

## Feasibility study on data harvesting using INSPIRE infrastructure

INSPIRE MIG-T, 24.-25.10.2018

Darja Lihteneger

## Outline

- Finding services and data
- Services
- Downloaded data
- Referencing spatial objects
- Findings and input for Reportnet 3.0

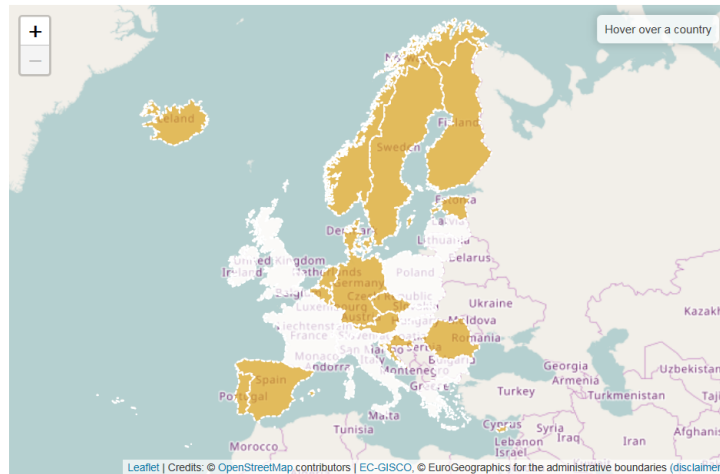
# Reportnet 3.0 / Feasibility study on data harvesting

- **Purpose:**  
To assess the applicability of harvesting national INSPIRE services to automate the collection of geospatial data sets falling under reporting obligations
- **Scope:**
  - Use case 1: data harvesting
  - Use case 2: referencing spatial objects
- **Thematic domain:**
  - Natura 2000 sites (INSPIRE theme “Protected Sites”)

The study will address **INSPIRE download services** providing **Natura 2000 spatial data sets**

# Identification of data and services – needs & means

## Legislation - Directive 92/43/EEC



Show:  Downloadable  Viewable

### INSPIRE Geoportal Data Set Statistics

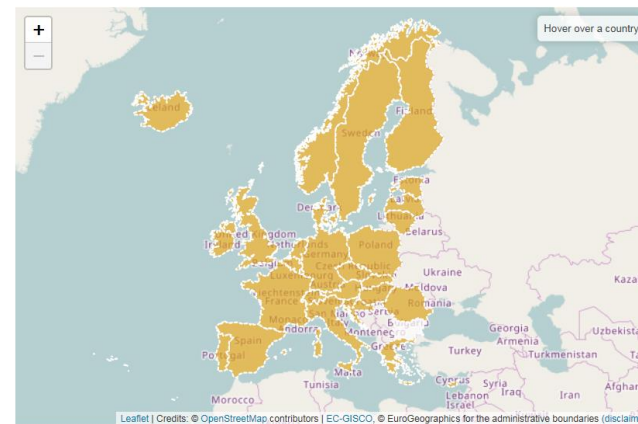
94

Metadata records

43

Downloadable Data Sets

### Data Sets - Protected sites of Europe ?



Show:  Downloadable  Viewable

### INSPIRE Geoportal Data Set Statistics

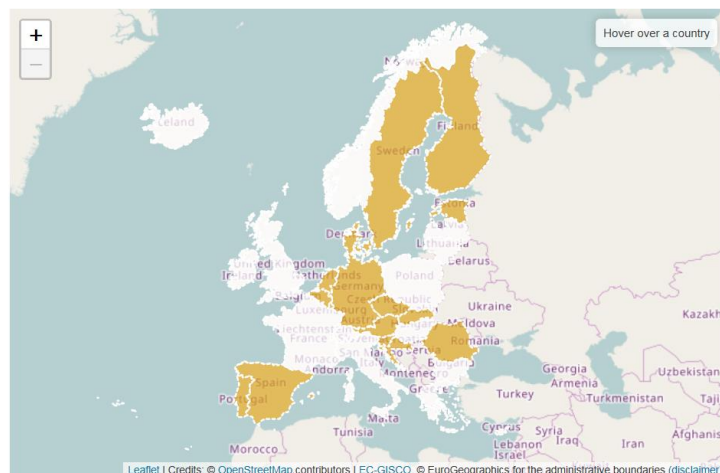
2511

Metadata records

257

Downloadable Data Sets

## Legislation - Directive 2009/147/EC



Show:

### INSPIRE Geoportal Data Set Statistics

59

Metadata records

26

Downloadable Data Sets

Images – INSPIRE Geoportal snapshot on 21.10.2018

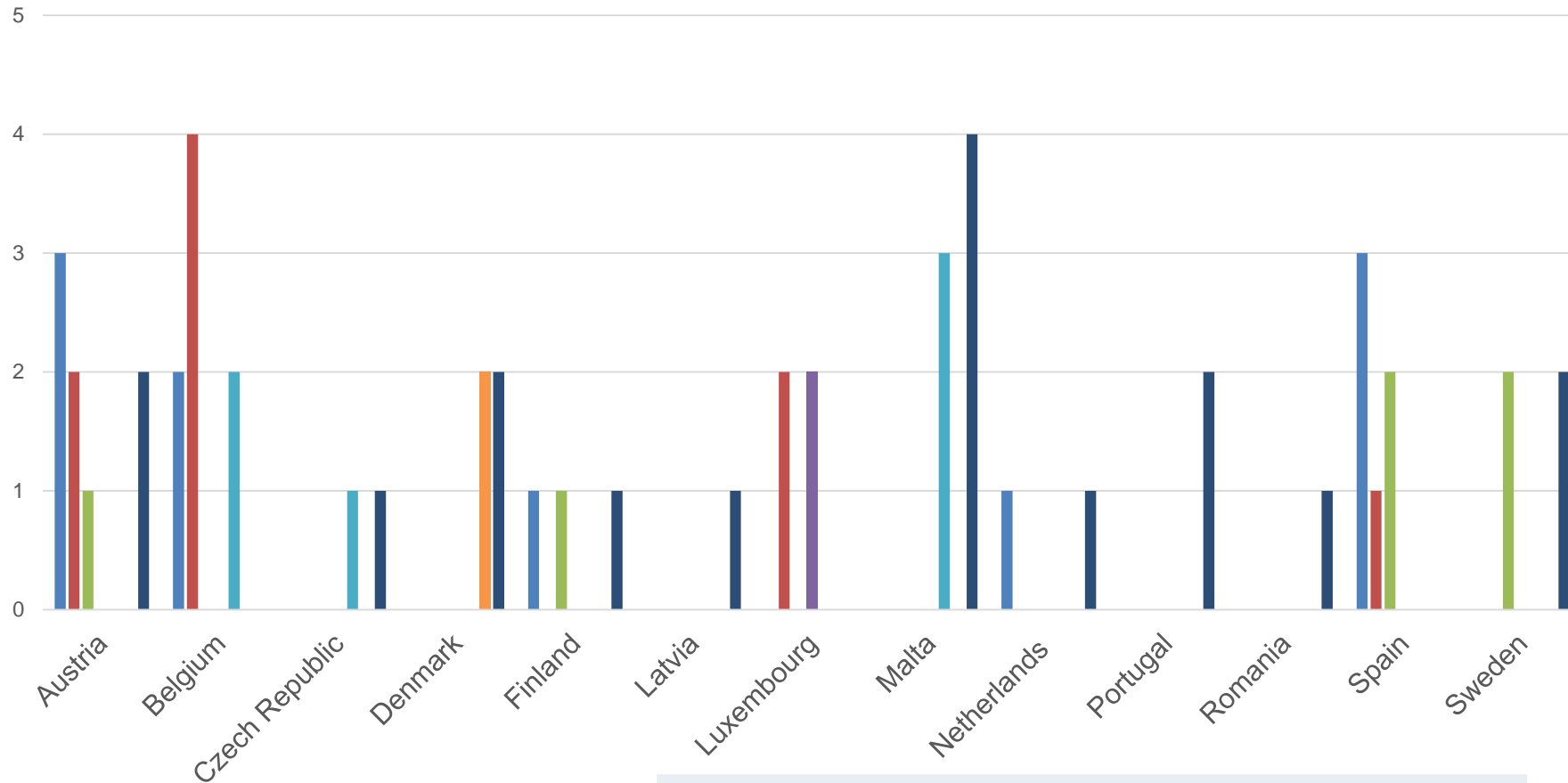
## Identification:

- Results of INSPIRE Geoportal PDS Viewer (23.08.2018)
- Additional manual search in INSPIRE Geoportal
- Additional manual refinement (selection of Natura 2000 data sets from the list of several data sets or from top Atom Feed, selection of predefined stored query in WFS)
- Creating WFS GetFeature requests



# Sample - initial list of services

Natura 2000 spatial data - number and type of INSPIRE download services per country



■ Atom feed

■ File download / compressed GML file (zip)

■ File download / compressed Shapefile (zip)

■ File download / geoJSON

■ File download / GML file

■ FTP / compressed file (ZIP)

■ WFS

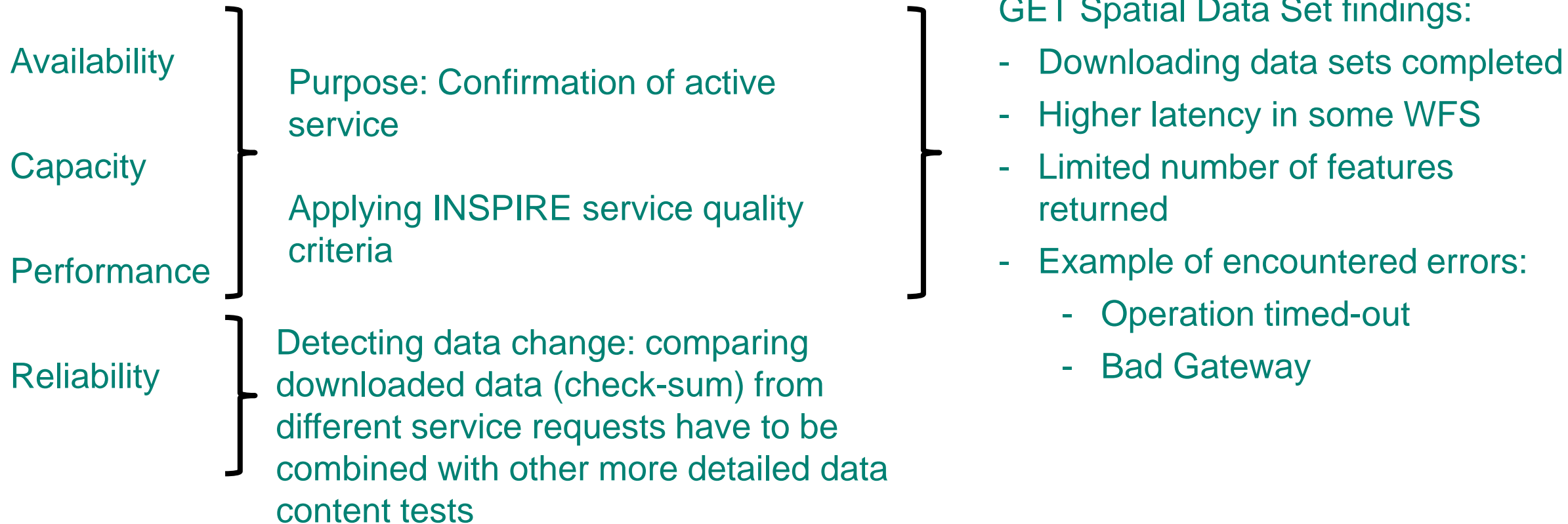
**Not always 1-1 service-country or data-service mapping**

