TELEIOS
Virtual Observatory Infrastructure for Earth Observation Data

3rd User Community Workshop
Kolympari Crete, 12-14 June 2013

Presenter: Manolis Koubarakis (NKUA)
Outline

• Why TELEIOS?
• State of the art in Earth Observation data centers
• Going beyond the state of the art: developing a Virtual Earth Observatory using TELEIOS technologies
• Technical work highlights
• Impact
Motivation

N. of Earth Observation satellites launched

<table>
<thead>
<tr>
<th>Date Range</th>
<th>Total</th>
<th>USA (NASA, NOAA)</th>
<th>China</th>
<th>India</th>
<th>Russia</th>
<th>Single State Agencies (Europe)</th>
<th>Other single State Agencies (RoW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1967-1980</td>
<td>120</td>
<td>10</td>
<td>5</td>
<td>12</td>
<td>17</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>1981-1990</td>
<td>43</td>
<td>24</td>
<td>13</td>
<td>11</td>
<td>17</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>1991-2000</td>
<td>80</td>
<td>28</td>
<td>13</td>
<td>11</td>
<td>17</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>2001-2010</td>
<td>109</td>
<td>91</td>
<td>20</td>
<td>27</td>
<td>11</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>246</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Motivation (cont’d)

Evolution of ESA’s EO Data Archives between 1986-2007 and future estimates (up to 2020)

~20 PB

Big Data!
Estimated directly affected inhabitants Japan, Honshu 2011

Damage Assessment
Earthquake Haiti, 2010

Motivation (cont’d)
• A STREP project funded under the 5th call of FP7/ICT under the Strategic Objective ICT 2009.4.3 “Intelligent Information Management”.

• Emphasis of this objective was on “scalability” and “systematic empirical testing” of systems for “very large amounts of data” (i.e., big data).
Consortium

1. National and Kapodistrian University of Athens

2. Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V.

3. Deutsches Zentrum für Luft- und Raumfahrt e.V.

4. Centrum Wiskunde & Informatica

5. National Observatory of Athens

6. Advanced Computer Systems

With support from
State of the Art in EO Data Centers
Example

• Can I pose the following query using EOWEB?

Find images taken by the MSG2 satellite on August 25, 2007 which contain fire hotspots in areas which have been classified as forests according to CORINE land cover, and are located within 2km from an archaeological site in the Peloponnese.
Example (cont’d)
Example (cont’d)

- Well, only partially.

Find images taken by the MSG2 satellite on August 25, 2007 which contain fire hotspots in areas which have been classified as forests according to CORINE land cover, and are located within 2km from an archaeological site in the Peloponnese.
Example (cont’d)

• But why?

• All this information is available in the **satellite images** and other **auxiliary data sources** of EO data centers or **on the Web**.

• However, EO data centers today do not allow:
  • **the mining of satellite image content** and
  • **its integration with other relevant data sources** so the previous query can be answered.
The TELEIOS Earth Observatory: Concept View

Scientific Database and Semantic Web Technologies

Linked Geospatial Data
Semantic Annotations
Ontologies
Knowledge Discovery and Data Mining

Web Portals
Rapid Mapping

GIS Data
Derived Products
Metadata
Features

Raw Data
Ingestion
Processing
Cataloging
Archiving
Content Extraction

DATA
KNOWLEDGE
Technical Highlights

• Community forming
• User requirements capture
• Theory (data models, query languages, knowledge discovery algorithms, software architecture)
• Development of individual software components
• The TELEIOS software architecture and infrastructure
Community Forming

- Organized two TELEIOS user community workshops:
  - October 13, 2010 in Frascati (ESA/ESRIN)
  - May 10-12, 2012 in Darmstadt (Fraunhofer)
- Attended by:
  - Users
  - Data providers
  - Scientists
  - Technology providers
- Disseminated TELEIOS results
Capturing User Requirements

- We collected detailed requirements for the two TELEIOS use cases:
  - **DLR Use Case:**
    - A Virtual Observatory for TerraSAR-X data
  - **NOA Use Case:**
    - Real-time fire monitoring based on continuous acquisitions of EO images and geospatial data.
Community Forming (cont’d)

• We are today in the 3rd **TELEIOS user community workshop**.

