BIG DATA FROM SPACE

5–7 June 2013 | ESA-ESRIN | Frascati (Rome), Italy

EVENT REPORT
1 INTRODUCTION

Big Earth observing data can be defined in terms of volumes, their degree of diversity and complexity - including streaming of data from presently available and upcoming satellite capabilities, and ubiquitous ground devices - the unpredictable value added derivable from their innovative analyses and fusion. This event’s main rationale is that Earth observing data are growing in size and variety at an exceptionally fast rate, posing challenges and opportunities for their access and application. Increasing diversity of space data, increasing combined use of diverse space missions data, increasing integration of satellite born data with non-space ones, naturally lead to an unprecedented opportunity to serve new types of user applications, impacting the way such data are collected, referenced, disseminated, processed and delivered. Their combination coupled with today’s Internet technologies open to new exciting opportunities, which may lead far from the original purposes such data were thought for. Innovative exploitation and the potential economic value and social return are drivers in this framework. Not all the fields of application serve people skilled in understanding Earth observing data. They increasingly serve a public focussed on issues rather than data, looking for rapid reliable answers to generally complex questions, bypassing time consuming processes and analyses, which may be necessary in the backend. Generally, people are good in finding information when it is reasonable to find them, when it is easy to locate the necessary information and easy to understand them. For this, data - including those from space - need to be discoverable, machine consumable, handled by interoperable services at the WWW. The emerging Internet of Things will allow combining large heterogeneous environmental information in ways unthinkable just a decade ago, and consumers of such information are readily there.

This document provides the report of the BIG DATA FROM SPACE event, held at ESA-ESRIN on 5-7 June 2013.

2 ORGANISATION

Two committees served the organisation of the BIG DATA FROM SPACE event. Participants are listed below for completeness. Organising committee:

Bargebini, P - European Space Agency
Cheli, S - European Space Agency
Desnos, YL - European Space Agency
Greco, B - European Space Agency
Guidetti, V - European Space Agency
Marchetti, PG - European Space Agency
Comparetto, C - Congrex c/o ESA Conference Bureau
Nativi, S - National Research Council of Italy
Sawyer, G - European Association of Remote Sensing Companies

Scientific committee:

Atkinson, M - University of Edinburgh
Baumann, P - Jacobs University
Bennett, V - Centre for Environmental Data Archival, Science and Technology Facilities Council
Burger, A - Joint Research Centre
Busswell, G - Climate and Environmental Monitoring from Space
Casey, K - NOAA National Oceanographic Data Center
Cocco, M - Istituto Nazionale di Geofisica e Vulcanologia
Corlett, G - University of Leicester  
Dactu, M - German Aerospace Center  
Donlon, C - European Space Agency  
Fuentes, M - European Centre for Medium-Range Weather Forecasts  
Garelo, R - Fellow IEEE  
Gough, E - Centre Maritime Research and Experimentation  
Guidetti, V - European Space Agency  
Mikusch, E - German Aerospace Center  
Pearlman, J - IEEE Committee on Earth Observation  
Percival, G - Open Geospatial Consortium  
Piolle', J-F - Institut Français de Recherche pour l'Exploitation de la Mer  
Robida, F - Bureau de Recherches Géologiques et Minières  
Steenmans, C - European Environment Agency  
Trieschmann, O - European Maritime Safety Agency  
Wilson, H - European Organisation for the Exploitation of Meteorological Satellites

Participation to the event was opened to decision makers and technical representatives from all organisations active in using or delivering large complex data sets of Earth observations, including:

- Space agencies and satellite operators
- Agencies/institutions with any R&D/operational requirement for using large Earth data volumes
- European industrial operators providing services running large Earth data volumes
- Earth and computer scientists and professionals, as well as students in those areas.

