Presenting Quality Information: From Dataset Quality to Individual Sample.

INSPIRE Workshops.
GeoViQua project

Lucy Bastin (ASTON), Eva Klien (Fraunhofer),
Lorenzo Bigagli (CNR), Joan Masó (CREAF)

Room: 4
Sunday 23rd June
16:00–18:00
GeoViQua provides a set of scientifically developed software components and services that facilitate the **creation**, **search** and **visualization** of quality information on EO data integrated and validated in the GEOSS Common Infrastructure.
• Information about data quality is regularly created by official producers of geospatial data and by scientists who process and derive new datasets. Unfortunately this information takes a number of different forms (e.g., quality reports, scientific papers, informal comments...) whose formats are inhomogeneous, making comparison difficult. Current or extended standards for data quality descriptions (ISO19115, ISO19157, etc) can be used or developed to define ‘quality indicators’, including quality measures and provenance parameters but these become too verbose to represent sample-level quality.

• **More tools are needed.** From the producer side, there is a need to simplify the creation of quality descriptions without compromising detail. From the user side, the identification of datasets that fit a user’s purpose can be achieved by allowing metadata inter-comparison, quality labels, data discovery using quality thresholds refinement of search results and common visualization techniques. Additionally, user experiences need to be collected in a structured way so that comments, citations, discovered issues and ratings are captured, aggregated and exposed.

• This workshop will discuss a INSPIRE perspective in data quality by analyzing the outcomes of the GeoViQua FP7 project which aims to add rigorous data quality representations to existing search and visualization in the GEO Portal functionalities of the Global Earth Observation System of Systems (GEOSS)

• Conclusions of the workshop will serve to improve quality descriptions in both the INSPIRE and GEOSS initiatives.
We are presenting:

- **A quality framework** that enhances producer metadata, and proposes the addition of user feedback. The producer model builds on existing ISO standards (19115 and 19157) adding reference dataset information, citations, traceability of quality statements and discovered issues. The user model informs the database structure for a feedback server from which comments, citations, discovered issues, ratings and reports of usage may be stored and retrieved.

- **A quality-aware discovery service**, namely a quality-aware extension of the OGC Catalog Service for the Web (CSW-Q), which could cope with quality-constrained search. This will be included in the GEOSS Discovery and Access Broker.

- A standards-based approach for the **visualization of quality / uncertainty information** in 2D, developed using the OGC Web Map Service (WMS), reusing concepts in UncertML.

- A **GEO Label** as a graphical representation of a dataset in the GEOSS (or other data portals and clearinghouses) based on the quality information that is available for that dataset.

- A **user feedback catalogue** where users can introduce comments, citations, discovered issues, ratings and reports of usage. This information can then be retrieved by the Discovery and Access Broker.

- **Some enhancements in metadata presentation** such as: metadata side by side comparison; rubric metadata completeness assessment; provenance visualization, etc.

Agenda

• 16:05 Introduction Joan Masó
• 16:20 Quality models Lucy Bastin
• 16:35 Producer Quality Model Lucy Bastin
• 16:50 User Feedback Model Lucy Bastin
• 17:00 KML-Q Eva Klien
• 17:10 WMS-Q Joan Masó
• 17:20 GeoLabel Lucy Bastin
• 17:30 GeoLabel demo Lucy Bastin
• 17:40 Discovery and Access Broker Lorenzo Bigagli
Resources

- http://twiki.geoviqua.org/twiki/bin/view/GeoViQua/INSPIRE2013Workshop

- Inspire tutorial for Producer Quality in ISO, User feedback and GeoLabel
  - http://uncertgeo.aston.ac.uk/INSPIREtutorial

- GeoLabel generation system
  - http://www.geolabel.net/geolabel.html

- WMS-Q integrated client
  - http://www.ogc.uab.cat/geoviqua/wmsq

- DAB-Q demo client
Introduction to GeoViQua activities and results

Joan Masó
16:05
The context: GEOSS versus INSPIRE

• Same basic services and architecture
  – Same metadata standards and catalogues
  – Web services for viewing and accessing and transforming

