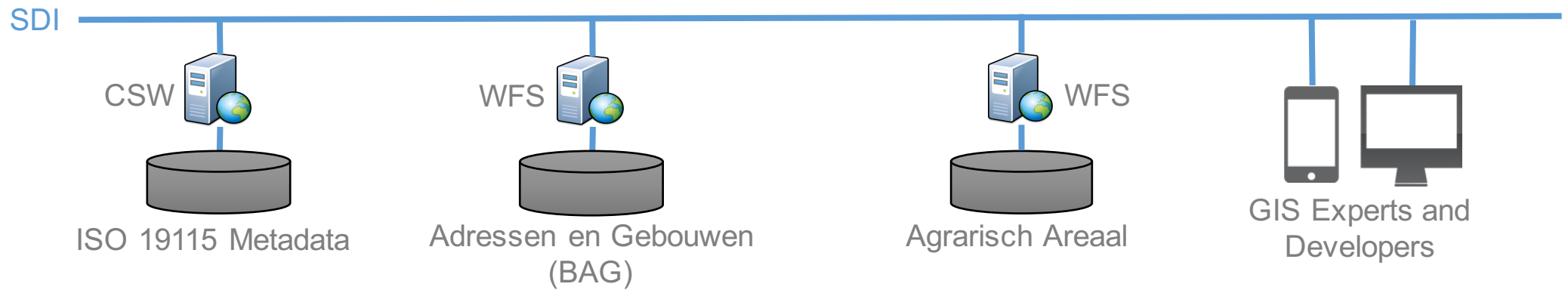


Geonovum Testbed – Topic 4  
“Spatial data on the Web using the current SDI”

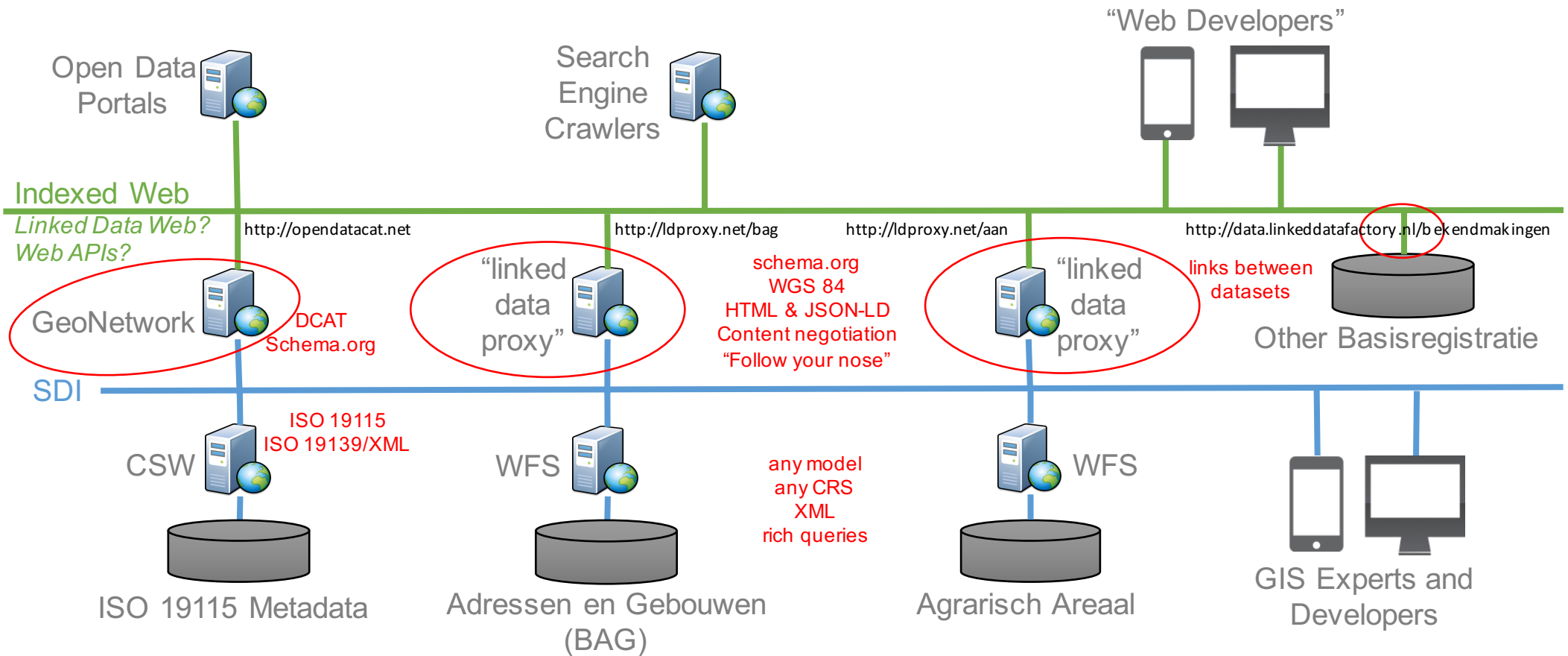
**Spatial Data on the Web Best Practices – How are we doing?**



