Earth Observation Image Librarian

EOLIB SCENARIOS

Prepared By	Mario Winkler (Digital unterschrieben von Mario Winkler DN: cn=Mario Winkler, o=DLR, ou=DFD-IT, email=mario.winkler@dlr.de, c=DE Datum: 2012.05.24 16:11:22 +02'00')	2012-05-24
	Mario Winkler (DLR-DFD)	Date
	Mihai Datcu (DLR-MF)	
	Katrin Molch (DLR-DFD)	
Released By	Dietmar Preuß (Digital unterschrieben von Dietmar Preuß DN: cn=Dietmar Preuß, o=DLR, ou=DFD-IT, email=dietmar.preuss@dlr.de, c=DE Datum: 2012.05.24 16:28:46 +02'00')	2012-05-24
	Dietmar Preuß (DLR-DFD)	Date
Approved By		
	Jordi Farres (ESA)	Date

DOCUMENT CHANGE RECORD

Issue	Date	Change
1.0	2011-09-29	Draft for project kick-off
	2012-01-20	Restructured document, added scenarios and use cases
	2012-02-07	Added introductory parts and scenarios from Stakeholder Workshop hand-out
2.0draftE	2012-02-10	Draft release for ESA review
2.0	2012-04-02	Update in response of comments from Michele Iapaolo: • Added description of System Context diagram [Figure 2-4] • Epitome is generated in step 8 of SCE-Generate-Semantic-Catalogue [§3.1] • Updated Use Case Diagram [Figure 4-1] • Added the UC-Automatic-Annotation actor [§4.4.2] • Added the epitome to UC-Export-Results [§4.5.2]
2.1	2012-05-23	Updated user scenarios provided by Mihai Datcu: • Added scenario Epitome usage • Added scenario Generation of Content Taxonomies and Application/Domain Ontologies • Added scenario Quantitative Semantic Queries • Added scenario Support for Training Image Analysts • Added scenario Automatic Image Processing in CGS's value-adding Production System • Added scenario Collaborative Ground Segment Tool • Added scenario Support for Image Quality Assessment • Added scenario Collaboration with Scientific and User Communities Renamed user scenarios to improve top-down understanding of EOLib system: • Renamed scenario SCE-Generate-Semantic-Catalogue to Image Processing and Epitome Generation • Renamed scenario SCE-Semantic-Queries to Image Retrieval using Semantic Queries • Renamed scenario SCE-Knowledge-Discovery to Image Analysis and Knowledge Discovery • Renamed scenario SCE-Rapid-Mapping-Support to Support for Rapid Mapping Applications • Renamed scenario SCE-EOLib-Evaluation to EOLib Evaluation
	2012-05-24	Included review responses by Katrin Molch, Mihai Datcu and Christoph Reck

TABLE OF CONTENTS

LIST OF FIGURES

1. Introduction

The Earth Observation Image Librarian (EOLib) project serves to setup the next-generation Image Information Mining (IIM) system implementing novel techniques for image content exploration. EOLib utilizes information about Earth Observation (EO) product contents which is usually hidden in raster data, image time series and metadata, and enables content-based search in very large archives of high resolution EO data. EOLib will be interfaced to and operated in the DLR Multi-Mission Payload Ground Segment (PGS) of the DLR Remote Sensing Data Center.

1.1 Purpose and Scope

This document presents the EOLib scenarios and use cases.

Note: To insure the self-consistency of the document, several paragraphs from the EOLib SoW and Project Proposal are repeated.

1.2 References

The following project documents contain provisions which, through reference in this text, become applicable to the extent specified in this document. For versioned references, subsequent amendments to, or revisions of, any of these publications do not apply.

[1] EOLib Statement of Work — EOEP-EOMD-EOPG-SW-10-0002 Issue 0.9, 2011-03-28

[2] EOLib Technical and Management Proposal — EOLIB-DLR-PR-002 Issue 1.0, 2011-05-25

[3] EOLib Image Librarian Glossary — EOLIB-LST-DLR-005 Issue 1.2, 2012-04-17

[4] Entity DIMS Operations: Entity Identification Process — QMH-EOC-E.DIMS-E.ID 1.1

[Popescu2012] — Anca A. Popescu, Inge Gavat, and Mihai Datcu: Contextual Descriptors for Scene Classes in Very High Resolution SAR Images, *IEEE Geoscience and Remote Sensing Letters* Vol. 9, No. 1, January 2012

1.3 Terms and Abbreviations

The *EOLib Image Librarian Glossary* [3] provides a general glossary for terms, definitions, and acronyms used in all the EOLib project documentation.

1.4 Document Overview

This document has the following chapters:

Chapter 1 Introduction outlines the purpose of this document and lists the reference documents

Chapter 2 EOLib Overview describes the problem space addressed by the EOLib system and introduces the actors

Chapter 3 User Scenarios presents the scenarios

Chapter 4 Use Cases details the use cases

2. EOLib Overview

2.1 Content of Very High Resolution EO Images

With the increased resolution offered by modern space borne imaging sensors, the amount of available information calls for new mechanisms and processing techniques to handle, understand, and discriminate information.

In current high-resolution optical satellite imagery individual objects such as buildings, vehicles, or bridges are depicted separately and oftentimes by more than one image pixel. Classification approaches applied to lower resolution optical multi-spectral imagery, which were traditionally based on pixel-based analysis of multi-spectral signatures, are replaced by object-based analysis methods in high and very high resolution image data. Rule-based context information provides additional input for correctly classifying objects in their environment.

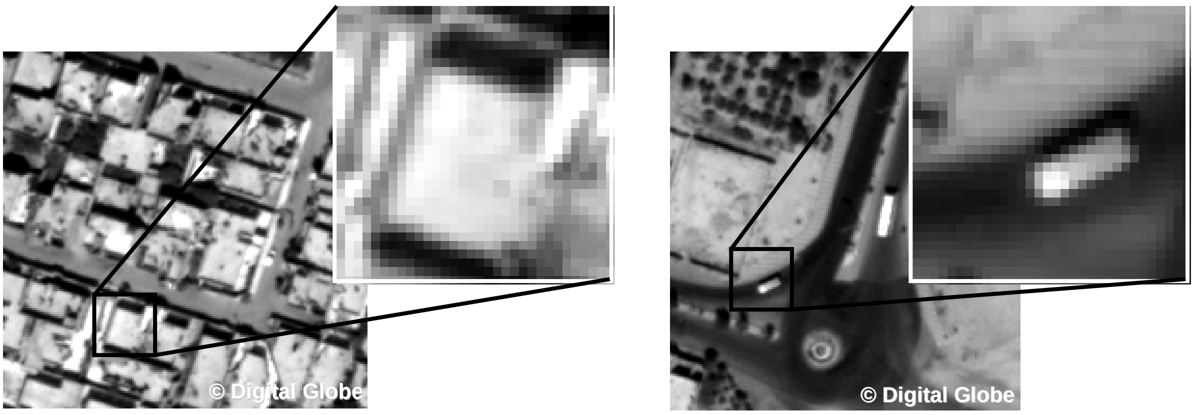

Figure 2-1: Content of optical panchromatic satellite image at half meter resolution

Modern SAR imaging systems expose a number of individual objects, which are comparable to the number of objects distinguishable by the human eye, and therefore, a solid comprehension of a high resolution (HR) SAR scene demands that hundreds of object/scene classes be defined and possibly identified.