• Different governance
  – INSPIRE is a European directive
    • Services and INSPIRE profiles will be followed by member states so services will be more that interoperable: they will function in a plug’n play mode.
    • Data topics list is well defined and shares the same data model
  – GEOSS is global voluntary effort
    • Data provides contribute what the have
    • Brokering efforts are done to harmonize differences in services and models

• More details about this differences in the GIGAS project
Quality models

- From requirements process and user interviews

Producer quality

- ISO 19115 19157
- Small Extensions
- Dataset level in metadata
- Pixel level with UncertML

Consumer quality

- User feedback
- More flexible
- Not present in GEOSS
- New model
Embedding quality in all phases of the geospatial data exploitation

**Discover**
- **DAB-Q**: Query with quality information

**Evaluate**
- **WMS-Q**: Extension to relate data with spatialized quality layers
- **GEOLabel**: Graphical picture that helps to evaluate data

**Access**
- **WCS-Q**: GMLCov extension for spatialized quality layers (in project)

**Use**
- **NetCDF-U**: A way to supply spacialized quality layers (UncertWeb)

Presenting Quality Information: From Dataset Quality to Individual Sample.
Presenting Quality Information: From Dataset Quality to Individual Sample.
Provenance visualization (2/2)

- Process:
  - ps_228: The Twice-Weekly 50 km Nighttime SST Anomaly is derived from the Twice-Weekly 50 km Nighttime SST analysis and 50 km nighttime SST climatologies. The SST anomaly is calculated by using the daily SST climatology for the last day of the twice-weekly period against the twice-weekly SST analysis. To obtain the SST climatology for a specific date, the linear interpolation method is applied on the two SST monthly mean climatologies that are closest to that date, with the assumption that a monthly mean SST climatology is more accurate on the 15th of the month. The linear interpolation is done on the number of days away from the 15th of the first month’s climatology. The SST anomaly is calculated by subtracting the interpolated daily climatology from the SST. The SST anomaly data are updated twice a week as the twice-weekly SST analysis.

  - Process Characteristics:

- Process:
  - ps_238: The Coral Bleaching HotSpot is the excursion of SST above the Maximum Monthly Mean SST Climatology. At each pixel, the coral bleaching hotspots value is calculated by subtracting the climatology value from the SST value and then setting all the negative values to zero. The HotSpot data are updated twice a week as the twice-weekly SST analysis.

  - Process Characteristics:

- Process:
  - ps_231: The Twice-Weekly Coral Bleaching Degree Heating Weeks (DHW) value at each pixel is calculated as the accumulation of Coral Bleaching HotSpots values that is at least one degree Celsius, over the most recent 12 weeks up to and including the current twice-weekly HotSpots. There are totally 24 consecutive twice-weekly HotSpot data files, covering a consecutive 12-week period, included in each DHW calculation. The DHW data are updated twice a week as the twice-weekly SST analysis.

  - Process Characteristics:

Mission Start

- Mission Completion

- Process Characteristics:

Processor:

- Organization Name: [Organization Name]
- Individual Names: [Individual Names]
- Role: [Role]
- Contact: [Contact Information]
- Processing Identifier: [Processing Identifier]
- Software Reference: [Software Reference]
- Run Time Parameters: [Run Time Parameters]
- Date of: [Date of]
- Procedure Description: [Procedure Description]
- Algorithms: [Algorithms]

Source:

(1978-10-13T00:00:00) TIROS-N launched. Instruments: AVHRR and TOVS.
### Metadata comparison. GeoViQua Project

#### Bounding Box

![Bounding Box Maps](image)

