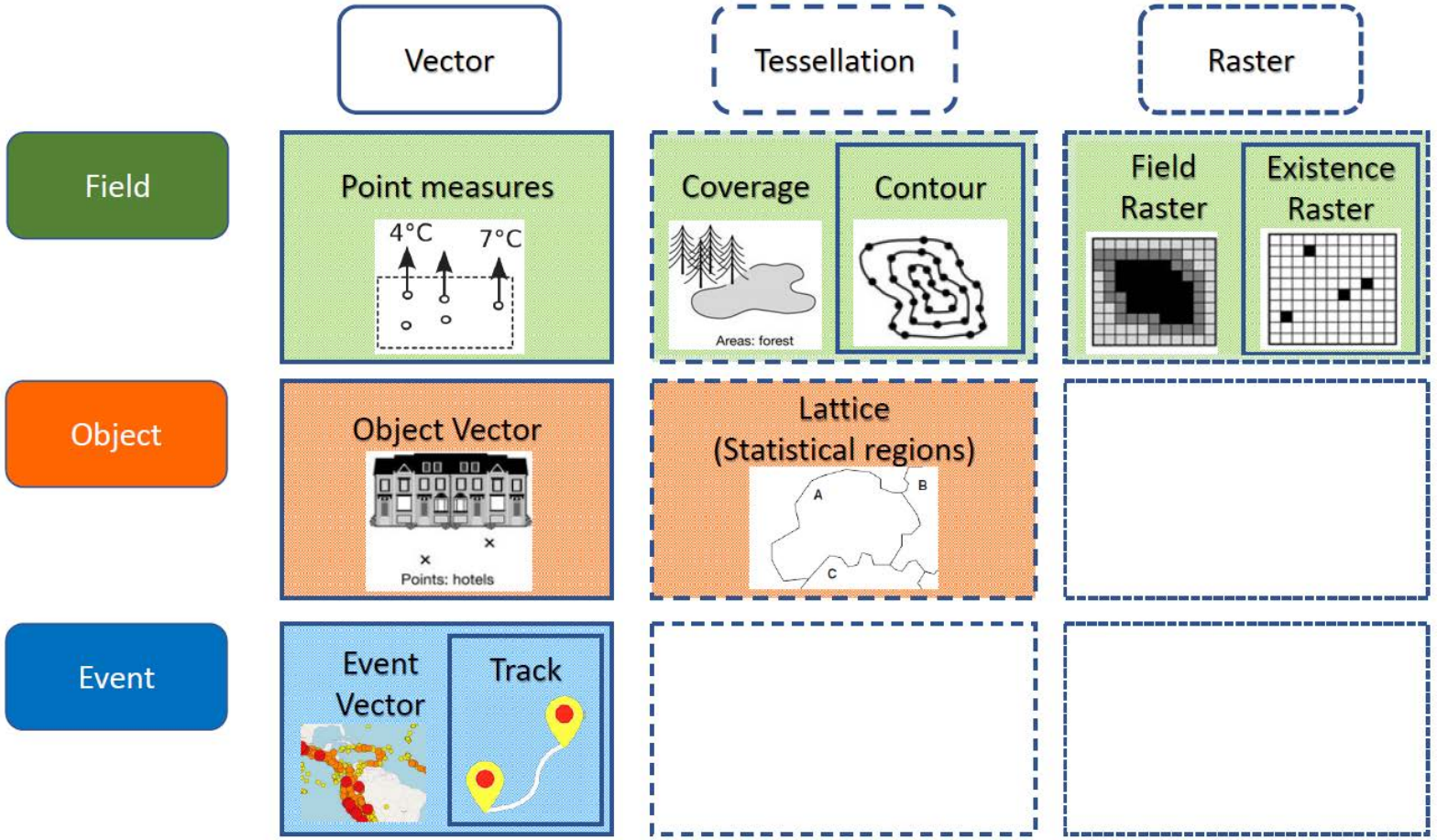
(Galton 2004, Scheider et al 2019)

1) Typing geodatasets using core concepts



Spatial data types
≠ Core Concepts!

