RARE

(*Rapid Response Support Server*)

Executive Summary Report

Title: RARE Executive Summary Report
Author: Bernard Valentin (SpaceApps)
Contributors: Horacio Brizuela (Epistematica)
Checked by: Leslie Gale (SpaceApps)
ESA Contract No: 4000102859/11/I-NB

European Space Agency

Contract Report

The work described in the report was done under ESA contract. Responsibility for the content resides with the author or the organization that prepared it.

Technical Officer:	Philippe Mougnaud	ESRIN
Contract Officer:	Nathalie Boisard	ESRIN
Distribution:	Philippe Mougnaud	ESRIN
	Andrea Della Vecchia	ESRIN
	Richard Aked	Space Applications Services
	Rani Pinchuk	Space Applications Services
	Bernard Valentin	Space Applications Services
	Horacio Brizuela	Epistematica
	Project File	Space Applications Services

ESA STUDY CONTRACT REPORT		
ESA Contract N^{o} 4000102859/11/I-NB	SUBJECT RARE Executive Summary Report	
ESA CR (..)N^{o}	STAR CODE	

ABSTRACT:

In most domains, for potential users (so-called "end-users") the access to Earth Observation products, processes and services is still not straightforward. End Users know which information they want, can describe it in their own terminology but cannot identify the EO resources they want because the resources are not described in terms they are familiar with. The Rapid Response Support Server (RARE) overcomes this problem by introducing two important user oriented features. Firstly, by enabling users to express their needs in terms they are familiar with, and secondly, by allowing them to use natural language structures to more easily identify the resources.

The RARE project has achieved this user oriented approach by applying semantic resolution of textual user queries to map user requests to protocol specific search criteria to find the most appropriate Earth Observation resources meeting the requirements of the user. The project has resulted in the development, testing and delivery of a prototype system.

The RARE prototype is a complex, distributed system that involves Web-based technologies (incl. HTML, HTTP, JavaScript, CSS, AJAX, and portlets), service-oriented technologies (incl. SOAP, REST, servlets, and RDBMS), semantic technologies (incl. SPARQL, SKOS, SWRL), and geospatial technologies (incl. OGC standards and software tools).

This document summarizes the results of the work performed in the RARE project.

The work described in the report was done under ESA contract.

Responsibility for the content resides with the author or the organization that prepared it.

Names of authors: Bernard Valentin, Horacio Brizuela

NAME OF ESA STUDY MANAGER Philippe Mougnaud DIV: DIRECTORATE:	ESA BUDGET HEADING

Document History

Date	Version	Changes
1-Sep-2014	1.0.0	RARE AR Issue

Table of Contents

1 Scope and Applicability

1.1 Purpose of the Document

This **Executive Summary Report** has been produced for the *Rapid Response Support Server* project **RARE** under the European Space Agency's contract number **4000102859/11/I-NB**. It is part of the result of the work performed in WP 1.0 "Management".

In this document a summary is given of the work performed during the project.

The work has been performed by Space Applications Services, S.A., in collaboration with Epistematica S.r.l. and contributions from ERDAS/Intergraph Belgium.

1.2 Document Organization

Apart from this first chapter, this Executive Summary Report is structured as follows:

- Chapter 2 "Motivation", page 9, introduces the motivation and the context which lead to the execution of the RARE project;
- Chapter 3 "The RARE System", page 10, provides an overview of the key elements of the RARE system, including the user interface, the multi-domain ontology, the reasoning rules, the disambiguation mechanism, and the search engine.
- Chapter 4 "Evaluation and Conclusions", page 20, summarizes the outcome of the evaluation activity, commenting on the users' feedback, and provides a conclusion.

1.3 Acronyms

AD	Applicable Document
AJAX	Asynchronous JavaScript and XML
API	Application Programming Interface
ATP	Acceptance Test Plan
CSS	Cascading Style Sheet
CSCDA	Copernicus Space Component Data Access (formerly GSCDA)
CSW	Catalogue Service for the Web
EO	Earth Observation
ESA	European Space Agency
ESRIN	European Space Research Institute
IAM	Installation and Administration Manual
ICD	Interface Control Document
OGC	Open Geospatial Consortium
OTS	Ontology and Terminology Service
RARE	Rapid Response Support Server
RD	Reference Document
RDBMS	Relational Database Management System
REST	Representational State Transfer
RRC	Rapid Response Client
RRS	Rapid Response Server
SKOS	Simple Knowledge Organization System

SOAP	Simple Object Access Protocol
SPARQL	SPARQL Protocol and RDF Query Language
SUM	Software User Manual
W*S	Web Map Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS), Web Processing Service (WPS)
XML	Extensible Mark-up Language

1.4 Applicable Documents

This document incorporates, by dated or undated reference, provisions from other publications. These normative references are cited at appropriate places in the text and publications are listed hereafter. For dated references, subsequent amendments to, or revisions of, any of these apply to this document only when incorporated in it by amendment or revision. For undated references, the latest edition of the publication referred applies.

