Geonovum Testbed – Topic 4

“Spatial data on the Web using the current SDI”

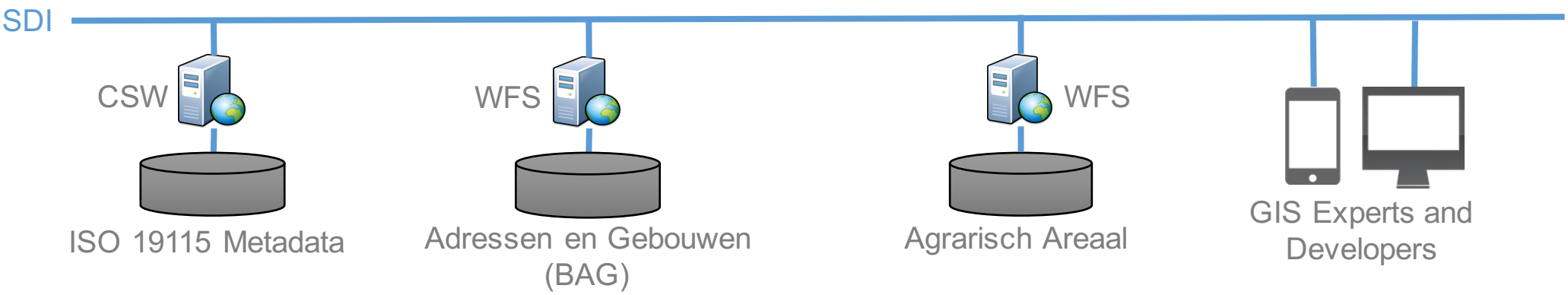
Spatial Data on the Web Best Practices – How are we doing?