Service type	
Atom feed	10
File download (http, https, ftp)	25
WFS	17
<b>Total</b>	<b>52</b>



# Service connection and monitoring / testing

- All services tested to build a better understanding of service performance and use for reporting



**Service monitoring and statistics is important in the period of official reporting data flow: documentation of data harvesting process, identification and awareness of service errors, trigger for notifications**

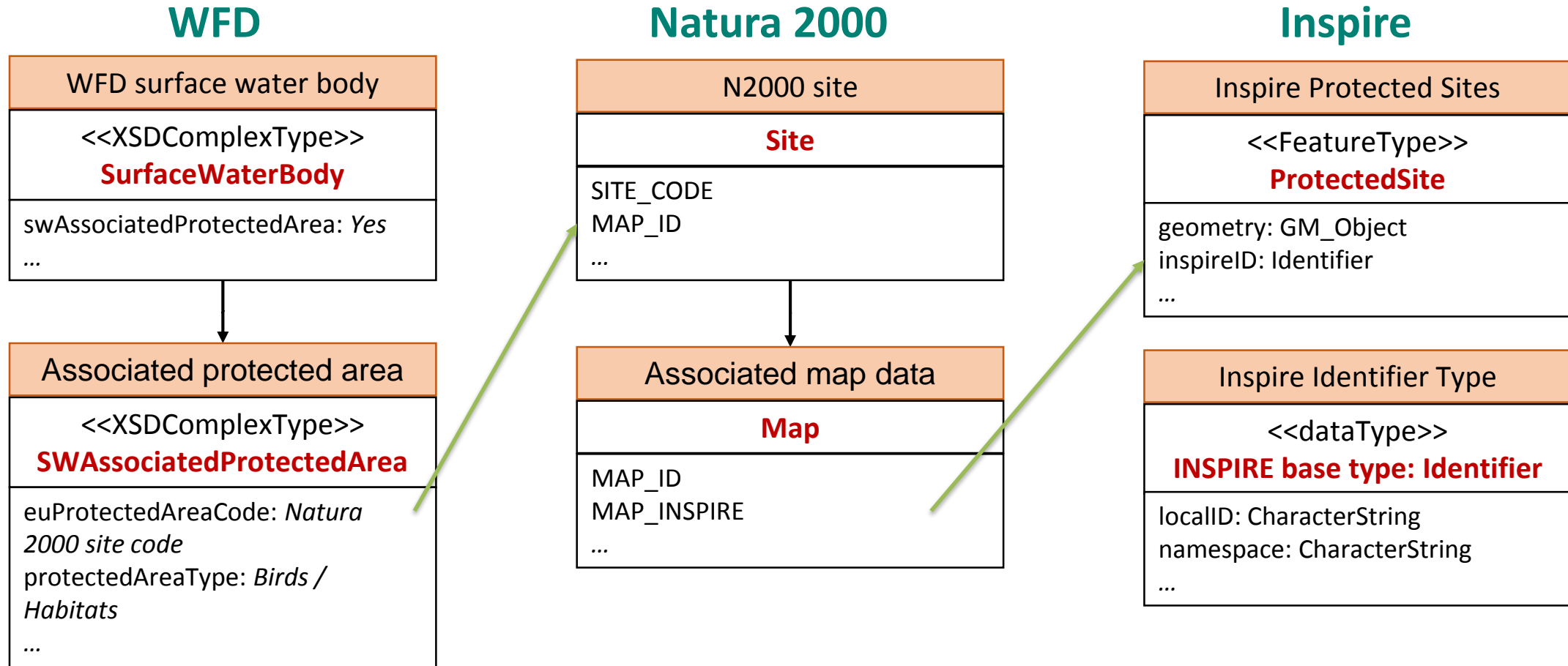
# Spatial data quality control: is it Natura 2000 data set?