<table>
<thead>
<tr>
<th>Identification</th>
<th>Dataset 1</th>
<th>Dataset 2</th>
<th>Dataset 3</th>
<th>Dataset 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title</strong></td>
<td>Digitale Topographische Karte 1:10 000 - Vorläufige Ausgabe 4150-NW Leipzig</td>
<td>Digitale Topographische Karte 1:10 000 - Vorläufige Ausgabe 3452-NW Gneisenau</td>
<td>Hagerman Fossil Beds NM Administrative Boundaries</td>
<td>Administrative Basins 01-03</td>
</tr>
<tr>
<td><strong>Abstract</strong></td>
<td>Die vorläufige Ausgabe besteht aus georeferenzierten Rasterdaten der gesamten Topographischen Karte 1:10 000 (TK10). Zunächst soll die DTK aus dem Basis-DLM abgeleitet werden. Die Daten sind in Ebenen (Einzellagen, im wesentlichen den Druckfarben entsprechend) oder kombiniert aus diesen als Summenlayer verfügbar. Sie liegen in 5 Inhaltsbezeigen: Situation schwarz, Waldflächen, Gewässerflächen mit Gewässerschiff, Gewässerläufe, Relief</td>
<td>Die vorläufige Ausgabe besteht aus georeferenzierten Rasterdaten der gesamten Topographischen Karte 1:10 000 (TK10). Zunächst soll die DTK aus dem Basis-DLM abgeleitet werden. Die Daten sind in Ebenen (Einzellagen, im wesentlichen den Druckfarben entsprechend) oder kombiniert aus diesen als Summenlayer verfügbar. Sie liegen in 5 Inhaltsbezeigen: Situation schwarz, Waldflächen, Gewässerflächen mit Gewässerschiff, Gewässerläufe, Relief</td>
<td>Hagerman quadrangle 1:24k Digital Line Graph administrative boundaries.</td>
<td>Idaho Department of Water Resources administrative basin boundaries - basin 01-03. Shapefile should be used as a guide since basins 01-03 are defined as diversions from the Snake River along specific reaches. Basin 01 is diversions from the Snake River (Eastern Idaho Border) to Milner Dam. Basin 02 is from Milner Dam to the Oregon border. Basin 03 is from where the Snake River reenters the</td>
</tr>
</tbody>
</table>
ISO 19115 Completeness Report with Quality extension

This report identifies ISO 19115 metadata elements that are documented in a specific metadata record. The Robo-CQ version is based on the original NOAA stylelint developed by NIES. Some differences with NOAA Robo-CQ Quality Extensions and that some restrictions are not used as they are not ISO mandates but NOAA recommendations (the attribute "sal" equal to "boundingBag" for extent elements; specific keyword types not included in ISO code list as "dataCenter" or "project", etc.). Additionally, to the elements tested on the original Robo-CQ, the Robo-CQ extends the report to include Quality information, as one of the main aims of the FP GeoViQua (ENV2019 4.1.2.2 nr. 203179) is to add rigorous quality specifications to the Global Earth Observation System of Systems (GEOSS) spatial data in order to improve reliability studies and policy decision making. The ISO 19115 Standard recommends Core Elements for inclusion as metadata. This tool tests also conformance with those recommendations. The structure of the Robo-CQ report presents at the top of the report the summary of the results. Each spiral is represented by a row in the rubric. The columns show the % of the elements that are documented in the record. Click the spiral name for more details. Moreover, a second section shows the quality metadata information included on the metadata description of the elements.

This report is produced using this stylelint. Please contact GeoViQua team if you have questions or suggestions. The original NOAA stylelint that is the base for this report is here.

Title: NOAA Coral Reef Watch Operational Twice-Weekly Near-Real-Time Global 50-km Satellite Nighttime Sea Surface Temperature Anomaly Product

Resource URL:
http://coralreefwatch.nco.nos.noaa.gov/satellite/current_products.html
http://data.nodc.noaa.gov/satellite/satellite_current_product_summary.html
http://data.nodc.noaa.gov/crs/935/stn/sat_repos.html
http://www.esrl.noaa.gov/psd/nws/satellite/virtual_stations/index.html
http://coralreefwatch.nco.nos.noaa.gov/satellite/geo_qua.html

Total Spiral Score: 21/39 (grid)
Summary of the results. Each spiral is represented by a row in the rubric. The columns show the % of the elements that are documented in the record. Click the spiral name for details.