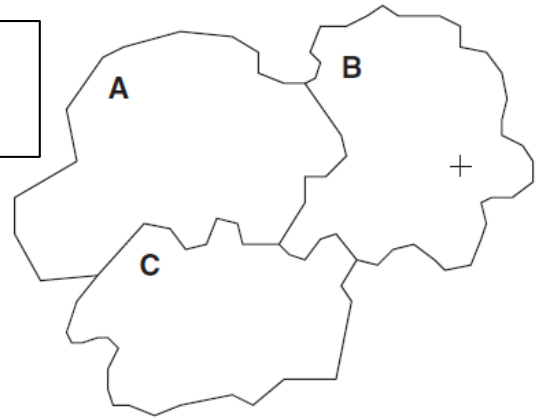
1) Typing geodatasets using core concepts



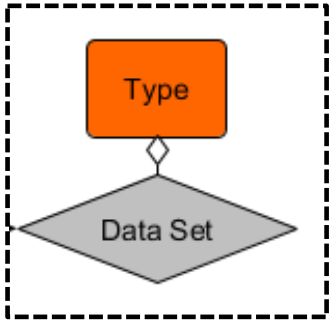
Scheider et al 2019: Semantic data type signatures for representing spatial core concepts in GIS operations on spatial layers

1) Typing geodatasets using core concepts

Polygon
Vector data



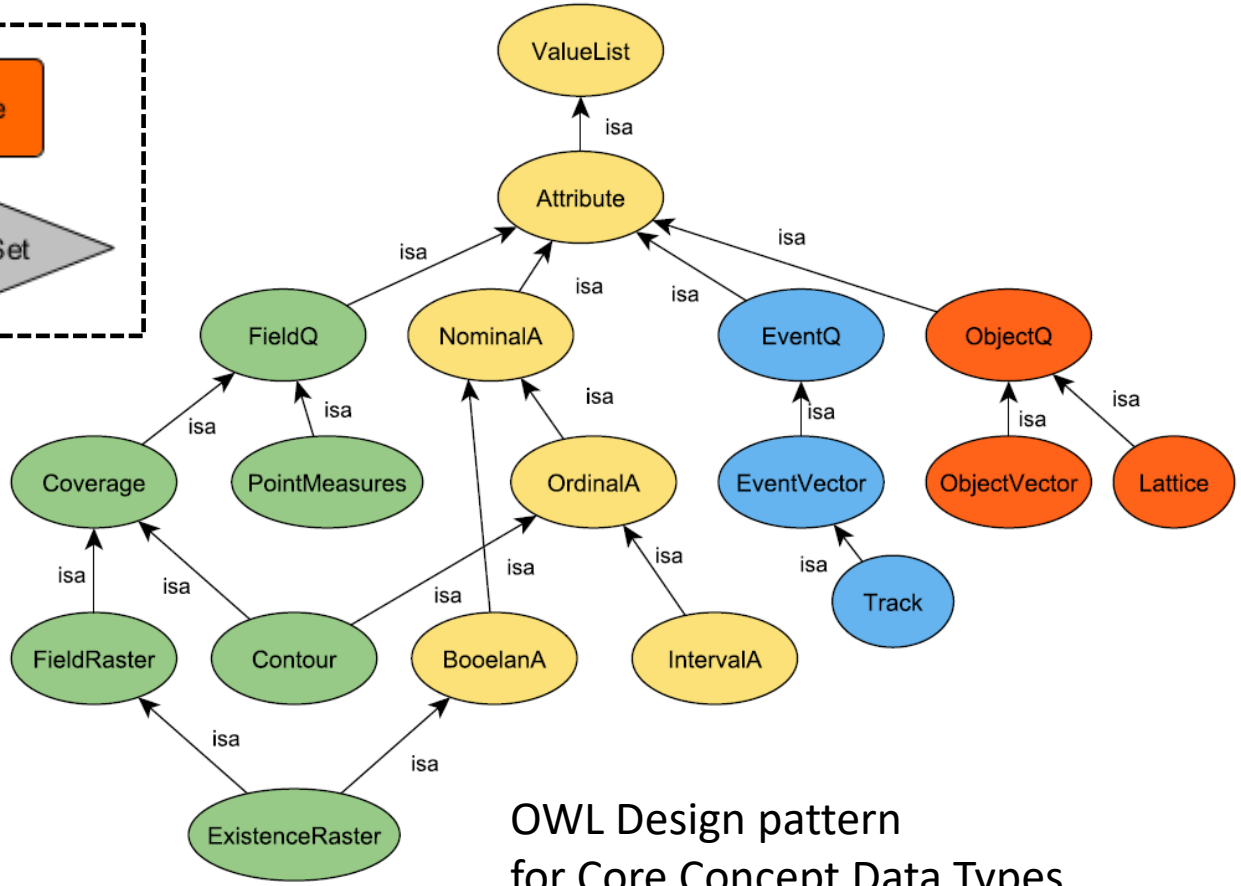
Linked data:



Coverage?
(self-similar)

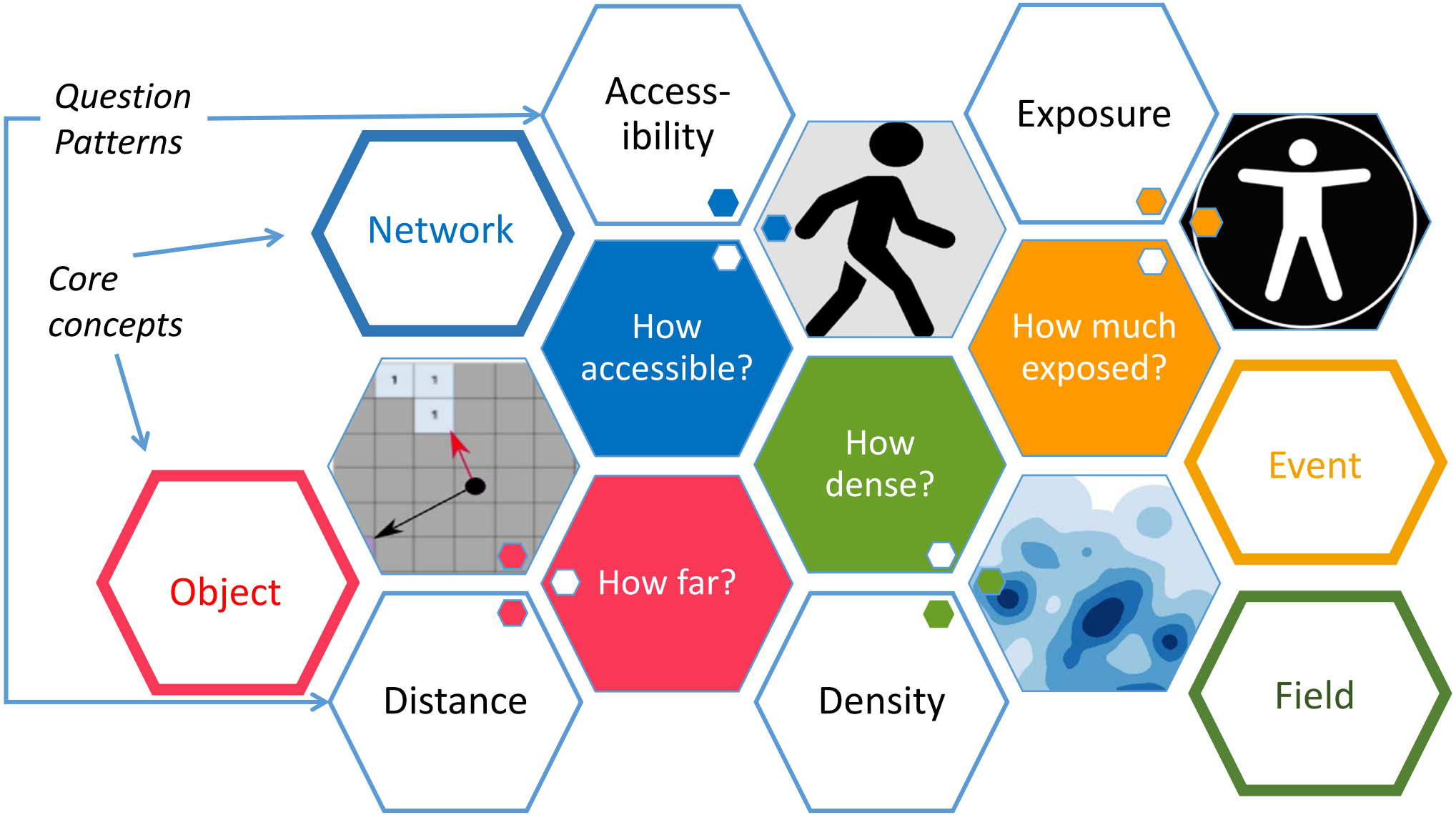
Lattice?
(object boundaries)

	land cover type	average elevation (m)
A	Forest	631
B	Urban	220
C	Water	42
+	Urban	