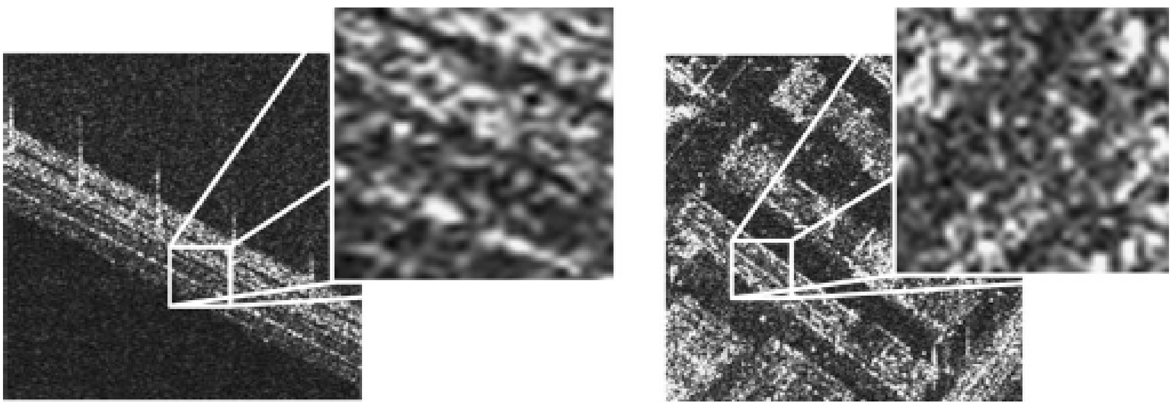

Figure 2-2: SAR Image Content at Meter Resolution

The figure above demonstrates that if context is ignored (the analysis windows being used are relatively small as compared to the object scale), very different scene classes may be confused. The example shows a bridge (left) and buildings (right) which are comprehensible only when the window size is sufficiently large to incorporate relevant context. Otherwise, both scene classes “look” similar and can be confused.

The HR SAR scenes used for this analysis contain complex arrangements of numerous and various objects, mostly man made. It is very unlikely that a scene shall contain two or more identical objects, and in addition, the same object may have different meanings in different scenes or arrangements. Context becomes an important source of information. The most important issue to be considered is the choice of the information extraction method to be employed. Complex structures are usually a mixture of regions and cover many pixels; therefore, a pixel level analysis is preferred once an object is identified and outlined. Moreover, different distributions of the same objects can have different semantic meanings. At meter resolution, two very different objects can have regions that are similar within a small analysis window as compared to the object size. Figure 2-2 gives an example of such a situation: the signature of a bridge at very small scale may be very similar to the signature of a building.

The signatures can be differentiated at larger scales. Moreover, two objects belonging to the same scene class can be similar on a global, small scale, but differ significantly on a large scale (e.g., have different structure details). By scene class, we define a group of objects that share the same semantic (e.g., bridge) or geometrical, radiometrical, and texture properties.

In this frame, we propose a different usage of the term image patch, justified by the need to account for the scene context. We employ patches that cover areas of approximately 200 × 200 square meters. The patch dimension was chosen after testing various sizes of image patches, ranging from 30 to 250 pixels with TerraSAR-X data of 1 m resolution, the best results being attained for a patch size of 200 × 200 pixels. Such a patch can comprise both large and relatively small man-made structures, and is also useful for texture computation within homogeneous areas (e.g., green areas, sport yards, etc.). Our aim is to detect and classify scene classes based on their intrinsic content, and not individual objects, as a scene class may be composed of different objects of diverse size.

Another factor affecting meter resolution images in particular is the very large number of categories/classes, which is increasing as larger and larger geographic areas are being observed.

For instance, for two sites of 10 x 10 km, the number of categories can be higher than 30. In Figure 2-3we present a selection of different classes [Popescu2012]

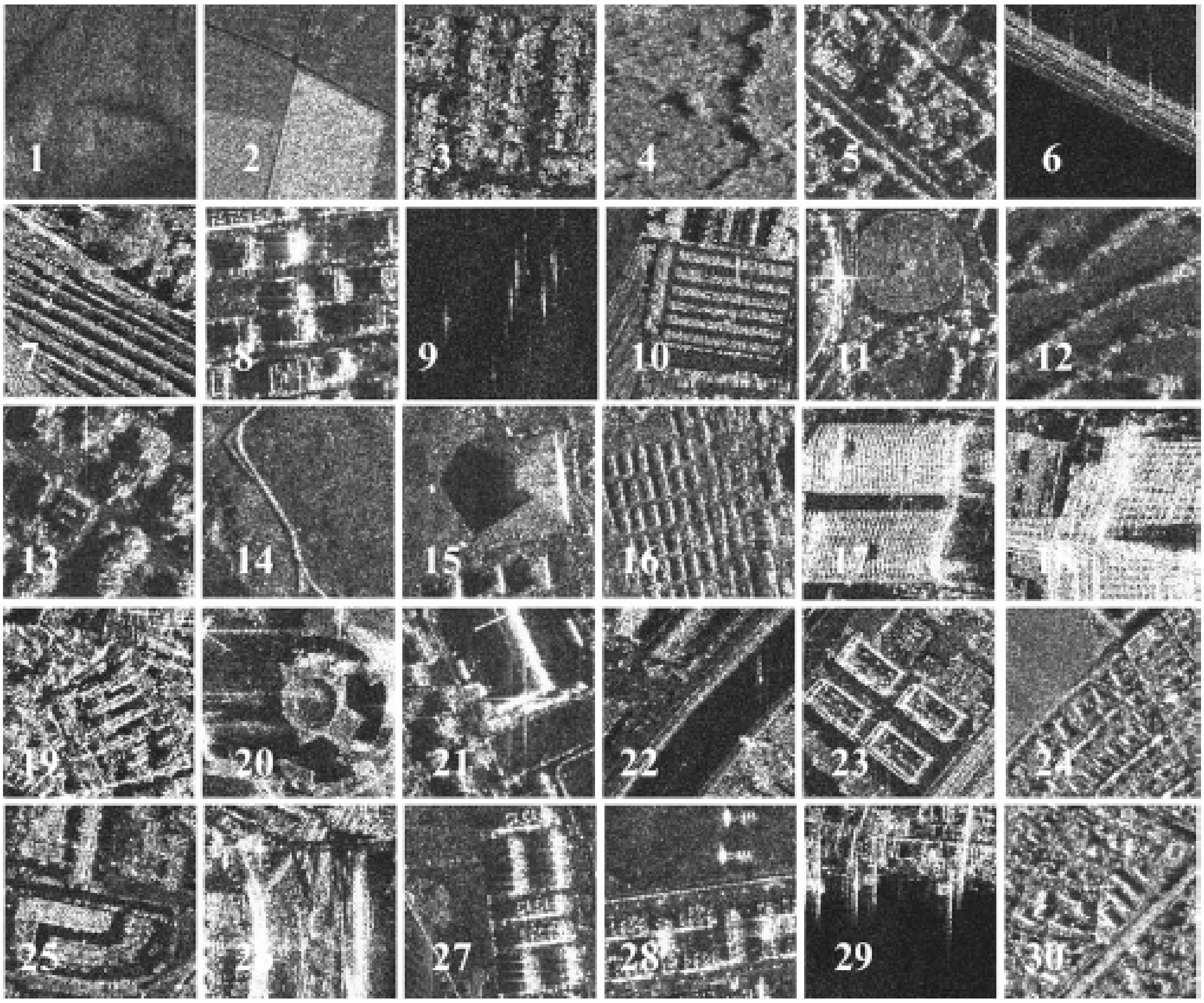

Figure 2-3: Scene Classes

The examples above show the variety of structures and details in a high resolution SAR image, ranging from large homogeneous areas (classes 1, 2, 9) to structures with clear textural patterns (classes 7, 16) and classes with a high level of detail (classes 17, 18, 20). In this example, our database consists of 6916 patches, with a class cardinality that ranges between ten and several hundreds of patches per class.

