Spatial Data on the Web

Tools and guidance for data providers
W3C Data on the Web Best Practices

7. Best Practices Summary

Best Practice 1: Provide metadata
Best Practice 2: Provide descriptive metadata
Best Practice 3: Provide structural metadata
Best Practice 4: Provide data license information
Best Practice 5: Provide data provenance information
Best Practice 6: Provide data quality information
Best Practice 7: Provide a version indicator
Best Practice 8: Provide version history
Best Practice 9: Use persistent URLs as identifiers of datasets
Best Practice 10: Use persistent URLs as identifiers within datasets
Best Practice 11: Assign URLs to dataset versions and series
Best Practice 12: Use machine-readable standardized data formats
Best Practice 13: Use locale-neutral data representations
Best Practice 14: Provide data in multiple formats
Best Practice 15: Reuse vocabularies, preferably standardized ones
Best Practice 16: Choose the right formalization level
Best Practice 17: Provide bulk download
Best Practice 18: Provide Subsets for Large Datasets

Best Practice 19: Use content negotiation for serving data available in multiple formats
Best Practice 20: Provide real-time access
Best Practice 21: Provide data up to date
Best Practice 22: Provide an explanation for data that is not available
Best Practice 23: Make data available through an API
Best Practice 24: Use Web Standards as the foundation of APIs
Best Practice 25: Provide complete documentation for your API
Best Practice 26: Avoid Breaking Changes to Your API
Best Practice 27: Preserve identifiers
Best Practice 28: Assess dataset coverage
Best Practice 29: Gather feedback from data consumers
Best Practice 30: Make feedback available
Best Practice 31: Enrich data by generating new data
Best Practice 32: Provide Complementary Presentations
Best Practice 33: Provide Feedback to the Original Publisher
Best Practice 34: Follow Licensing Terms
Best Practice 35: Claim the Original Publication
4. Best Practices Summary

This document contains a variety of best practices related to the publication and usage of spatial data on the Web. First, it contains several more in-depth introductions on Spatial Things and geometry, coverages, spatial relations, coordinate reference systems, linked data, and Spatial Data Infrastructures. After that, the best practices themselves are described.

The following best practices can be found in this document:

**Best Practices Summary**

<table>
<thead>
<tr>
<th>Best Practice 1</th>
<th>Use globally unique persistent HTTP URIs for Spatial Things</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Practice 2</td>
<td>Make your spatial data indexable by search engines</td>
</tr>
<tr>
<td>Best Practice 3</td>
<td>Link resources together to create the Web of data</td>
</tr>
<tr>
<td>Best Practice 4</td>
<td>Use spatial data encodings that match your target audience</td>
</tr>
<tr>
<td>Best Practice 5</td>
<td>Provide geometries on the Web in a usable way</td>
</tr>
<tr>
<td>Best Practice 6</td>
<td>Provide geometries at the right level of accuracy, precision, and size</td>
</tr>
<tr>
<td>Best Practice 7</td>
<td>Choose coordinate reference systems to suit your user's applications</td>
</tr>
<tr>
<td>Best Practice 8</td>
<td>State how coordinate values are encoded</td>
</tr>
<tr>
<td>Best Practice 9</td>
<td>Describe relative positioning</td>
</tr>
<tr>
<td>Best Practice 10</td>
<td>Use appropriate relation types to link Spatial Things</td>
</tr>
<tr>
<td>Best Practice 11</td>
<td>Provide information on the changing nature of spatial things</td>
</tr>
<tr>
<td>Best Practice 12</td>
<td>Expose spatial data through 'convenience APIs'</td>
</tr>
<tr>
<td>Best Practice 13</td>
<td>Include spatial metadata in dataset metadata</td>
</tr>
<tr>
<td>Best Practice 14</td>
<td>Describe the positional accuracy of spatial data</td>
</tr>
</tbody>
</table>

5. Namespaces

This section is accumulative.
Search, evaluation and use of data in an SDI

• Open the geoportal in a web browser

Prior-knowledge about geoportals required, most people search with a search engine
Search, evaluation and use of data in an SDI

- Open the geoportal in a web browser
- Navigate to search

Prior-knowledge about geoportals required, most people search with a search engine
Search, evaluation and use of data in an SDI

- Open the geoportal in a web browser
- Navigate to search
- Enter search text and potentially structured search criteria
- Browse the metadata records and select a result you want to investigate further

Prior-knowledge about geoportals required, most people search with a search engine
Search, evaluation and use of data in an SDI

- Open the geoportal in a web browser
- Navigate to search
- Enter search text and potentially structured search criteria
- Browse the metadata records and select a result you want to investigate further
- Review of the metadata

INSPIRE (ISO 19115) metadata is for GIS specialists, hard to understand for most
Search, evaluation and use of data in an SDI

- Open the geoportal in a web browser
- Navigate to search
- Enter search text and potentially structured search criteria
- Browse the metadata records and select a result you want to investigate further
- Review of the metadata
- Follow the link to the “download service”

Prior-knowledge required, what a download service is and what to do with a WFS capabilities XML document
Search, evaluation and use of data in an SDI

- Open the geoportal in a web browser
- Navigate to search
- Enter search text and potentially structured search criteria
- Browse the metadata records and select a result you want to investigate further
- Review of the metadata
- Copy the WFS GetCapabilities URL
- Open a WFS client, connect to the WFS and access the data