<table>
<thead>
<tr>
<th>Spiral</th>
<th>None</th>
<th>1-33%</th>
<th>34-66%</th>
<th>67-99%</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Spiral</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connection</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distribution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lineage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Total ISO Core Score: 18/22

Note: The Total ISO Core Score does not count toward the Total Spiral Score

<table>
<thead>
<tr>
<th>Spiral</th>
<th>None</th>
<th>1-33%</th>
<th>34-66%</th>
<th>67-99%</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Core</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core Mandatory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core Conditional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core Optional</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Additional Information

This report identifies ISO metadata elements described as spirals of documentation development described as Creating Good Documentation. The elements are listed by name and are followed by M, C, or O if they are Mandatory, Conditional or Optional. They are followed by UDD (attribute name) if they are included in the NetCDF Attribute Conventions for Dataset Discovery.

Identification Score: 8/9
The Identification Spiral sets the stage for discovery using text search engines. It includes a unique identifier for the metadata, a title, an abstract, theme keywords and contact information for the metadata and the data set.

<table>
<thead>
<tr>
<th>Score</th>
<th>Attribute (Count)</th>
<th>Description</th>
<th>Best Practices</th>
<th>Path</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Metadata Identifier: UDD title</td>
<td>A unique phrase or string which uniquely identifies the metadata title. Each metadata record shall include a character string as a unique identifier. There are two general approaches to ensuring uniqueness for these identifiers:</td>
<td>use a universal unique identifier (UUD) to distinguish it from other resources. For example: go帙urn; class: AERO10. In this case go帙urn; class is a namespace and AERO100 is a code guaranteed to be unique, in that namespace. It seems likely that the upcoming revision of ISO 19115 will support MD_ Identifiers as metadata identifiers. More...</td>
<td>geo:MD_Metadata/ood/fileIdentifier</td>
</tr>
</tbody>
</table>
INSPIRE 2013: THE GREEN RENAISSANCE

Multilayer product: new uncertainty and quality indicators

Mean: 56%  Range: 7-92%

Pixel level

Cat1-Classification  Probability of success (%)  Second Classification  Third Classification  Fidelity

Second Fidelity  Third Fidelity  Promiscuity  Majority classes  H Entropy  Uncertainty

F, F2, F3, R, R2, R3, Probability of success: %
P, Q: number of thematic categories
I: fraction of unity (0/1)
Comparative mode of Carbon Fluxes Data

- **Synchronized maps comparison**
- **Currently, 11 models (atmospheric inversions): recompilation by CEA-LSCE.**

More information in the future.
GEOSS GeoPortal Integration
Quality models
Producer Quality Model
User Feedback Model

Lucy Bastin
16:20
Theory and practice!

• Assessing fitness-for-purpose – what is missing?

• GeoViQua models and tools:

  • The producer quality model
    – Build and publish your own enriched document

  • The user quality model
    – Try out our feedback server

• The GEO label
  – Summarising and interrogating rich metadata
Fitness-for-use... What do people want?

- peer and expert review
- better traceability and provenance information
- information on citations and usage of a dataset
- warnings about problems identified with a dataset and potential workarounds,
- ‘soft knowledge’ from data producers (e.g. recommendations for use which are not easily encoded using the existing standards)

Chrisman, 1988: asked for ‘user experience’ to be included in data quality specs
Epstein et al. 1998: disclaimers only cover ‘reasonably foreseeable use’
Comber et al., 2006: identify mismatches between producer/user ontologies
Devillers et al., 2007: tools needed to help users understand quality information.
Boin & Hunter, 2007: users want a simple summary that they can interrogate for more detail as appropriate.
Lush et al., 2011: Interviews identified that these needs are still not being met.
How can we deliver it?

**Producer quality model**
- ISO 19115 / 19157
- Extensions to ISO (adopted by GEOSS):
  - Citations
  - Reference datasets
  - Traceability
  - Discovered Issues
  - Pixel and dataset level – UncertML
  - INSPIRE-recommended identifiers and codelist

**User Quality model**
- User feedback
- More flexible
- New model - not present in GEOSS
- Ratings
- Examples of use in named application domains
- ‘Soft knowledge’ from producers
Presenting Quality Information: From Dataset Quality to Individual Sample.

http://schemas.geoviqua.org/GVQ/4.0/UML/Producermodel.jpg
Presenting Quality Information: From Dataset Quality to Individual Sample.
Traceability
### Traceability

#### Traceability of this quality report

**Evaluation method:** 1. Analysis published in a peer-reviewed journal article with full description of methodology. 2. Original data and reference dataset published via FAO and JRC.