2.2 Project Objectives

Through EOLib, ESA and DLR aim at enlarging the IIM frame for a more complete exploration of EO data sources establishing large scale Information Mining functions within the multi-mission Payload Ground Segments which currently is and will be operated for missions such as TerraSAR-X/TanDEM-X, Sentinel-1/-2, and similar high resolution SAR and optical missions.

The EOLib system will allow users to find EO products of interest for their specific application based on information image content, to prototype EO applications by locally applying Information Mining tools to selected products, and to identify unusual and yet undiscovered patterns in large EO datasets including time series. EOLib aims at developing and implementing a tool for sustainable long term and efficient utilisation of EO data content.

The objectives of the EOLib project are stated in the *EOLib Technical and Management Proposal* [2] and can be summarized as:

- Analyse the scientific and technical Data and Information Mining achievements for their use in an operational PGS context on well-defined very large EO datasets,
- Analyse Information Mining use cases for their uniqueness and expected success as opposed to (or in collaboration with) other approaches of accessing large EO data holdings,
- Extend the IIM concepts to methods for mining heterogeneous EO sources, e.g., EO images, image Time series, their metadata, and other related geo-information.
- Develop relevant state-of-the-art and elaborate new Information Mining algorithms up to operational maturity,
- Elaborate a multi-mission multi-temporal Image Information Mining system architecture ensuring PGS integration, and
- Develop, demonstrate and evaluate EOLib as a system implementing selected IIM and KDD functions for selected use cases on selected datasets.

2.3 System Context

In order to define the EOLib system scope it is important to place the EOLib system in its system context, thus to identify the external interfaces of EOLib. Due to the tight integration of EOLib into the PGS, two viewpoints can be adopted to define the EOLib context: the viewpoint of the external user, not differentiating between PGS and EOLib functions, and the PGS viewpoint, distinguishing the PGS-supported functions and the original EOLib functions. Figure 2-4 EOLib System Context adopts the second viewpoint here, especially in order to demonstrate the relationship between PGS-supported functions and original EOLib functions, and to show the interfaces towards the missions, the EO experts, and scientific IM developments and the external users.

The following figure depicts the expected context of the EOLib system.

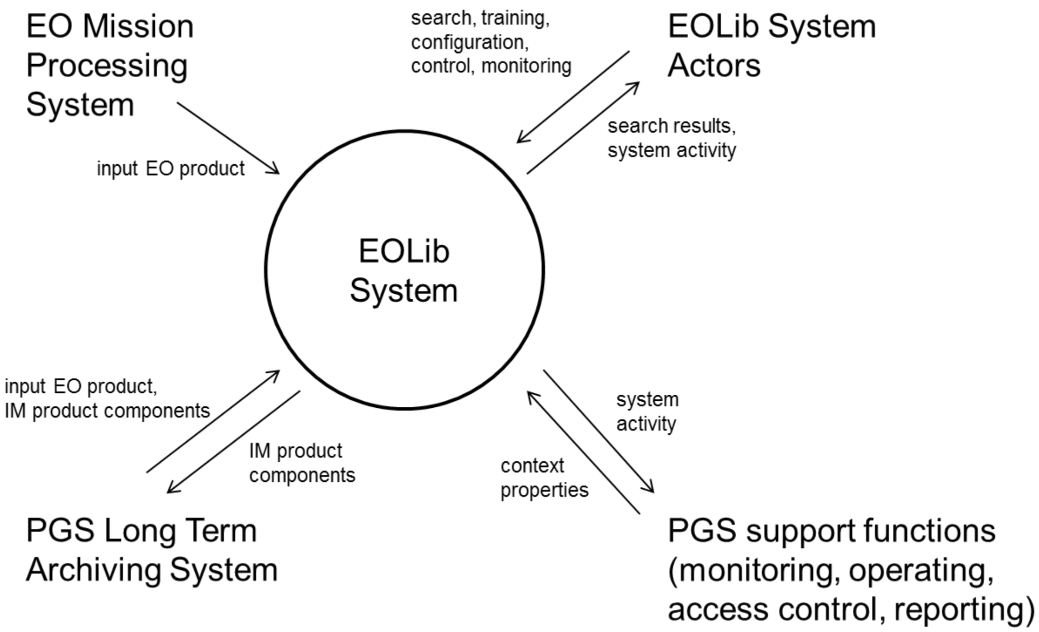

Figure 2-4 EOLib System Context

The EOLib system is interfaced to the PGS **EO Mission Processing Systems**, either directly coupled in the mission processing pipeline or fed with input data by the PGS Long Term Archiving System.

The **PGS Long Term Archiving System** is used to query the information mining data set and archive the extracted basic features of the products.

The **EOLib System Actors** split into several user groups. The EOLib User will use the system as search engine specifying queries for content based images retrieval. The Information Mining Operator will train the system with new data to explore and assign semantic annotations to the data or to refine them in iterative training sessions. Finally the System Operator and Administrator are responsible for the configuration, extension and monitor and control of the system’s operation.

Since EOLib will be part of the PGS it can use several infrastructural **support functions** like monitoring or reporting. Following the PGS development and deployment guidelines which define structural and environmental settings for the deployment of a software component (context properties), allows that e.g. the system activity can be monitored by the DIMS Operating Tool.

2.4 EOLib Concept

EOLib is a concept and system (search engine), to be integrated as an organic component within an EO PGS system. It aims at interpreting image content, associate it with other information, create categories, and understand user' inquiries. It will interact with the user to refine his expression of needs, comments and informs users about data content, and suggests the most appropriate products including alternative interpretations. EOLib is a proposal for establishing the next generation EO search engine which adds value in comparison with current systems by integrating “live” products and image and geo-information intelligence.

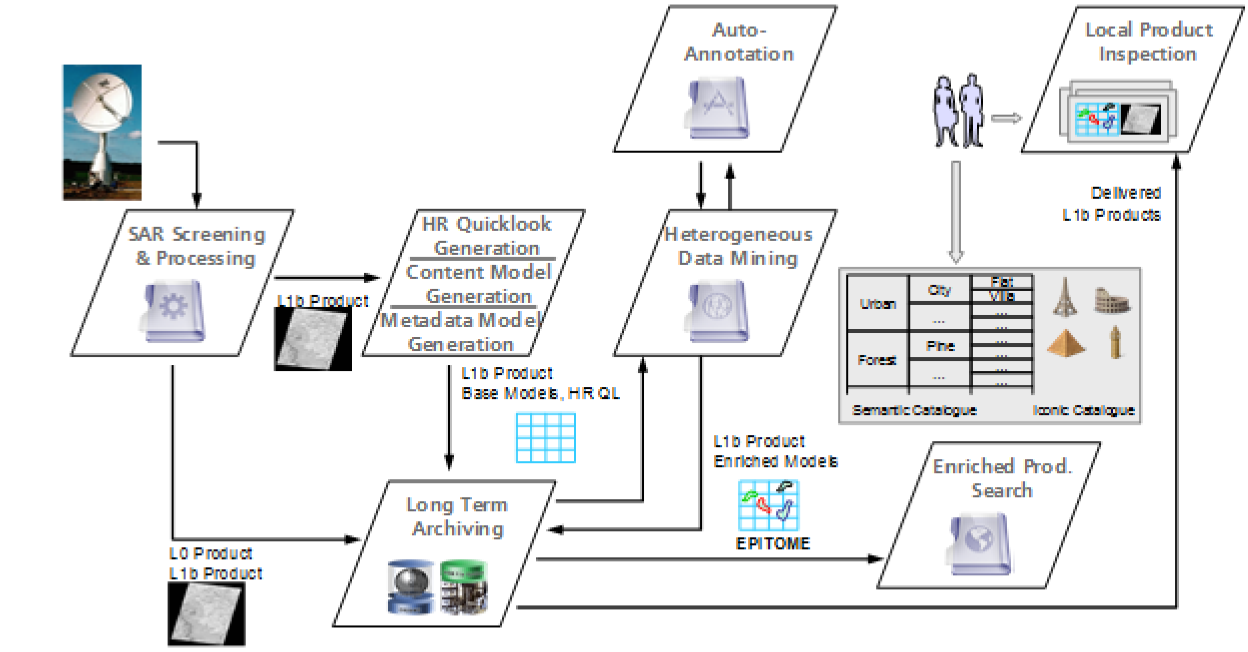

Figure 2-5 Sample EOLib Product Data Flow for TerraSAR-X

The EOLib system will analyze the received images and their associated metadata, identify image parts at different scales and resolutions, find their spatio-temporal context and their description by other metadata and generate a “data model”. The data model will be an “epitome” of the image content, i.e. a compact representation of classes of scene structures or their evolution. The epitome will be the basic component for:

- an enriched index for the product catalogue,
- a base for a new type of interactive value added EO products,
- the image and data representation for further auto-annotation and data mining processing.

Further, the EOLib system will help the human operator, in a separate interaction loop, to use mining functions for the aggregation of the obtained image and data elements and also to add semantics with alternatives depending on the application context. The next step is the generalization of the induced semantics to all image products by an auto-annotation process. EOLib will also suggest to the operator the semantic annotation of the data, supporting the use and generation of bodies of knowledge and using elements of taxonomy and potentially a simple ontology.

The results of semantic annotation or discovery of EO archive content will be made available directly to the users via a new EOWEB interface. It will include new functionalities, such as novel queries by example, iconic catalogues, and semantic queries.

2.5 System Actors

The EOLib system is expected to have the following actors. They all are assumed to be human individuals excluding explicitly machines and software processes as actors. Most of them were already identified in the *EOLib Technical and Management Proposal* [2].

2.5.1 ACT-User

EOLib *users* are persons or entities with access to EOLib who need information mining capabilities for the EOLib corpus of earth observation products. Possible categories are scientific, industrial, military, and private users *EOLib Statement of Work* [1].

2.5.2 ACT-Information-Mining-Operator

The *information mining operator* is an EO expert. He has the necessary knowledge to train the system, build a taxonomy/ontology and assign semantic annotations to images or image patches.

2.5.3 ACT-System-Operator

The *system operator* is the supervisor and the configuration manager of the EOLib system, including related DIMS components.

2.5.4 ACT-System-Engineer

The *system engineer* is responsible for the integration, deployment and update of the EOLib system’s feature extraction and information mining modules.

2.6 Functional Domain (Logical Model)

Figure 2-6 EOLib Functional Domain (Logical Model), EOLib consists of several functional (logical) units that together fulfil the requirements base. Their choice was further inspired by the constraints of the existing Data and information Management System (DIMS) in use with the DLR PGS. The individual components as well as the complete logical model, however, are generic enough to be used within other PGS architectures as well. The functional units will be mapped in a later step to the architectural components of EOLib.

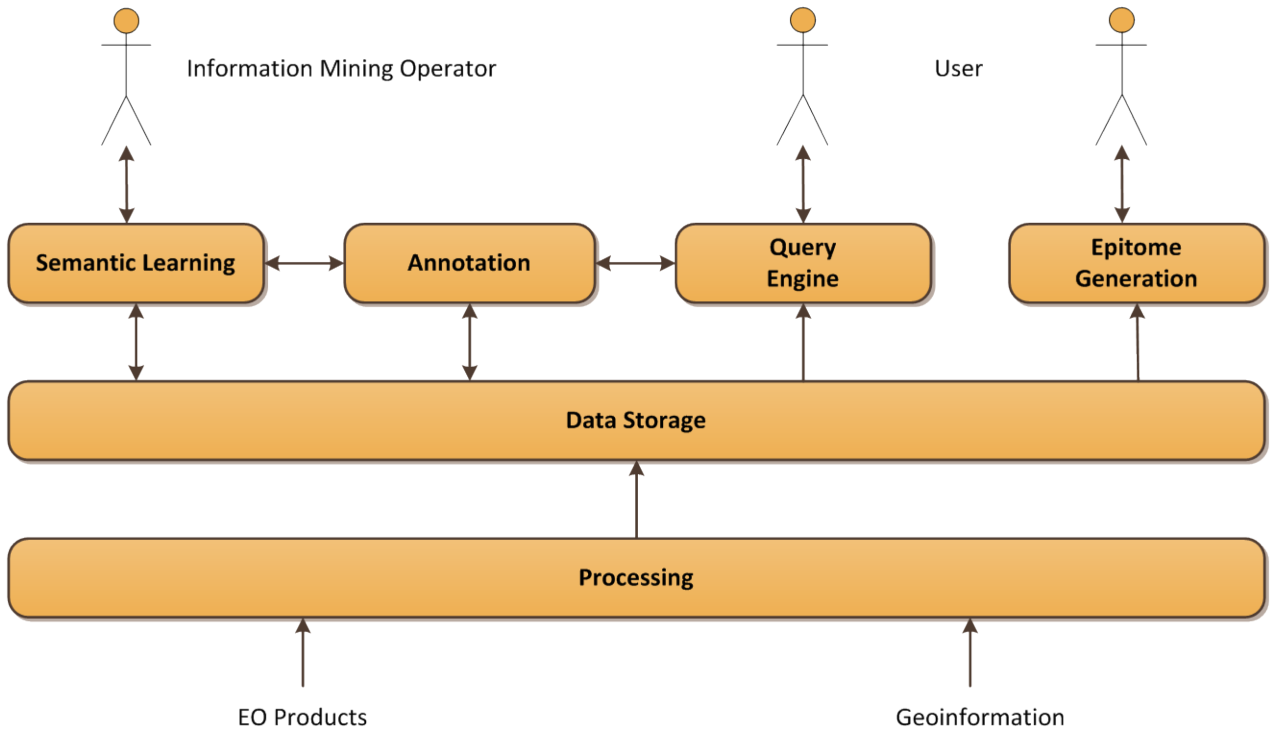

Figure 2-6 EOLib Functional Domain (Logical Model)

Semantic Learning

The Semantic Learning module allows the system to learn, possibly guided by an operator, the parameters for an automatic feature detection and classification. Interactive image information mining and knowledge discovery happen at this level.

Annotation

The *Annotation* performs heterogeneous data mining as clustering, classification, dimension reduction and similarity metrics.

Query Engine

The *Query Engine* offers the query processing and retrieval functionality for information retrieval in EOLib.

Epitome Generation

The *Epitome Generation* produces the Semantic Catalogue for the user to browse and download semantically enhanced EO products.

Data Storage

Data Storage is offered by the underlying multi-mission PGS, ensuring the consistency of mission data with basic feature and epitome items.

Processing

The *Processing* unit provides the processing functionality for EO products, including scaling, patching and feature extraction. A data model will be generated suitable for the image information mining.

3. User Scenarios

This section introduces the application-level user scenarios of EOLib. They define the high level sequences of interaction of end users for a given main purpose of EOLib, such as the ad-hoc semantic search for EO products. The application use scenarios shall identify and illustrate the *concrete* capabilities of EOLib for certain EO applications.

