The GEOSS User Requirements Registry (URR): A Component of the GEOSS Common Infrastructure in Service of Work Plan Implementation and a User-Driven GEOSS

Contact author:
Hans-Peter Plag, IEEE and University of Nevada, Reno, Nevada, USA (hpplag@unr.edu)
(*for information on authors, contributors, and reviewers, see end of document*)

DRAFT

ABSTRACT

The User Requirements Registry (URR) is a versatile component of the GEOSS Common Infrastructure (GCI) for the collection and quantification of user-related information, which provides key functionality needed to ensure a user-driven development of GEOSS. The URR is inherently linked to other components of the GCI, including the Discovery and Access Broker, the Semantic Registry, the Best Practices Wiki, and the Standards and Interoperability Registry. The URR provides the functionality for three types of analyses: (1) prioritization of observational requirements; (2) gap analyses; and (3) determination of dataset/product relevance. The first of these can be carried out entirely within the URR. The second analysis, which is initiated by a user of the URR, requires a dialog between the URR and the Discovery and Access Broker (DAB). The third analysis, which is initiated by a user of the GEO Web Portal, requires a dialog between the GEO Web Portal and the URR. The URR has a number of controlled vocabularies, and these vocabularies are being aligned with the GCI Semantic Registry. Any semantic issues arising from differences in the URR controlled vocabularies and similar vocabularies used outside of the URR and GCI are handled by the GCI Semantic Registry. The URR can play a key role in the implementation of the GEO Work Plan. It provides the means to link the individual Task activities with applications, users, requirements, and needs for infrastructure, research, technology and capacity building. By doing so, the individual Task activities can better assess the relevance of their outputs and thus get support for the prioritization of activities within a task. The URR also supports the identification and specification of Essential Variables for the Societal Benefit Areas (SBAs). All Work Plan Tasks are expected to conduct gap analyses, and the URR provides the functionality to store the relevant user-related information collected by each Task. With increasing comprehensiveness of the information published in the URR, this enables a gap analysis across Task boundaries and cross-cutting through all SBAs.

1 INTRODUCTION

The GEOSS User Requirements Registry (URR) is an integral part of the GEOSS Common Infrastructure (GCI), which is used to publish user-related information and to provide analysis tools that allow answering questions related to a user-driven design and functionality of GEOSS. The URR provides the means for the collection of user-related information, such as the user types, their applications and activities, the requirements of the applications in terms of observations and other

products, and the needs in terms of research, infrastructure, technology, and capacity building that would enable or improve applications. The URR collects this information in standard formats and nomenclature across disciplines and Societal Benefit Areas (SBAs), and makes the information accessible for analysis. The URR captures the interconnectivity of user types, applications, requirements and needs, and with this supports the prioritization of applications, requirements, and needs, the gap analysis in terms of requirements not being met by available datasets or products, and the determination of the relevance of a given product.

The URR is described in more detail in Plag et al. (2010, 2011). More information on the URR data model is also provided in Plag et al. (2012a). The Concept of Operations for the URR is detailed in URR Team (2011).

This document first reflects on the purpose of the URR and introduces the three primary types of analyses enabled by the URR as part of the GCI. In Section 3, we discuss the role of the URR for the implementation of the GEO Work Plan. Section 4 presents an overview of the methodology for the population of the URR. Sections 5 to 7 consider in detail each of the three analysis types. Section 5 discusses the prioritization of entries in the URR based on their relevance, which is an analysis that can be conducted within the URR alone. Sections 6 and 7 consider gap and relevance analyses, respectively, which require a dialog between the URR and other parts of the GCI. The remaining sections provide details on technical aspects of the URR. Section 8 specifies the data model of the URR. Section 9 provides an overview of the controlled vocabularies of the URR. Section 10 provides some initial thoughts on the design and architecture of an URR that is fully integrated into the GCI. In Section 11, we discuss priorities for the URR implementation and population, and the integration of the URR into the GCI. Section 12 gives an overview of the available documentation of the URR, including tutorials for users.

2 PURPOSE OF THE URR

The URR is the place in the GCI for the specification and quantification of user needs and user-related information such that this information can be utilized in analyses that give guidance to the development of a user-driven GEOSS. Development of a user-driven GEOSS requires an understanding of (1) what has priority for users; (2) which of these priorities are not met by the available datasets and products; and (3) what is the user-related relevance of existing or proposed datasets. The URR provides the tools and information to conduct three primary types of analyses corresponding to these questions.

The user-related information collected in the URR enables the three generic types of analyses:

(1) prioritization of user needs and observational requirements;
(2) gap analyses;
(3) relevance analyses.

Key for all three analyses is information on the interconnectivity between specific entries in the URR, which is captured by the entries in the Links relation (see Section 8). In the URR, Links entries describe the relationship between two specific entries in any of the URR relations. Links can be defined between entries in the same or in two different relations.

With this versatile concept to capture interconnectivity, the data model of the URR (see Section 8) allows for the prioritization to be carried out solely within the context of the URR. Prioritization of

specific applications, observational requirements, and needs in research, infrastructure, technology and capacity building can be performed on the basis of their relevance for other entries either in the same relation or the cross links to other relations published in the URR. A measure for the relevance of an entry A can be based on the number of other entries depending on entry A and the strength of these dependencies (Plag et al., 2012). In order to lend credibility to this measure, it will be important that relevant GEO entities (in particular, the Implementation Boards) endorse the definition of the measure. The algorithms for the determination of importance will depend on the definition of the measure. The functionality to determine the importance of an entry will be provided by the Analysis menu of the URR.

The information published in the URR is required for one of the five approach to gap analyses recommended by the Gap Analysis Task Force (see GEO Gap Analysis Task Force, 2011). Based on the interconnectivity captured by the published Link entries, it is possible to answer the question of whether an application, which depends on another application, can be carried out or whether the link is not implemented. Similarly, infrastructure, technology, capacity building and research needs, together with relevant link information, are often indicative of gaps to be filled.

A gap analysis in terms of observational system gaps is based on the core question of whether a specific requirement entry can be met with existing data or products (Figure 1, Query 1). This question cannot be answered based only on the information published in the URR. Answering this question necessitates a search for data that meet, or exceed, the characteristics specified for the requirement entry. This search needs to be performed by other components of the GCI, specifically the Discovery and Access Broker (DAB).

In a relevance analysis, the core question is whether a specific or generic dataset or product (existing or proposed) corresponds to one or more requirements (Figure 1, Query 2). A measure for relevance based on the number and relevance of the requirements met by a dataset is under development. In the most simple case, a 0/1-measure could be used, where 0 means that there is no requirement in the URR that needs this dataset, and 1 means that there are one or more requirements in the URR that are met by this dataset. Of course, it is desirable to develop more advanced measures of relevance. It is noted here that the implementation of Query 2 is not part of the current work assignment for the URR development, and it will have to be implemented by other GEOSS support entities.