OWL Design pattern
for Core Concept Data Types

2) Posing spatial questions using core concepts



2) Posing spatial questions using core concepts

Question patterns correspond to core concept data types:

"Which trees exist in New York?"

QuestionPattern: "Which [ObjectVector]s exist in [Place]?"

Semantic Query:

```
PREFIX ccd: <http://geographicknowledge.de/vocab/CoreConceptData.rdf#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?dataset WHERE {
    ?dataset a ccd:ObjectVector;
    rdfs:label ?name; ccd:extent ?MBB.
    ?place rdfs:label ?pname; ccd:extent ?PBB.
    ?PBB geo:sfOverlaps ?MBB
    FILTER regex('trees', ?name)
    FILTER regex('New York', ?pname)
}
```

2) Posing spatial questions using core concepts

Question patterns correspond to core concept data types:

"How dense/far is green around/from Tom?"

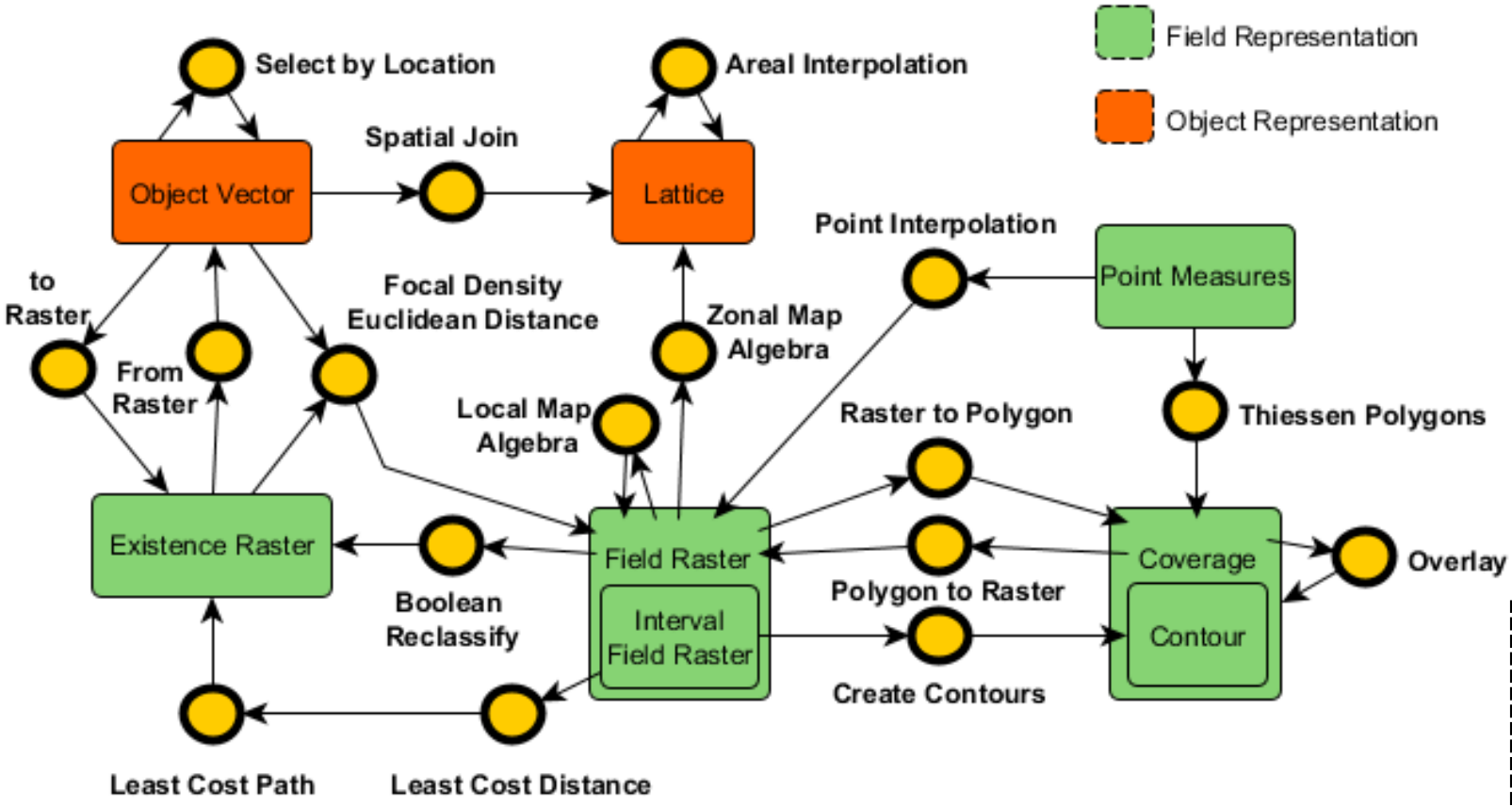
QuestionPattern: "How [FieldRaster] is [DataSet] around [ObjectVector]?"

