Linking from Earth Observation Data and Products to executable web-based Algorithms based on Metadata

Jonas Eberle, PD Dr. Soeren Hese, Prof. Dr. Christiane Schmullius

Department for Earth Observation
Institute of Geography
University of Jena
Germany
Siberian Earth System Science Cluster (SIB-ESS-C)

• Siberian
  – focus area: Siberia
  – Earth observation products from several research projects

• Earth System Science
  – Not only focussing on a specific topic
  – Interdisciplinary research products

• Cluster
  – Computing Infrastructure
  – Data Processing and Distribution based on open standards

www.sibessc.uni-jena.de
SIB-ESS-C: Current infrastructure

SIB-ESS-C Web Portal

Data Discovery
- Gl-cat
- Federated Catalogue

Data Access
- GeoServer

Data Products
- Collection of derived data products

Data Analysis
- PyWPS
- Processing Services

Processing Back-end
- R Statistical Toolbox

Metadata
- ISO 19115/19139

www.sibessc.uni-jena.de
Data Discovery & Access

CLIENT TIER

SIB-ESS-C Web Portal

External Client

SERVICE TIER

GI-cat: Federated Catalogue and Broker

SIB-ESS-C Accessor

XML File Accessor

External Resources
(e.g. WCS, WFS, CSW, OPeNDAP, THREDDS)

METADATA TIER

Metadata DB
ISO 19115 PostgreSQL

XML Metadata
ISO 19139

www.sibessc.uni-jena.de
Data Processing & Analysis

CLIENT TIER
SIB-ESS-C Web Portal / External Client

SERVICE TIER
PyWPS

WPS
getCapabilities
describeProcess
Execute

WCS
getCapabilities
describeCoverage
getCoverage

WFS
getCapabilities
describeFeature
getFeature

BACKEND TIER
R

Process 1
Process 2
Process ...

www.sibessc.uni-jena.de
Linking Data and Algorithms
Linking Data and Algorithms

- Connection between Data, Algorithms and Products
- Each algorithm should know which data can be used as input
- Automatic descriptions of process chains
## Describing Data with ISO 19115: Basic information

<table>
<thead>
<tr>
<th>Data Information</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>Contact persons</td>
</tr>
<tr>
<td>Extent</td>
<td>Title</td>
</tr>
<tr>
<td>Abstract</td>
<td>Publication date</td>
</tr>
<tr>
<td>Purpose</td>
<td></td>
</tr>
<tr>
<td>Data Format</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Lineage steps</th>
</tr>
</thead>
<tbody>
<tr>
<td>View-Service (WMS)</td>
<td>Raw input data</td>
</tr>
<tr>
<td>Download-Service (WFS, WCS)</td>
<td>Description of applied algorithms</td>
</tr>
<tr>
<td>View and Download restrictions</td>
<td></td>
</tr>
</tbody>
</table>
Describing Data with ISO 19115: Detailed information

<table>
<thead>
<tr>
<th>Content information</th>
<th>MD_CoverageDescription (MD_ContentInformation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band Information</td>
<td>• <strong>contentType</strong> image</td>
</tr>
<tr>
<td>Cloud Cover Percentage</td>
<td>• <strong>MD_ImageDescription</strong></td>
</tr>
<tr>
<td>Coverage Content Type</td>
<td>• <strong>cloudCoverPercentage</strong> 30.0</td>
</tr>
<tr>
<td>Imaging Condition</td>
<td>• <strong>imagingCondition</strong> cloud</td>
</tr>
<tr>
<td>Processing Level Code</td>
<td>• <strong>processingLevelCode</strong></td>
</tr>
<tr>
<td></td>
<td>• <strong>imageQualityCode</strong> Quality information</td>
</tr>
<tr>
<td></td>
<td>• <strong>MD_RangeDimension</strong></td>
</tr>
<tr>
<td></td>
<td>• <strong>MD_Band</strong></td>
</tr>
<tr>
<td></td>
<td>• <strong>descriptor</strong> RED</td>
</tr>
<tr>
<td></td>
<td>• <strong>maxValue</strong> 0.69</td>
</tr>
<tr>
<td></td>
<td>• <strong>minValue</strong> 0.63</td>
</tr>
<tr>
<td></td>
<td>• <strong>Units</strong> μm</td>
</tr>
<tr>
<td></td>
<td>• <strong>scaleFactor</strong> 1.0</td>
</tr>
</tbody>
</table>
Describing Coverage Data with ISO 19115-2

<table>
<thead>
<tr>
<th>Acquisition information</th>
<th>Processing information</th>
<th>Algorithm information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirements</td>
<td>Software Reference</td>
<td>Citation</td>
</tr>
<tr>
<td>Conditions</td>
<td>Steps / Levels</td>
<td>Description</td>
</tr>
<tr>
<td>Platform</td>
<td>Runtime Parameters</td>
<td></td>
</tr>
</tbody>
</table>

**MI_AcquisitionInformation**
- environmentalConditions
  - MI_EnvironmentalRecord
    - AverageAirTemperature
    - maxRelativeHumidity
    - maxAltitude
    - meteorologicalConditions
- Platform
  - MI_Platform
    - Identifier
      - Aqua
    - Description
- Instrument
  - MI_Instrument
    - Identifier
      - MODIS
    - Type
    - description
### Describing Coverage Data with ISO 19115-2

