

Labeling Geo-Referenced Information in Support of Data Sharing and the Facilitating of Societal Benefits of Earth Observations

Hans-Peter Plag
Nevada Bureau of Mines and Geology and Nevada Seismological Laboratory,
University of Nevada, Reno, NV, USA; hpplag@unr.edu.



A CIVILIZATION DEPENDING ON GEO-REFERENCED DATA AND SERVICES

Geo-referenced information is increasingly important for many scientific and societal applications. The availability of reliable and applicable spatial data and information is fundamental for addressing pressing problems such as food, water, and energy security; disaster risk reduction; climate change; environmental quality; pandemics; economic crises and wars; population migration; and, in a general sense, sustainability. Today, more than 70% of societal activities in developed countries depend directly or indirectly on geo-referenced information. The rapid development of analysis tools, such as Geographic Information Systems (GISs) and web-based tools for viewing, accessing, and analyzing of geo-referenced information, and the growing abundance of openly available Earth observations (e.g., through the Global Earth Observation System of Systems, GEOSS) likely will increase the dependency of science and society on geo-referenced information.

THE MAGIC OF DATA COMBINATION

Increasingly, tools such as GISs and web-based cyber infrastructure allow the access, viewing and combination of geo-referenced data sets from various sources. Improvements of interoperability, promoted particularly by GEOSS, will strengthen this trend and lead to more tools for the combinations of data from different sources. What is currently lacking is a service-oriented infrastructure helping to ensure that data quality and applicability are not compromised through modifications and combinations. Most geo-referenced information comes without sufficient information on quality and applicability.

How can we ensure data integrity if we can easily combine data of unknown quality and useability from different sources?

TOWARD LABELING AND CERTIFICATION

A valuable service-oriented infrastructure would be a trusted globally available GEO Label that communicates easy to understand information on the quality, usability and relevance of geo-referenced data. In many cases, certification attached to the label would further enhance the importance of such an infrastructure. Many users of geo-referenced data ranging from reinsurances to large infrastructure construction companies have indicated that certified off-the-shelf products would greatly ease their business.

MAKING THE RIGHT DECISION

In designing a service-oriented architecture that could support a GEO Label, it is important to understand the impact of the goals for the label on the design of the infrastructure. Design, concept, implementation, and success of a label depend on the goals, and these goals need to be well-defined and widely accepted. Strong labels are generally those that are unique in their field and accepted by an authoritative body in this field. A label requires time to get accepted, and once established the key characteristics normally can not be changed. Therefore, an informed decision on a labeling for geo-referenced data is crucial for success.

A GOAL-BASED APPROACH TO LABELING

There is a wide range of potential goals for a GEO Label (Table 1) including:

- (1) provide an attractive incentive for involvement of Science and Technology communities by giving recognition for contributions; enabling credits for providers (attribution); and supporting forward traceability (usage);
- (2) promote data sharing by signaling data availability and conditions;
- (3) inform users by increasing trustworthiness; characterizing quality; characterizing applicability; ensuring backward traceability (data sources);
- (4) inform providers (and their funders) by providing information on relevance (meeting user needs); and provide information on usage.

Table 1: Portfolio of goals for the labeling of geo-referenced data sets.

Community/Area	Impacts
Science and Technology	attractive incentive for involvement of S&T communities: <ul style="list-style-type: none"> - recognition for contributions (citation); - enable credits for providers (attribution); - support forward traceability (usage).
Users	inform users: <ul style="list-style-type: none"> - increase trustworthiness; - characterize quality; - characterize applicability; - ensure backward traceability (data sources).
Data Sharing	promote data sharing: <ul style="list-style-type: none"> - communicate data availability and conditions; - support backward traceability (data source).
Providers	inform providers (and their funders): <ul style="list-style-type: none"> - provide information on relevance (meeting user needs); - provide information on usage; - provide user feedback on applicability

A GEO LABEL PROPOSAL

Here we proposed a dynamic, web-based multi-faceted label that combines objective and subjective components in a distinguishable way. These components are:

- **quality:** an objective assessment measuring quality, reliability, acceptability, interoperability, etc.
- **acceptance:** a subjective assessment scaling user perception, fit-for-usage, and user-related practical aspects.
- **relevance:** a combined assessment objectively measuring the match between the entity and somewhat subjective user needs published in a user requirement registry.

A Quality Measure: The basis for this facet of a label would be a range of measures which provide an assessment of the quality of the data or information provided by a system, including the meta data. These measures would have to be open for validation by an outside evaluator. Designing a label certifying a certain quality is a difficult goal. Much of data quality is dependent upon the particular application in a way that is difficult to measure. The one area where it might be possible to measure quality at a quantitative level is the completeness of the documentation or meta data. This should, however, not be simply an assessment of how many fields in a standard scheme are completed, but should assess scientific context and fundamental stewardship issues as well.