You need a specific application to access and review the data as the WFS capabilities document does not include links to the data itself – non-specialists will typically do not have one and the XML does not help to find one
Search, evaluation and use of data in an SDI

- Open the geoportal in a web browser
- Navigate to search
- Enter search text and potentially structured search criteria
- Browse the metadata records and select a result you want to investigate further
- Review of the metadata
- Copy the WFS GetCapabilities URL
- Open a WFS client, access the data
- Analyse the dataset → Use

Spatial data is often hard to understand for those not familiar with the dataset – additional thematic and technical documentation is needed to understand it – but no links to these documents are available
Searching and using data: typical expectations

• Enter search criteria for the desired data in the address/search bar of the browser
• Browse through the first results and check, if one of them seems to provide the desired data or refers to it
• Browse through the dataset to determine, if it has the desired information
• If yes, download the complete dataset or study the online API documentation and examples for accessing the data
• Use the data in your application using existing and well documented libraries or APIs

→ Use the standard tools of the target platform – in this case: the Web (web browser, HTTP, delegation to applications using media types, search engines, Web APIs, etc.)
Datasets and distributions

Example:
• National Road Network dataset

Distributions of the dataset (examples):
• File Download (national format)
• WMTS (national schema/layers)
• WFS (national schema)
• National e-Government Data API
• INSPIRE View Service (WMS)
• INSPIRE Download Service (ATOM)
• INSPIRE Download Service (WFS)

Issues:
• Geoportals typically return several results for a single dataset (one result per dataset or distribution), but their relationship is unclear
  • Default should be only the dataset; then navigate to distributions from the dataset
• ISO 19115/GMD is not compatible with this
• GeoDCAT-AP narrows distributions to files and uses a new resource type for services
• Some data providers in INSPIRE consider a distribution of a national dataset that conforms to the INSPIRE data specifications as a new dataset, not as another distribution
Assessment and recommendations for each best practice

**DWBP 19: Use content negotiation for serving data available in multiple formats**

Use content negotiation in addition to file extensions for serving data available in multiple formats.

**Key statements:**

A resource, such as a dataset, can have many representations. The same data might be available as JSON, XML, RDF, CSV and HTML. These multiple representations can be made available via an API, but these should be made available from the same URL using content negotiation to return the appropriate representation.

**Assessment:**

In general, this is not implemented. OGC web services typically do not support content negotiation, but support this capability using other mechanisms only, for example, as a parameter “outputFormat”.

**Links to related recommendations:**

- Section 6.7: Data access API

Identified key aspects of the Best Practices

• Presence on the Web of data
  ▪ crawlability and linkability, i.e. make each resource, including each feature, available via a persistent URI and ensure that all resources can be reached via links from a “landing page” for the dataset distribution

• Harmonisation of data discovery
  ▪ classification of the resources using vocabularies supported by the main search engines on the Web
  ▪ discovery of both spatial and non-spatial data by the same search engine

• Data access based on current Web practices
  ▪ representations of data for consumption by humans (HTML), developers (GeoJSON, GML, JSON-LD) and search engine crawlers (HTML with structured data annotations), accessible via HTTP(S)
  ▪ return data in chunks fit for use in Web applications
  ▪ return data as useful sets of information
  ▪ provide a Data API consistent with the Web architecture

• Connecting data with other data on the Web
  ▪ establishing and maintaining links between data

→ Recommendations
General approach

- Assumption: Changes to INSPIRE will be evolutionary
  - Continue to build on the existing legal and guidance framework
  - At the same time explore and test new technical options

- Prioritize gaps that
  - can be explored in experiments based on the current infrastructure
  - are expected to have the potential to increase the use and utility of data from the INSPIRE service network
Implementing the Best Practices on top of INSPIRE

Key practices implemented in the proxy:

- **Best Practice 1**: Use globally unique persistent HTTP URIs for Spatial Things
- **Best Practice 2**: Make your spatial data indexable by search engines
- **Best Practice 4**: Use spatial data encodings that match your target audience
- **Best Practice 5**: Provide geometries on the Web in a usable way
- **Best Practice 12**: Expose spatial data through 'convenience APIs'
Idproxy

- An open-source tool developed to experiment with implementing the Best Practices

- Originally developed in a Geonovum testbed
  - [http://geo4web-testbed.github.io/topic4/](http://geo4web-testbed.github.io/topic4/)

- Documentation
Idproxy Demo

- Landing pages of INSPIRE datasets (DE: GN, NL: PS)
- Browse a dataset
- Show schema.org in Google Structured Data Testing Tool
- Use spatial and thematic filters
- Get feature also as GML, GeoJSON, JSON-LD
- Show API definition and test API
- Code generation from API definition
Impact on OGC and ISO/TC 211

Revision of WFS in progress in OGC and ISO based on the Best Practices, plus:

- Developer-driven process
- Open to everyone
- Modular approach, simple core
- API building blocks rather than monolithic services
- Support for OpenAPI as an alternative to capabilities
- No mandatory encoding
- No schema required
- Security considered in the design

https://github.com/opengeospatial/WFS_FES
Related OGC Innovation Program initiatives

Next Generation OGC RESTful Web service interfaces
Complex Feature Handling
WFS 3.0 Compliance