**Statement:** The thematic validation procedure was carried out using design-based inference which conforms to the CEOS Cal-Val recommendations. The resulting research was published in a peer-reviewed journal, and the component datasets generated for the validation may be obtained from JRC.

**Process step 1**

**Description:** Stratified sampling of ground truth blocks.

**Rationale:** Systematic sampling to obtain an irregular stratification with different sampling rates for each stratum.

**Process step 2**

**Description:** Interpretation by regional experts of 3x3km sample boxes at specified sample sites. Regional interpreters used ancillary data like aerial photographs, thematic maps and NDVI profiles at coarse resolution in support to the Landsat interpretation.

**Rationale:** Local knowledge required to verify exact landcovers at each ground truth point.

**Process step 3**

**Description:** Cross-validation of independently interpreted sites against GLC2000 data to generate confusion matrix and Kappa statistic.

**Rationale:** Identification of classes prone to confusion, and overall assessment of the accuracy of the classification.
Discovered issues

Problem: Legend issues with South East Asia

<table>
<thead>
<tr>
<th>Workaround</th>
<th>Legend mapping available from FAO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alternative Dataset</td>
<td><strong>Title:</strong> South East Asia legend mapping: regional classes to GLC2000</td>
</tr>
<tr>
<td></td>
<td><strong>Identifier:</strong> jrc.europa.eu: GLC2000_SEA_mapping</td>
</tr>
<tr>
<td>Alternative Dataset Abstract</td>
<td>This table gives a full hierarchical mapping from all regional SEA landcover categories to those used in GLC2000</td>
</tr>
</tbody>
</table>
Discovered issues

This table gives a full hierarchical mapping from all regional SEA landcover categories to those used in GLC2000.
Reference datasets
Reference datasets

<table>
<thead>
<tr>
<th>Report Type: Thematic Classification Correctness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evaluation Method</strong></td>
</tr>
<tr>
<td><strong>Evaluation Description</strong></td>
</tr>
<tr>
<td><strong>Reference document</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Reference Dataset</strong></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

A two-part unique identifier: codespace and code = the newest version of MD_Identifier
Presenting Quality Information: From Dataset Quality to Individual Sample.

Full statistical summaries: UncertML

- <gmd:DQ_QuantitativeAttributeAccuracy>
  - <gmd:result>
    - <gmd:DQ_QuantitativeResult>
      - <gmd:valueType>
        - <gco:RecordType xlink:href="http://www.uncertml.org/distributions/normal">Value of the vertical DEM accuracy</gco:RecordType>
      - <gmd:valueUnit>m</gmd:valueUnit>
      - <gmd:value>
        - <gco:Record>
          - <un:NormalDistribution>
            - <un:mean>1.2</un:mean>
            - <un:variance>3.6</un:variance>
          </un:NormalDistribution>
        </gco:Record>
    </gmd:value>
  </gmd:DQ_QuantitativeResult>
</gmd:result>
</gmd:DQ_QuantitativeAttributeAccuracy>

- 

- Explicit recognition that errors acceptably fit a Normal distribution with mean 1.2
- An overall positive bias was observed (a difficult feature to convey by traditional means)