3.1 Image Processing and Epitome Generation

The scenario demonstrates the use of EOLib for the creation of an actionable catalogue for the EO products, based on the image content and analysis of image metadata, geo-information, and other documents or available information resources (e.g., Internet). Important data types are Digital Elevation Models (DEMs) to be used together with external GIS data as "base layers" for the purpose of stable reference information. The scenario includes the following steps:

1. Loading and ingesting EO images and Image Time Series and their metadata
2. Primitive Feature Extraction, (spectral, texture, etc.) and metadata processing for descriptor extraction
3. Interfacing with geo-data (e.g. Corine Land Cover, Urban Atlas, DEMs, or street maps) and usage of their content as geospatial information
4. Using Heterogeneous Data Mining to generate semantic annotations of images and ITS interactively and storing them into the semantic catalogue.
5. Updating also the PGS product catalogue with annotation enabling users to perform queries at semantic level
6. Operating the KDD module for Product Quality Control or further discovering of groupings, categories, and rules
7. New image ingestion and generalization of the previous annotation to the new data
8. Encapsulating the data model generated in step 4 into a standard, actionable value-added EO product, called the epitome[1]. The product epitome will be provided with an interactive browser for product visualization and exploitation, active learning and product ordering.

This scenario is covered by the following use cases:

- o UC-Product-Ingestion
- o UC-Semantic-Annotation

3.2 Epitome usage

This scenario describes how the epitome generated in 3.1 can be used for a quick evaluation of EO image content using a high-resolution quicklook (QL) image.

1. Load the Epitome on the user workstation (even a notebook)
2. Open the Epitome
   - 2.1. Open the Epitome with standard EO image analysis tools, e.g. ENVI
     - 2.1.1. Fast inspection of the scene using the High resolution QL which is optimized for visualization (both SAR and multispectral data)
     - 2.1.2. Evaluate the PF of the product.
     - 2.1.3. Evaluate the PF of the product.

[1] The epitome is an output product of the EOLib system. It is generated on demand based on product information at the PGS Product Library and the semantic annotation of the semantic catalogue. A more detailed description of the epitome product will be done during the design phase.

2.1.4. Use the PF for further prototype/evaluation classification

2.1.5. Evaluate the encapsulated external information

2.2. Open the Epitome with the Epitome browser

2.2.1. Fast inspection of the scene using the High resolution QL which is optimized for visualization (both SAR and multispectral data)

2.2.2. Evaluate the PF of the product.

2.2.3. Use the PF for further prototype/evaluation classification

2.2.4. Evaluate the encapsulated external information

This scenario is covered by the following use case:

- UC-Export-Results

3.3 Image Retrieval using Semantic Queries

This scenario enables the user to express a complex sentence, or a question composed of words and numeric descriptors. The syntax is predefined as a set of simple operators (e.g. less than, North off). These powerful semantic queries can be performed using a query language which operates on semantic annotations describing image content which have been previously extracted from the image archive and compiled in the product epitome. The use of semantics will help the user to better understand the image content.

Three types of queries may be distinguished:

- *Semantic Labels*: The user can enter a “simple” label in the form of text or select an item from the available labels in the catalogue to perform the query (as for example “forest”). It has to be notet that these labels are pre-defined labels previously obtained as results of the image annotation process.
- *Query language and ontologies*: Here, the queries can be performed either using only semantics or semantics and spatial content in the form of text or numerical entries. The queries based on spatial content are performed by using the image descriptors. The query language can rely on a query template in order to help the user and to avoid spelling errors.
- *Query Builder* enabling queries by using semantics, topological relations, different operators, and numerical descriptors.

The figure below shows an example of a GUI for queries based on semantics and spatial content. The left hand side shows the possible queries while the right hand side displays the results. The result is a list of images and a map, showing their location.

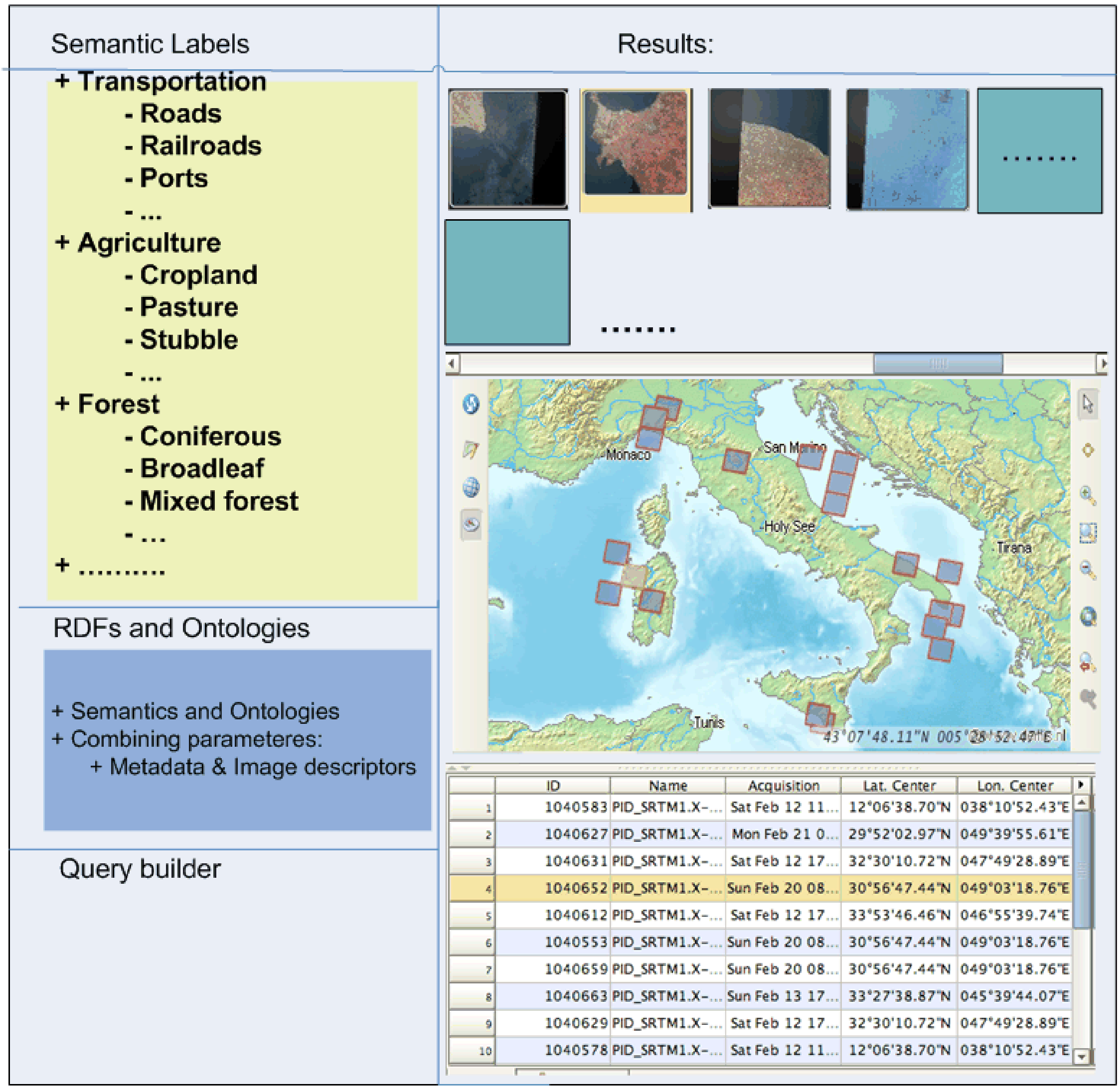

Figure 3-1 Example of a GUI for queries based on semantics and spatial content

This scenario is covered by the following use case:

- UC-Semantic-Search

3.4 Image Analysis and Knowledge Discovery

This scenario describes more complex exploratory searches, performed in an active learning loop. Here IIM and KDD functions are involved to providing the query modes:

- Query by example,
- Active learning and
- Visual data mining.

These individual query types or modes can be combined and enable interactive querying such as in relevance feedback systems.