ID	Reference
AD01	**ECSS-E-ST-40C** Software Engineering – Software Issue 3, dated 6-Mar-2009
AD02	**ECSS-Q-ST-80C** Space Product Assurance – Software Produce Assurance Issue 3, dated 6-Mar-2009
AD03	**RARE Technical Proposal** Reference: RARE-PROP-001 Issue 1.0.0, dated 29-Sep-2010
AD04	**RARE Negotiation Meeting Minutes** Reference: MOM-RARE-ESRIN-001 Issue 1.0, dated 22-Dec-2010
AD05	**MOM 2: Architecture Requirements Collection** Reference: RARE-MOM-02 Issue 1.0, dated 11-Mar-2011
AD06	**MOM 3: Use of Remaining Work Hours from SEPR** Reference: RARE-MOM-03 Issue 1.0, dated 28-Jul-2011
AD07	**MOM 4: Extensions To Epistematica's Activities** Reference: RARE-MOM-04 Issue 1.0, dated 22-Sep-2011
AD08	**Project Management Plan** Reference: RARE-SA-PMP Version 1.0.0, dated 21-Dec-2010
AD10	**Semantic Architecture Review Technical Note** Reference: RARE-TNO-SemArch Version 3.0.0 dated 5-Dec-2011
AD11	**Standards for Ontology Integration Technical Note** Reference: RARE-TNO-OntInt Version 4.2.0 dated 13-Jul-2013

ID	Reference
AD12	**Software System Specification** Reference: RARE-SA-SSS Version 2.0.0, dated 12-Dec-2011
AD13	**Software Requirement Specification** Reference: RARE-SA-SRS Version 3.1.0, dated 21-Dec-2012
AD14	**Software Design Document** Reference: RARE-SA-SDD Version 2.5.0, dated 1-Sep-2014
AD15	**Interface Control Document** Reference: RARE-SA-ICD Version 2.5.0, dated 1-Sep-2014
AD16	**Acceptance Test Plan** Reference: RARE-SA-ATP Version 3.0.0, dated 4-Mar-2014
AD17	**Resources Survey Technical Note** Reference: RARE-TNO-ResSurvey Version 2.2.0, dated 1-Sep-2014
AD18	**Software User Manual** Reference: RARE-SA-SUM Version 1.2.0, dated 17-Dec-2013
AD19	**Installation and Administration Manual** Reference: RARE-SA-IAM Version 1.3.0, dated 25-Feb-2014
AD20	**User Evaluation Report** Reference: RARE-SA-UER Version 1.0.0, dated 1-Sep-2014
AD21	**Software Release Document** Reference: RARE-SA-SRelD Version 1.0.0, dated 1-Sep-2014
AD22	*Reserved to refer to this Executive Summary Report in other documents*
AD23	**Final Report** Reference: RARE-SA-FR Version 1.0.0, dated 1-Sep-2014
AD30	**EOP-G Technical Baseline** Reference: OSMV-OPMT-EOPG-TN-10-0001 Issue 1.1 for approval, dated 2010
AD31	**RSS Change Management Process** Reference: OSMV-OPMT-LOGI-PR-11-0699 Issue 1.0, dated 1-Apr-2011
AD32	**RARE Contract Change Notice 1** Reference: RARE-SA-PROP-AddActivities Issue 1.3.0, dated 6-May-2011

ID	Reference
AD33	**RARE Contract Change Notice 2** Reference: RARE-SA-PROP-AddActivities2 Issue 1.0.1, dated 22-Feb-2012
AD34	**ERDAS SSOI Test Validation Report** Operation and Maintenance of the EO Payload Data Systems Reference: SWENG-ERDAS-SSOI-TVR Issue: 1.0, dated 31-Aug-2012
OGC-SemAnnot	**Semantic Annotations in OGC Standards** Reference: OGC 08-167r2 Version 2.0 dated 10-Oct-2012 http://portal.opengeospatial.org/files/?artifact_id=47857

2 Motivation

For end users in most domains, access to Earth Observation products, processes and services is still not straightforward. End Users know which information they want, can describe it in their own terminology but cannot identify the EO resources they want because the resources are not described in terms they are familiar with.

RARE overcomes this problem by introducing two important user oriented features. Firstly, by enabling users to apply the terms they are familiar with, and secondly, by allowing them to use natural language structures to more easily identify the resources. RARE does this by establishing vocabularies of the various Copernicus (formerly GMES) main thematic areas[1], applying existing metadata and terminology standards, performing natural language processing and word sense disambiguation to understand users requests and developing reasoners that allow the accurate linking of user terms to products characteristics.

A depth of knowledge and experience in semantics and semantic technology is necessary to obtain a system that behaves adequately. For this reason Space Applications Services NV, which is knowledgeable of Earth Observation missions, performs European natural language and semantic research, develops software solutions incorporating semantics and is experienced with ECSS software engineering, has teamed with Epistematica Srl which has developed precursor components of RARE and has an exemplary record in knowledge engineering as well as designing and developing working semantic systems for ESRIN.

Recognising that achieving a high quality solution is necessary if end users are to be satisfied with RARE, we have involved directly end user representatives from the EC's Copernicus core services projects and to ensure that RARE will be interoperable with existing and future EO resource providers and consumers we have involved consultants expert in OGC and related standards.

[1] Copernicus main thematic areas: http://www.copernicus.eu/pages-principales/services/.

3 The RARE System

RARE is a complex, distributed system that involves Web-based technologies (incl. HTML, HTTP, JavaScript, CSS, AJAX, and portlets), service-oriented technologies (incl. SOAP, REST, servlets, and RDBMS), semantic technologies (incl. SPARQL, SKOS, SWRL), and geospatial technologies (incl. OGC standards and software tools).