The Call for Abstracts was issued early 2013, proposing a list of topics: applied multivariate analysis, data mining; computing power and storage scalability; costs and weighting factors; data access and use policies, licensing of derivative work; data capturing and description; data interoperability, retrieval, navigation; data protection, and trustworthiness; data delivery timeliness, distribution services, network capacity; data slicing, subsetting, extraction; data variety, fusion, correlation; data visualisation, rendering, video streaming; peak data processing; performance indicators for big Earth data services; systematic data processing; spatial on-line analytical processing systems; sustainability of big Earth data services.

The appointed scientific committee reviewed the received abstracts, selecting about fifty talks and thirty posters. ESA produced the event Abstract Book, including the selected abstracts.

3 OBJECTIVES

The intended objectives of the event were advertised at the website, as follows:

- To examine current solutions, practices and role of big Earth data services, and identify a common ground.
- To examine issues associated with data organisation and provision, and the associated costs.
- To identify scenarios of data-intensive services, traditional and innovative with respect to new form of processing, enabling additional information derivable from navigation, analytics and correlation of large Earth data sets, and integration across heterogeneous resources.
- To identify challenges, barriers, opportunities for such scenarios, and attempt to define a baseline of activity to make the identified scenarios actionable.
- To critically review current working methods and approaches with respect to the baseline proposal and its application.
3.1 Program

The resulting program has included more than fifty talks over 2.5 days. Oral presentations have been divided into four areas, organised into eight sessions. European Commission’s Directorates General Connect, Enterprise and Industry, Research and Innovation, and representatives from European Environment Agency, National Oceanic and Atmospheric Administration and Open Geospatial Consortium were chairs to the program sessions. NASA opened the event together with the European Commission and ESA.

The selected oral presentations were organised into four areas, Examples of Application Scenarios, (Big) Best Practices, Examples of Infrastructure Tools, Some Views in Prospect, as follows below for completeness.

**EXAMPLES OF APPLICATION SCENARIOS**

<table>
<thead>
<tr>
<th>Example</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big data from an EEA perspective</td>
<td>P Kjeld, European Environment Agency</td>
</tr>
<tr>
<td>Big data initiatives in Ireland</td>
<td>B Fennell, International Technology Programmes</td>
</tr>
<tr>
<td>Big Data Challenges in the Maritime Domain</td>
<td></td>
</tr>
<tr>
<td>A global, near real-time map of ocean activity – the challenges and</td>
<td></td>
</tr>
<tr>
<td>opportunities of space based AIS</td>
<td></td>
</tr>
<tr>
<td>APPS4GMES - development of operational services for GMES using cloud</td>
<td></td>
</tr>
<tr>
<td>computing technology</td>
<td></td>
</tr>
<tr>
<td>Integrating Satellite EO with UAV and In-Situ for Real-Time Visual</td>
<td></td>
</tr>
<tr>
<td>Situational Awareness</td>
<td></td>
</tr>
<tr>
<td>Applications of large-scale Floating Car Data collection from consumer</td>
<td></td>
</tr>
<tr>
<td>navigation systems</td>
<td></td>
</tr>
<tr>
<td>Enabling EO Data Exploitation – the GeoFarm approach of DLR</td>
<td></td>
</tr>
<tr>
<td>Exploring of Heterogeneous Earth-Observation Data Archives using</td>
<td></td>
</tr>
<tr>
<td>Enriched Metadata and Data Mining Tools</td>
<td></td>
</tr>
<tr>
<td>Near-real time operational processing of soil moisture indicators for</td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td></td>
</tr>
<tr>
<td>Ten years of EO data analyses on active volcanoes: lesson learnt</td>
<td></td>
</tr>
<tr>
<td>Global Deforestation through Time: Big Data Meets Scalable Visualizations</td>
<td></td>
</tr>
<tr>
<td>Google Earth Engine: A Global-Scale Geospatial Analysis Platform</td>
<td></td>
</tr>
<tr>
<td>The US Big Data Initiative</td>
<td></td>
</tr>
<tr>
<td>Using Earth Observation data for finding impact craters - the</td>
<td></td>
</tr>
<tr>
<td>Kamil crater as example</td>
<td></td>
</tr>
<tr>
<td>Pertinence of Processing Heterogeneous Legacy Earth Observation Data</td>
<td></td>
</tr>
<tr>
<td>Developing EO Service Markets (the importance of data Policy)</td>
<td></td>
</tr>
</tbody>
</table>

