

Introduction

The Global Earth Observation System of Systems (GEOSS) is estimated to contain more than 28 million dataset records and is constantly growing. To tackle the problem of data quality assessment and dataset selection decision making, our project – **GeoViQua** – is undertaking active research to define, develop and evaluate a GEO label. The GEO label will visually summarise and allow interrogation of key informational aspects of geospatial dataset records upon which users rely when selecting datasets for use.

GEO Label Studies

To date, we have conducted **3 user studies** to **(1)** identify the informational aspects of geospatial datasets upon which users rely when assessing dataset quality and trustworthiness, **(2)** elicit initial user views on a GEO label and its potential role and **(3)**, evaluate prototype label visualisations.

Geospatial Data Quality:

We conducted an initial investigation into how geospatial data users evaluate the comparative quality and trustworthiness of geospatial datasets. Using a series of face-to-face and telephone interviews our intention was to uncover initial information about dataset selection, use and production in order to *inform* design and development of the GEO label.

The results revealed eight informational facets that could potentially be a part of the GEO label function:

- the reputation of the data producer;
- producer comments on dataset quality;
- dataset compliance with international standards;
- community advice;
- dataset ratings;
- links to dataset citations;
- expert value judgments; and
- side-by-side metadata records comparison.

GEO Label Function:

We conducted an investigation into geospatial data producers' and users' views on the concept of a GEO label and the role it should serve. Our study investigated: whether a GEO label is relevant to geospatial data; whether users and producers want a single "one-for-all" label or separate labels; the function that would be most relevant for a GEO label to carry; and the functionality that could be transferred from common rating and review systems.

Results indicated that users and producers support an all-in-one drill-down interrogation facility that would combine:

- the reputation of the data producer;
- producer comments on dataset quality;
- dataset compliance with international standards;
- community advice;
- dataset ratings;
- links to dataset citations;
- expert value judgments; and
- quantitative quality information.

GEO Label Representation:

On the basis of our findings from Phase I, we developed and evaluated three prototype GEO label visualisations:

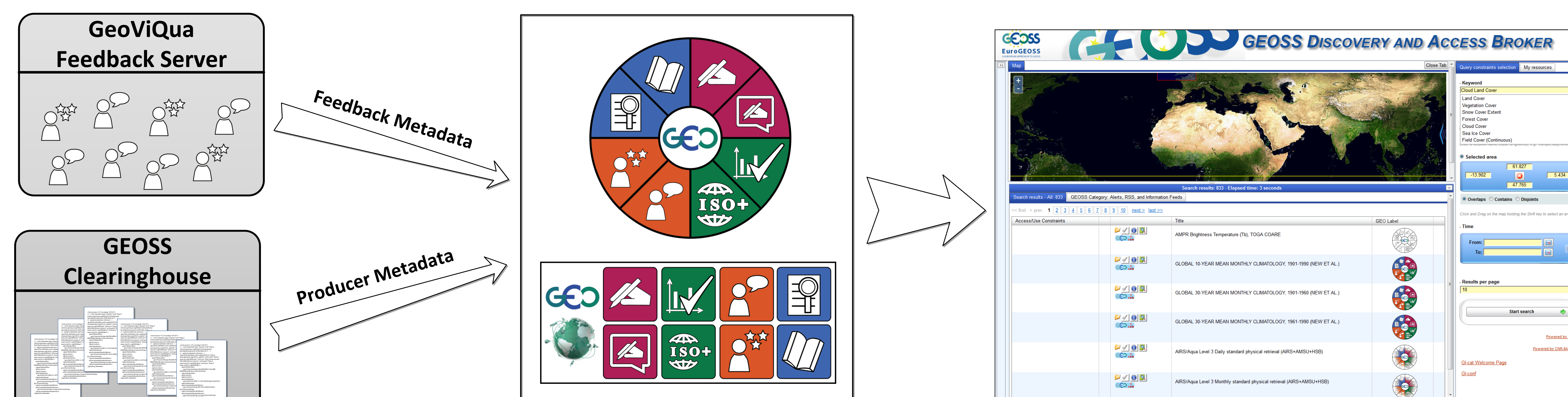


Based on results, we developed updated user-dictated proposals for the GEO label. The visualisations have not yet been validated; their evaluation/validation will be carried out as part of Phase III of our research.



Developing a GEO Label for Integration in the GEOSS

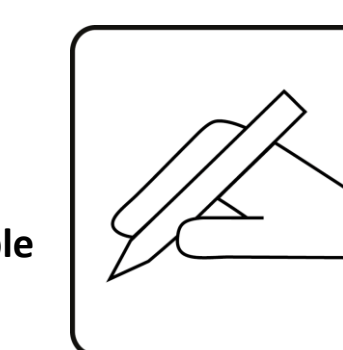
When integrated in the GEOSS, an individual GEO label will be provided for each dataset in the GEOSS clearinghouse (or other data portals and clearinghouses) based on its available quality information. Producer and feedback metadata documents are being used to dynamically assess information availability and generate the GEO labels. The producer metadata document can either be a standard ISO compliant metadata record supplied with the dataset, or an extended version of a GeoViQua-derived metadata record, and is used to assess the availability of a producer profile, producer comments, compliance with standards, citations and quantitative quality information. GeoViQua is also currently developing a feedback server to collect and encode (as metadata records) user and producer feedback on datasets; these metadata records will be used to assess the availability of user comments, ratings, expert reviews and user-supplied citations for a dataset. The GEO label will provide drill-down functionality which will allow a user to navigate to a GEO label page offering detailed quality information for its associated dataset. At this stage, we are developing the GEO label service that will be used to provide GEO labels on demand based on supplied metadata records.