The following sections describe the key components of RARE. They are thus not meant to exhaustively cover all the RARE components.

3.1 The User Interface

The Web-based user interface allows searching, disambiguating, inspecting and accessing the search results in a single page, much like the familiar Google Search page.

The initial version of the user interface followed the "wizard" mechanism: a series of individual Web pages were used to progress in the search process. This was considered too complex and not intuitive enough for most of the users. A new search page was designed and implemented that combines the different steps of the wizard: disambiguation, search, inspection and access.

The combined search page exploits the AJAX technique to communicate with the server and update the display without freezing the user interface and without requiring the whole page to be reloaded.

The page has been designed to hide the complexity of the internal communications and processing as much as possible.

Figure 1 shows the page as it is initially presented: it contains a unique input field, highlighted in yellow, drop-down lists allowing to restrict the search to specific resource types (EOP, WMS, WFS, ISO Data, etc.) and resource providers, a link allowing to insert application terms using the Multi-Domain Ontology Browser (as shown on Figure 3) and, in the left margin, a "Multi-Criteria …" button giving access to another version of the page that makes use of traditional, non-semantic search techniques.

The "Search" button must be used to initiate a search.

As can also be deduced on the figure the RARE user interface has been implemented as a Liferay portlet. This allows integrating the RARE in any pre-existing or newly deployed Liferay portal.

Figure 1 – RARE Combined Search Page (initial state)

Figure 2 shows the page as it appears after a search has been performed.

When the user presses the "Search" button, the textual query is disambiguated and the interpretation is displayed right below the input field. In the figure, the search term "disaster" has been found in the "GSCDA" ontology[2]. In this case, the interpretation is not considered as ambiguous by the system, the matching resources are thus searched for automatically and displayed. When an interpretation is ambiguous, the user is asked to select among proposed meanings the once that suits his needs before resuming the search.

The matching resources are listed directly below the disambiguation information. A number of controls provided in the left margin allow filtering and sorting the entries in the page. No new search requests are issued while using these controls.

The content of each result entry depends on the information available about the corresponding resource. The entries represented in the figure contain a title, a description, keywords and timestamps, source links and a footprint drawn on an interactive map.

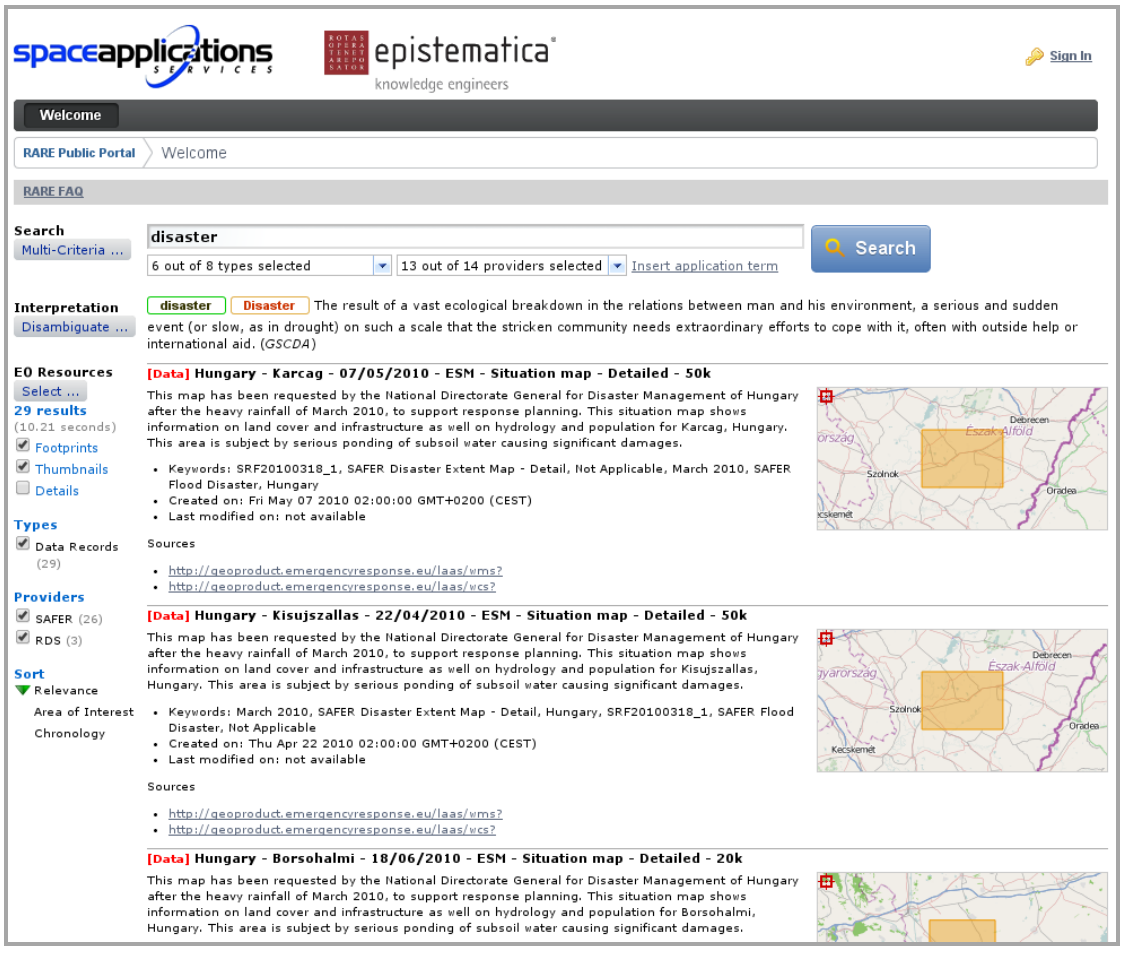

Figure 2 – RARE Combined Search Page (after a search)