30
<table>
<thead>
<tr>
<th>Result Scope</th>
<th>Level: dataset</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Result Type</th>
<th>Confusion matrix recording counts from ground truthing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>558 0 7 0 0 0 0 0 0 10 0 0 0 0 0 10 0 0 0 0 0</td>
</tr>
<tr>
<td>2</td>
<td>0 104 0 0 0 0 0 0 0 4 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>3</td>
<td>19 56 19 0 0 0 0 0 0 21 0 10 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>4</td>
<td>0 14 0 174 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>5</td>
<td>0 0 0 0 10 19 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>6</td>
<td>7 50 0 111 10 113 0 0 0 4 10 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>7</td>
<td>27 0 0 7 0 0 14 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>8</td>
<td>10 0 0 0 0 0 10 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>9</td>
<td>0 0 0 0 0 0 0 14 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>10</td>
<td>10 0 0 0 0 0 0 0 10 0 35 14 45 0 16 0 0 0 0 0</td>
</tr>
<tr>
<td>11</td>
<td>0 23 52 7 0 0 0 0 0 27 0 109 82 95 10 19 0 0 4 0 0</td>
</tr>
<tr>
<td>12</td>
<td>0 10 7 0 0 0 0 0 0 21 0 44 112 0 0 19 0 31 0 0 0</td>
</tr>
<tr>
<td>13</td>
<td>0 0 0 0 0 0 0 0 0 0 14 62 181 0 14 0 0 22 0 0 0</td>
</tr>
<tr>
<td>14</td>
<td>10 19 0 10 0 0 0 0 0 4 0 33 7 0 10 0 0 0 0 0</td>
</tr>
<tr>
<td>15</td>
<td>0 0 0 11 0 0 0 0 0 0 0 50 7 0 442 10 21 27 10 0 0</td>
</tr>
<tr>
<td>16</td>
<td>19 10 10 23 0 7 0 0 0 0 0 0 0 0 39 87 16 14 0 0 0</td>
</tr>
<tr>
<td>17</td>
<td>0 10 0 0 0 0 0 0 0 0 0 10 0 0 0 31 0 0 0 0 0</td>
</tr>
<tr>
<td>18</td>
<td>0 0 0 0 0 0 0 0 0 0 0 7 30 0 10 0 4 999 0 0 0</td>
</tr>
<tr>
<td>19</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 353 0 0 0</td>
</tr>
<tr>
<td>20</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 18 0 0 0</td>
</tr>
<tr>
<td>21</td>
<td>0 0 0 7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>
Populating the producer model

• Tools such as Geonetwork can be adapted to create, edit and publish GeoViQua-compliant metadata
• Harvested records can be used for discovery that considers quality.
Populating the models: Producers

Reliability diagram for a validation:

Predicted probability of classifying a pixel as the most probable class, against observed frequency that it was the most probable class.

Red dots: bootstrapped samples to provide confidence interval.

Dataset: rice fields
http://uncertgeo.aston.ac.uk/INSPIREtutorial/
User feedback in ISO 19115

MD_Usage
– Brief description of ways in which the resource is currently or has been used
Fitness-for-use…What do people want?

• peer and expert review

• information on citations and usage of a dataset

• warnings about problems identified with a dataset and potential workarounds

• ‘soft knowledge’ from data producers (e.g. recommendations for use which are not easily encoded using the existing standards)
Presenting Quality Information: From Dataset Quality to Individual Sample.

http://schemas.geoviqua.org/GVQ/4.0/UML/Feedbackmodel.jpg
User feedback model

```
class User Feedback model (simplified)

GVQ_FeedbackTarget
  + parent :GVQ_FeedbackTarget
  - resourceRef :MD_Identifier

GVQ_FeedbackFocusType
  + primaryFocus 1
  + secondaryFoci 0..*
  + supplementaryFoci 0..*

GVQ_FeedbackTarget
  + parent :GVQ_FeedbackTarget
  - resourceRef :MD_Identifier

GVQ_FeedbackGroup
  + timestamp :CI_Date
  - user :GVQ_UserInformation
  - roles :GVQUserRoleCodeEnum [1..*]

GVQ_FeedbackItem
  + identifier :MD_Identifier

GVQ_Rating
  + ratingValue :int

GVQ_UserComment
  + comment :String
  + mime-type :String = text/plain

GVQ_QualityOverride
  + alternativeDataQualityEstimate :DQ_DataQuality

GVQ_ExternalFeedback
  + resourceURL :String
  + mime :String

GVQ_UsageReport
  + usagePurpose :GVQ_ReportAspectCode [0..*]
  + Citation :CI_Citation [0..1]
  + usageDescription :string
  «XSDelement»
  + alternativeDatasets :MD_Identifier [0..-1]

GVQ_UserInformation
  + user :CI_ResponsibleParty [0..1]
  + applicationDomain :string [0..*] [ordered]
  + expertiseLevel :int
```