3 ROLE OF THE URR FOR GEO WORK PLAN IMPLEMENTATION

The implementation of the 2012-2015 GEO Work Plan (GEO, 2011) is monitored through Task Sheets made available at http://www.earthobservations.org/geoss_imp.php. The structure of the Task Sheets includes several areas with user-related information, which can be published in the URR and thus made available for more comprehensive analyses. This presents an opportunity during 2012 for the Institutions and Development Board of GEO to examine these analyses. The areas and the benefits of the URR for these areas are:

- Infrastructure: infrastructure needs can be published in the URR and linked to a number of applications;
- Capacity building: capacity building needs can be published in the URR and linked to applications, user types, and other needs;
- Science and Technology: research needs as well as needs for new technology can be published

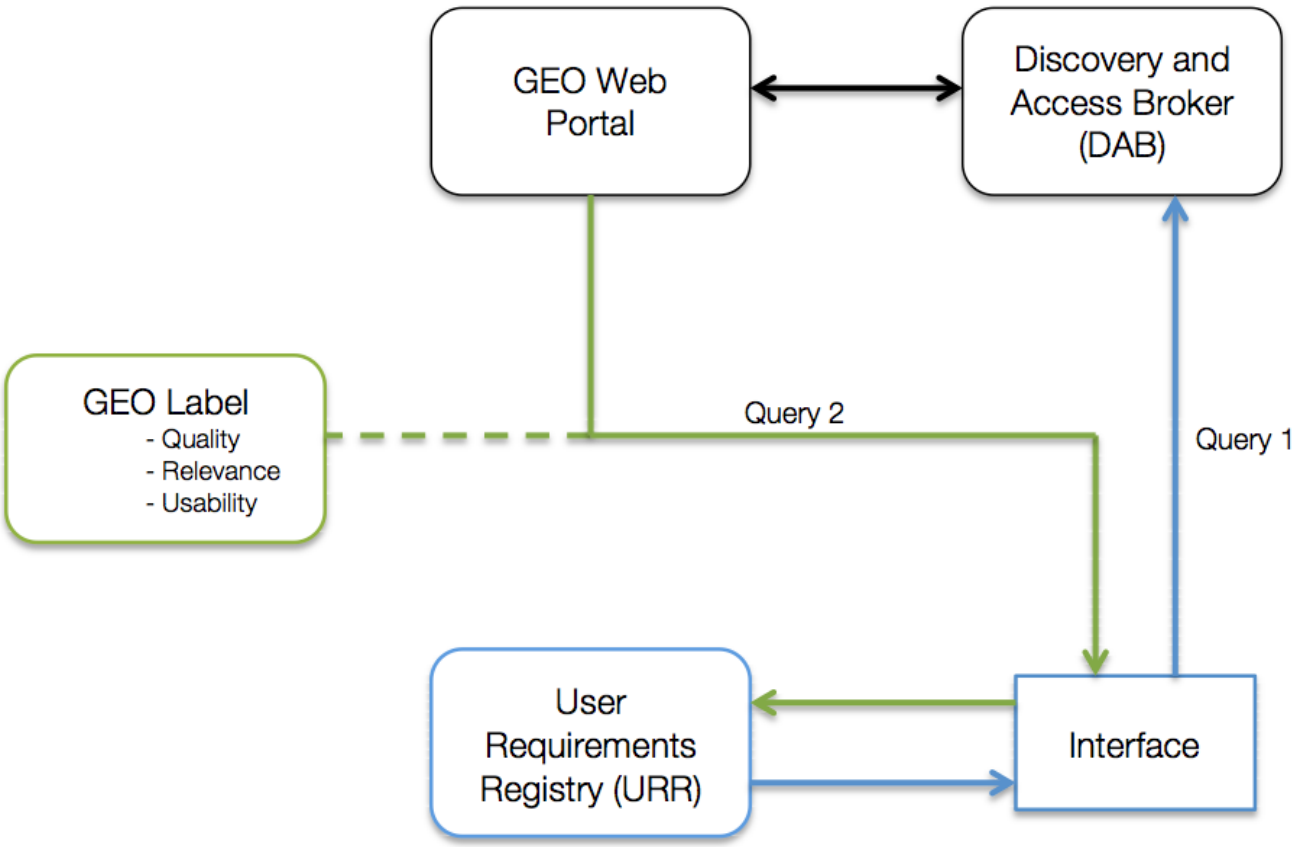

Figure 1: Queries requiring linkages between the URR and other GCI components. Gap analyses and the measuring of relevance require a dialog between the URR and other GCI components. Query 1 goes from the URR to the DAB and poses the question: "Which datasets/products meet requirement A?" (i.e., gap analysis). An interface will manage this dialog and ensure correct semantic interpretation of the query by the DAB and of the response by the URR. Query 2 goes from the GEO Web Portal to the URR and poses the question: "Which requirements are met by dataset/product B?" (i.e., relevance measure). Again, the interface handles the semantic issues. Once the GEO Label is established, the relevance measure can be obtained as part of the GEO Label for a dataset/product, and in this case, the GEO Label would use the Interface to communicate with the URR. From Plag et al., 2012b. For more details on the integration of the URR into the GCI see Section 10.

and their importance for enabling new applications or improving existing ones can be captured;
- User engagement: user types can be linked to applications, requirements, and other needs, thus capturing the relevance of GEOSS for the users;
- Gap analysis: the URR provides the tools to carry out customized gap analyses serving the needs of each individual Work Plan Task, and the results can be published in the URR;
- Societal benefits: the URR offers several ways of linking datasets, products, and activities performed by the Work Plan Tasks to societal benefits and thus supports the prioritization of requirements and activities based on these benefits.

The URR provides a unique way of interconnecting all activities performed by the Work Plan Tasks and to explore several cross-cutting themes. The interconnectivity captured in the link concept of the URR provides a basis to assess the relevance of a Task activity for the overall success of GEO and GEOSS, and its importance for societal applications and users. The GEOSS Strategic Targets (GEO, 2009) are included in the URR as applications, which can be linked to Task activities. For example, the Strategic Target for the Disasters SBA is published in the URR as application "GEOSS Strategic Target Disasters." Figure 2 shows an example of how Tasks and Strategic Targets can be included in the URR and linked to societal applications and other URR entries.

A number of initiatives have analyzed the 2009-2011 GEO Work Plan from different perspectives (e.g., the Science and Technology Committee, STC, with respect to science and technology aspects; the Architecture and Data Committee, ADC, with respect to the interrelations of Tasks), but there is no tool available that would allow one to answer in a comprehensive and well-defined way questions such as "Which end users depend on this Task activity?" "Which application will benefit from this Task output?" "Who will benefit from this research activity?" "Which applications would be enabled by this capacity building activity?"