The disambiguation service is configured with the ontology that covers the CSCDA thematic domains. Any searched term that match one of the ontology concepts is automatically recognized (associated the corresponding meaning) and given the highest priority. Identified

[2] GSCDA refers to GMES, the former name of Copernicus. GSCDA must now be referred to as CSCDA, the Copernicus Space Component Data Access.

toponyms and time expressions are given a lower priority and the named entities and common nouns are given the lowest priority. The system then uses the priorities to compute and propose a default query interpretation.

In this process, when a user wants to search about a domain concept it is important that the system identifies it correctly. The user may type in directly the terms in the input field, which is prone to mistakes, or use the multi-domain ontology browser to navigate in the ontology and select the appropriate concept.

Clicking on the "Insert application term" link located below the search input field accesses the ontology browser. The ontology browser, represented in Figure 3, below, is displayed. This shows an interactive and animated tree-like representation of the ontology. Information about the selected concept (here: "Oil Pollution") is displayed below. Clicking the button "Use Selected Term" in the top right corner allows copying the concept label into the query string.

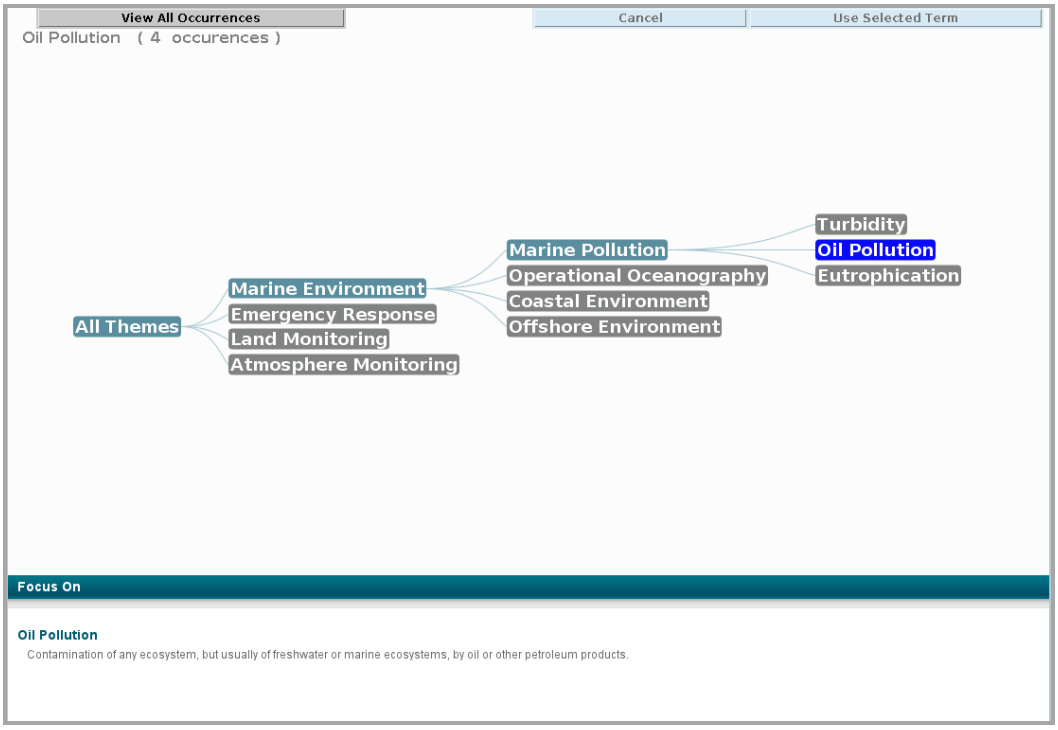

Figure 3 – Ontology Navigation Page: Oil Pollution Application Term

The multi-criteria search page mentioned above shows traditional search fields that allow searching for specific property values in the resources metadata (see Figure 4). The values entered in the first three fields are searched in respectively the title, the description and the keywords of the resources. The resources types and resources providers lists allow reducing the scope of the search, as already described above. An area of interest (bounding box) may be specified in the "Where" field to filter out the resources whose footprint does not overlap. The "Where" field allows to limit the search to the resources produced within a given date range (time of interest).