Semantic Query:

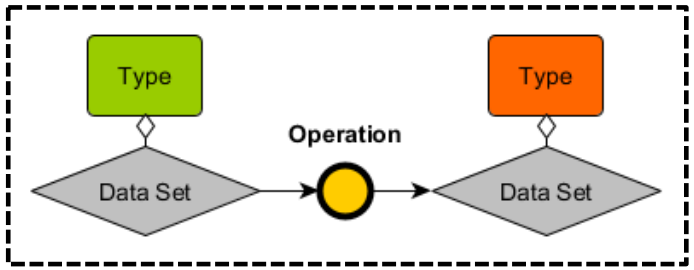
```
PREFIX ccd: <http://geographicknowledge.de/vocab/CoreConceptData.rdf#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?dataset WHERE {
    ?dataset a ccd:FieldRaster.
    _:Tom a ccd:ObjectVector.
    ?exposedto a ccd:DataSet.
    ?exposedto rdfs:label ?name
    FILTER regex('green', ?name).
}
```

3) Finding answers using core concepts

Using Core Concept Data Types as operation signatures (ArcGIS):



Linked data model:

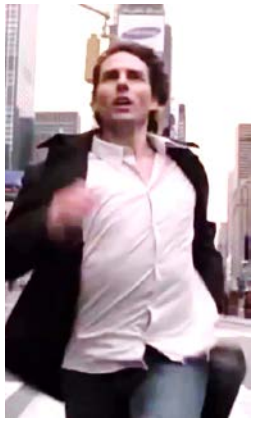


3) Finding answers using core concepts

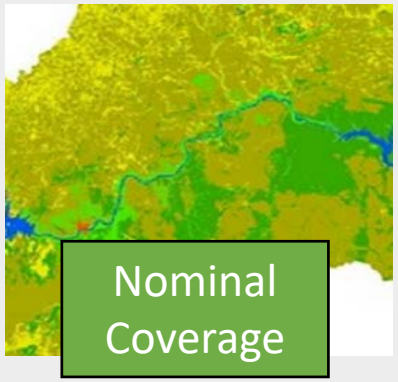
Map of urban trees

"How dense is green around Tom?"

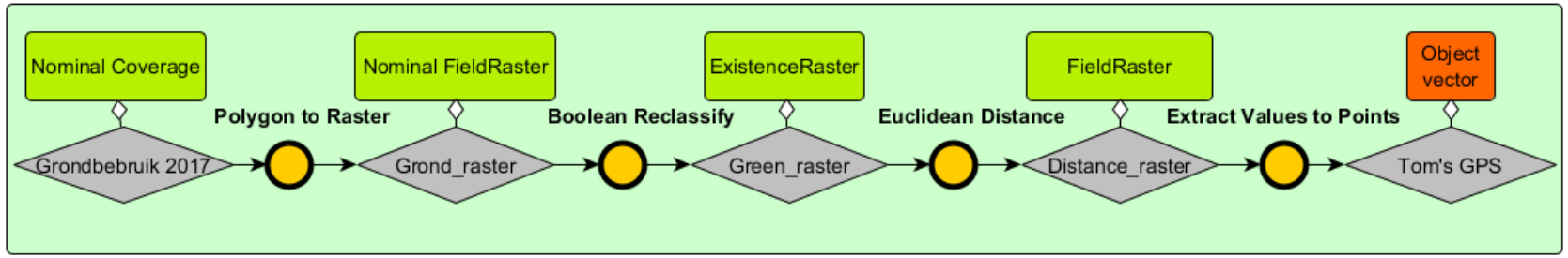
Automated Workflow Composition (Lamprecht 2013)