# 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

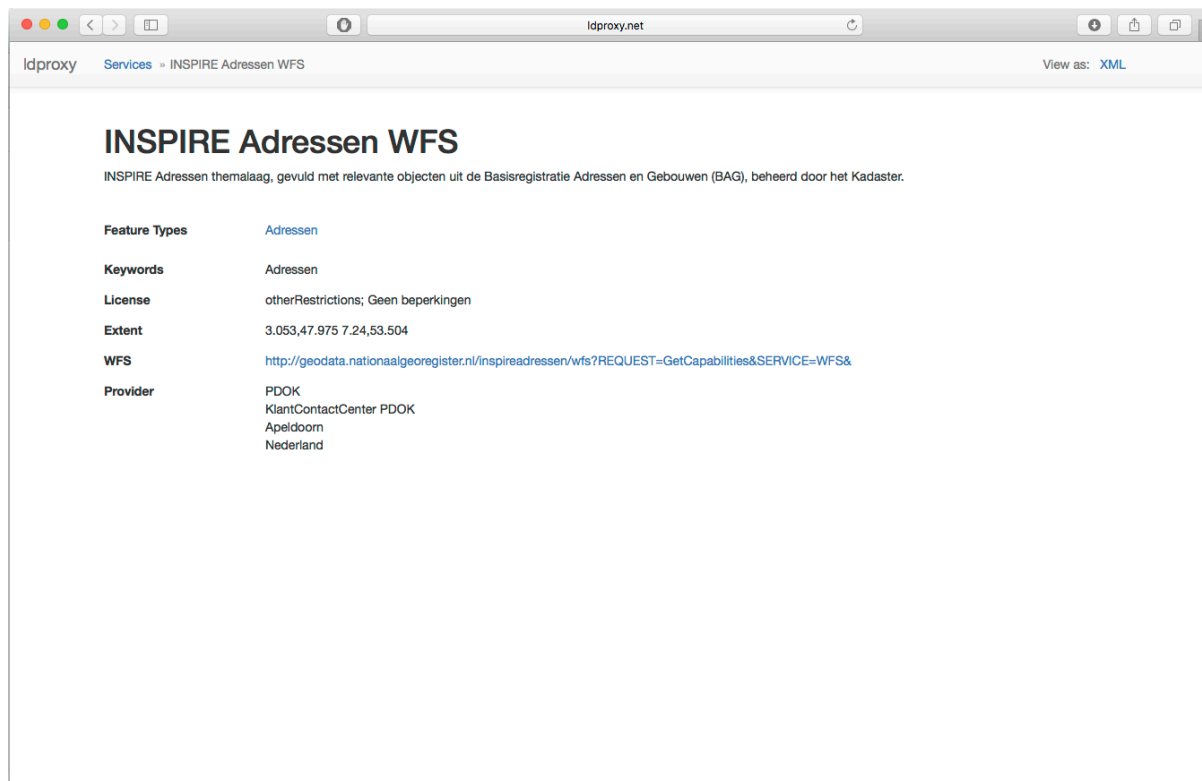
- [BP2](#): Reuse existing (authoritative) identifiers when available
- [BP10](#): How to describe positional (in)accuracy



# 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



The screenshot shows a web browser window with the address bar displaying 'ldproxy.net'. The page title is 'INSPIRE Adressen WFS'. Below the title, there is a description: 'INSPIRE Adressen themalaag, gevuld met relevante objecten uit de Basisregistratie Adressen en Gebouwen (BAG), beheerd door het Kadaster.' The page features a table with metadata:

<b>Feature Types</b>	<a href="#">Adressen</a>
<b>Keywords</b>	Adressen
<b>License</b>	otherRestrictions; Geen beperkingen
<b>Extent</b>	3.053,47.975 7.24,53.504
<b>WFS</b>	<a href="http://geodata.nationaalgeoregister.nl/inspireadressen/wfs?REQUEST=GetCapabilities&amp;SERVICE=WFS&amp;">http://geodata.nationaalgeoregister.nl/inspireadressen/wfs?REQUEST=GetCapabilities&amp;SERVICE=WFS&amp;</a>
<b>Provider</b>	PDOK KlantContactCenter PDOK Apeldoorn Nederland

The browser interface includes standard navigation buttons (back, forward, home, search) and a 'View as: XML' link in the top right corner.