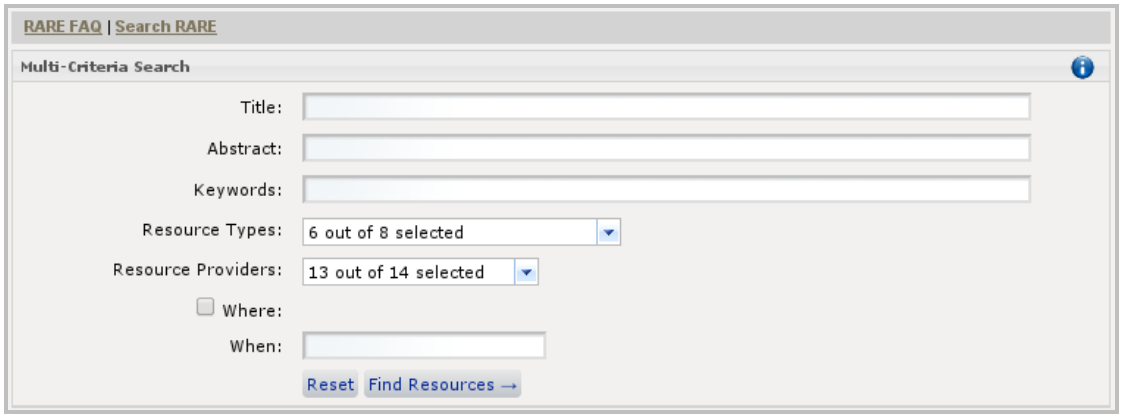

Figure 4 – Multi-Criteria Search Form

A Web-based management interface implements a number of functions strictly restricted to the user authenticated in the system and having the administrator role in their profile. The following management functions are available:

- RARE front-end and back-end services testing.
- Management of the user profiles and generation of usage statistics.
- Management of the RARE catalogue directory, scheduling of the harvesting of remote services, and configuration of the catalogue broker GI-cat.
- Management of the disambiguation modules (ontology, toponym, and time recognizers, named entity recognizer, etc.): enable, disable and re-weight the modules.
- Validation and deployment of new versions of the multi-domain ontology.
- Validation and deployment of new versions of the reasoning rules.

The end-user interface is described in details in the RARE Software User Manual (SUM) while the administration interface is covered in the RARE Installation and Administration Manual (IAM).

3.2 The Multi-Domain Ontology and the Terminology Service

The Multi-Domain Ontology is a knowledge base integrating the ontologies of various application domains. It has been set up by the OTE, OTEG and SEPR projects. It has been extended in RARE to include the concepts required to perform the use cases.

The Ontology and Terminology Service has been built to hold the Multi-Domain Thesaurus and to allow performing remote queries on it. The Ontology and Terminology Service is a major constituent of the semantic architecture of RARE.

The Multi-Domain Ontology

The structure of the Multi-Domain Ontology was initially encoded in an OWL file while the properties of the ontology concepts were stored in an Excel table. This encoding of the ontology facilitates the modifications made by people not familiar with semantic technologies. The main disadvantage is however that the format is not standard and thus does not allow using existing advanced tools such as ontology editors and navigators.

During the RARE project an analysis of new standard languages was carried out. It was determined that SKOS is the most popular and performing language. Access to the ontology

is faster, thanks to the OWL ART API libraries[3] (*Open Source libraries developed by ART Group at Tor Vergata Rome University*) that index the ontology in memory and can query the SKOS language. Another reason to choose SKOS is that it became a standard for Thesauri representation shortly before the start of the project.

The new and extended version of the Multi-Domain Ontology prepared in RARE is thus encoded in that SKOS format.

The **Thesaurus Sanity Checker** is an independent Java library that implements a series of operations which allow checking the consistency and the integrity of a Multi-Domain Ontology. This is used by RARE to make sure that new uploaded versions of the ontology are valid. If errors are detected in an ontology, its deployment is aborted.

The Ontology and Terminology Service

The Ontology and Terminology Service (OTS) is a Web Service previously developed in the OTEG project that uses the Multi-Domain Ontology as a knowledge base to provide matching operations between searched terms and applications terms (concepts). It also provides operations allowing to navigate between related application terms and to obtain their textual description.

The Ontology and Terminology Service has been given a double SPARQL / OpenSearch interface: the SPARQL interface allows navigation in the Ontology, while the OpenSearch interface allows executing textual queries on it.

3.3 The Query Disambiguation Service

The user query disambiguation mechanism, which has already been mentioned in the description of the user interface, is implemented as a service known in the project as the Query Analyzer. The service exposes a custom SOAP-based interface which receives a textual query, processes it, and returns an XML-encoded document that contains the generated disambiguation data.

Internally, the Query Analyzer invokes a series of modules, each of which targets a specific type of semantic interpretation (see Figure 5). For instance the "Application Term searcher" module tries to interpret words from the query string by searching these words in the multi-domain ontology. Figure 5 shows the external interactions that take place between the client applications (Rapid Response Client – RRC, Rapid Response Server – RRS) and the Query Analyzer as well as the internal interactions between the Query Analyzer and each of its modules.

The returned document (identified as the Normalized Query in the figure) contains all the meanings identified by the modules for each search term or combination of terms. For example, "English Channel" is a combined search term that would be identified as a place name.

[3] OWLART API Libraries: http://art.uniroma2.it/owlart/.