- **Data quality control on data provided as GML:**
  - Using INSPIRE ETF validator
  - Result: high level of passed tests
- **Some encountered errors:**
  - Test not completed (Timeout on test)
  - GML schema errors
  - Protected site feature was not in dataset
- **Differences in content but not necessarily errors:**
  - Certain content is not included, e.g. SPA, SCI
  - Other content is included (non-Natura 2000 designation scheme, no protected sites)

Input for data quality control tests in Reportnet at the level of data acceptance

# Use case 2: Referencing mechanism

**Challenge:** For WFD water dependent protected areas that correspond to Natura 2000 sites find their related spatial data !



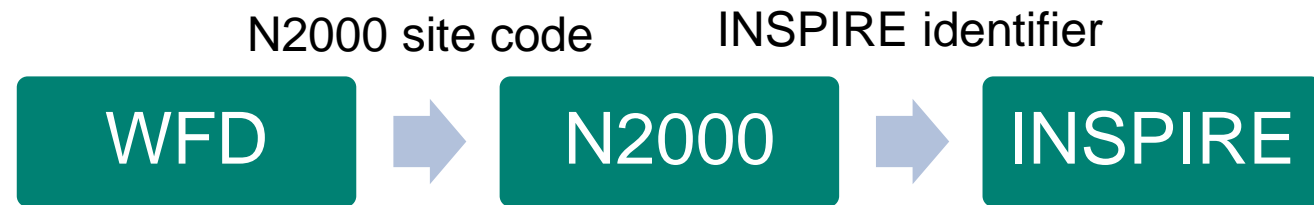
Each resource follows own life cycle / maintenance



# Referencing mechanism findings

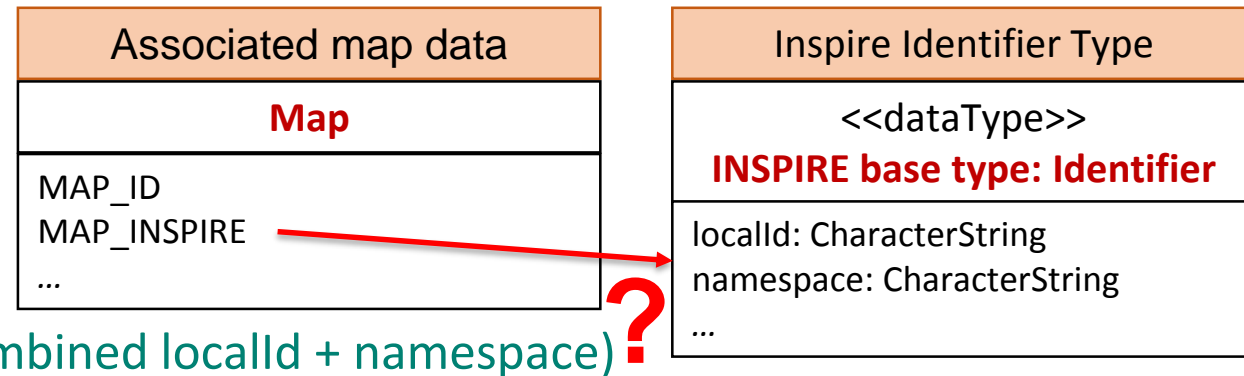
## Positive patterns:

- **WFD -> Natura 2000:** link gives no issues
- **N2000 - INSPIRE:**
  - Inspire ID is provided
  - N2000 site code is part of Inspire ID
- **Referencing of spatial data:** through WFS requests based on Inspire identifier



## Issues:

- **Natura 2000 SDF reporting:**
  - Inspire ID often not filled in
  - Inspire ID provided in different ways (localId, combined localId + namespace)
- **WFS – spatial objects provided according to different schemas:**
  - NO STANDARDISATION > customised matching needed
  - Alternatives for linking if N2000 <> Inspire / possible if Natura 2000 site code is provided in INSPIRE spatial data