interactive instruments

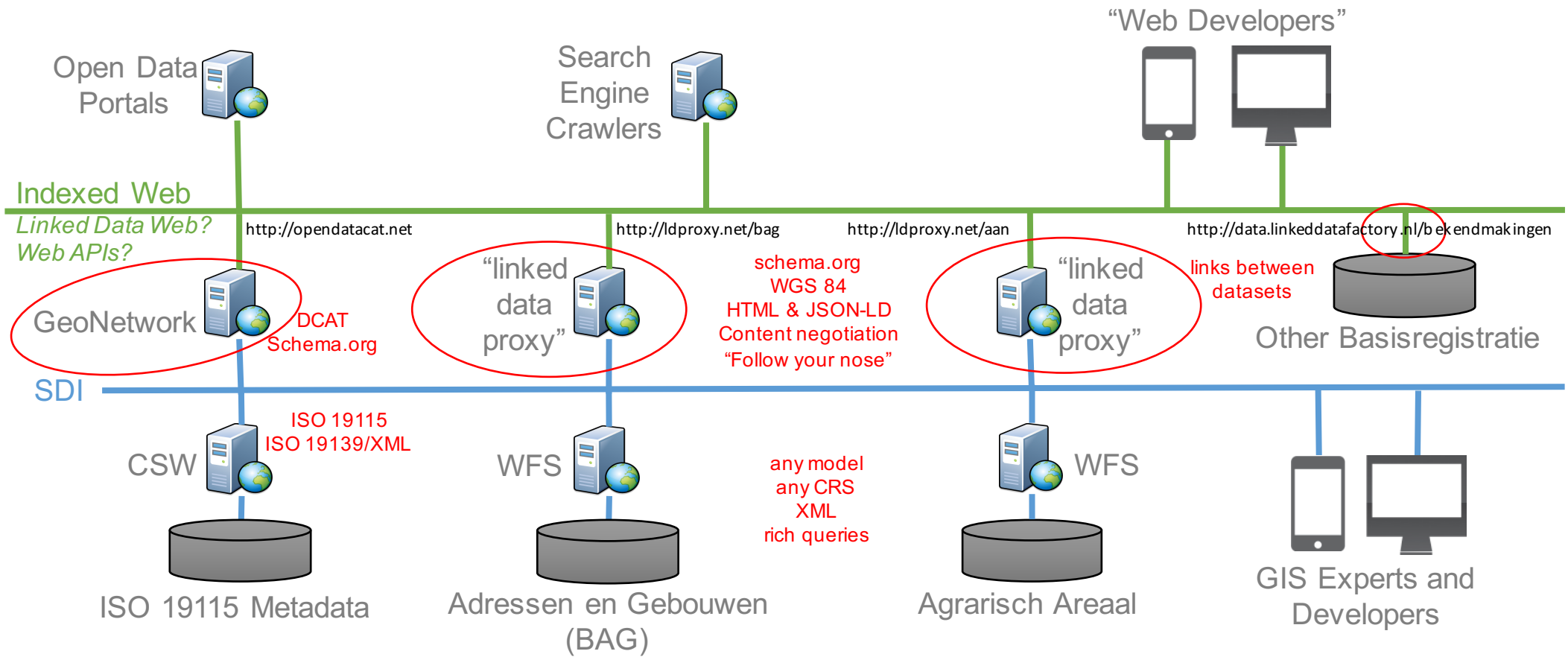
GEOCAT
GOVERNMENT GEOGRAPHIC DATA PUBLISHING

Linked Data Factory

The starting point



What we are building



What are we trying to do?

- crawlability and linkability, i.e. making each resource hosted by a WFS or CSW available via a persistent URI and ensure that all resources can be reached via links from a “landing page” for a data set (= presence on the web of data)
- classification of the resources using vocabularies supported by the main search engines on the web (= harmonisation of data discovery)
- representations of data for consumption by humans (HTML), developers (JSON-LD, GeoJSON, GML) and search engine crawlers (HTML with structured data annotations) (= data access based on current web practices)
- establishing and maintaining links between data (= connecting data with other data on the web)
- discovery of both spatial and non-spatial data by the same search engine (= harmonisation of data discovery)

Best Practices – Implemented

- BP1: Use globally unique HTTP identifiers for entity-level resources
- BP3: Convert or map dataset-scoped identifiers to URIs
- BP6: Provide a minimum set of information for your intended application
- BP7: How to describe geometry
- BP8: Specify Coordinate Reference System for high-precision applications
- BP12: Use spatial semantics for Spatial Things
- BP19: Make your entity-level links visible on the web
- BP20: Provide meaningful links
- BP25: Make your entity-level data indexable by search engines
- BP26: Include spatial information in dataset metadata
- BP27: Publish data at the granularity you can support
- BP28: Expose entity-level data through 'convenience APIs'

Best Practices – Implementation in progress

- BP5: Provide identifiers for parts of larger information resources
- BP13: Assert known relationships
- BP21: Link to spatial Things
- BP23: Link to related resources
- BP30: Include search capability in your data access API

Best Practices – Not used (but could be implemented)

- BP22: Link to resources with well-known or authoritative identifiers
- BP24: Use links to find related data
- BP29: APIs should be self-describing