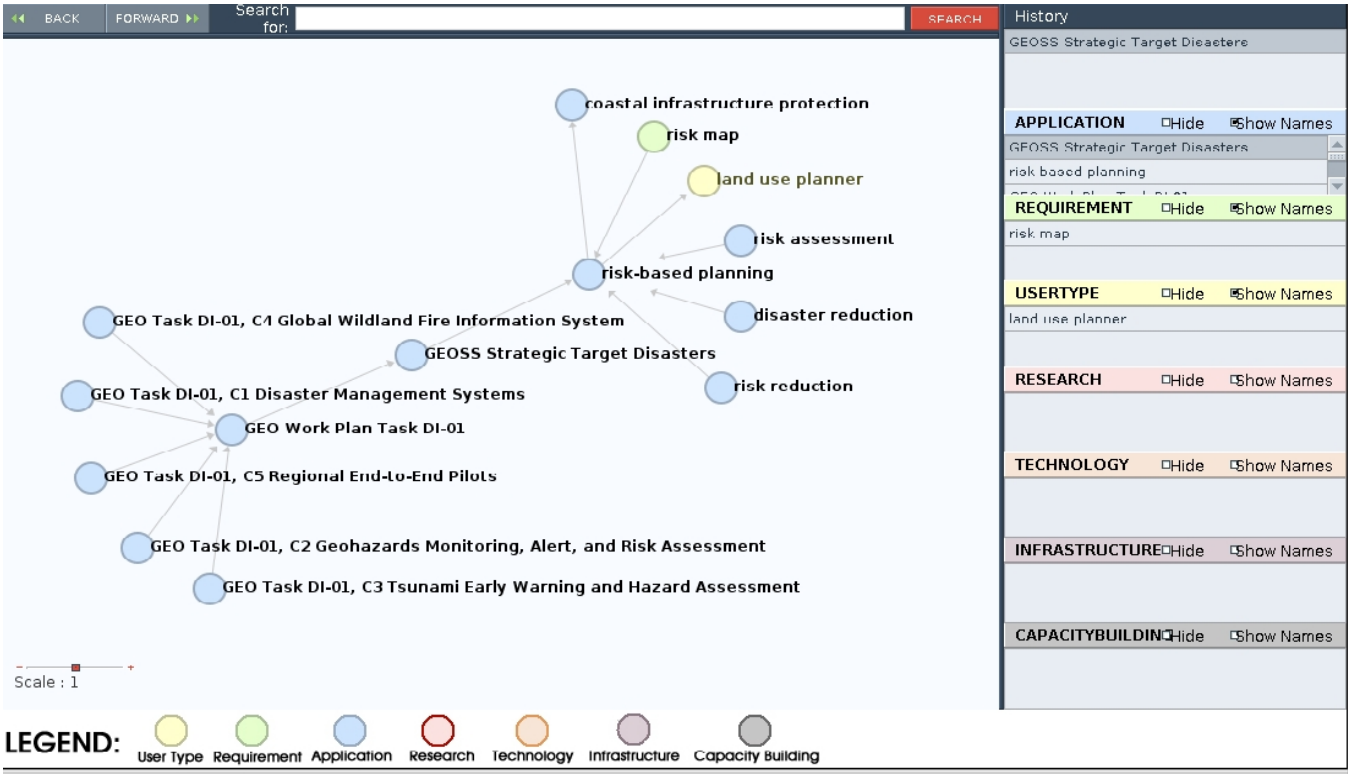

Figure 2. Graphical interface of the URR. The example shows in the center the application "GEOSS Strategic Target Disasters" and its second-degree links. This example illustrates how GEO Work Plan Tasks, their components and GEOSS Strategic Targets can be integrated in the URR to visualize connections between Tasks and Targets as well as between Targets and societal applications. In the example, the components of GEO Work Plan Task DI-01 are linked to the Task DI-01, which in turn is linked to the Strategic Target for the Disasters SBA. The Strategic Target in the example is linked to the application "risk-based planning," which in turn serves many other applications (not all shown). Links between individual Task components of DI-01 and other Tasks could also be added. The interconnectivity of Task components, Tasks, Strategic Targets of the Work Plan could be published in

the URR, thus providing graphical and analytical tools to consider links between Task Components and activities as well as links between GEOSS and societal applications.

The URR offers some unique cross-cutting capabilities that are currently not available within GEO. Several of the Work Plan Tasks, for example, particularly in the Infrastructure and Institutions and Development areas, address cross-cutting issues with relevance for many other Tasks. The data model of the URR allows us to capture and describe the cross-cutting nature of Priority Activities defined for each Task Component. Using this approach, the URR supports individual Task Components and their Priority Activities by making other Tasks aware of these activities, and the URR allows an overall evaluation of the cross-cutting issues and their relationship to the Strategic Targets.

In 2011, a Work Plan Task was proposed with the goal to conduct a general gap analysis within the field of Earth observations (EOs). It was decided to delete this Task from the Work Plan and instead require that all Work Plan Tasks would conduct their own specific gap analysis. This is reflected in the Task Sheet structure. The URR provides the only means within the GCI to generate a synthesis of the results reached by the individual tasks. If the results of the individual gap analyses are published in the URR, the URR can generate a synthesis report that does not only report the gaps, but also evaluates their societal relevance. This synthesis would support a prioritization of the gaps by the Implementation Boards and GEO, and it would help to direct resources to those gaps with the highest societal relevance.

In summary, the URR is a place for the publishing of many aspects of the Work Plan Task components. By doing so, a comprehensive picture of the work carried out in the Task can be generated and the relevance of the individual Task components can be seen in the context of the overall plan. With its distinctive ability to carry out prioritizations, gap analyses, and relevance analyses, the URR offers support to the Task components needed to accomplish the cross-cutting goals of the Work Plan. It also can support the Implementation Boards in assessing the interconnection between the Work Plan Tasks and the Strategic Targets.

4 POPULATING THE URR

The value of the URR depends strongly on the extent and quality of the information published in the URR. We first address the quantity.

Four generic types of approaches have been used to populate the URR:

(1) Expert input based on dedicated assessment reports;
(2) Open peer contributions (Wikipedia approach);
(3) Harvesting existing registries;
(4) Interviews.

In approach (1), assessment reports of user needs in a specified field are used by one or more experts to populate the URR. This approach has been used for several fields in the Health SBA and the Disaster SBA. The assessment reports produced by the GEO Task US-09-01a were used for infectious diseases, air quality, and aero-allergens. The requirement documents produced by the Geohazards Community of Practice (GHCP) and the Coastal Zone Community of Practice (CZCP) provided a basis for publishing information on geohazards and integrated coastal zone management (ICZM). Other Communities of Practice (CoPs) are being asked to participate in the population of the URR using approach (1),

including the Water Community of Practice (WCP). Additional communities to be explored are the Task Teams and a number of externally funded projects supporting the implementation of GEOSS, which in many cases have considerable user-related information or the goal to collect such information. This information could be published using either approach (1) or (2). In fact, the URR could provide a tool for capturing and analyzing the user-related information of a Task or a project.
In approach (2), anyone can register as a user of the URR and publish new entries or edit existing ones. By promoting this open, Wikipedia-like approach in GEO communities, a larger expert base can participate in adding or improving URR contents than would be reached through approach (1). Potentially, this approach also can lead to contributions from user communities not formally connected to GEO, or not known to GEO.

Approach (3) leverages the fact that a number of agencies have established user requirements databases in their specific fields. In most cases, the data models of these databases differ significantly from the data model of the URR, and in all cases, these data models only cover a part of the comprehensive URR data model. In a demonstration case for this approach, the Observing Requirements Database (ORD, see http://www.wmo-sat.info/db/) of the World Meteorological Organization (WMO) is being used to explore the viability of this approach, and to build an interface that addresses the semantic issues arising from differences between the controlled vocabularies. WMO's ORD is updated with the "Rolling Review of Requirements," (RRR) and the challenge of keeping the URR subspace defined by WMO's ORD consistent with the contents of ORD is addressed by this interface.