# A feature type

ldproxy Services » INSPIRE Adressen WFS » Adressen View as: [GeoJson](#) | [GML](#)

## Adressen

INSPIRE Adressen afkomstig uit de basisregistratie Adressen, beschikbaar voor heel Nederland

**Keywords** Adressen

**Extent** 3.053,47.975 7.24,53.504

« 35925 35926 35927 35928 35929 »

**Dalfsen, Schildersweg 24**

**streetAddress** Schildersweg 24  
**addressLocality** Dalfsen  
**postalCode** 7721KR

**Dalfsen, Schildersweg 26**

**streetAddress** Schildersweg 26  
**addressLocality** Dalfsen  
**postalCode** 7721KR

**Dalfsen, Schildersweg 28**

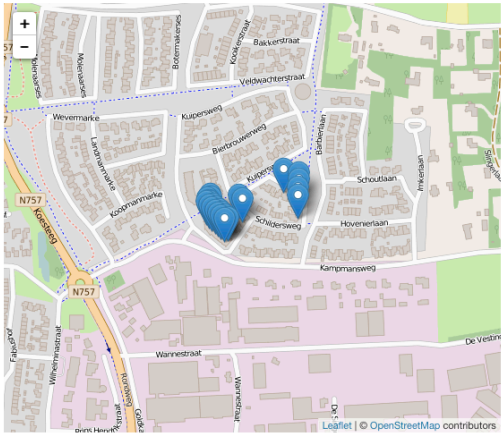
**streetAddress** Schildersweg 28  
**addressLocality** Dalfsen  
**postalCode** 7721KR

**Dalfsen, Schildersweg 30**

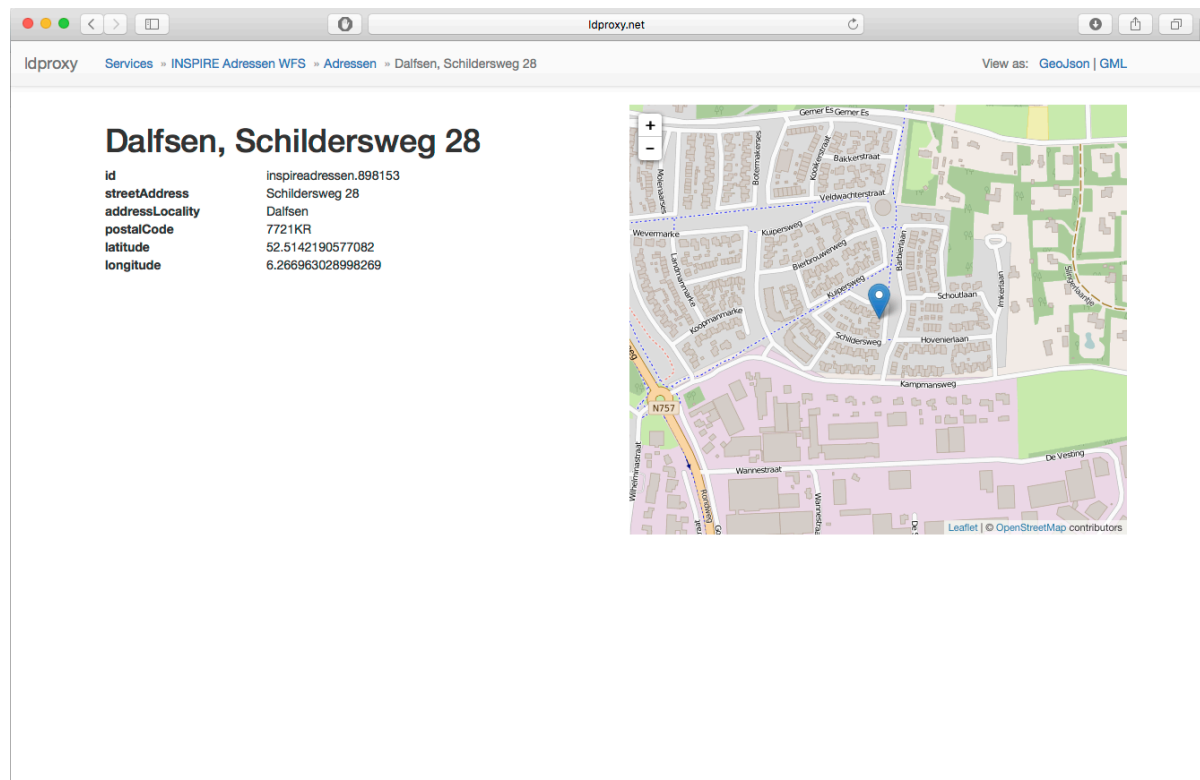
**streetAddress** Schildersweg 30  
**addressLocality** Dalfsen  
**postalCode** 7721KR

