Linking and Finding Earth Observation (EO) Data on the Web

MACS-G20 Workshop: Linked Open Data in Agriculture
Berlin, September 27-28, 2017

Dr. Uwe Voges
u.voges@conterra.de
Introduction

- Earth Observation (EO) Datasets (incl derived Products) play an important role for decision making processes
  > New satellite programs recently setup: EU/ESA (Copernicus), EUMETSAT (Meteosat, MetOp)…
  > worldwide observations with high temporal repetition rates -> real Big Data!

- EO Datasets published as Open Data - but often through special Portals - stored in Archives
- For users not knowing about the different portals it is hard to find, discover and get access to it
Introduction

- Definition: EO Collections represent set of EO Datasets sharing a common specification

- For usage of product it can either be downloaded...
  > Requires: register, order, download, process (format conversions, re-projections,...) the data

- OR - for an easier access - data is provided by Access and View Services, e.g. OGC WMS

- Collections, Products and Services are usually described and indexed based on specific metadata models and encodings (e.g. ISO 19115 or 19119)

EUMETSAT SEVIRI data & Forest Fires WMS’ used with Esri ArcMap
Introduction – Linked Open Data

- Metadata indexed and searched via specific Catalogue Services (e.g. OGC CSW)
  - to Google & Co: metadata hidden
    - Require specific clients
  - Access to metadata just following web links is mostly not possible.

- For improving the situation: technologies related to Linked Open Data are promising
  - starting at a piece of Linked Data, and follow links to other pieces
  - link EO data and mainstream web data
  - Can be used on top of existing SDI
Linked Open Data for Earth Observation

- **Basic technologies and best practices for LOD in EO:**
  - unique persistent HTTP URIs for EO Datasets
  - provision of EO data (upon de-referencing) in different formats
    - ATOM / (Geo)JSON metadata
      - using well known (e.g. GeoNames) or well-designed vocabularies
    - HTML page
    - binary data
    - maps in tiled form suitable for display in zoomable maps (WMTS)
  - Link EO data to other resources to integrate with the web of data.
Searching with mainstream Search Engines

- Alternative to getting to data via links: searching with Google & Co
  - Search engines index web pages (HTML)
    - should provide metadata and links to service endpoints
    - schema.org annotated: to improve search results /SEO
  - landing page: to enable crawlers to find web pages
    - Set of EO collections as landing page:
      - point to EO datasets
        - paging should be provided
    - Set of links to web pages can also be added to a sitemap

From: EUMETSAT Product Navigator
From: ESA Sentinel Data Hub
Searching with mainstream Search Engines

- But in case of Mio of similar datasets (new permanently coming in) … approach is no more an option:
  
  > not possible to force Google to index data as needed
  > indexing takes too long: months / years or not possible…
  > Domain specific vocabulars to index/search not supported

- Approach:
  
  > let just collections be indexed by Google
  > use **specialized search engine** for indexing / filtering Datasets
OpenSearch

- Possible Solution: OpenSearch (OS)
  - REST-Style, web compatible API
  - providing key-value parameters to constrain the search
  - OS engines provide description documents (OSDD) used by clients
    - include URL template per response format with parameters to be replaced by client
    - usable by browser plugIns via Autodiscovery

    <link rel="search" type="application/opensearchdescription+xml" href="http://...osdd.xml" title="EUMETSAT Product Navigator..." />

  - ...or by Search clients which can easily be created from OSDD
OpenSearch – Responses

> Results: can be returned as HTML, Atom, RDF, KML, JSON....

> Resultset: container with search entries

> Container includes information about current search and means for pagination

> Entries represent EO Datasets
  - element values mapped from original metadata
  - links provide means for linking to external information, e.g. to metadata details (@rel="via")

> Results can be aggregated by clients and e.g. rendered in browser
OpenSearch – OGC Extensions for Geo/EO

- OpenSearch is extendable:
  - by query parameters (in addition to those provided in OS 1.1 namespace, e.g. searchTerm)
  - by additional search response formats

- OGC 10-032r8: OGC OpenSearch Geo and Time Extensions
  - parameters and operators to define geographic and temporal constraints
    - Example URL template: http://server...?q={searchTerms}&g={geo:geometry?}&r={geo:relation}
    - Example request: http://...?q=wind&g=POLYGON((0.582 40.496, ...., 0.582 40.496))&r=disjoint
  - Atom response format defining constraints for the elements
OpenSearch – OGC Extensions for Geo/EO

- OGC 13-026r9: OGC OpenSearch Extension for Earth Observation (OpenSearch-EO)
  > Search for EO Collections and EO Datasets  (build upon 10-032 - last slide)

  > EO specific parameters, e.g. platform, orbitType, sensorType, cloudCover, ...
  > semantic defined by mappings to EO MD models (e.g. OGC 10-157)
  > Defines Atom response format with constraints - plus:
    > links with defined semantic (“rel” attribute):
      > e.g. 2-step-search, source of entry, alternate MD formats, docu, images, ...
    > service or online content offering for entries targeting OGC clients

  > Used by Aggregators: CEOS/CWIC, ESA/FEDEO, EUM/EOP, NASA...
GeoJSON / JSON-LD

- Problem: Atom not ideal for web clients and developers, poor mass market adoption
- Better: GeoJSON
  > JSON: text-based, lightweight, key/value pairs, derived from JavaScript
  > GeoJSON: format to encode simple geogr. features using JSON
    > supported by many js-libraries
  > But: No commonly agreed vocabulary and not LD

- Solution: JSON-LD (JSON-based format to serialize Linked Data)
  > way to combine web-based environments and linked data
  > implements: 5 star SDW-BP
GeoJSON / JSON-LD

> In addition to JSON, JSON-LD introduces:

> Semantically enriched properties => identified by an IRI

> defines the vocabulary

> @id keyword: refer to JSON object on different site by IRI

> JSON-LD introduces a @context

> addresses issue of verbose documents by mapping terms to IRIs

> a way to associate datatypes with values (e.g. dateTime)

> serializable as RDF: can be used with SPARQL

> next generation REST for hypermedia driven systems (HATEOAS)
OpenSearch – OGC GeoJSON / JSON-LD response formats for EO

- **OGC 17-074**: OGC OpenSearch-EO GeoJSON (-LD) Response Encoding Standard
- **OGC 17-003**: OGC EO Dataset Metadata GeoJSON(-LD) Encoding Standard
  > both define GeoJSON encoding as compaction through a context of the proposed JSON-LD encoding
  > define properties with domain, range, namespace,…
  > widely adopted vocabularies used DC, OGC, Atom,…
  > **17-003**: EO Vocabulary defined with RDF Schema:
OpenSearch – Outlook

- Problems with OpenSearch OSDD…
  - XML based
  - URL template not structured, string based
  - Limited options especially for response definitions
- Need for a better concept for a service description:
  - Better aligned with web clients and developers
  - structured and more flexible
  - existing tool support
- Possible Solution: OpenAPI 3.0
  - Some points open: e.g. how to link parameters to predefined EO parameter definitions…
Thank you!