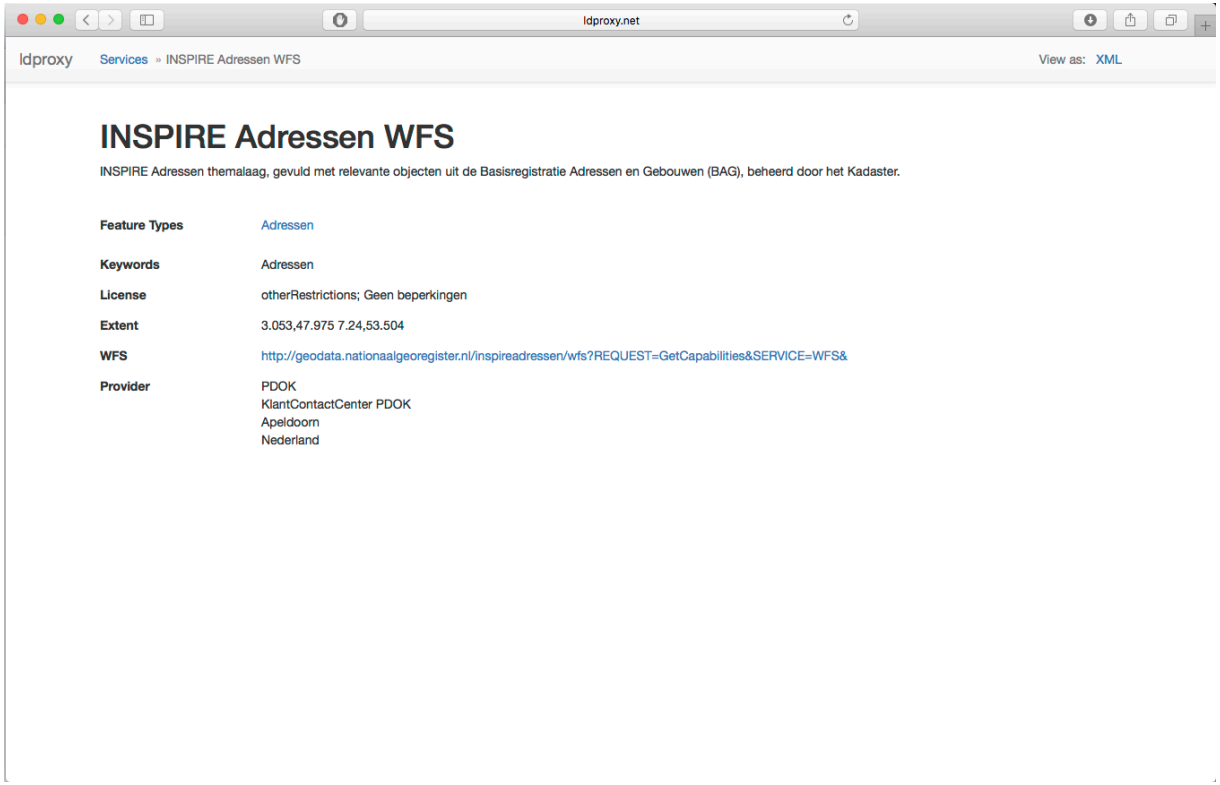
Best Practices – Not Implementable

- BP2: Reuse existing (authoritative) identifiers when available
- BP10: How to describe positional (in)accuracy