<table>
<thead>
<tr>
<th>Acquisition information</th>
<th>LE_ProcessStep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requirements</td>
<td>• output</td>
</tr>
<tr>
<td>Conditions</td>
<td>• LE_Source</td>
</tr>
<tr>
<td>Platform</td>
<td>• processedLevel</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>• resolution</td>
</tr>
<tr>
<td></td>
<td>1km</td>
</tr>
<tr>
<td>Processing information</td>
<td>• report</td>
</tr>
<tr>
<td>Software Reference</td>
<td>• LE_ProcessStepReport</td>
</tr>
<tr>
<td>Steps / Levels</td>
<td>• name</td>
</tr>
<tr>
<td>Runtime Parameters</td>
<td>MOD14 Logfile</td>
</tr>
<tr>
<td></td>
<td>• description</td>
</tr>
<tr>
<td></td>
<td>MODIS MOD14 SPA</td>
</tr>
<tr>
<td></td>
<td>• fileType</td>
</tr>
<tr>
<td></td>
<td>ascii</td>
</tr>
<tr>
<td>Algorithm information</td>
<td>• processingInformation</td>
</tr>
<tr>
<td>Citation</td>
<td>• LE_Processing</td>
</tr>
<tr>
<td>Description</td>
<td>• identifier</td>
</tr>
<tr>
<td></td>
<td>MOD14</td>
</tr>
<tr>
<td></td>
<td>• softwareReference</td>
</tr>
<tr>
<td></td>
<td>MODIS MOD14 SPA</td>
</tr>
<tr>
<td></td>
<td>• runTimeParameters</td>
</tr>
<tr>
<td></td>
<td>• algorithm</td>
</tr>
<tr>
<td></td>
<td>• LE_Algorithm</td>
</tr>
<tr>
<td></td>
<td>• citation</td>
</tr>
<tr>
<td></td>
<td>Justice et al., 2006</td>
</tr>
<tr>
<td></td>
<td>• description</td>
</tr>
<tr>
<td></td>
<td>Modis Fire Products</td>
</tr>
</tbody>
</table>
Metadata: Describing Algorithms

• OGC Web Processing Service (WPS)
  – Service-based processing as OGC specification
  – Metadata-Elements:
    • Title
    • Abstract / Description
    • Data type / format
    • Further Metadata can be specified as X-Link-URL
Metadata: Describing Algorithms

- **Example Input-Description within DescribeProcess response**

```
<Input minOccurs="1" maxOccurs="1200">
  <ows:Identifier>wcs_input1</ows:Identifier>
  <ows:Title>WCS Input 1</ows:Title>
  <ows:Abstract>
    For a single WCS request the following parameters are required: WCS version=1.1.1, BoundingBox=xmin, ymin, xmax, ymax, urn:ogc:def:crs:OGC:1.3:CRS84, format=image/tiff;subtype="geotiff", GridBaseCRS=EPSG:4326. A sample WCS request encoding would look like: http://localhost:8080/geoserver/ows?service=WCS&version=1.1.1&request=GetCoverage&BoundingBox=111,69,116,74,urn:3Aogc:3Adef:3
  </ows:Abstract>
  <ows:Metadata about="Description" xlink:title="Title of linked file" xlink:href="http://argon.geogr.uni-jena.de/wcs_input1.xml"/>
  <ComplexData>
    <Default>
      <Format>
        <ows:MimeType>multipart/mixed</ows:MimeType>
      </Format>
    </Default>
    <Supported>
      <Format>
        <ows:MimeType>multipart/mixed</ows:MimeType>
      </Format>
    </Supported>
  </ComplexData>
</Input>
```
Metadata: Describing Algorithms (NDVI calculation)

- **Input RED Band:**
  - MD_CoverageDescription
  - contentType: image
  - MD_RangeDimension
    - MD_Band
      - descriptor: RED
      - maxValue: 0.69
      - minValue: 0.63
      - Units: µm

- **Input NIR Band:**
  - MD_CoverageDescription
  - contentType: image
  - MD_RangeDimension
    - MD_Band
      - descriptor: NIR
      - maxValue: 0.90
      - minValue: 0.76
      - Units: µm

- **Output NDVI Index:**
  - MD_CoverageDescription
  - contentType: thematicClassification
  - MD_RangeDimension
    - MD_Band
      - descriptor: NDVI
      - maxValue: 1.0
      - minValue: 0.0
Matching Data and Algorithms

• Matching:
  – Metadata for each dataset
  – Metadata for each algorithm and their in- and outputs
  – Both Metadata are described with the same elements (ISO 19115)

• Executing algorithm:
  – Must prove conditions (same resolution, same acquisition date)
  – Specific conditions based on algorithm
Summary & Outlook

• Summary:
  – We need precise Metadata for data
  – We need additional Metadata for each input of an algorithm to reach the same level of preciseness
  – Outputs of an algorithm have to be also described for further tasks

• Outlook:
  – Create further test cases for new algorithms with specific data requirements
  – Define Metadata for different data types
  – Implementation of matching within SIB-ESS-C infrastructure
Thank you for your attention!

Questions? Visit me at the Poster Session today!

Siberian Earth System Science Cluster - A web-based Geoportal to provide user-friendly Earth Observation Products for supporting NEESPI scientists

Tuesday, 17.30 – 19.00

EGU2012-8350
Poster Area BG
BG65

Contact information:

Jonas Eberle
Friedrich-Schiller-University
Institute of Geography
Department for Earth Observation
Loebdergraben 32
07743 Jena, Germany

phone: +49 3641 94 88 89
email: jonas.eberle@uni-jena.de
http://www.eo.uni-jena.de/7020.0.html