IIM activities at ESA
Status and perspectives

Pier Giorgio Marchetti
European Space Agency
Earth Observation Programme
Research and Ground Segment Technology Section
1. Motivation
2. Overview of past activities in the IIM field
3. Systems and services for EO data exploitation
4. New activities in support of next generation of EO data
5. Strategy and vision for the future
Motivation

1. Foster the use of IIM and derived technologies in support of the EO PDGS infrastructures

2. Develop state-of-the-art data processing for improving access and dissemination of future EO data (e.g. Sentinels mission)

3. Implement systems and services for supporting the “scientific exploitation” of EO data

4. Investigate new approaches and methodologies to exploit data from all available missions and archives (LTDP)
1. The evolution of image information mining and related tools

2. The KIM-KEO Prototype: *user-centred information mining*

3. Grid-Processing On Demand

4. Multi-temporal Evolution Analysis

5. Primitive Feature Extraction

6. E-Collaboration

6. EO Ontologies (RARE, SMAAD and P-TREE projects)
User-centred information mining

1. New methodologies and tools for “mining information” in EO images:
   - Intelligent and effective access to information content in large EO data repositories
   - Improved exploration and use of EO data for science
   - Extraction of relevant information for different applications (change detection, global monitoring, disaster management)
   - Implementation, integration and validation of services derived from IIM methods

2. Joint effort of a wide range of experts / users:
   - Agencies (ESA, CNES, DLR, EUSC, JRC)
   - Universities and research institutes (ETHZ, …)
   - Industry
   - Mission owners and data providers
The EO Image Librarian: new concept for IIM and CBIR

KEO Prototype @ ESRIN (KIM+CPE)

KEO @ ROSA

SRoKEO

KLAUS

PIMS-DLR

IIM-TS

KEI

MIMS

KEO

KIMV

KES / KES-B

KIM Prototype @ ESRIN

KIM @ DLR


EOLib

Reference Baseline
1. The KIM prototype permits to interactively apply “Probabilistic Information Mining” techniques to search an entire image collection for specific user-defined areas

2. KIM consists of three main components:
   - Ingestion Software
     - Extraction of “primitive features”
     - Clustering
   - Database
   - Interactive Client Application
     - System training and definition of “semantic rules”
     - Application of training (rules) to the entire collection
     - Definition of “semantic labels” for extracted information
     - Store for successive re-use
KIM Architectural Elements

Input
- EO images

Output
- Identifiers of searched images
- Feature Maps / Thematic maps
KIM permits to inspect a collection of images...

...interactively define “semantic features” using the “primitive features” extracted by the system...

...search for the defined feature within the entire collection...
...and extract Feature Maps or Thematic Maps

Cloud masks  Flooded areas  Forest Monitoring
### KIM

**Primitive features**

<table>
<thead>
<tr>
<th>Spectral</th>
<th>Spectral signature</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Texture</strong></td>
<td>Structural information extracted with the Gibbs Marcov Random Fields (GMRF) model S0 - full resolution images; S1 - sub-sampled images</td>
</tr>
<tr>
<td><strong>DCT</strong></td>
<td>Discrete Cosine Transform: transforms signals and images from the spatial domain to the frequency domain</td>
</tr>
<tr>
<td><strong>EMBD</strong></td>
<td>Enhanced-Model-Based-Despeckling: performs a high quality despeckling of SAR images</td>
</tr>
<tr>
<td><strong>Area</strong></td>
<td>Area of the objects detected with the segmentation process</td>
</tr>
<tr>
<td><strong>Compactness</strong></td>
<td>Compactness of the objects detected with the segmentation process</td>
</tr>
<tr>
<td><strong>Spectral Mean</strong></td>
<td>Mean value of the radiometric information of the image inside the closed area detected by the segmenter</td>
</tr>
<tr>
<td><strong>Spectral Variance</strong></td>
<td>Variance of the radiometric information of the image inside the closed area detected by the segmenter</td>
</tr>
<tr>
<td><strong>Hu Moments</strong></td>
<td>Hu-Moment Invariants: shape information conveyed by the contour points. Hu moments are invariant to scale, rotation and translation (the first 4 out of 7 invariant moments as shape descriptors have been used).</td>
</tr>
<tr>
<td>Sensor</td>
<td>Product</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Envisat MERIS</td>
<td>RR, FR, FRS (.N1)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Landsat 5 TM / Landsat 7 ETM</td>
<td>CEOS Geotiff</td>
</tr>
<tr>
<td>ERS-1/ERS-2 SAR Envisat ASAR</td>
<td>GEC Geotiff</td>
</tr>
<tr>
<td>SPOT 4-5 HRG</td>
<td>DMAP Geotiff</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Ikonos Quickbird</td>
<td>Panchromatic Geotiff Multiband</td>
</tr>
<tr>
<td></td>
<td>Geotiff</td>
</tr>
<tr>
<td>Other sensors</td>
<td>Geotiff</td>
</tr>
</tbody>
</table>
KIM Validation