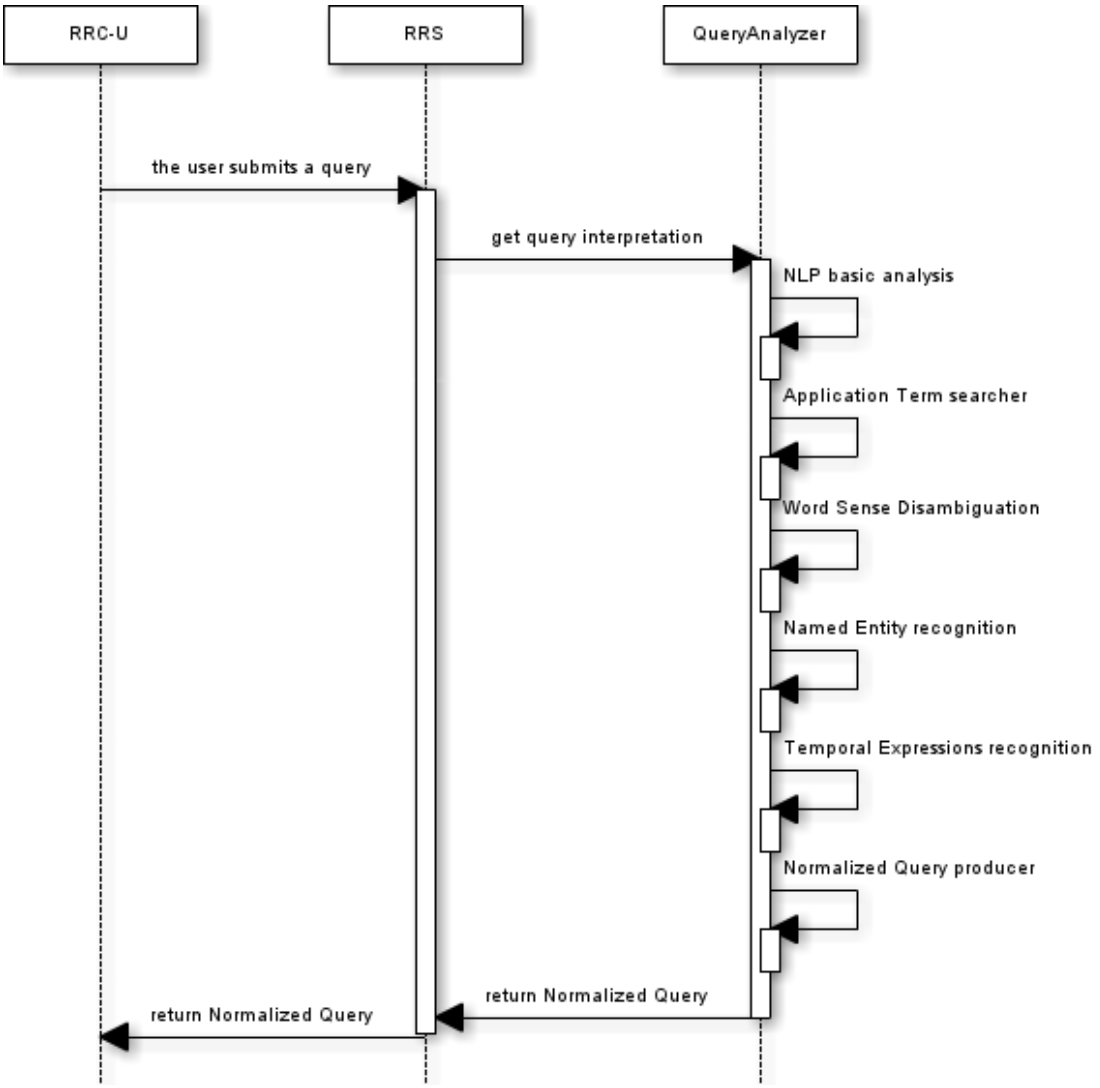

Figure 5 – Query Analyzer Main Loop

The modular nature of the Query Analyzer makes it possible to enhance the existing modules, and to create and integrate new modules supporting new types of interpretations. This also allows the modules to be enabled, disabled or re-weighted through the management interface mentioned earlier.

A separate Java library has been implemented that is meant to be integrated in the client applications. This library provides utility functions that may be used to process the normalized queries obtained from the Query Analyzer and extract from each of then the interpretation that has the highest combined weight. This is the interpretation proposed by default on the user interface.

3.4 The EO Resources Reasoners and Reasoning Rules

The EO Resources Reasoners

The EO Resources Reasoners are Web Services implemented in Java which provide the mapping between Application Terms (as described in the Multi-Domain Ontology) and EO Resources (i.e., products and collections) Characteristics.

Given an Application Term, an EO Resources Reasoner provides resource characteristics, i.e., information that can be used to create a query to OGC CSW catalogues, thus retrieving all the relevant EO resources.

Three reasoners have been implemented to allow searching for EO resources. Each reasoner provides in its response the search criteria applicable to a given type of resource. The following resource types are supported:

- the EOP extension package of the ebRIM application profile,
- SensorML catalogue entries, and
- any kind of entry annotated with traditional keywords.

The Reasoning Rules

The Reasoners are configured with a collection of RIF reasoning rules mapping each Application Term to a set of Application Requirements (e.g. Band and Resolution values). The rules are common to all three Reasoners.

Each Reasoner also contains a conversion table that defines the mapping between Application Requirements to Resources Characteristics. These conversion tables are used to obtain Resource Characteristics in all the supported formats: EO GML, SensorML, and keywords.

The **EO Reasoning Rules Sanity Checker** is an independent library that implements a series of operations which allow checking the consistency and the integrity of a set of reasoning rules and the accompanying translation tables. This is used by RARE to make sure that the new version of a reasoning rules package is valid. If errors are detected, the deployment of the new rules is aborted.