• Being attended by 34 participants (16 outside TELEIOS). Thank you for accepting our invitation!

• **Goals:**
  • Present our achievements and discuss them with you.
  • Introduce you to our systems
  • Get your feedback!
Technical Contributions

• EO data as multidimensional arrays: the SciQL query language.
• Data vaults.
• Implementation in the column store MonetDB.
What is SciQL?

- **SciQL** is an SQL-based query language for scientific applications with **arrays** as first-class citizens.

- **SciQL advantages:**
  - Express low level image processing and image content analysis in a **high-level declarative language**.
  - **Uniform access to image data** (arrays) and **metadata** (tables) for knowledge discovery.
Why SciQL?

```c
void firesClassify(double *t039, double *t108, int rows, int columns, int *fire) {
    int win, win_2, r, c, i;
    double mean039, std039, mean108, std108;
    double thr[] = {0.0,310.0,2.5,8.0,2.0,310.0,4.0,10.0,2.0};

    win=3;
    win_2=win>>1;

    for(i=0;i<rows*columns;i++) fire[i]=0;

    for(r=win_2;r<rows-win_2;r++) {
        for(c=win_2;c<columns-win_2;c++) {
            temperatureMeanAndStd(t039, columns, win, r, c, &mean039, &std039);
            temperatureMeanAndStd(t108, columns, win, r, c, &mean108, &std108);

            /* Fire Test */
                fire[r*columns+c]=2;

            /* Potential Fire Test */
                fire[r*columns+c]=1;
        }
    }
}
```
Why SciQL? (cont’d)

```c
void temperatureMeanAndStd(double *T, int columns, int windowSize, int r, int c,
                     double *mean, double *std) {

double s, ss, d;
int i, j;
int win_2, win2;

win_2 = windowSize >> 1;
win2 = windowSize * windowSize;

s = ss = 0.0;
for (i = 0; i < win; i++) {
    for (j = 0; j < win; j++) {
        d = T[(r - win_2 + i) * columns + (c - win_2 + j)];
        s += d;
        ss += d * d;
    }
}

*mean = s / win2;
*std = sqrt(ss / win2 - (*mean) * (*mean));
}
```
CREATE ARRAY hrit_T039_image_array
    (x INTEGER DIMENSION, y INTEGER DIMENSION, v FLOAT);

CREATE ARRAY hrit_T108_image_array
    (x INTEGER DIMENSION, y INTEGER DIMENSION, v FLOAT);
SELECT [x], [y],

CASE
  WHEN v039 > 310 AND v039 - v108 > 10 AND v039_std_dev > 4 AND v108_std_dev < 2
  THEN 2
  WHEN v039 > 310 AND v039 - v108 > 8 AND v039_std_dev > 2.5 AND v108_std_dev < 2
  THEN 1
  ELSE 0
END AS confidence
FROM (SELECT [x], [y], v039, v108,
     SQRT( v039_sqr_mean - v039_mean * v039_mean ) AS v039_std_dev,
     SQRT( v108_sqr_mean - v108_mean * v108_mean ) AS v108_std_dev
FROM (SELECT [x], [y], v039, v108, AVG( v039 ) AS v039_mean,
     AVG( v039 * v039 ) AS v039_sqr_mean, AVG( v108 ) AS v018_mean,
     AVG( v108 * v108 ) AS v108_sqr_mean
FROM (SELECT [T039.x], [T039.y], T039.v AS v039, T108.v AS v108
     FROM hrit_T039_image_array AS T039
     JOIN hrit_T108_image_array AS T108
     ON T039.x = T108.x AND T039.y = T108.y
     ) AS image_array
     GROUP BY image_array[x-1:x+2][y-1:y+2]
     ) AS tmp1;
) AS tmp2;
What is a data vault?

• The **data vault** framework addresses the problem of ingesting scientific data of various formats into tables or arrays.

• **Data vault advantages:**
  
  • Make the **DBMS “aware”** of external file formats.
  