Scaling Acceptance Based on Individual Perception: Subjective assessments by individual users of parameters such as their acceptance, perception of fit-for-usage, and social utility of the entities provide valuable information to other potential users. A user rating system could be a feasible mechanism for gathering such data on the subjective evaluation of geo-referenced entities. The subjective rating of individual users is different in nature from the quantitative scale of the objective quality tool. Individually, this facet of a GEO Label could be expressed, for example using a five-star rating system.

Towards Objective Measures of Relevance: Measuring the societal relevance of a datasets is a difficult task (Plag et al., 2012a). The extent to which a geo-referenced entity meets requirements combined with the societal relevance of these requirements provides a basis for a quantitative measure of relevance. In order to implement such a measure, a comprehensive database of user needs and observational requirements is needed.

A Tri-Faceted Label: The three facets introduced above could be the basis for three different individual labels, which could selectively be used to indicate the quality, user perception, and relevance of an entity. Alternatively, these three facets could be combined into a join GEO-Label, were for each of the three facets, the measure or rating would be projected onto a scale from, for example, 0 to 5. The integrated tri-faceted label could be presented using a star as a base (Figure 1). The three tips of the star could represent (counterclockwise from the left tip) the quality, perception of individual, and relevance to consolidate user needs. Axes extending from the center to the tips of the star could also be included and provide an exact scale. Since in many cases assessments will not be available for all three facets of the label, the label design would have to account for the absence of some of the criteria in a clearly discernible way.



Figure 1: Design study for a multi-faceted GEO Label. The two left symbols show the star for a high score in quality but low scores in perception and relevance, and a high score in perception and relevance but a low score for the quality facet. For the design of a GEO label, the star symbol could be embedded in a circular object with resemblance to the Earth, and the tips of the star could be labeled (right symbol).

GEOSS Infrastructure Interactions VERSION GC12-4B

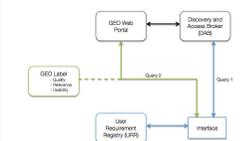
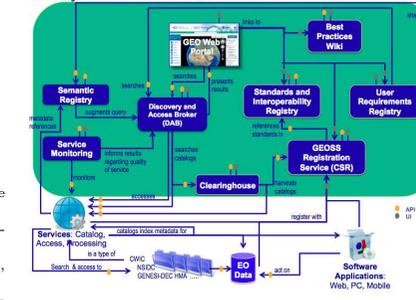


Figure 2: Integration of the GEO Label into the GCI. The Quality and Acceptance facets of the Label will have to be determined based on meta data and a to-be-designed user feedback system, respectively. The Relevance facet will be provided by the URR.

GEO-LABEL: TRUSTED BRAND FOR GEOSS

The Group on Earth Observations (GEO) has embarked on establishing a GEO Label. The responsibility for the development of a GEO Label concept rests with the GEO Work Plan Task ID-03. While the original idea (STC, 2010) was a voluntary label establishing a “trusted brand” focused on assessing the scientific relevance, quality, and acceptance of entities, the label concept emerging now recognizes a broader need of users for information on quality and suitability of data sets and services. It also is clear that a “trusted brand” cannot be based on a voluntary label.

IMPLEMENTING THE GEO-LABEL

The GEO Label will be implemented as a web-based, dynamic, exportable tag within the GEOSS Common Infrastructure (GCI, Figure 2). The quality facet will use quantitative measures established by GEO’s Quality-Assurance for Earth Observations (QA4EO) initiative. An important indicator of quality is the level to which products were reviewed. Developing a well-defined and widely accepted process for data reviewing is of vital importance, also as part of the data citation (McCallum et al., 2012). A user rating system embedded in the GCI will provide a subjective evaluation of products made accessible through GEOSS. Based on the GEOSS User Requirement Registry (URR), which compiles information on user types, their applications, and the requirements for observations and products (Plag et al., 2012b), the match of products with the needs and requirements published in the URR provides a measure for the relevance facet.

REFERENCES

GEO Science and Technology Committee (STC), 2010. The GEO Science and Technology Road Map. Available at <http://www.earthobservations.org>.
 McCallum, I., Plag, H.-P., Fritz, S., 2012. Data Citation Standard: A means to support data sharing, attribution, and traceability, *Geophysical Research Abstracts*, 14.
 Plag, H.-P. et al., 2012a. A Measure for Relevance of User Needs, Observational Requirements, and Earth Observations Data.
 Plag, H.-P. et al., 2012b. The GEOSS User Requirement Registry (URR): A component of the GEOSS Common Infrastructure in service of Work Plan Implementation and a user-driven GEOSS.