**Dalfsen, Schildersweg 32**

**streetAddress** Schildersweg 32  
**addressLocality** Dalfsen  
**postalCode** 7721KR




# A feature



# Structured data using schema.org

[Produkte](#) > [Search](#) > [Structured Data](#) > [Structured Data Testing Tool](#)

 **Structured Data Testing Tool**

ABRUFEN UND VALIDIERENAbbrechenKurzlink

1

<!DOCTYPE html>

2

<html>

3

<head>

4

<meta charset="utf-8">

5

<meta http-equiv="X-UA-Compatible" content="IE=edge">

6

<meta name="viewport" content="width=device-width, initial-scale=1">

7

<title>Agrarisch Areaal</title>

8

<meta name="description" content="Agrarisch Areaal Nederland (AAN). Geografische afbakening van landbouwgrond in Nederland (grond die wordt gebruikt als bouwland, blijvend grasland of de teelt van blijvende gewassen).">

9

<meta name="keywords" content="landbouw, landgebruik, veeteelt">

10

<link rel="next" href="aan/?page=2"/>

11

<link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.6/css/bootstrap.min.css" integrity="sha384-lq8m7JOASx811Au+a5WDVnPi2lkFfwwEAa8hDDdjZlpLegxhjVME1fgjWPGmkzs7" crossorigin="anonymous"/>

12

<link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.6/css/bootstrap-theme.min.css" integrity="sha384-FLW2N011MqjakBkx1l/M9Eahwp8fSfeNv63J5ezn3u8ZapT0u7EYsXhJQV+0En5z" crossorigin="anonymous"/>

13

<link rel="stylesheet" href="http://cdn.leafletjs.com/leaflet/v0.7.7/leaflet.css">

14

<link rel="stylesheet" href="/app/css/app.css">

15

<script src="https://ajax.googleapis.com/ajax/libs/jquery/1.11.3/jquery.min.js" defer></script>

16

<script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.6/js/bootstrap.min.js" integrity="sha384-0mSbJDEhialfmuBQ6A4Qrprq6OVFW3FRR3j5ELqxs1yVqOtnepnRVP9aJ7xS" integrity="sha384-0mSbJDEhialfmuBQ6A4Qrprq6OVFW3FRR3j5ELqxs1yVqOtnepnRVP9aJ7xS">

17

18

Ergebnisse - [Nach Anwendungsbeispiel filtern](#)

Dataset (1) Keine Fehler

|                     |   |
|---------------------|---|
| Dataset             |   |
| name:               | Agrarisch Areaal  |
| description:        | Agrarisch Areaal Nederland (AAN). Geografische afbakening van landbouwgrond in Nederland (grond die wordt gebruikt als bouwland, blijvend grasland of de teelt van blijvende gewassen). |
| url:                | http://www.ldproxy.net/aan/aan/   |
| keywords:           | landbouw landgebruik veeteelt   |
| isPartOf [Dataset]: |   |
| url:                | http://www.ldproxy.net/aan/   |
| spatial [Place]:    |   |
| geo [GeoShape]:     |   |
| box:                | 3.253,50.734 7.239,53.5   |

Place (25) Keine Fehler

|         |                                       |
|---------|---------------------------------------|
| Place 1 |                                       |
| url:    | http://www.ldproxy.net/aan/aan/aan.1/ |
| name:   | 1356489                               |
| Place 2 |                                       |
| url:    | http://www.ldproxy.net/aan/aan/aan.2/ |
| name:   | 1356493                               |
| Place 3 |                                       |
| url:    | http://www.ldproxy.net/aan/aan/aan.3/ |
| name:   | 1356495                               |
| Place 4 |                                       |
| url:    | http://www.ldproxy.net/aan/aan/aan.4/ |

# 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