A) Query by example

1. The user selects an image and its content is used as example for the query.
2. The system activates the data mining methods and retrieves the relevant images ranked according to specific metrics as for example “relevance”.

B) Active learning

1. Input the training data sets obtained interactively containing a list of images marked as positive or negative examples
2. Running "Data Mining methods" in order to search in the entire image database and pick out the relevant images
3. System outputs a verification of the Active Learning loop and the semantic annotation from in the data base catalogue
4. User gives further positive and negative examples and restarts the Active Learning loop until the targeted image category is learned as accurately and as exhaustively as possible
5. Generation and enrichment of the semantic catalogue based on the discovered image categories; in an auto-annotation process, all similar images in the database are automatically being assigned the new sematic label.

C) Visual Data Mining.

1. Images are displayed grouped by similarity metrics
2. User selects different images and/or image content in 2-D or 3D space
3. System searches in the EO database to retrieve the results
4. User adapts the result to his conjectures and searches again

This scenario is covered by the following use cases:

- UC-Semantic-Annotation
- UC-Semantic-Search

3.5 Generation of Content Taxonomies and Application/Domain Ontologies

Traditionally new EO sensors/missions are preceded by a preparatory phase. Typically airborne, or satellite data with similar properties are used to prototype and demonstrate possible applications. However, the true potential of the new sensor data begin to be known only after several years of operation when many users or researchers had the opportunity to work on large data volumes and also on broad diversity of information content.

This scenario proposes the use of EOLib to make a fast and exhaustive analysis of large amounts of sensor data from the early beginning of the mission, i.e. the commissioning phase, when already large and diverse data sets are being acquired. This will provide users with a deep insight of the data information content, content taxonomies and application/domain ontologies, thus enabling a broader use of the data.

In most applications EO images are mainly used to generate classification maps, extract physical parameters, or to index image structures/objects. Many types of EO images are also "not visual" i.e. SAR or multispectral observations, thus evaluation of the content becomes even more difficult.

The goal is to provide a sound basis for future investigations and applications that should benefit from a clever representation of descriptors derived from the full image information available, assess the information content extracted by various PF descriptors and semantic annotations built upon these descriptors. EOLib opens the way towards an automated analysis of hidden dependencies among the various image information facets.

Data features which are not immediately visible will be highlighted in relation to their meaning.

1. – 4. Follows the scenario 3.4 (Image Analysis and Knowledge Discovery) part B (Active learning), steps 1 to 4.

5. Generate a catalogue relating semantic annotations to the PF, resolutions, scales used to generate it, define application taxonomy and provide an analysis and quantitative metrics for the image information. Enable users to assess the data potential for their applications before ordering new data. Apply the methodology for the multi-sensor cases. Support the generation of taxonomies and elements of ontology for the data content.

This scenario is covered by the following use cases:

- UC-Semantic-Annotation

- UC-Ontology-Definition

3.6 Quantitative Semantic Queries

This scenario follows the scenario 3.4 (Image Analysis and Knowledge Discovery) part B (Active learning).

1. – 4. As in scenario 3.4 (Image Analysis and Knowledge Discovery).
5. The results of the active learning have been recorded in the database model. Using query language i.e. query by numerical and textual descriptors to extract structured or statistical information like percentage of urban vs. rural areas in given geographic region or percentage of building types, vegetation, water bodies in a specific city.

This scenario is covered by the following use case:

- UC-Semantic-Search

3.7 Support for Training Image Analysts

The training of Image Analysts mainly for the interpretation of VHRSAR images is a long process requiring the analysis of very large data sets and additional information.

EOLib can be a support tool for the training, by ingesting specific SAR and optical data sets of the same areas acquired in almost the same time interval and other supporting information or knowledge. The optical data can be used both for visual verification and as support for understanding the SAR signatures

1. – 4. Follow the scenario 3.4 (Image Analysis and Knowledge Discovery) part B (Active learning), steps 1 to 4.
5. Use KDD for discovering and learning the SAR signatures of various types of scene structures, verify them with optical images and the external information. Generate a systematic catalogue of objects, targets, and other information or knowledge.

This scenario is covered by the following use case:

- UC-Semantic-Annotation

3.8 Automatic Image Processing in CGS's value-adding Production System

Based on the scenario 3.4 (Image Analysis and Knowledge Discovery), use KDD quantitative analysis to identify those PF, combinations, resolutions, or scales, which give the most robust results in scene recognition and propose them for systematic processing.

This scenario is covered by the following use case:

- UC-Automatic-Annotation

3.9 Collaborative Ground Segment Tool

Using scenarios 3.4 (Image Analysis and Knowledge Discovery) and/or 3.8 (Automatic Image Processing in CGS's value-adding Production) for specific regions, thus capturing particular signatures in the images, EOLib can support the elaboration of products tailored to particular coverage/regions or services:

- higher level products than produced by the core ground segment
- products/algorithms tailored to a particular coverage or region, services or user community
- generation of local/regional data sets with correction, projection, calibration, merging, etc. different to the standardized one offered by DLR's Core Ground Segment.

The EOLib system can be used as collaborative tool. For example, the processing data from different regions could help to understand the common and the different features with regional specificity.

This scenario is covered by the following use cases:

- UC-Semantic-Annotation
- UC-Semantic-Search

3.10 Support for Rapid Mapping Applications

The scenario demonstrates the support of EOLib in rapid mapping activities to help the faster and more complete generation of risk, damage, or situational assessment maps. A rapid mapping session can be split into 3 main activities:

1. pre-event data discovery, compilation and analysis
2. post-event data compilation and analysis
3. joint pre- and post- event data processing and map generation

The EOLib system will contribute to this scenario by means of content-based fast image or image time series retrieval. The user will perform a fast search of existing images, GIS datasets, maps, or other documents annotated in the EOLib catalogue which carry a reference to the same site and/or to similar types of events. Additionally, he will load and ingest data and information from external archives and internet resources referring the same site and/or similar types of events to create new or refine existing semantic annotations of EO products. Finally the user will export the results for further external analysis and map creation.

This scenario is covered by the following use cases:

- UC-Semantic-Search
- UC-Export-Results

3.11 EOLib Evaluation

This scenario is based on the concept of "relevance". This concept generates a metric for the evaluation of information retrieval systems, the "precision recall" schemes. These are achievements from the area of "text" retrieval. The notion of "relevance" is linked to "semantics", thus it is almost independent from the area of applications or use cases. Thus, it results in very efficient but also very similar systems.

This is not at all the case for multimedia and is an important challenge for the case of EO data. The concept of "relevance" is manifold, vaguely defined, ambiguous, subjective, and very much linked to the applications. One of the key issues for future of research and applications in the area of information retrieval is the elaboration of exploration methods beyond ranking - and independent of the type, form, content or purpose of the data. That research has to be common to all application domains without the weakening effect of partitioning it for specific and particular applications areas.

This use case aims at better defining what "relevant information" is in relation to the notions of "data content" and "context," and for a broader class of EO data.

To achieve a more comprehensive picture of "information retrieval" performance and user needs, both system and user centred evaluations are needed. The analysis is based on knowledge scenarios with different degrees of structure; an evaluation strategy should consider the following aspects:

- An appropriate dataset for evaluation, which should be general enough to cover a large range of semantics from a human point-of-view and also should be large enough for the evaluation to be statistically significant.
- Ground-truth for relevance. Ground-truth is a very subjective issue, since people usually associate a given observation with a wide range of high-level semantics.
- An appropriate metric and criteria for evaluating competitive approaches. The evaluation criteria should try to model human requirements from a population perspective.
- Due to the high complexity, the evaluation should be performed in terms of the performances of the individual modules as an objective approach and/or taking into account the user criteria as subjective approach.