In approach (4), an interviewer, often an expert in a particular field, conducts an interview with users using a pre-defined set of questions. The goal of this interview is to capture a representation of the local environment of the user in terms of applications, related users, requirements, and needs. For end users close to societal decision making, this environment often is not directly linked to observational requirements, and this link emerges through several steps along the value chains captured in the URR. However, such end users are experts in their own environment providing details that often cannot be captured by those closer to the provider end of the value chain.

In terms of quality, experience has shown that the different approaches lead to information of variable quality. Within each of the four approaches, quality also varies, and several approaches display a wide range of quality.

Experience with approaches (1) and (2) revealed that there were considerable issues in understanding the data model of the URR (see Section 9). In particular, an understanding of the intended contents of entries for the relations Applications, User Types, Requirements, and Links could not be assumed. In reaction to the problems experienced by experts and other users in fully understanding these categories, detailed on-line tutorials were developed to lead novice users through all steps of publishing in the URR. These tutorials were tested and reviewed by members of the User Interface Committee (UIC). For several user communities assembled at their workshops, dedicated tutorial WebEx sessions were organized. Nevertheless, a large difference in the quality of the entries remained, underscoring the need for expert reviewing and editing of entries. The approach to reviewing and editing will have to be discussed by the Boards, with Task ID-04 being the best candidate for the development of a concept.

Another challenge for the open approach (2) is spam and willful overwriting of entries with false input. In the period between 28 September and 10 October 2011, five URR entries were visibly overwritten. In response to this, the requirement to register and login as user was introduced for publishing. Viewing

and analyzing the URR content is still possible without login. The simple user registration and the concept of user authority will be further developed in coordination with the user concept for the GCI.

A challenge for all approaches is the fact that user-related information is often described using structural concepts that differ significantly from the structural concepts underlying the URR data model. This requires in most cases an analysis of these differences prior to publishing in the URR. However, the URR provides helpful functionality for this analysis (see the tutorials for examples).

An important step towards a comprehensive population of the URR is the promotion of the URR in relevant user communities. Although those user communities already associated with GEO can be reached through Boards, Work Plan Tasks, and CoPs, user groups not linked to GEO pose a challenge. Dedicated outreach efforts are on-going to familiarize new communities with the services provided by the URR.

5 PRIORITIZATION

The URR contents can be used to support the prioritization of observational requirements, applications, and needs in infrastructure, research, technology, and capacity building. The interconnectivity between all of these entities captured in the URR can be used to construct answers to the generic question “How relevant is entry A?”

The metric to measure the relevance of an entry is under development (Plag et al., 2012c). It will be important to achieve a GEO-wide consensus on this measure. Initial steps will include the discussion of the measure by the three Implementation Boards. The measure will be composed of a number of sub-measures, which con be consider individually or combined. Consistency of the sub-measures will be used to assess the uncertainty of a relevance value. One sub-measure will take into account the number of dependencies and the importance of the linked dependent entries, especially the applications, and the strength of the links between entries. Another sub-measure particularly for requirements could be based on the prominence of the requirement in other requirement registries or peer-reviewed high-level documents.

A measure of relevance can also be deduced from the status of an entry in the mathematical network constituted by the entries and the links between them. It will be studied to what extent network theory can be used to support a measure of importance. Factors such as pervasiveness, uniqueness, difficulty, completeness, and uncertainty will be considered.

Originally, prioritization focused on requirements. However, there is a significant benefit in being able to prioritize the applications, as well as the needs mentioned above. For example, an answer to the question of “How relevant is this research need?” could guide researchers and support them in soliciting funds for their research. Answering the question “How relevant is this infrastructure need?” would help to prioritize efforts and secure funding for the implementation of important infrastructure.

6 GAP ANAYSES

The core question of the gap analysis considered here is “Is there a dataset/product that meets this requirement?” For a complete gap analysis, a search has to be conducted for each individual requirement entry published in the URR to discover the matching datasets. It can be expected that the

URR will soon have many thousands requirement entries, and each of them constitutes a search criteria. Thus, for a complete gap analysis, thousands of searches will have to be performed and then analyzed. Of course, a gap analyses can be constrained to subsets of the requirements published in the URR (e.g., the prioritized requirements, or those in a specific SBA).

In order to perform an analysis of the gaps in the observational system as compared to the requirements published in the URR, it is necessary to:

(1) collect information on existing datasets and products that can be discovered through GEOSS;
(2) develop algorithms that can compare prioritized requirements and the information on data availability.

The gap analysis will be implemented as part of the URR Analysis menu (together with the prioritization described in the previous section). For the gap analysis, a two-way dialog between the URR and the DAB is required (Query 1 in Figure 1).

The gap analysis does not provide information on specific datasets or products. It would be sufficient to display the result of a gap analysis by indicating which requirements are met and which are not (e.g., “Yes,” “No,” and “Maybe”). This presentation of the results will not constitute an alternative way of data discovery and access competing with the GEO Portal. However, a user performing a gap analysis within the URR Analysis menu for a set of requirements, who gets the information that a certain requirement is met, might want to actually find the datasets that meet the requirement. Therefore, within the Analysis menu of the URR, users will be able to ask: “What datasets satisfy this requirement?” without having to type the search into the GEO Web Portal or the DAB. Users who use this option will be transferred to the GEO Portal for the response.

7 RELEVANCE ANALYSES

The core question for a relevance analysis is “What is the societal relevance of this dataset?” As discussed in Section 4, a measure of relevance needs to be defined.

In order to perform a meaningful relevance analysis, it is necessary to:

(1) formulate for each known dataset or product a search in the URR;
(2) develop algorithms that can convert the result of this search in the URR, which returns all URR entries depending on this datasets or product, into a measure for relevance.

It is reasonable that the search is initiated by the GEO Web Portal, where users discover datasets and could have interest in the relevance of a given dataset (Query 2 in Figure 1). In response to the search, the URR could return this measure of relevance (i.e., the URR would determine the measure), or the URR could return all entries that depend on the dataset and then the DAB/GEO Web Portal would have to determine the measure. Since the measure will have to be based on the data model of the URR, it makes sense to give the URR the responsibility to evaluate the relevance according to the measures accepted by the Infrastructure Board.

One measure for relevance of a dataset or products can be derived from the relevance of the requirements (see Section 5) that are met by this dataset or product. However, other criteria also might lead to a measure for relevance. For example, the uniqueness of a dataset and the continuity (in case of

a time series) can be used to derive relevance. Similar to the relevance for requirements, a number of sub-measure will have to be considered individually and collectively to get an estimate of the uncertainty in a relevance value.
It is mentioned here that the question “What is the relevance of this dataset” will be of central importance for the GEO Label (see the extended documentation of Task ID-03 at http://www.geo-tasks.org/id03/ for more information on the GEO Label), because the GEO Label will most likely reflect the relevance of a dataset as one of the criteria. Thus, the implementation of the GEO Label will benefit from a relevance analysis being developed and implemented now (see Figure 1).