# Findings – requirements for focus groups

ReqID	Priority	Reportnet	Reporting data and service providers (MS)	Natura 2000 SDF	INSPIRE Protected sites	INSPIRE
F-024	M	X				
F-026	M	X				
F-028	M		X	X	X	
F-029	S					X
F-030	S		X			
T-010	S	X	X			
F-025	C	X				
F-027	C	X				
F-031	C		X			X
T-012	C	X				
T-013	C	X	X			
T-014	C	X	X			
T-011	W	X	X			

# Requirements - Reportnet

ReqID	Priority	Reportnet
F-024	M	Reportnet should provide the means to collect INSPIRE datasets and services, all relevant datasets have to be available (complete geographic coverage, topic)
F-026	M	If the Inspire service is indicated in the reporting data flow, testing the availability of that service should be part of the reporting workflow. Secondly, it must be tested if the returned data follows the expected schema
T-010	S	Atom feed: Supplied Atom feeds should be datasets feeds (not top feeds). The Reportnet should include quality procedures to test the service and to provide the notification on findings.
F-025	C	If an Inspire dataset contains features related to different reporting obligations, make sure it is possible to download only the part of the dataset related to that reporting.
F-027	C	Test of each reported entry can be reached in the Inspire dataset through the given service(s). The Reportnet should include quality procedures to test the data and matching and to provide the notification on findings.
T-012	C	Reportnet workflows should continue to accommodate manual uploading of datasets where services are not available.
T-013	C	WFS: Reportnet must be able to check if stored queries exist, therefore the ListStoredQueries feature is mandatory.
T-014	C	Filenames in archived datasets should be unique. The Reportnet should include quality procedures to test the service and to provide the notification on findings.
T-011	W	Atom dataset feeds supplied for harvesting could contain only the entries for the datasets under the specific reporting obligation. The Reportnet should include quality procedures to test the service and to provide the notification on findings.

# Requirements – Reporting data and service providers

ReqID	Priority	Reporting data and service providers (MS)
F-028	M	Improve the implementation of inspire ID in Natura 2000 SDF.
F-030	S	Ensuring the spatial data (downloaded through INSPIRE download services) match the reported thematic data. The responsibility to ensure this synchronisation lies with the MS. However, while checking the data, there should be a mechanism in Reportnet to indicate the discrepancies to the MS reporting contact person (reporter).
T-010	S	Atom feed: Supplied Atom feeds should be datasets feeds (not top feeds).
F-031	C	WFS: It should be easy and straightforward to extract a single feature using INSPIRE identifier or thematic identifier.
T-013	C	WFS: WFS Download Service must provide the ListStoredQueries feature for further interrogations.
T-014	C	Compressed files: Filenames in compressed files (archived datasets) should be unique.
T-011	W	Atom feed: Atom dataset feeds supplied for harvesting could contain only the entries for the datasets under the specific reporting obligation.

# Requirements – Data models, data provision, INSPIRE

ReqID	Priority	Natura 2000	INSPIRE Protected sites
F-028	M	The way the InspireID in the Natura 2000 SDF should be provided must be detailed, tackling the namespace/localID issue (one attribute, standardisation), or Natura 2000 SDF might change	Standardisation for using Inspire ID for thematic data sets, e.g. Natura 2000 SDF

ReqID	Priority	INSPIRE	
F-029	S	In general, it would be very useful if the Inspire data models would include thematic identifiers, that would allow to relate the features to the corresponding reporting obligations. For the purpose of the Natura 2000 reporting, the thematic ID (Natura2000 siteCode) should be added to the INSPIRE Protected sites (PS) schema.	
F-031	C	It would be much more useful to allow a standardised query on the Inspire ID or, if it is available in the Inspire dataset, on the thematic identifier. As a possible solution, making available a stored query returning a feature with the specific Inspire ID would be very useful. Alternatively, a getFeatureByThematic ID could be used	