**BIG BEST PRACTICES**

<table>
<thead>
<tr>
<th>Example</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Data for a Big Ocean – Big Data Efforts at the US National</td>
<td></td>
</tr>
<tr>
<td>Oceanographic Data Center</td>
<td></td>
</tr>
<tr>
<td>Experience with managing a multi-petabyte meteorological archive</td>
<td></td>
</tr>
<tr>
<td>Nephelae: a platform for data intensive science on oceans</td>
<td></td>
</tr>
<tr>
<td>Experience of processing of ESA data</td>
<td></td>
</tr>
<tr>
<td>First Results from CEMS (A)ATSR Hosted Processing Pilot Projects</td>
<td></td>
</tr>
<tr>
<td>The Fourth Paradigm: Data-Intensive Scientific Discovery</td>
<td></td>
</tr>
<tr>
<td>Climate Data: Challenges &amp; Opportunities, with focus on the ESA Climate</td>
<td></td>
</tr>
<tr>
<td>Change Initiative (CCI)</td>
<td></td>
</tr>
</tbody>
</table>

Examples of Infrastructure Tools, Some Views in Prospect, as follows below for completeness.

- **Example 1:**
  - **Description:**
  - **Presenter:**
  - **Platform:**
  - **Note:**

- **Example 2:**
  - **Description:**
  - **Presenter:**
  - **Platform:**
  - **Note:**

- **Example 3:**
  - **Description:**
  - **Presenter:**
  - **Platform:**
  - **Note:**

- **Example 4:**
  - **Description:**
  - **Presenter:**
  - **Platform:**
  - **Note:**

- **Example 5:**
  - **Description:**
  - **Presenter:**
  - **Platform:**
  - **Note:**

- **Example 6:**
  - **Description:**
  - **Presenter:**
  - **Platform:**
  - **Note:**

- **Example 7:**
  - **Description:**
  - **Presenter:**
  - **Platform:**
  - **Note:**

- **Example 8:**
  - **Description:**
  - **Presenter:**
  - **Platform:**
  - **Note:**
Combining uncertainty information and server based data exploration and visualisation to analyse large datasets
Collaborative Data Analysis using Large Visual Display Screens
Near Real-Time Satellite Image Services
Performance analysis of GPU-based software solutions for processing of large volumes of SAR data
ESRI’s Living Planet best practices

EXEMPLARY OF INFRASTRUCTURE TOOLS
Earth Observation User Exploitation Platforms
Processing and making available pan-European very high resolution Earth observation data to the public
The European Plate Observing System (EPOS): Integrated Services for solid Earth Science
Let’s Talk About Big Data: a Language Based Approach
EarthServer for Geological applications: opening up access to big data using OGC web services
Analyzing Big Terrain Data from Space
CloudEO - An Open Cloud Based EO-Services Production Platform and Marketplace
Value-adding with hosted algorithms on shared Caualus/Hadoop clusters
Collaborative Data Infrastructure Services: New Ways to handle Large Earth
Big Data: methods and technology for content access
ESA LDCM Data Repository, Processing and Dissemination System
Semantic research and advanced display of large collections of Earth Observation data

SOME VIEWS IN PROSPECT
The Challenges of Securing Big Data from Space
Big data only for big users? New challenges of the Sentinel-3 SLSTR and SYN Land PAC
Bringing the infrastructure of the web to the science of Earth Observation
Global climate monitoring
From Big Data to Big Information - Geoscience Australia’s approach to the challenges and opportunities of ‘Big Data from Space’
SWOT mission data downlink and processing: the challenge
A data e-infrastructure for the International Space Station