More information on the priorities for the population of the URR and implementation of the three types of analyses described above are contained in Section 11.

8 THE DATA MODEL OF THE URR

To inform the discussion of URR services, requirements, and interoperability within the GCI, the data model is briefly introduced here. The core of the data model of the URR consists of seven relations that describe the users, applications, requirements, and various needs related to decision making and support through EOs (Figure 3). The basic relations of the URR are:

Applications: processes and activities that use EOs or derived information to produce new information, arrive at decisions, or execute decisions.
User Types: generic users who are involved in applications, benefit from these, or contribute to them.
Requirements: specifications of EOs or derived products.
Research Needs: research tasks to be performed to enable applications that are currently not possible due to a lack of knowledge.
Technology Needs: descriptions preferably of observational requirements that cannot be met because the technology to carry out the observations is not available, and of what this technology would be.
Infrastructure Needs: descriptions of requirements that cannot be met or applications that cannot take place because of the lack of infrastructure, and of the infrastructure necessary to enable the applications.
Capacity Building Needs: descriptions of problems that cannot be solved satisfactorily today because of a lack of capacity in terms of organizational or human resources and a description of the necessary capacity building that would enable applications addressing these problems.

An additional relation is used to capture the interconnectivity between entries in each of these relations:
Links: the entries in this relation connect a source entry and a target entry in two different relations or in the same relation. This concept is a novel and versatile way of capturing interconnectivity. Information on the societal relevance and the implementation status of a Links entry enables the analysis of value chains from EOs to end applications.

Two auxiliary registries represent a lexicon of all terms used and the references to documents that provide additional information on the entries in the main forms:
Lexicon: collects all terms used in the URR; allows for the definition of acronyms and abbreviations; and specifies units of EOs or other quantities. The Lexicon also includes several controlled vocabularies.
References: a relation that collects all references to documents (e.g., publications, reports, web pages, etc.) that are used to link entries to more detailed background information.

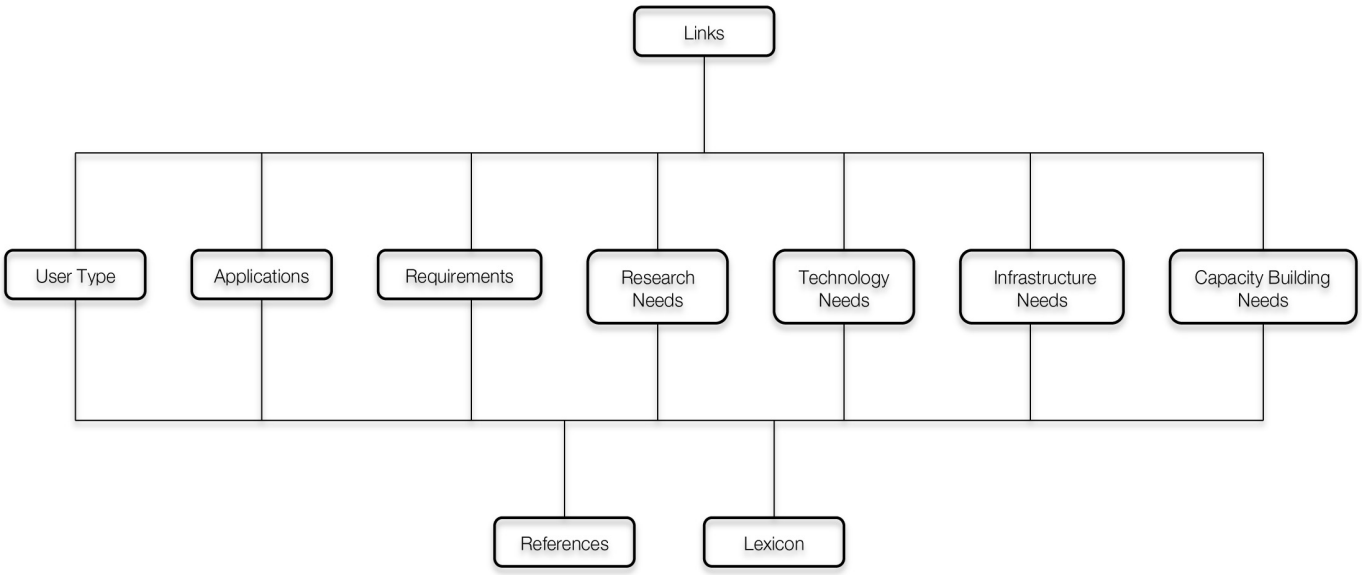

Figure 3: Data model of the URR. The data model can be separated into three parts: (1) seven relations to capture information on user needs and applications (middle); (2) the Links relation to capture connectivity (top); (3) a number of auxiliary relations to capture relevant information used in the other relations (bottom). Note that this figure does not show the complete relational data model of the URR. From Plag et al. (2012a).

9 CONTROLLED VOCABULARIES IN THE URR

The controlled vocabularies of the URR, which can be changed by authorized users, include:
Keyword: entries can be linked to keywords. The keywords are defined in the Lexicon.
Earth Observation Parameter: requirements can be specified for EO parameters, which first need to be entered in the Lexicon and there specified as EO parameters.
Attribute: in some cases, it is not the EO parameter itself but rather a derived quantity that is needed (e.g., the concentration). Thus, derived quantities are denoted as attributes and specified in the Lexicon.
Medium: similarly, some observations are only needed in a specific medium (e.g., soil moisture in soil); available mediums are specified in the Lexicon.
Unit: units of measurements (of EO parameters) are defined in the Lexicon. For all S.I. units, the S.I. naming is used.
Geographical Area: this vocabulary is used to characterize the geographical area for which the requirement is specified. The current list consists of Africa, Americas, Antarctica, (AquaCover), Arctic, Areas of known occurrence, Asia, Australia, Coastal Zones, Continent, Europe, Ice-covered areas, Impacted Areas, Land areas, (Landcover,) Oceans, Polar regions, (Population,) Relevant Geographical Features, Rural areas, Surface water, Tropical region, and Urban areas.

The controlled vocabularies of the URR, which currently can only be changed by the system administrator are:
Societal sector: a short list of societal sectors potentially useful for searches. Applications and User Types can be associated with one or more societal sectors. The current list includes: Academic, Commercial, Educational, Governmental, Private, and Others.
Societal Benefit Area: the list of nine SBAs defined by the 2nd Earth Observation Summit in 2004. The

vocabulary used is: Agriculture, Biodiversity, Climate, Disaster, Ecosystems, Energy, Health, Water, and Weather. Applications and User Types can be associated with one or more SBAs, and this association can be used to constrain searches.
Entry Status: identifies the status of an entry. The list consists of Incomplete, Preliminary Draft, Final Draft, In Review, and Accepted. Although publishers are urged to complete an entry, it is not mandatory. In some cases, it will be necessary to save an incomplete or preliminary draft. In cases where data is harvested from other databases, entries also may be incomplete.
Pre-fix: pre-fixes for units such as milli, centi, kilo, and so forth are used to specify the units in which requirements are detailed. The vocabulary is yotta (y), zetta (Z), exa (E), peta (P), tera (T), giga (G), mega (M), kilo (k), hecto (h), deca (da), no prefix, deci (d), centi (c), milli (m), micro (μ), nano (n), pico (p), femto (f), atto (a), zepto (z), and yocto (y).
Reference frame: the reference frame attribute of requirements specifies the spatial scale of the frame. The vocabulary is global, local, national, and regional.
Requirement Type: indicates the type of a requirement; the vocabulary includes Unknown; Target; and Threshold.
Sampling Type: the required type of sampling. The current list includes: once, event-triggered, repeated, and time series.
Link Type: The type of link currently can have the values Weak, Strong and Crucial.