KIM has been tested and validated with different datasets:

1. MERIS RR / MERIS FR
2. ERS / ASAR
3. SPOT
4. Landsat
5. Maps (Level 2 / Level 3 products)

- Large number of collection created
- Low number of significant semantic features identified
KIM for Information Extraction

1. Flood Detection (SAR data)
2. Cloud Detection (MERIS RR)
3. Long-term Forest Monitoring (Landsat)
4. Rapid Mapping / Damage Assessment (VHR optical data)

Potentialities of the tool have been highlighted and confirmed in different contexts.

End-users expectations not always achieved.
KEO
Knowledge centred Earth Observation

KEO is a distributed component-based processing environment (CPE)
   a. Create & semantically identify internal/external *Processing Components*
   b. Graphically chain *Processing Components* into processing chains
   c. Create *Processing Components* from IIM components (KIM training)
   d. Export and store outputs into Web Servers (WFS, WMS, WCS)

KEO also provides some relevant Reference Data Sets
   a. Heterogeneous data and information, growing with external contributions (images, documents, DEMs, photos, processors, etc.)
   b. In support of various applications: Classification, Time Series Analysis, Ortho-rectification, Urban Monitoring, Interferometry, etc.)
1. KEO currently provides more than 300 *Processing Components*
2. Set of Processing Components extensible by users
3. From “basic” processing to “high-level” applications
4. Easy creation of very complex processing chains
5. Supporting various applications:
   a. change detection / time series analysis
   b. biomass estimation
   c. fire detection
   d. land cover classification
   e. snow mapping
   f. etc.
Analysis of new CBIR tools
LIRE / Caliph-Emir suite

1. Evaluation of the tool in terms of capabilities and performances
   a. Use of various image descriptors for indexing images
   b. Performance evaluation using the precision-recall framework

2. Extension of the tools to take into account specific EO metadata
during the annotation and search process

3. Extension for semantic annotation of EO image

4. Export of search results into mpeg-7
Grid- Processing On Demand (G-POD)
Objectives and concept

1. Promote use of EO data
2. Offer on-line access to products and computing infrastructure / tools
3. Assist in the generation of “scientific added value products”
G-POD Services

✓ FAIRE 2
  o Flood crisis / damage mapping service
  o Used for Charter & GMES RESPOND

✓ MGVI on-demand
  o 1km vegetation on user area
  o User-defined aggregation period

✓ Imager
  o Multi-mission imaging tools
  o Selection, processing, geocoding, rendering / visualisation

✓ MIRAVI Geo-toolbox
  o Geocoding of MERIS full resolution images

✓ BEAMARITHM
  o BEAM user-defined pixel computations
  o Aggregation in Level-3
G-POD Services

- MERIS Level-3 Products
  - ACRI (France), JRC/Ispra (EC) and Brockmann Consult (BEAM)
  - 11 daily & monthly products on-line

- Daily ASAR GM mapping of Antarctica
  - Daily Generation of 400-m resolution mosaics
  - Publishing on ESA Web Map Server

- Volcanoes Monitoring by Infrared
  - Real-time Extraction of AATSR thermal anomalies over > 300 volcanoes
  - Long-Term database on line

- River and Lake Processor
  - Accurate River and Lake heights measurements in NRT from satellite altimetry (RA2)
  - Products published online
Multi-temporal analysis of HR / VHR products:

1. Select Multi-temporal Applications that might benefit from such extension
2. Design, implement and integrate the Automatic / Multi-temporal Algorithms to support the selected applications
3. Create the needed HR/VHR Reference Data Sets and Evolution Models
4. Develop standard interfaces between the different systems for common exploitation of ingested data and processing capabilities
5. Integrate algorithms and Evolution Models provided by other independent projects
6. Validate (with the support of a Validation Group) the Automatic Multi-temporal algorithms and Evolution Models
Multi-temporal Evolution Analysis (MEA)  
Automatic Semantic Information Mining (ASIM)