The standard metrics used in information retrieval include Mean Average Precision (MAP), Binary Preference (Bpref), Mean Reciprocal Rank (MRR) and Geometric Mean Average Precision (GMAP). However, some tasks are unsuited for evaluation using these measures. In this case, we indicate the evaluation metric for the specific initiative.

This scenario is covered by the following use case:

- UC-Evaluation

3.12 Support for Image Quality Assessment

This scenario will be demonstrated using TerraSAR-X and TanDEM-X data in the view of preparing for Sentinel 1 and 2. TerraSAR-X and TanDEM-X are missions which contribute enormously to the generation of next generation, very high accuracy and precision, global DEMs. The methods used are InSAR, multiple InSAR processing coherent or un-coherent. The data acquisition and the InSAR processing itself may be affected by artifacts induced by the scattering mechanism, imaging mode, or InSAR DEM generation processor. To insure the high quality of the DEM in condition of processing enormous data volumes, a new data quality approach is needed. The KDD functions of EOLib are proposed for a new approach for artifacts detection and indexing.

1. – 4. Follow the scenario 3.4 (Image Analysis and Knowledge Discovery) part B (Active learning), steps 1 to 4.
5. Generate a catalogue of various types of artifacts and quantitate metrics for the data quality.

This scenario is covered by the following use cases:

- UC-Semantic-Annotation
- UC-Semantic-Search

3.13 Collaboration with Scientific and User Communities

From Scenario 3.4 (Image Analysis and Knowledge Discovery) propose a selection of data and its PF, annotations, metadata and exogenous relate data, accompanied with the performance metrics for DM and KDD functions and provide them freely to the scientific and user communities for further evaluation of new algorithms, benchmarking, tests, etc. Thus a more systematic and traceable monitoring of the progress in the sensor data use and evolution of algorithms and applications is expected.

This scenario is covered by the following use case:

- UC-Export-Results

4. Use Cases

The system-level use cases define the complete set of interactive use cases of all actors (chapter 2.5) interacting with the EOLib system, including the EO content experts and system operators. The system use cases shall identify all system capabilities required by the different system actors as a basis for the requirements engineering and the elaboration of the system architecture.

4.1 Conventions

We use in this document the following template in order to describe use cases. The meaning of the different fields is explained in the template.

4.1.1 UC-Template

Description	A description of the use case. Additional details may define the target system, in-put, output etc.
Actors	The name of the actors that are involved in this use case.
Assumptions	Use cases whose output is needed or other prerequisite external conditions.
Steps	Detailed description of a more complex action.
Variations	Variations to this use case s.a. other actors or action paths.
Notes/Issues	Well,

4.2 Overview

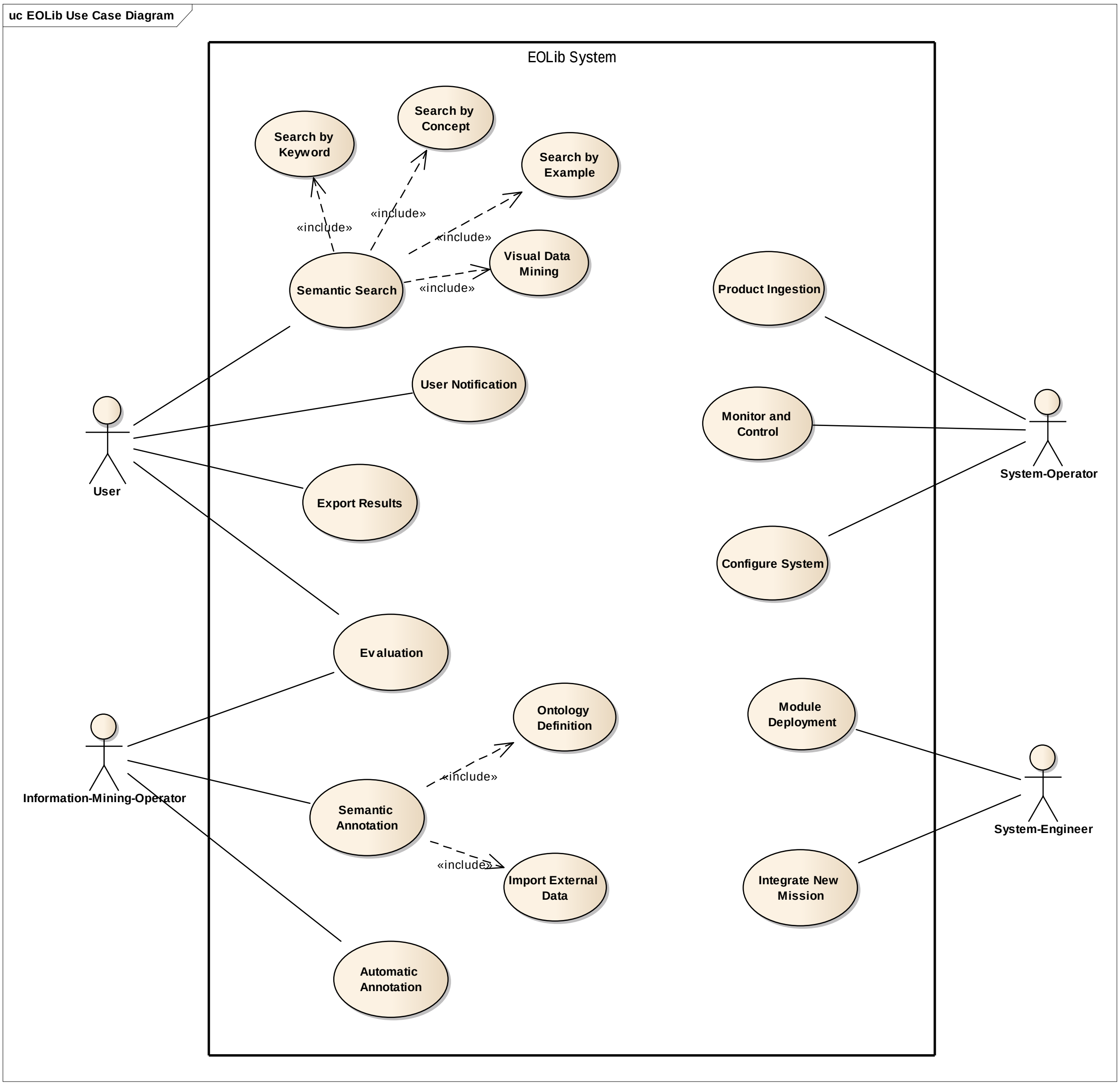

Figure 4-1 Use Case Diagram

4.3 Ingestion

4.3.1 UC-Product-Ingestion

Description	After a new EO product arrives in the PGS it has to be ingested into the EOLib system. Its main input is the original EO product in full resolution at a certain processing level, either directly available within a processing pipeline, or forked into this processing element by a previous processing element, or retrieved from the PGS archive, the Product Library, e.g. in case of reprocessing. The output is archived into the Product Library either as component of the mission EO product, or as distinct IM product with link to the original EO product. • Scaling: The information mining system reduces the resolution of an earth observation product to one or more predefined resolution levels in order to provide uniform input to EOLib. • Patching: The information mining system cuts an earth observation product into patches for further processing. • Basic Feature Extraction processes all the mission data in the information mining data set and generates a high-resolution quick-look as well as the basic feature vectors for spatial or Image Time Series description on data and metadata per product patch, applying the generic reference model for EO data patches defined by EOLib.
Actors	ACT-System-Operator
Assumptions	UC-Module-Deployment, UC-Configure-System
Steps	1. new EO product arrives in PGS Product Library 2. product is processed by system including scaling, patching, primitive feature extraction, metadata descriptor generation 3.all generated information is added to EO product and stored
Variations	Step 2 may also be triggered manually by an operator.
Notes/Issues	None

4.3.2 UC-User-Notification

Description	The user receives a notification if a new image that fulfills certain pre-defined criteria is ingested into the system.
Actors	ACT-User
Assumptions	The user must have performed a subscription to this service beforehand. UC-Product-Ingestion
Steps	1. after arrival, processing and auto-annotation of a new arrived EO product the semantic annotation is checked against the users criteria 2. if they match, the system will send an notification to the user (preferably an email)
Variations	None
Notes/Issues	If new data meets the requirements but does not meet the user's clearance level (sensitives, SatDSiG), no notification is send.