In addition to the original nine SBAs, the GEO Work Plan now also includes several cross-cutting SBAs such as Oceans, Land Cover, and Forests. Consideration will be given to how these themes can be added to the controlled vocabulary for the nine SBAs.

It is also being consider to introduce an additional controlled vocabulary for Essential Variables (EVs). A set of EVs has been identified for Climate (ECVs), and a discussion is taking place to identify similar EVs for all other SBAs. Observational requirements for such EVs could be included in the URR and linked to the depending applications. The URR would allow to identify those EVs that are of relevance in several SBAs.

A key requirement for interoperability is the consistency of all relevant controlled vocabularies embedded in the URR Lexicon with the corresponding vocabularies used elsewhere in the GCI. Any inconsistencies would have to be addressed in the interface between URR and the other components of the GCI. Alignment of the controlled URR vocabularies to the controlled vocabularies used elsewhere in the GCI is under way. The GCI will include a component for semantic services, and once this component is available, full alignment of the controlled vocabularies in the URR will be possible.

Harvesting of other user-related databases requires a translation of the vocabularies used in these databases into URR vocabularies. As a first test case, the WMO Requirement Database is integrated into the URR.

10 DESIGN AND ARCHITECTURE CONSIDERATIONS

The development of GCI and URR has evolved considerably over the last six months. Figure 4 reflects the current status of the GCI and includes a number of new components developed over the last six months or still under development. One result of the development is that the URR is an integral part of the GCI interacting with a number of other GCI components (Figure 4). The URR utilizes the services of these components to maintain controlled vocabularies (Semantic Registry), to perform gap analyses

(DAB), and to link applications to best practices (Best Practices Wiki) and standards (Standards and Interoperability Registry). Access to the URR will be controlled through the GEOSS Registration Service, once this service is fully implemented. The URR also provides services to other GCI components. For the GEO Label (not shown in Figure 4), the URR will provide the relevance of datasets and products. Users of the URR will be able to augment the controlled vocabularies and these augmentations will be integrated into the Semantic Registry.

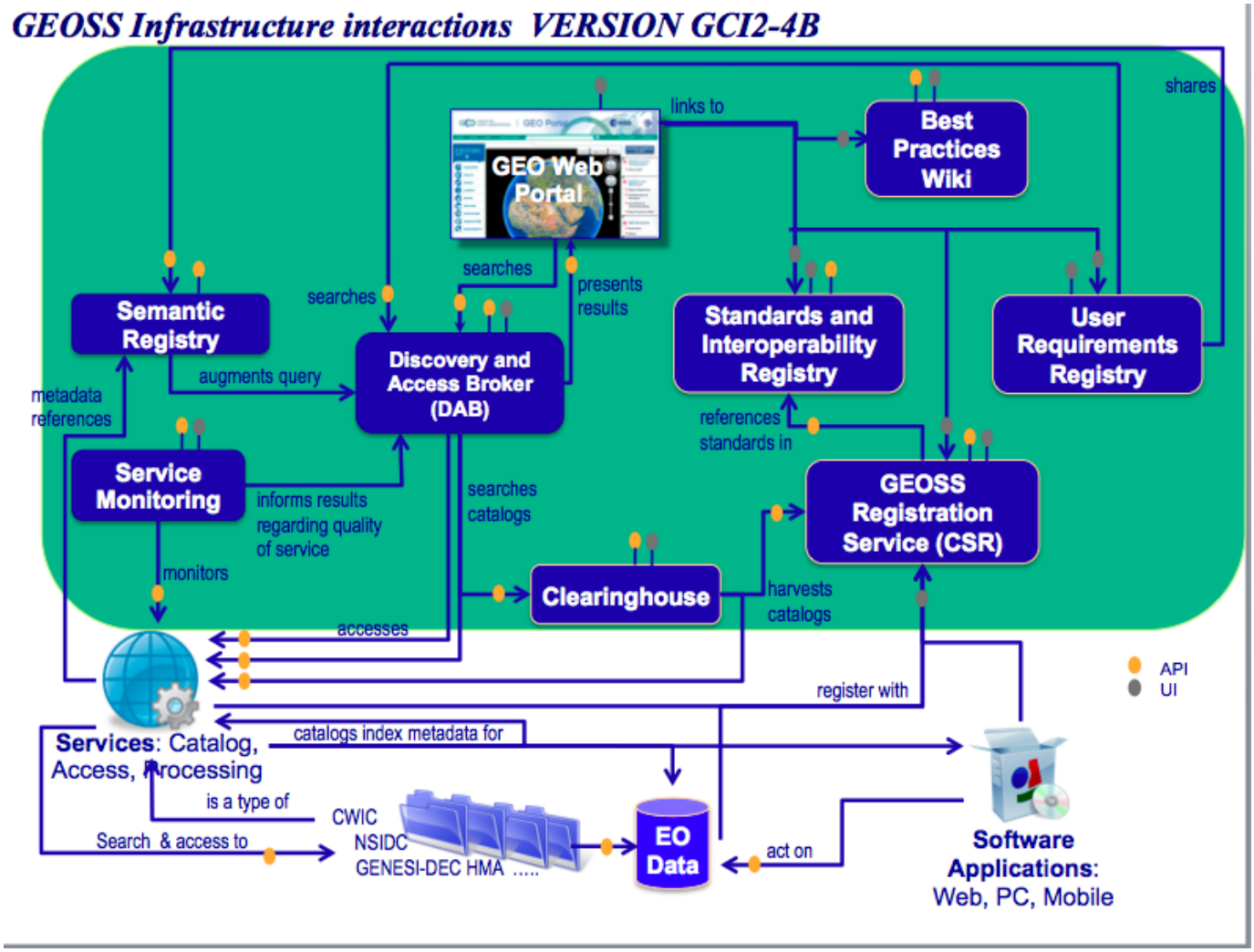

Figure 4: Overview of the GCI and the interactions between GCI components. The URR will align its controlled vocabularies to those maintained in the Semantic Registry. Through the User Interface (UI) of the URR, users are able to determine the importance of requirements, applications, and other needs (see Section 5) as a basis for prioritization, and to conduct gap analyses. For the latter, the URR engages in a dialog with the DAB (Query 1 in Figure 1) through an Application Programming Interface (API). The GEO Web Portal, through the GEO Label (not shown) will allow measurement of the relevance of a dataset or product based on the URR contents (Query 2 in Figure 1). Best Practices and Standards will be linkable to applications in the URR. From Pearlman (2012, personal communication).