3.5 The RARE Catalogue

The RARE catalogue implements a service that stores locally the resources metadata harvested from remote data, processing and cataloguing services. The RARE catalogue is powered by the OGC-compliant commercial product GeoRegistry from Intergraph[4].

The data models supported directly from the shelf include (omitting the version information): EO O&M, EO GML, SensorML, ISO Data and Services, OWS Common, WMS, WFS and WCS capabilities.

GeoRegistry provides a Java-based API that permits to implement extended features and add support for more data models. In particular, the API allows developing specific modules called *handlers*. These handlers communicate with remote services using their formats and protocols, extract the useful information from the request responses, and store the resulting data and metadata in the catalogue registry.

The following two key features required in RARE were missing in GeoRegistry:

- GeoRegistry is capable to harvest remote cataloguing services, that is, store locally the capabilities and other characteristics of these services, but not to duplicate locally the records present in the remote catalogues. It thus also means GeoRegistry does not automatically discover the services registered in remote catalogues.

  The RARE catalogue must be capable to retrieve the records registered in remote catalogues and, if appropriate, ask GeoRegistry to ingest the corresponding services.

[4] GeoRegistry was known as ERDAS HMA Component in the recent years. ERDAS has been integrated in Intergraph and it has now been re-shaped as Hexagon Geospatial.

This mechanism is required to discover the layers, the feature types, the coverages and the processes supported by remote OGC W*S services.

- GeoRegistry contains handlers for importing the metadata of WMS, WFS and WCS services, EOP records and SensorML records. It does not support WPS services. This is however needed by RARE to discover the processes

Two GeoRegistry extensions have thus been implemented during the project: a catalogue harvester and a WPS handler. These are described shortly hereafter.

Catalogue Harvesters

The purpose of the Catalogue Harvesters is to gather metadata records from remote catalogues, and prepare them for storage in the RARE Catalogue. The storage itself is then performed by the appropriate GeoRegistry handlers. For example, the built-in WMS handler is activated by the harvesters on each record representing a WMS service.

In the course of the project, harvesters have been implemented for supporting OGC CSW ISO catalogues and OGC SOAP/CSW catalogues.

A catalogue harvester performs the following actions each time it is invoked for harvesting a given catalogue service:

1. Retrieve the configuration of the catalogue service to be harvested (service type and version, end-point, etc.)
2. Communicate with the remote catalogue service (OGC CSW, ISO, or others) to obtain metadata records. When supported by the remote service, only metadata records modified since the last harvesting are retrieved.
3. Parse the gathered metadata records, and prepare them for the storage in the RARE Catalogue.
4. Identify the type and data model of each record and invoke on each of them the suitable GeoRegistry handler. The handlers are responsible for fetching the original metadata, converting them into an ebRIM metadata structure, and for storing the result in the registry.
5. Annotate the harvested metadata records with two additional properties – a Priority and possibly a Provider. This is done through the GeoRegistry Java API.

The harvesters have been successfully used to ingest records hosted by MyOcean, VITO, FedEO, RDS/CSCDA, and Disasters Charter. Because the related projects evolve and reach completion at some point, the remote catalogues may not be accessible anymore at the time of reading this report. This is in particular the case of Disasters Charter which replaced its OGC CSW ebRIM catalogue with an OpenSearch one.

A separate RARE component, not covered in this summary report, integrates a scheduler to initiate the harvesting of remote services at regular interval. This permits to synchronize the records, detecting the additions, changes and deletions with a small delay.

WPS Handler

As introduced above, the WPS handler is an additional module implemented for RARE that adds support for ingesting WPS and processes metadata into GeoRegistry.

The WPS handler performs the following actions each time it is invoked for ingesting a given processing service:

1. Use the service end-point URL to retrieve the capabilities information of the processing service. This information includes metadata about the service itself (type, version, name, etc.) as well as the name of each implemented process.
2. Retrieve specific information about each implemented process, including the description, the input and output parameters.

3. Prepare the data for storage in the RARE Catalogue.
4. Add the associations between the service record, the process records and the parameter records in compliancy with the WPS data model.
5. Store the structure in the RARE Catalogue.

Ingested WPS processes are searchable through the RARE Web interface. A proof-of-concept mechanism has also been implemented that allows activating a process on user-defined input and delivering the outputs to the requesting user.

3.6 The Search Engine

The RARE search engine is implemented as a service that sits in the component named Rapid Response Server (RRS).

The engine exposes two key operations which map to the behaviour of the combined search page described in section 3.1, on page 10. Concretely, the first operation receives a user query, invokes the query disambiguation service (described in section 3.3), and returns the disambiguation information. The second operation receives a disambiguated query, that is, a series of terms and meanings, transforms this query into technical search criteria, uses the search criteria to look for Earth Observation resources in the RARE catalogue, and returns the metadata of the matching resources.

Search Query Disambiguation

This step is entirely performed by the query disambiguation service already described in section 3.3. The task of the search engine is to act as a proxy, verifying the input parameters, invoking the disambiguation service, and returning the results.

The engine is also responsible for logging the search queries in an internal database. Search queries are linked to the profile of the authenticated users. This mechanism allows:

- the users to inspect their search history;
- the administrators to generate statistics on the search queries and terms.

Statistics are important as they permit to identify the popular search terms that deserve to be treated differently, for example by adding them to the multi-domain ontology and the reasoning rules.