Presenting Quality Information: From Dataset Quality to Individual Sample.
User feedback

• A FeedbackItem contains:
  – Mandatory information on **user role** (e.g., ‘Commercial data producer’) **application domain, expertise level** etc.
  – Information qualifying the **feedback**, such as its **subject**, **application domain** and **keyword** tags
  – Other optional information such as **rating, comments**, a **quality override** (superseding producer quality information), a **report of usage** or a **citation**.
  – **Focus** - **spatial, temporal** or **other** subsets to which feedback pertains
Populating the user model
Including data quality in search

Presenting Quality Information: From Dataset Quality to Individual Sample.

<table>
<thead>
<tr>
<th>Evaluation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation method</td>
<td>directExternal</td>
</tr>
<tr>
<td>Evaluation method description</td>
<td>Calculation of omission and commission statistics from sample polygons.</td>
</tr>
<tr>
<td>Reference document title</td>
<td>Global land cover mapping from MODIS: algorithms and early results</td>
</tr>
<tr>
<td>Reference document category</td>
<td>journalArticle</td>
</tr>
<tr>
<td>Reference document doi</td>
<td>10.1109/TGRS.2006.864370</td>
</tr>
</tbody>
</table>

**Result 1**

<table>
<thead>
<tr>
<th>Value type</th>
<th>Percentage accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Result value</td>
<td>55</td>
</tr>
</tbody>
</table>

**Quality report traceability**

1. Analysis published in a peer-reviewed journal article with full description of methodology.
2. Original data and reference dataset published via FAO and JRC.

The thematic validation procedure was carried out using design-based inference which conforms to the CEOS Cal-Val recommendations. The resulting research was published in a peer-reviewed journal, and the component datasets generated for the validation may be obtained from JRC.

**Process step 1**

<table>
<thead>
<tr>
<th>Description</th>
<th>Stratified sampling of ground truth blocks.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rationale</td>
<td>Systematic sampling to obtain an irregular stratification with different sampling rates for each stratum.</td>
</tr>
</tbody>
</table>
http://uncertgeo.aston.ac.uk/INSPIREtutorial/
Quality in OGC Web Map Service
WMS-Q

Joan Masó
17:10
Easy Spacialized Quality Visualization

• Our aim was to introduce spacialized quality layers in the evaluation phase of the geospatial data selection.
  – “quality-enabled” Web Map Service profile (“WMS-Q”)

• “Quality” means different things:
  – Completeness, consistency, accuracy, lineage …

• We focused on two main aspects of data quality:
  – Visualizing thematic accuracy, expressed as uncertainties
  – Linking to further information recorded in metadata documents

• We considered quality information at various levels:
  – Dataset, variable and sample level

• We aimed to avoid modifying OGC WMS 1.3.0, restricting ourselves to specializations of the spec
Semantic groupings of WMS Layers

- We need a method to convey that individual Layers are related semantically
  - E.g. one Layer represents the variance of another Layer

- We use Layer **nesting** + coupled with **keywords** from the UncertML vocabulary

- See fragment of Capabilities document (right, simplified)
  - In this examples uncertainties are normally distributed

- Also could be applied to other kinds of semantic groupings
  - E.g. components of a velocity field
Styling of Layers

There are many different ways of representing uncertainties visually:
- Contours, textures, shading, transparency, bivariate colour maps...

Different methods suit different datasets and users.

WMS provides two methods:
- Named Styles
  - Simple but inflexible
- Styled Layer Descriptors and Symbology Encoding
  - More flexible but still rather basic for raster data

ncWMS provides some simple extensions to WMS.

None of these meet the use cases for visualization of uncertainty.

Hence we have developed a new XML language for specifying styles for raster data.
- Named styles can map to XML definitions for backward compatibility.