The MEA-ASIM system aims at providing:
1. Advanced tools for Land Use / Land Cover change analysis
2. Level-2 EO products for real time exploitation
3. Interfaces to external systems (G-POD, KEO, data providers, etc.)
4. Access to data via standard WCS OGC interface
5. Native support for Sentinel-2 datasets
Multi-sensor Evolution Analysis

Pixel / coverage analysis

time-series analysis
single-/multi-plot functionality
Multi-sensor Evolution Analysis
Pixel / coverage analysis

cross-comparison of EO products
1. Prepare the ground for a systematic feature extraction of EO data
2. Demonstrate the usefulness of extracted features in the data selection process (both for scientific and operational use)
3. Extend benefits of data and extracted features also to external communities (e.g. ASAR data on Supersites)
4. Identify suitable datasets and algorithms for feature extraction for HR/VHR SAR and Optical products, with special attention to Time Series exploitation
5. Outline and demonstrate utilization scenarios
6. Enrich with extracted information EO data ("enriched product")
PFA Process (II)

Data Repositories

Content Based Image Retrieval

Unsupervised Classification

Time series analysis

AUX
Develop and implement “collaborative platforms” for:

1. Fostering the scientific exploitation of EO data
2. Automating the creation data mining and information extraction experiments and algorithms
3. Supporting the creation of EO-based applications and services
4. Supporting the scientific research process:
   a. Addressing specific scientific challenges and tackling new research problems in a “parallel and collaborative way”
   b. Generation of reproducible results that can be easily shared and validated
E- Collaboration for Earth Observation
Architectural sketch
Ontology (RARE) Semantic access to EO catalogues

Centralized service interfacing a number of on-line resources that were not integrated in the past. These resources include:

1. a terminology service supporting the navigation facility between related application terms
2. a query analyzer that augments the knowledge with unforeseen concepts and relations extracted from the Internet
3. gazetteers used to resolve place names
4. various reasoners mapping the application terms selected by the users to product categories
5. a centralized catalogue service that collects the properties of on-line resources registered in the ESA internal registries and other remote catalogues.
Ontology (P-TREE)
Enriching EO ontology services

Prepare the use of CF-NetCDF for forthcoming ESA missions and high level products from existing missions:

1. Taking into account the Climate Change Initiative community and Guidelines for Data Producers

2. Propose an enriched CF-NetCDF metadata model (on the basis of ISO 19115 and EOP-O&M), permitting the EO product to be as much self-describing as possible (product content, processing chain, application domain, etc.)

3. Develop open-source libraries (reusing available OSS), permitting to handle enriched CF-NetCDF products (for visualization, metadata queries, annotation, XML export, adding new tags, etc.)

4. Develop open-source libraries (reusing outcomes from ESA activities on ontology, RARE, SMAAD) to create a sharable vocabulary from the enriched CF-NetCDF metadata model and map it with existing ontologies (e.g.: GSCDA, GCMD, GEMET, GEOSS, etc...)

5. Set-up a demonstrator for discovery and download of enriched CF-NetCDF products, exploiting available ontologies and verify it through test cases.
1. Research support

**Scientist/Research Centre**
- Question definition
- Available information and resources study
- Hypothesis formulation
- Perform experiment and collect results
- Analyse data
- Data interpretation and conclusions
  - Hypothesis confirmation
  - Possible new hypothesis
- Results publication

**ESA’s RSS service**
- Question definition
- Hypothesis formulation
- RSS processes EO data and delivers processing results
- RSS supports data analysis
- Also via RSS

**RSS resources**
- e-collaboration (Join&Share)
- Reference Data Sets
- Web (Map/Feature/Coverage) Server
- Support Algorithm development
- Algorithm integration
- Processing
- Product access (processing output)
- Reference Data Sets
- Processors and tools
- e-collaboration (Join&Share)
- Reference Data Sets
- Web (Map/Feature/Coverage) Server
1. Involvement of new industry and agency partners to increase the relevance of IIM activities in Europe

2. Promote the use of IIM for management and exploitation of very large archives (PB of data)

3. Foster the IIM role in the context of future missions (e.g. Sentinels) and Long Term Data Preservation

4. Define, develop and implement a new generation of “enriched EO products” tailored to user-defined needs

5. For information and support visit rssportal.esa.int