In the future, this information may also be used as input to the Query Analyzer to derive search contexts, automatically augment search queries, and refine the prioritization mechanism of the senses.

Search for Earth Observation Resources

Disambiguated user queries contain a series of meanings associated to each search term or combination of terms. A particular combination of meanings is referred to as an Interpreted Query. A default interpreted query is automatically proposed by the system when the disambiguation data allows identifying the most probable meaning for each term. Otherwise, the user is invited to select the appropriate meanings, as described in section 3.1 dedicated to the user interface.

The search engine uses the information stored in an interpreted query to search for Earth Observation resources. Concretely, the following actions are performed each time the search operation is called:

1. Reception of an interpreted query and associated search parameters (including the list of resource types and resources providers to be considered).
2. For each Application Term (multi-domain ontology concept) found in the interpreted query, invocation of the Resources Reasoners for obtaining protocol-specific search constraints (see section 3.4, page 15).

3. For each keyword (plain text), location (bounding box) and time range included in the query, creation of a list of search constraints matching the query element.
4. Generation of a set of OGC Filters based on the search constraints created in step 2 and step 3.
5. Submission of the filters to the CSW catalogues. Two catalogues are queried: the RARE Catalogue (see section 3.5) and the GI-cat catalogue broker.

   GI-cat allows RARE to communicate with non-standard external services. It exposes an OGC CSW ISO interface to client applications and allows querying various types of services in the background (including OpenSearch, CF-netCDF and GeoRSS services). Information about GI-cat may be found on-line (see http://essi-lab.eu/do/view/GIcat). It is not further described in this summary report.
6. The resource records are collected and merged in a unique collection of heterogeneous entries. This heterogeneous collection allows the service to manipulate resources of different types in a generic manner.
7. For each entry, computation of sort keys related to different sort criteria (e.g. relevance, geographical proximity and chronology).
8. The collection is returned to the calling application. In the context of RARE, the result list is displayed in the Web user interface as shown in Figure 2 on page 11.

The search engine implemented as a service makes it possible to integrate it in other client applications.

4 Evaluation and Conclusions

The main achievements of the RARE project are:

- A multi-domain ontology covering the main Copernicus thematic areas has been defined and implemented.
- A service for searching and navigating in this ontology has been implemented.
- A modular query disambiguation service has been implemented. This makes use of plug-ins to disambiguate search terms, assign possible meanings to each term and combination of terms, and compute the more probable interpretation of the original query.
- Reasoning rules have been designed and implemented. These rules map application terms (ontology concepts) to protocol-specific search constraints.
- Services for applying the reasoning rules have been implemented.
- A modular search engine implemented as a Web service provides a centralized interface for disambiguating queries and searching for EO resources. The modularity of the engine will permit to extend the component to add support for new types of data sources and new types of resources.
- A Web application has been developed which makes use of recent technologies to expose a dynamic and responsive user interface. Moreover, this application implemented as a Liferay portlet may be integrated in pre-existing or newly deployed Liferay portals.
- An heterogeneous OGC CSW ebRIM catalogue product has been integrated and extended to support the harvesting of different types of remote services, including CSW ISO (data and service entries), CSW ebRIM (EOP, CIM, SensorML EPs), SOAP/CSW ebRIM, WMS, WFS, WCS, and WPS.
- Products may be generated asynchronously when requested by a user by executing a process implemented in an OGC WPS service. This capability has been implemented and tested for a single process accepting simple arguments. The process execution engine is however generic and may thus be extended by means of plug-ins to add support for new services and processors.

4.1 Evaluation

At the end of the project, an evaluation questionnaire has permitted to collect comments and suggestions from end-users. In general, the users have considered the user interface as intuitive and easy to operate. The users are satisfied with the default interpretations of their queries and the resources listed in the result list are relevant to the original query.

Reported annoyances and weaknesses mostly concerned the following topics, which thus require a particular attention for the future works:

- The Java applet technology used to implement the multi-domain ontology browser has caused issues in particular because it requires a Java runtime environment and a browser plug-in to be properly installed on the client system.
- The relation between the search queries and the displayed resources is not always understood. This is due to the rule-based mapping performed internally by the system and hidden to the users.
- The lack of flexibility of the disambiguation service has also been mentioned. Synonyms and derived terms should be added in the future.
- Users sometimes consider that the system takes too much time to return the search results. This is mainly due to the amount of internal catalogue queries that need to be issued to answer a single user query.

4.2 Conclusions

The obtained results are promising. During testing users found the resources proposed by the RARE system relevant but experienced some difficulty in understanding the relation between their original queries and the proposed resources. This implicit and non-trivial relation is due to the internal semantically-based processing of the queries performed by the system and that the user is not aware of. Providing to the users feedback from the reasoning process in a user understandable form may mitigate this issue.

Using the currently available catalogues displayed results do not always suit the users' needs. There is a very limited amount of resources registered in the catalogue, and most of these resources are derived products. This dependency between the size of the catalogue and the relevance of the results is particularity inherent to any search engine. Increasing the amount of entries registered in the catalogue is expected to have a positive impact on the relevance of the best ranked resources.

There is sufficient interest to warrant extending RARE to include delivery of products based upon the results of the user’s query.

Additional enhancements are foreseen to increase the value to the user. In particular the extension to additional catalogues.

RARE is available under a free-of-charge license to ESA project partners.