4.4 Processing

4.4.1 UC-Semantic-Annotation

Description	Train the system with target images and generate the semantic annotation of the image.
Actors	ACT-Information-Mining-Operator
Assumptions	UC-Product-Ingestion
Steps	1. select image to be learned 2. select/define class to be learned 3. start the search and let the system retrieves the results 4. start Active Learning loop in selecting positives and negatives from the previously retrieves results and restart the search 5. repeat until all classes of image are learned
Variations	None
Notes/Issues	None

4.4.1.1 UC-Ontology-Definition

Description	Part of the semantic training of the system. Subsidiary to UC-Semantic-Annotation
Actors	ACT-Information-Mining-Operator
Assumptions	None
Steps	1. EO data selection 2. Define semantic definition terms and their relations, representing the semantic concepts expected to be identified in the EO data, taking into account the sensor capabilities, the sensing characteristics, the spatial/temporal data resolution and the interests of application domains.
Variations	None
Notes/Issues	None

4.4.1.2 UC-Import-External-Data

Description	Interface with geo-data (e.g., Corine Land Cover, Urban Atlas, DEMs, or street maps) and use their content as geospatial information in EOLib. To avoid duplication the geo-data will be only stored temporarily for analysis and content extraction purposes. Subsidiary to UC-Semantic-Annotation.
Actors	ACT-Information-Mining-Operator
Assumptions	None
Steps	1. Locate and connect to sources of GIS/text data 2. external data is displayed as new layer over existing data
Variations	None
Notes/Issues	None

4.4.2 UC-Automatic-Annotation

Description	The semantic annotation established for a small training data set is applied to the whole PGS product catalogue. The auto-annotation matches the basic feature vectors with the trained semantic annotation definitions and assigns the semantic annotations to patches together with a calculated level of confidence.
Actors	ACT-Information-Mining-Operator configures triggering rules to automatically start the annotation processing on the corresponding products.
Assumptions	UC-Product-Ingestion, UC-Semantic-Annotation
Steps	1. system is triggered to perform auto-annotation on specified EO products 2. system retrieves basic feature component of product from Product Library 3. system generates generic EOLib data model 4. system starts auto-annotation 5. Created semantic annotation is stored into semantic catalogue.
Variations	The process may also identify feature cluster/semantic classes which are new to the system. They will also be stored but require manual analysis and annotation during a learning phase. The process may also be started manually be the operator.
Notes/Issues	None

4.5 Retrieval

4.5.1 UC-Semantic-Search

Description	Find images based on their content.
Actors	ACT-User
Assumptions	UC-Product-Ingestion, UC-Semantic-Annotation
Steps	1. enter search criteria and start the search 2. system retrieves and presents the result images
Variations	The search results may be refined in changing the search criteria.
Notes/Issues	None

4.5.1.1 UC-Search-By-Keyword

Description	The user can enter a “simple” label in the form of text or select an item from the available labels in the catalogue to perform the query (as for example “forest”).
Actors	ACT-User
Assumptions	UC-Product-Ingestion, UC-Semantic-Annotation
Steps	1. user enters keyword and starts the search 2. system retrieves and presents the results
Variations	None
Notes/Issues	None

4.5.1.2 UC-Search-By-Concept

Description	The user can enter a semantic concept to perform the query.
Actors	ACT-User
Assumptions	UC-Product-Ingestion, UC-Semantic-Annotation
Steps	1. user enters the semantic concept and starts the search 2. system tries to match given ontology with semantics stored internally and retrieves and presents the results
Variations	None
Notes/Issues	None

4.5.1.3 UC-Search-By-Example

Description	Find images which are similar to a given example.
Actors	ACT-User
Assumptions	UC-Product-Ingestion
Steps	1. establish EO product working data set 2. select/provide image to be used as example 3. system retrieves and presents relevant results
Variations	This use case may also be used in similar way to UC-Semantic-Annotation for learning and generating semantic annotation
Notes/Issues	None

4.5.1.4 UC-Visual-Data-Mining

Description	Representation of the image data in the 3D space to explore content of data set used.
Actors	ACT-User
Assumptions	UC-Product-Ingestion, UC-Semantic-Annotation
Steps	1. Images are displayed grouped by similarity metrics 2. User selects different images and/or image content in 2-D or 3D space 3. System searches in the EO database to retrieve the results 4. User adapts the result to his conjectures and searches again
Variations	This use case may also be used in similar way to UC-Semantic-Annotation for learning and generating semantic annotation
Notes/Issues	None

4.5.2 UC-Export-Results

Description	EO products and epitome for the retrieved images can be exported for offline analysis.
Actors	ACT-User
Assumptions	UC-Semantic-Search
Steps	1. collect data set (images) to be saved 2. provide format and target information for the export 3. system retrieves the EO product(s) and epitome and delivers them in the specified way
Variations	None
Notes/Issues	None

4.6 System Management

4.6.1 UC-Monitor-Control

Description	The system operator monitors the correct operation of the EOLib system and takes corrective measures (configuration, powercycle, require support from systems engineer) as needed.
Actors	ACT-System-Operator
Assumptions	None
Steps	An administrative tool is used to perform various tasks.
Variations	None
Notes/Issues	None

4.6.2 UC-Module-Deployment

Description	Integrate, deploy and configure new (versions of) software modules for basic feature extraction. Integrate, deploy and configure new (versions of) software modules for information mining/auto-annotation.
Actors	ACT-System-Engineer
Assumptions	None
Steps	1. The new module is deployed into the system but not yet inserted into the processing chain of the automatic workflow. 2. The new module is evaluated with a small well known data set. 3. At approval the module is activated for automatic processing.
Variations	None
Notes/Issues	None

4.6.3 UC-Integrate-New-Mission

Description	The EOLib system is extended to process the data of a new EO mission.
Actors	ACT-System-Engineer
Assumptions	None
Steps	1. EOLib system is prepared. All necessary configuration and extension of the system are made. 2. a subset of the new mission data is established 3. proper operation of EOLib system is evaluated with test data 4. new mission data is activated for automatic processing
Variations	None
Notes/Issues	None

4.6.4 UC-Configure-System

Description	The operation of the EOLib system must be ensured by the EOLib System Operator. He configures controls and monitors the EOLib system components. New BFE processors and KDD functions are integrated and deployed by the EOLib System Engineer.
Actors	ACT-System-Operator
Assumptions	None
Steps	An administrative tool is used to configure the system.
Variations	None
Notes/Issues	None

4.7 Assessment

4.7.1 UC-Evaluation

Description	This use case is based on the concept of “relevance”. This concept generates a metric for the evaluation of information retrieval systems, the “precision - recall” schemes.
Actors	ACT-Information-Mining-Operator, ACT-User
Assumptions	None
Steps	1. an appropriate dataset for evaluation is selected 2. the ground-truth for relevance is defined 3. define an appropriate metric and criteria 4. perform evaluation of the individual modules as objective approach
Variations	Taking into account the user criteria as subjective approach.
Notes/Issues	None