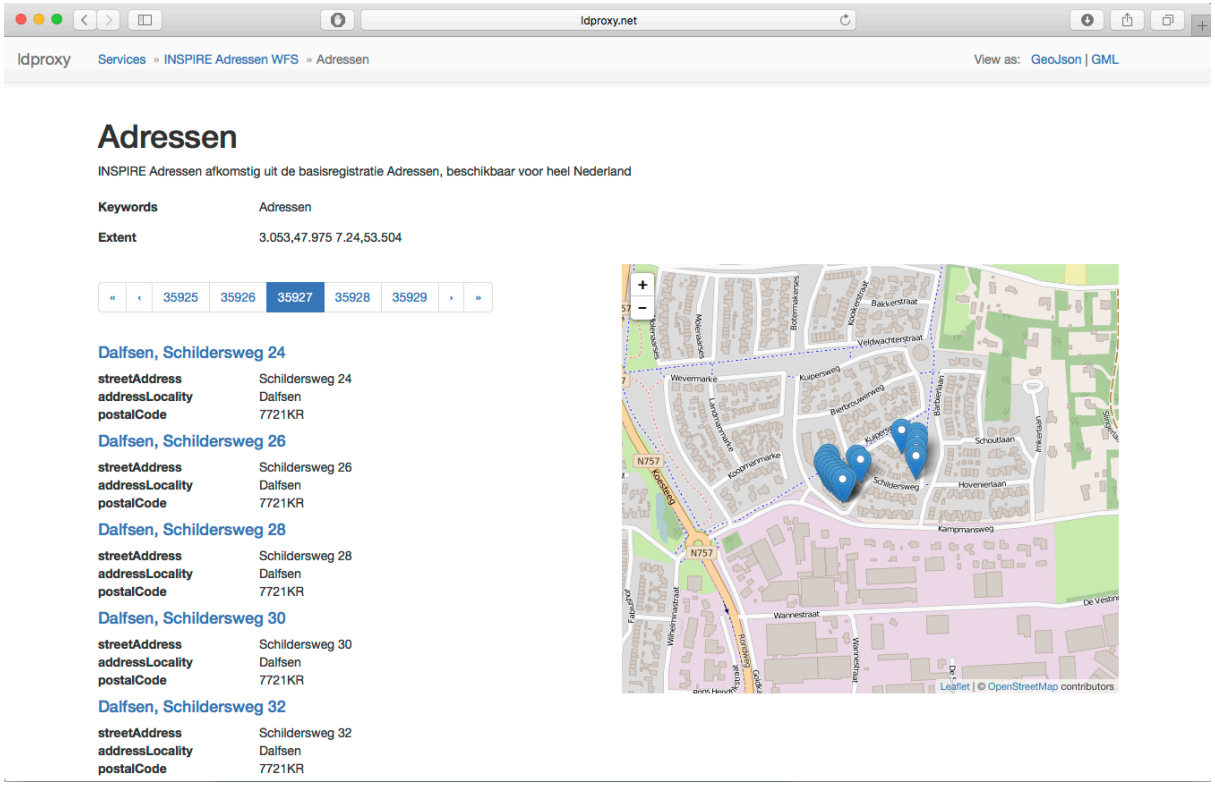
Best Practices – Not Applicable

- BP4: Provide stable identifiers for Things (resources) that change over time
- BP9: How to describe relative positions
- BP11: How to describe properties that change over time
- BP14: Provide context required to interpret observation data values
- BP15: Describe sensor data processing workflows
- BP16: Relate observation data to the real world
- BP17: How to work with crowd-sourced observations
- BP18: How to publish (and consume) sensor data streams