Contours, textures, shading
MiraMon WMS-Q client and server

- [http://www.ogc.uab.cat/geoviqua/wmsq](http://www.ogc.uab.cat/geoviqua/wmsq)
- Three different layers with several quality indicators

- Ebro delta rice fields flooding maps
- Iberian annual air temperature
- Catalan Landsat land use classification

- **Service GetCapabilities**
  - [http://www.ogc.uab.cat/cgi-bin/GeoViQUA/WMSQ/MiraMon.cgi?REQUEST=GetCapabilities&VERSION=1.3.0&SERVICE=WMS](http://www.ogc.uab.cat/cgi-bin/GeoViQUA/WMSQ/MiraMon.cgi?REQUEST=GetCapabilities&VERSION=1.3.0&SERVICE=WMS)

Presenting Quality Information: From Dataset Quality to Individual Sample.
Greenland: Visualisation of DATA quality / uncertainty

http://giv-uw.uni-muenster.de/vis/v2/
GeoLabel
GeoLabel demo

Lucy Bastin
17:20
Visualisation of **METADATA** quality: GEO Label

- a quality **indicator** for GEOSS geospatial data and datasets
- assist in **searching** - provide users with visual clues of dataset quality / relevance.
- provide **accreditation, provenance, monitoring**
- increase **visibility** of EO data
- emphasize **open access** and **easy availability**
Producer statement:
ISO 19115 / GVQ

Feedback Collection

Styled / filtered feedback

Styled / filtered metadata

Presenting Quality Information: From Dataset Quality to Individual Sample.
Citations for this dataset

Title: Objective air temperature mapping for the Iberian Peninsula using spatial interpolation and GIS
DOI: 10.1002/joc.1462

Category: journalArticle

Title: Monthly precipitation mapping of the Iberian Peninsula using spatial interpolation tools implemented in a Geographic Information System
DOI: 10.1007/s00704-006-0264-2
http://www.springerlink.com/content/36463p2677318556/

Category: journalArticle

Discovery and Access Broker

Lorenzo Bigagli
17:40
Scientific and technical objective

• Enhance the GEOSS Common Infrastructure (GCI) with innovative quality-aware search and evaluation tools
  – Extend the GEO-Discovery & Access Broker (GEO-DAB)
GEO-DAB functional breakdown

Discover
- Brokering approach
- Semantic Discovery
- Web 2.0 support

Evaluate

Access
- Common grid data

Use

GeoViQua
DAB-Q: extension to support quality information

Presenting Quality Information: From Dataset Quality to Individual Sample.
Activities and goals

- Identify relevant properties (queryables) for quality-constrained queries
  - GeoViQua user and technical requirements
  - GeoViQua quality indicators
  - GeoViQua quality label

- Define and implement the quality extension to CSW-ISO (CSW-Q)

- Define and implement brokering logic (mapping and accessors) for GeoViQua services
  - Feedback Catalog
  - WMS-Q, SOS-Q
DAB-Q in context
CSW-Q extension

Producer Quality model queryables

- Count of data quality info elements
- Count of report elements
- Misclassification rate
- Number of missing items
- Domain conformance rate
- Positional accuracy value
- Count of publication elements
- Uncertainty level value
- Number of items noncompliant to the rules of the conceptual schema
  Measurement method
  Number of process steps cited
  Number of sources cited
- Records that uses a process with the following description or identifier
- Records that uses a source with the following title or identifier

Quality parameters
- Completeness
- Logical consistency
- Thematic accuracy
- Positional accuracy
CSW-Q extension

User Quality model queryables

- Rating score
- Presence/absence of feedback fields (e.g. user comment, usage, rating, citation,...)
  Any text
- Average of all rating score
  Category of the report
- Count of feedback items
  Domain in which the feedback is deemed relevant
  User domain
  Role of the user when submitting the item
URL’s of interest

• GeoViQua DAB

• Capabilities Document

• Demo portal
Demo portal

WP4 Demo portal

Presenting Quality Information: From Dataset Quality to Individual Sample.
Thanks!

l.bastin@aston.ac.uk
joan.maso@uab.cat

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2010-2013) under grant agreement no. 265178.