Landuse polygons



"How far is green from Tom?"



3) Finding answers using core concepts

ASK

"How much is Tom exposed to green space?"

Question based GIS request

How to generate

* Select Dropdown

Exposure

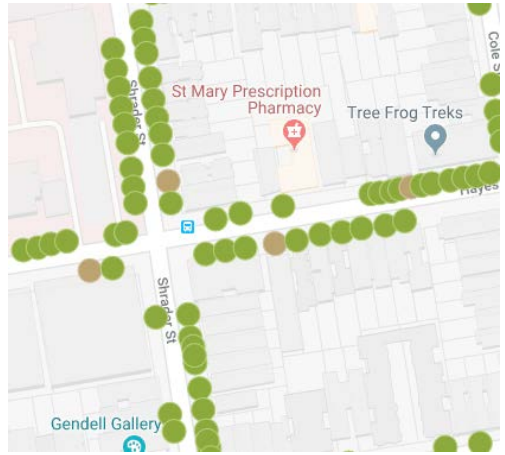
your result

MATCH Semantic query:

```
PREFIX ccd: <http://geographicknowledge.de/vocab/CoreConceptData.rdf#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?dataset WHERE {
    ?dataset a ccd:FieldRaster.
    _:Tom a ccd:ObjectVector.
    ?exposedto a ccd:DataSet.
    ?exposedto rdfs:label ?name
    FILTER regex('green', ?name).
}
```

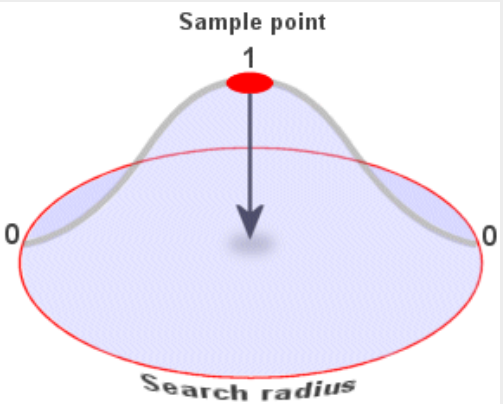
ANSWER Possible answer resources:

Map of urban trees



"Given this layer of trees,

ArcGIS: Kernel Density

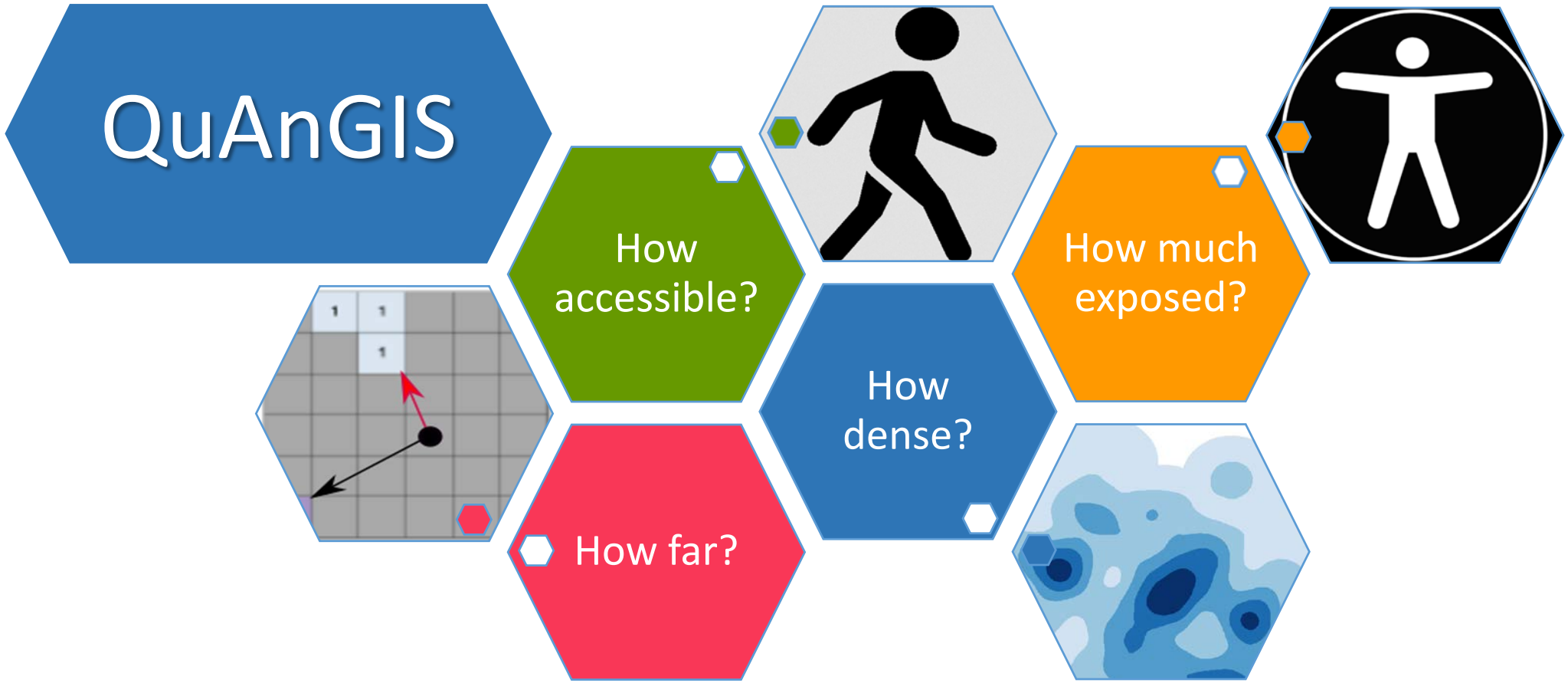


how densely located are trees around Tom?"

Project outcomes → future opportunities

Project outcomes	future opportunities
Theory of *spatial questions* used in spatial sciences	Core concepts of spatial information theory
Technology that makes spatial questions machine-readable	Natural ("Alexa") interfaces for data science
Web tool repository for registering/ requesting tools and data	Integrated Web repositories (tools + data)
Standard geoanalytic tools (ArcGIS, QGIS, R) *packaged* with standard Web data sources (OSM, Kadaster and CBS)	Boost GIS technology through data sciences

https://questionbasedanalysis.com/



References

- Werner Kuhn 2012: Core concepts of spatial information for transdisciplinary research. International Journal of Geographical Information Science, 26(12):2267–2276.
- Behzad Vahedi, Werner Kuhn, and Andrea Ballatore 2016. Question-based spatial computing. A case study. In Geospatial Data in a Changing World , pages 37–50. Springer,
- Ballatore, A., Scheider, S. and Lemmens, R., 2018: Patterns of consumption and connectedness in GIS web sources. In *The Annual International Conference on Geographic Information Science* (pp. 129-148). Springer
- Song Gao and Michael Goodchild 2013: Asking spatial questions to identify GIS functionality. In 2013 Fourth International Conference on Computing for Geospatial Research and Application
- Anna-Lena Lamprecht 2013: User-Level Workflow Design: A Bioinformatics Perspective, volume 405 8311. Springer
- Antony Galton 2004: Fields and objects in space, time, and space-time. Spatial cognition and computation, 4(1):39–68
- Dennis Diefenbach, Vanessa Lopez, Kamal Singh, and Pierre Maret 2010. Core techniques of question answering systems over knowledge bases: a survey. Knowledge and Information systems , 55(3):529–569
- Junwei Bao, Nan Duan, Ming Zhou, and Tiejun Zhao 2014: Knowledge-based question answering as machine translation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics
- Simon Scheider et al 2019: Semantic data type signatures for representing spatial core concepts in GIS operations on spatial layers