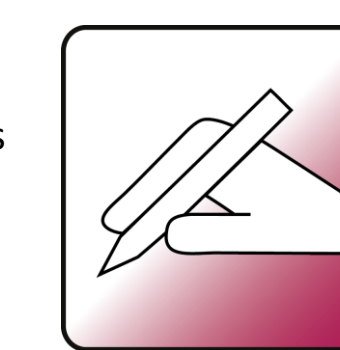
As the GEO label is intended to convey the availability of quality information for a given dataset, each informational facet can represent one of three availability states: **'available'**, **'not available'**, and **'available only at a higher level'** (indicates that information is not immediately available for the dataset, but is available for a parent dataset). These three information availability states will be expressed through varying the appearance of the facet icons.



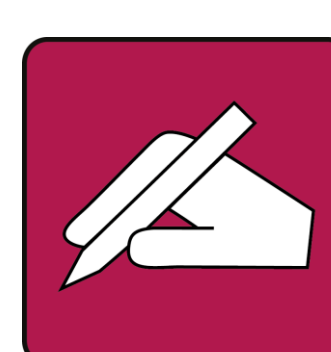
Fully filled-in background + white icon – indicates that information is **available** for this dataset.



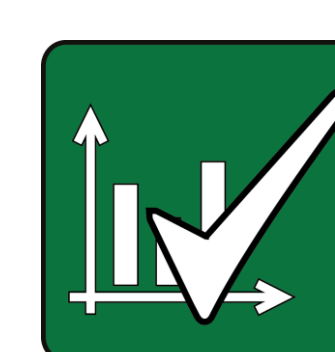
White background + icon outline – indicates that information is **not available** for this dataset.



Partially filled-in background + icon outline – indicates that information is **available only at a higher level** for this dataset.



Producer Profile facet conveys availability of information about the producer of the dataset, e.g., organisation or individual who produced the dataset, their contact information, etc. ISO 19115 compliant producer metadata document is used to assess its availability (<gmd:contact> element).



Quality Information facet conveys availability of formal quality measures of the dataset, e.g., uncertainty measures recorded in UncertML, errors, accuracy information, etc. GeoViQua-derived ISO 19115 producer metadata document is used to assess its availability (<gvq:dataQualityInfo> element).



User Feedback facet conveys availability of feedback and comments provided by the users of the dataset, e.g., general comments on dataset quality, identified problems, suggested use for the dataset, etc. GeoViQua-derived feedback metadata document is used to assess its availability (<gvq:userComment> element with <gvq:expertiseLevel> below 4).



Expert Reviews facet conveys availability of domain experts' comments on dataset quality, e.g., results of formal quality checks, expert suggestions on the dataset applications, etc. GeoViQua-derived feedback metadata document is used to assess its availability (<gvq:userComment> element with <gvq:expertiseLevel> being 4 or over).



Producer Comments facet conveys availability of any informal comments about the dataset quality as provided by the dataset producer, e.g., any identified problems, suggested use, etc. ISO 19115 compliant producer metadata document is used to assess its availability (<gmd:supplementalInformation> element).



Compliance with Standards facet conveys availability of information about dataset's compliance with international standards, e.g., compliance with ISO 19115, Dublin Core, etc. ISO 19115 compliant producer metadata document is used to assess its availability (<gmd:metadataStandardName> element).



User Ratings facet conveys availability of ratings of the dataset quality as provided by the users of the dataset (based on a 5-star rating system). GeoViQua-derived feedback metadata document is used to assess its availability (<gvq:rating> element).



Citations Information facet conveys availability of citations where the dataset was used and cited, e.g., formal reports on dataset quality checks, journal articles, etc. Both ISO 19115 producer metadata and GeoViQua-derived feedback metadata documents are used to assess its availability (<gmd:Citation> element).

Conclusion

Our three studies results indicate significant potential for the provision of visual dataset quality and trust indicators to help data users in dataset evaluation, intercomparison and selection. Specifically, we believe that a GEO label can be a step towards better visualization and communication of trust and quality information in the GIS domain. Our proposed GEO label will visually summarize and allow interrogation of user-derived key informational aspects of geospatial datasets, acting as a decision support mechanism in dataset selection. When integrated in the Global Earth Observation System of Systems (GEOSS), the GEO label will be used as a value and trust indicator for datasets accessible through the GEO Portal.

For additional information about the GEO label and its development, please contact:

Miss Victoria Lush,
Aston University,
Aston Triangle,
B4 7ET, UK.

Email: lushv@aston.ac.uk