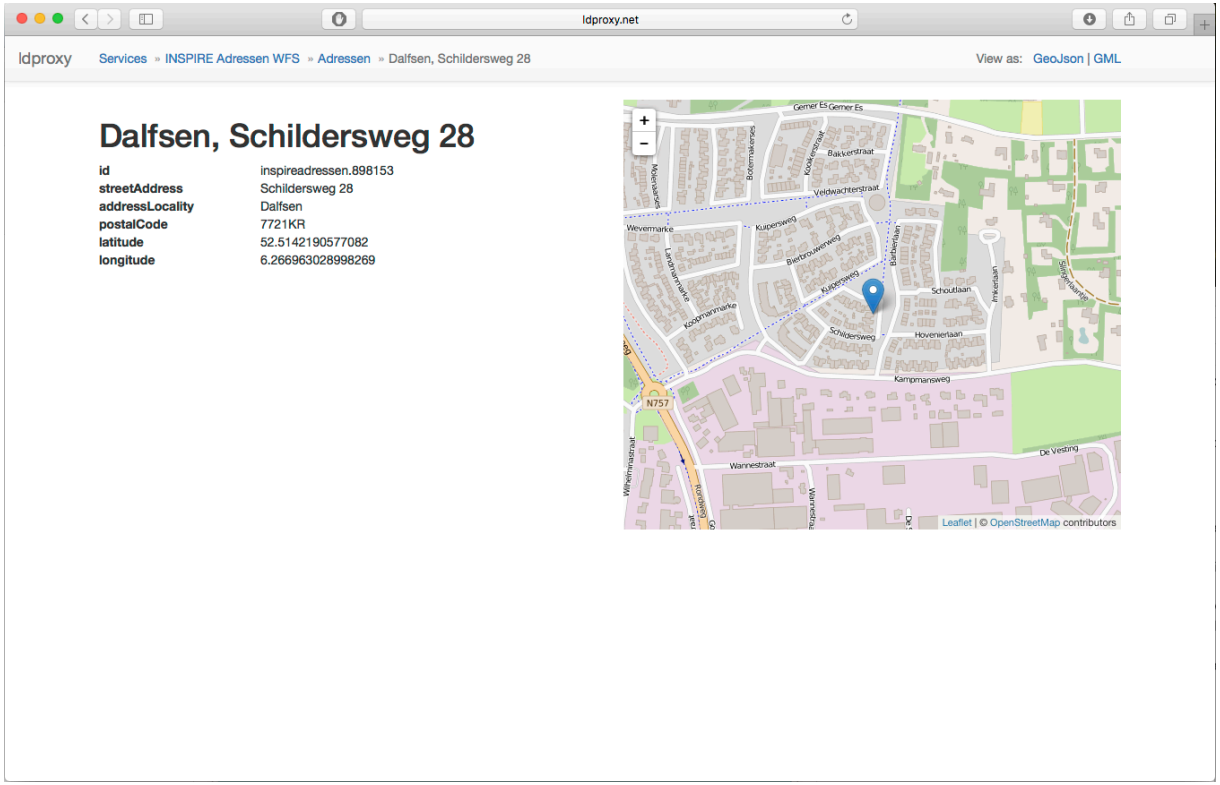
Dataset – the WFS landing page



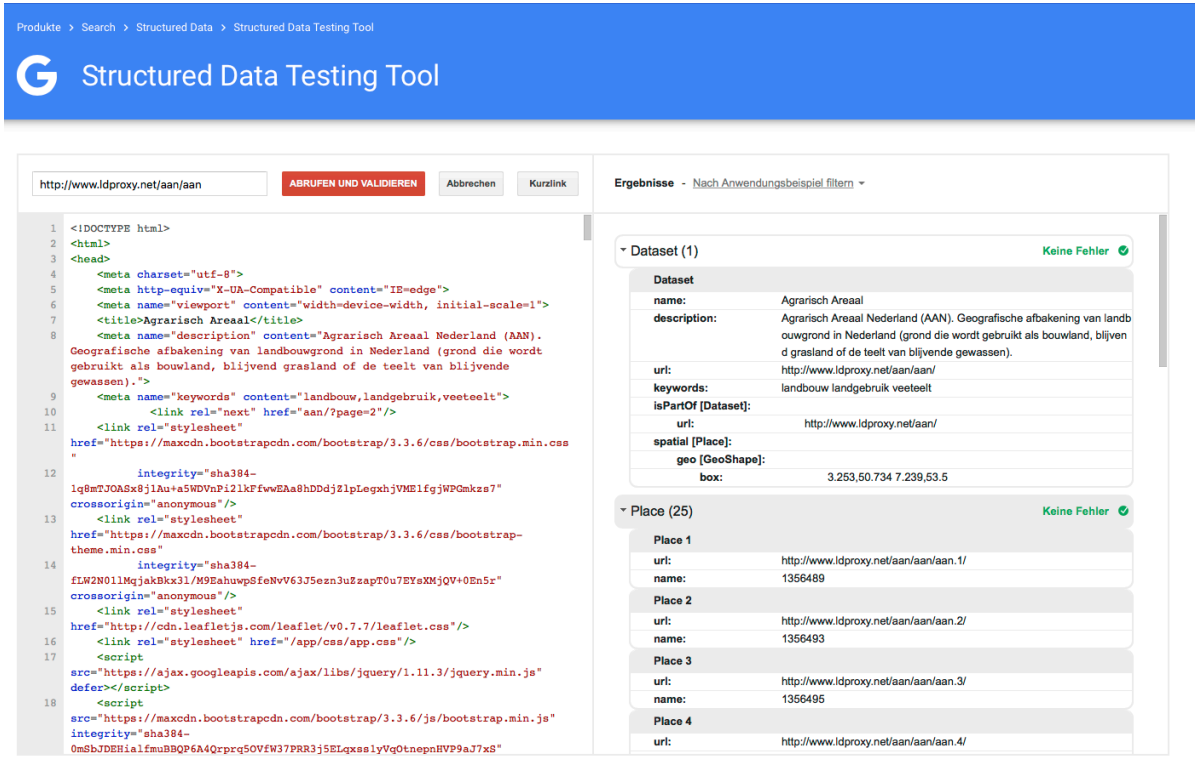
A feature type



A feature


ldproxy.net
ldproxy
Services » INSPIRE Adressen WFS » Adressen » Dalfsen, Schildersweg 28
View as: GeoJson | GML
Dalfsen, Schildersweg 28
id inspireadressen.898153
streetAddress Schildersweg 28
addressLocality Dalfsen
postalCode 7721KR
latitude 52.5142190577082
longitude 6.266963028998269
Leaflet | © OpenStreetMap contributors

Structured data using schema.org



Some challenges …

- Search engines are a black box
  - their use of spatial information in structured data
  - impact of URI strategies, formats, vocabularies, rdfa vs microformats vs JSON-LD, ... → topic 3
- Performance & data compactness
  - response time and size have an impact on search engine ranking and usability
  - existing services may not be performant enough & geometries may be large
- Content negotiation based on media types insufficient
  - for different use cases we want to provide multiple RDF representations using different vocabularies, e.g. metadata in DCAT or schema.org → separate URIs necessary
- Support for real time link discovery
  - often difficult or impossible due to inadequate information or inadequate structuring of information in existing datasets
- Metadata today is often incomplete and inconsistent
  - WFS capabilities and dataset metadata in CSWs often have incomplete and inconsistent information