Figure 5 provides more detail on the interactions between the URR and other GCI components, as well

as components outside the GCI. The interfaces between the URR and the GCI components are under development. A key question to be addressed is to what extent the linkages between the URR and the

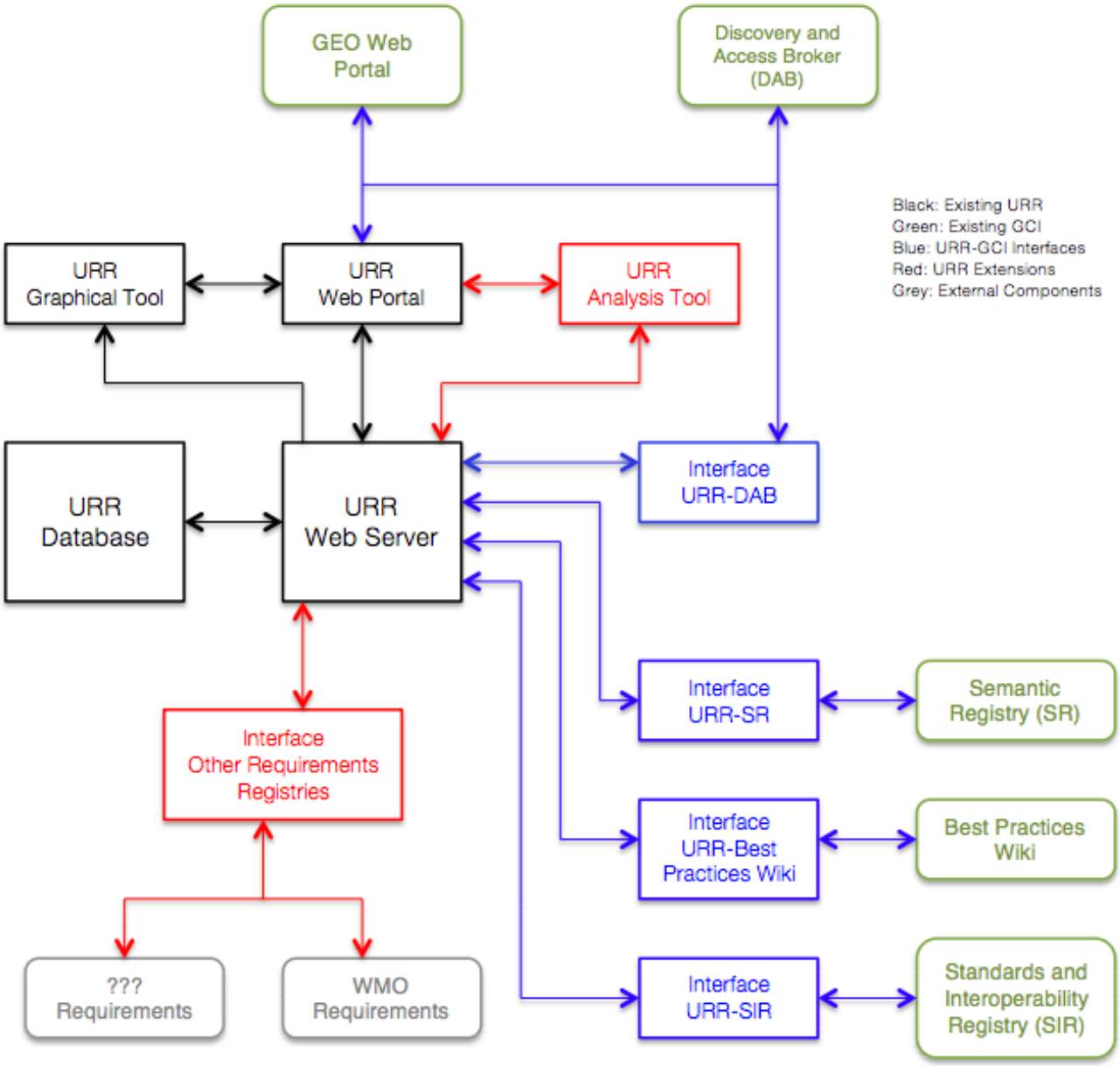

Figure 5: Detailed architecture around the URR. The core of the URR is a database accessible through a URR Web Portal. This Web Portal also provides access to a graphical tool and an analysis component. An interface to external user requirements registries facilitates the harvesting of these registries, which generally use different data models and vocabularies. The interfaces between the URR and GCI components communicate with the URR through a web server.

GCI components is dynamic with communication as needed or static with frequent updates. The interface between the URR and the DAB is a fully dynamic interface responding to queries (Query 1,

see Figure 1) as they are issued by the URR. Similarly, the interface will respond to queries from the GEO Web Portal or GEO Label (Query 2 in Figure 1) as they are issued. The interface between the URR and the Semantic Registry (SR) will use a static approach with frequent updates when changes occur in the SR or when changes are requested by URR users. Such changes include user-requested extensions of the controlled vocabularies. The interfaces between the URR and the Best Practice Wiki and Standard and Interoperability Registry also will use static approaches with updates when changes in these components require such updates. All of these interfaces communicate with the URR database through standard web server protocols.

As discussed in Section 4, harvesting existing user requirements databases is an important approach to populating the URR. However, existing user requirements registries use their own data models and vocabularies not necessarily aligned to the EO vocabulary of the GCI and URR. The interface between the URR and external registries therefore has to transform the contents of the external registries into the URR data model, and it has to mediate the semantic issues that result from differences in the vocabularies. Ownership of the information harvested from external resources will be maintained, and in general, harvested information will not be open to further edits. Since external registries can change, the interface allows for frequent updates through push or pull.

In designing the various analysis tools for prioritization, gap analysis, relevance analysis, and, ultimately, the GEO Label, it is important to ease the access for users to the relevant GCI components. The analysis tools will provide maximum benefits for the users contingent on complete information sharing between the URR and the other components of the GCI. At the same time, the architecture provides for a utilization of the expert knowledge associated with each component. Therefore, prioritization of requirements and the determination of the societal relevance of a dataset is being carried out by the URR with expert knowledge of the URR data model, but in both cases, it will be possible to initiate requests for these actions from the GEO Web Portal. Gap analysis, which requires knowledge of the URR data model to constrain the gap analysis to subsets of the URR requirement entries, is made available through the URR Analysis menu, but other UIs in the GCI may also provide access to gap analyses.

The most important interface to be fully implemented is the one between the URR and the DAB, which will enable the gap analysis. The SR of the GCI will need further development before the URR Lexicon and all controlled URR vocabularies can be fully integrated into this GCI component. Until the full capability of the SR has been developed, the URR will employ a temporary ontology form to mediate any differences between URR vocabularies and vocabularies used elsewhere in the GCI.

The existence of a URR Web Portal is consistent with most of the GCI components having their own UIs (see Figure 4). However, as far as possible, the URR Web Portal will be restricted to functionalities directly related to URR tasks. A user of the URR Web Portal will be able to ask the question “Which specific dataset meets this URR requirement entry?” However, the question will be referred to the DAB/GEO Web Portal and the answer will be presented there.

