Open Tools for Integrating Geospatial Statistical Analysis into Spatial Data Infrastructures

UNECE/UN-GGIM Europe Joint Workshop on Integrating Geospatial and Statistical Standards, 08/11/2017

Benedikt Gräler¹, Christoph Stasch¹, Benjamin Pross¹, Olav Peeters², Simon Jirka¹
1: 52° North Initiative for Geospatial Open Source Software GmbH
2: Belgian Interregional Environment Agency (IRCEL-CELINE)
Who is 52° North?

- Open Research and Innovation Network
- Applied research in (geo)informatics
  - Spatial Data Infrastructures/Standardization:
    - Sensor Web
    - Geoprocessing/Geostatistics
  - Semantics & Linked Open Data
  - Earth Observation
- 52° North GmbH is legal entity
  - National and international research projects, but also professional services projects
- Results of projects are published as open source
Background@52N

**Spatial Data Infrastructures:**
- Evaluation/application of existing standards (OGC, ISO, INSPIRE)
- Development of novel standards (OGC, ISO, INSPIRE)
- Open Source implementation of Spatial Web Services (OGC/INSPIRE) and various Web clients

**Spatio-temporal Statistics:**
- Usage and contribution to R Software
- Concepts for Spatio-temporal Aggregation
- Modeling Spatio-temporal Dependencies

Integrating Geospatial and Statistical Standards & Tools
Spatial Data Infrastructures

- **WMS** – Maps as Images (jpg, tiff, …)
- **WCS** – Coverages (geoTiff, netCDF, …)
- **WFS** – Vector Data (GML, shp)
- **SOS** – Observations (O&M, SweCommon, …)
- **WPS** – Geoprocesses, Simulations, …
Fusion of Geospatial and Stats
Three ways of fusion

Option 1:
Extension of Statistical DB

WFS – Vector Data (GML, shp) → Statistical Database

Option 2:
Extension of Geospatial Data

Statistical Database → WFS – Vector Data (GML, shp)

Option 3:
on-demand fusion

WFS – Vector Data (GML, shp) ↔ Statistical Database
Web Processing Service (WPS)

- OGC Standard since late 2007 (Version 1.0.0), version 2.0.0 was published in 2015
  http://www.opengeospatial.org/standards/wps

- Standardized service interface to publish and perform geospatial computing processes over the web

- No restriction on processes
  (any in-/outputs, any processing backend)

- More than 50 implementations available
  (Commercial & Open Source)

- Easy to use REST-binding for 52N-WPS implementation
Demo: WPS for Statistical Production

OGC WPS
- Data Fusion
- Spatial Transformations
- Statistical Analysis
- Integration of Legacy Systems
- ...

Statistics, e.g. nights spent

OGC Web Feature Service

Administrative Boundaries (GML)

OGC WPS

Web Apps

Desktop Apps

OGC Web Feature Service

Statistical Database

Statistics, e.g. nights spent

Administrative Boundaries (GML)
R Shiny web application

Fusion WPS

- Year:
  - 2013
  - 2014 [selected]
  - 2015

- Select a statistic:
  - nights spent

[Map of Germany with different shades indicating nights spent]
ArcMap Integration
Use case at IRCEL-CELINE

(a)

(b)

(c)

(d)
Benefits of WPS

- Standardised interface
  - Description format for processes with inputs/outputs
  - Integrates with other standards for spatial data infrastructures (OGC, ISO, INSPIRE)
- Web Service:
  - Remote processing, reusability, scalability, automatisation, ...
- Concepts for workflow composition (BPMN)
- On-demand fusion: retain DBs
Outlook

- SDMX and Simple Features/GML
- Usage of GSBPM for composition of statistical production workflows with geospatial information
  - Execution of GSBPM workflow using OGC WPS
  - OGC WPS as a module in a GSBPM workflow
- Handling of uncertainties (Uncert-ML)
- Derivation of secondary statistics (e.g. nights spent / km²)
Thank you!

More information and contact details:

http://52north.org

Dr. Benedikt Gräler
b.graeler@52north.org
Spatial Statistics & R

Dr. Christoph Stasch
c.stasch@52north.org
Sensor Web & Geoprocessing
Observations & Measurements (OM)

OM_Observation
- phenomenonTime: TM_Object
- resultTime: TM_Interaction
- validTime: TM_Period [0..1]
- resultQuality: DQ_Element [0..*]
- parameter: NamedValue [0..*]

+procedure 0..*

OM_Process
+observedProperty 1

GFI_PropertyType
+featureOfInterest 1

GFI_Feature

+relatedObservation 0..*

0..*

Any

+result
Representing Uncertainties

- Uncertainty: „parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used”


11 Meter or 11.3 ? or 10.8 ?
Uncertainty Model Language (UncertML)

- Discussion Paper at OGC (OGC 08-122r2)
- Defines a data model and different encodings (XML, JSON) for uncertainties
Many-to-many

Web Apps

Prediction Models

Desktop Apps

SDIs, SOA, Big Data

Geospatial Data Source

Statistical Database

External Data Sources

Sensor Sources

Scientific Data
Many-to-one

Web Apps

Prediction Models

Desktop Apps

SDIs, SOA, Big Data

Standardised Analysis

OGC  ISO  W3C  SDMX  GSBPMN

Standardised Data Exchange

OGC  ISO  W3C  SDMX

Geospatial Data Source

Statistical Database

External Data Sources

Sensor Sources

Scientific Data

UNECE Workshop 11/17 - Tools for Integrating Statistical Analysis in SDIs
WPS Tools

- **52° North WPS:**
  - Open Source Java Implementation
  - Support for default geo-operators:
    - Geotools, Sextante, GRASS GIS
  - Supports variety of different inputs/outputs:
    - GML, SHP, GeoTIFF, NetCDF, CSV, (Geo)JSON,...
  - Integration of custom processes: Java, R, Python, Executables, ...

- **52° North WPS4R:** Uploading of R Skripts and publication as WPS processes

- **R Shiny:** Framework for building Web applications with R

- **WPS-JS:** Java Script library