  • **Files are loaded** into tables/arrays only when needed.
Technical Contributions (cont’d)

• A framework for knowledge discovery from EO data, metadata and relevant geospatial data sets.
Large Scale Annotation of TerraSAR-X data

<table>
<thead>
<tr>
<th>Phase</th>
<th>No. of scenes / No. of patches</th>
<th>No. of semantic categories</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase I</td>
<td>39 scenes 40,307 patches</td>
<td>336 categories</td>
<td>• Support Vector Machine</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Relevance Feedback – ranking</td>
</tr>
<tr>
<td>Phase II</td>
<td>82 scenes 76,649 patches</td>
<td>688 categories</td>
<td>• Support Vector Machine</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Relevance Feedback – ranking</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– enhanced visualisation</td>
</tr>
</tbody>
</table>

*Semantic annotation for Phase I and Phase II*
Multi-level Category Hierarchies

Higher level nodes in the hierarchy correspond to more general concepts.

Intermediate level nodes in the hierarchy correspond to moderately general concepts.

Lowest level nodes correspond to "atomic" image concepts.

Patches of an image may consist of several "atomic" concepts.
Venice categories using CORINE land cover

Venice categories identified using CLC

Level 1
- NODATA
- UNCLASSIFIED LAND SURFACE
- Artificial surface
- Agricultural areas
- Forest and semi natural areas
- Wetlands
- Water bodies

Level 2
- Urban fabric
- Industrial, commercial, transport units
- Artificial, non-agricultural vegetated areas
- Pastures
- Heterogeneous agricultural areas
- Forests
- Open spaces with little or no vegetation
- Marine waters

Level 3
- Continuous urban fabric
- Discontinuous urban fabric
- Port areas
- Green urban areas
- Pastures
- Complex cultivation patterns
- Mixed forests
- Beaches, dunes, sands
- Salt marshes
- Sea and ocean
- Coastal lagoons
Venice categories using the KDD framework

Venice categories (using TerraSAR-X data) identified based on our methodology

- Bridge
- Port
- Airport
- Water and boats
- Buoy
- Water
- Water and vegetation
- Agriculture
- Vegetation
- Cemetery
- Railway tracks
- Urban
- Water and urban
- Breaking waves
- River deposit
- Beach area
- Vegetation and buildings
Technical Contributions (cont’d)

- Semantic annotations capturing the knowledge discovered: the model stRDF, the query language stSPARQL and their implementation in the system Strabon.
What are stRDF and stSPARQL?

- **stRDF** is an extension of the W3C standard RDF for the representation of **geospatial data that may change over time**.

- **stSPARQL** is the query language for stRDF.
SELECT ?IM

WHERE

  ?RD noa:isDerivedFromSatellite "METEOSAT9" .
  dbpedia:Peloponnese strdf:hasGeometry ?PGeo .
  ?SITE rdf:type dbpedia:Archaeological_sites .
  ?SITE strdf:hasGeometry ?SGEO .

FILTER(strdf:anyInteract(?RGeo,?HGeo) &&
       strdf:distance(?RGeo,?SGEO, geo:metre) < 2000 &&
       strdf:contains(?PGeo,?SGEO) &&
       strdf:during(?T, "[2007-08-25T00:00:00, 2007-08-25T23:59:59]"^^strdf:period))
}
Technical Contributions (cont’d)

• Helping users to formulate stSPARQL queries: a visual query builder.
Technical Contributions (cont’d)

• The TELEIOS **software architecture and infrastructure**

• **An implementation of the NOA use case** (fire monitoring, burnt scar mapping only partially).

• **An implementation of the DLR use case** (semantic catalogue for TerraSAR-X data, rapid mapping)
Impact

• Important societal and industrial impact
• ICT, Space, Environment and Security research

• European programs in Earth Observation
  – GMES (EO data management, Emergency Response Core Service)
  – ESA (KIM, KEO, KLAUS, EOLib etc.)

• Scientific data management
• Linked open EO data
• Linked open geospatial data
• Satellite image mining
Welcome again and thank you for your attention!

Questions?