11 PRIORITIES FOR THE IMPLEMENTATION AND POPULATION OF THE URR

The Version 3.0 of the URR provides the initial menus for the publishing of entries in all URR relations, the viewing of entries, the exploring of the relationships between entries as expressed in the entries in the Links relation, and the searching of the URR for specific contents. To fully utilize the

potential of the URR, improvements are needed in three main areas:

1. Full implementation of the URR infrastructure;
2. Population of the URR with comprehensive information;
3. Implementation of the interfaces to GCI components.

Developments are required independently in each of these areas to enable the use of the URR in support of Work Plan implementation. In each of these areas, a number of steps are needed, and here we discuss the priorities for each area separately.

1. Full implementation of the URR infrastructure: the parts of the URR infrastructure still to be implemented are the Analysis and Preferences menus, an improved search utility, an improved user registration, and an interface to external requirement registries. The analysis menu has highest priority and is under development. It is expected to have an initial version available by May 1, 2012. Milestones in the development of the Analysis menu are the definition of the metric for importance (of requirements, applications, and the other needs) as a basis for prioritization, and the metric for relevance (of datasets and products). These definitions need to be provided to the Implementation Boards for discussion as soon as possible. The availability of the gap analysis and the relevance analysis depends on the implementation of the interfaces between the URR and relevant GCI components (see point 3 below).

The improved search utility also has a high priority. The search criteria to be used are defined, and the utility is under implementation.

The development of the interface to external requirements registries is under way and expected to be completed by May 1, 2012 for one test case, WMO's ORD.

The functions to be provided by the Preferences menu are defined in URR Team (2011), and it is expected to have the Preference menu available by November 2012.

Implementation of an improved user registration is pending until the GEOSS Registration Service is fully implemented.

2. Population of the URR with comprehensive information: priority is on extending the information available in the URR and improving the quality of the available information. Important progress towards a more comprehensive population of the URR is made with the help of CoPs. The goal is to complete the information for at least two SBAs. The Water CoP is contributing entries in the Water SBA, and the Geohazards CoP is providing information on the Disasters SBA. The GEO Work Plan Tasks will be invited to contribute their relevant information, which will cover all SBAs, although not comprehensively. Harvesting WMO's ORD will add information in the Weather and Climate SBA.

As mentioned in Section 4, the quality of the entries provided through the different approaches requires some form of review and editing. High priority is on the definition of a process for the reviewing and editing of URR contents. A proposal will be submitted to ID-04 for discussion and further handling in the ID Board. Once a process has been defined, a thorough review of the contents will be initiated.

A focus is on the full development of one or two use cases to demonstrate the full potential of the URR. It is planned to integrate these use cases into the AIP-5. Candidates are a use case in the Disaster and/or Water SBAs.

3. Implementation of the interfaces to GCI components: Highest priority is on the implementation of the interface between URR and DAB, which will enable gap analyses. This interface will also be used for relevance analyses; however, these analyses require implementation of requests in the GEO Web Portal or as part of the GEO Label. Therefore, the priority for this capability is determined independent of the URR developers. The implementation of the interface between URR and SR is pending until a further development of the SR will enable a full integration of the URR Lexicon into the SR.

The interfaces between the URR and the Best Practice Wiki and SIR are currently pending. The main reason for these interfaces is the possibility to connect URR entries to best practices and standards, which has been requested by URR users. It is expected that this capability will be available by November 2012.

12 ACCESS TO THE URR AND DOCUMENTATION

The URR is accessible through a hyperlink on the GEO Portal at http://www.geoportal.org. The URR Home Page (currently at http://www.scgcorp.com/urr) provides an overview covering available documentation. The technical documentation consists of:

- Technical Specifications: in preparation;
- Concept of Operations (URR Team, 2011): describes the functionality of the URR;
- This document: describes the role of the URR for Work Plan Implementation.

The Tutorials, which are accessible through any of the URR pages, provide detailed guidance to all tasks related to publishing, viewing, searching, and analyzing URR contents. They also provide extensive background on technical aspects of the URR and on approaches to the analysis of user needs.

A number of publications provide additional information on the URR:

- Plag et al. (2010): a general overview at a non-technical level;
- Plag et al. (2011): focus on the dialog between users and GEOSS providers;
- Plag et al. (2012a): focus on cross-links between SBAs.

ACKNOWLEDGMENTS

The development of the URR and the population of the URR was supported by the United States Environmental Protection Agency (EPA).

REFERENCES

URR Team, 2011. Concept of Operations Document User Requirements Registry (URR), Tech. Rep., Available at http://www.scgcorp.com/urr/pdfs/URR_COO_FINAL.pdf.

GEO, 2011. GEO Work Plan 2012-2015. Available at http://www.earthobservations.org/documents/work plan/GEO 2012-2015 Work Plan_Rev1.pdf.

GEO, 2009. GEOSS Strategic Targets. Document 12(Rev1) as accepted by GEO-VI. Available at

http://www.earthobservations.org/documents/geo_vi/12_GEOSS Strategic Targets Rev1.pdf.

GEO Gap Analysis Task Force, 2011. Development of a GEOSS Gap Analysis Strategy. Document 13, presented to the 21st GEO Executive Committee Meeting, March 21-22, 2011.

Plag, H.-P., Ondich, G., Kaufman, J., Foley, G., 2010. The GEOSS User Requirement Registry - Supporting a User-Driven Global Earth Observation System of Systems. *Imaging Notes*, **25**(4), 28-33.

Plag, H.-P., Ondich, G., Kaufman, J., Foley, G., Pignatelli, F., 2011. The GEOSS User Requirement Registry: A Versatile Tool for the Dialog Between Users and Providers. Proceed. 34th Internat. Symp. Remote Sensing of the Environment, Sydney, April 10-15, 2011, available at http://www.isprs.org/proceedings/2011/ISRSE-34/211104015Final00579.pdf.

Plag, H.-P., Foley, G., Jules-Plag, S., Kaufman, J. and Ondich, G., 2012a. The GEOSS User Requirement Registry: Linking users of GEOSS across disciplines and societal areas. Invited abstract, IGARSS 2012, Munich, July 23-27, 2012.

Plag, H.-P., et al., 2012b. Prioritizing, identifying gaps, and assessing relevance of Earth observations: a user-based view. In preparation.

Plag, H.-P., Pearlman, J., Pearlman, F., 2012c. A measure for relevance of user needs, requirements, and Earth observations. In preparation.

AUTHORS, CONTRIBUTORS, REVIEWERS

Contact author:
Hans-Peter Plag, IEEE and University of Nevada, Reno, Nevada, USA (hpplag@unr.edu)

Contributing author:
Shelley Jules-Plag, Tiwah, Inc., Reno, Nevada

Contributors:
Gary Foley, EPA
Greg Ondich, SCG, Inc.
Justin Kaufman, SCG, Inc.
Ric Blackman, SCG, Inc.
Steve Browdy, IEEE

Reviewers:
Francoise Pearlman, IEEE
Jay Pearlman, IEEE
GCI Provider Team Members
Jennifer McCulley, SCG, Inc.