4 RESULTS

This event served to stimulate discussion between the different communities involved in the business of providing and manipulating very large-scale data and complex analyses of Earth observations. Selected talks provided a big data vision, critically covering aspects like typical order of data volumes, challenges of data access, including timeliness, policies for their dissemination, data capture, search, sharing, transfer, mining, analysis and fusion, and visualisation issues. Vertical examples touched a broad range of application domains, including: situational awareness, maritime and land cover monitoring, oceanography, climatology, meteorology, and geology. Questions and discussion topics were mainly focussed on attempting to identify the state of the art experience and lessons learned; which instruments presently lack to allow effectively capturing and understanding the value of heterogeneous large-scale Earth data sets;
which barriers hamper a most effective use of Earth observing data, space and ground based; concerted solutions, traditional models and businesses vs raising ones.

4.1 Attendance

Some 250 scientists, industry representatives, national delegates from Europe, United States, Australia, China and Africa convened at ESA-ESRIN to attend the BIG DATA FROM SPACE event.

4.2 Main points captured

The following is a summary of the main points captured during the event.

- The concept of big Earth observing data is commonly agreed as strictly relating to the capability of rapidly and reliably accessing and consuming them, on one side, and that of running performing analysis and visualisation to discover trends otherwise not accessible, on the other side. Big data applications are by nature borderless, regional and global: big data centres shall cooperate to ensure data are available to such applications.
- The data deluge will make it increasingly difficult to find data of relevance to a given issue. Earth observing satellite data are much bigger than before, leading to a revision of their download approach. Mass processing is more and more required. Downstream services do mass processing for certain category of users. Sophisticated users could run ad-hoc processing on user platforms.
- Taking the environment to users, bringing processing to where data are. Physical data aggregation or web processing services approaches; large data holdings associated to commercial hosting, private or public Clouds. Cloud based platforms are proposed as facilitating the production of qualified EO services.
- With the current solutions on the market, it is perceived that users need to be educated to use data facilities.
- Finding data correlations often requires cross disciplines expertise. Rise of the data scientist role. Data science will complement traditional domain knowledge.
- Users may need focussing on issues rather than data, looking for performing visualisation of trends from the most large and reliable basis of information possible (hosted remotely).
- User platforms, web tools and crowdsourcing models to structure mass collaboration to environmental monitoring are still at their infancy. Innovative technical solutions often rely on unstable poorly documented technology. Systems scalability and services operationalisation are generally limited by lack of resources.
- Technical development is envisaged in: data automatic acquisition, discovery and aggregation (cubes); intensive visualisation; security of data access, manipulation, transfer; data and algorithms (standard, machine-readable) description; peak loads enabled web services; spatial processing in Cloud environments; interoperability standards for Cloud based environments; archive policy; semantic mining of information.
- Business models to run public private partnerships still far from enabling sustainable big Earth data applications. The different degree of readiness and utilisation of the many solutions presented leads to the need of further characterisation in terms of common practices, organisational and business models.
- The big data business is perceived as to have a favourable economic impact.
5 CONCLUSIONS

A knowledgeable worldwide public of professionals and experts in all aspects dealing with Earth observing data and systems, satellite and ground based, convened at the BIG DATA FROM SPACE event. Selected talks captured the aspects advertised at the event website at different level of degree, with reference to specific application domains, meeting most part of the intended objectives. Main concluding remarks can be summarised as follows:

- There is a strong call for the ability to handle and use big Earth observing data, by the most disparate profiles in the public attending the event. More in terms of data diversity and so combined analysis, than volumes. This brings along new opportunities for research, requires new cooperation schemes – European and worldwide – including sound technical, programmatic and industrial coordination, new approaches and new types of service, new skills. Overall, a change in thinking for making Earth observing data more used is perceived as necessary, e.g. looking wider at what other information areas do.
- The spectrum of potential application use scenarios is wide, touching traditional as well as new application domains, with different scopes and objectives.
- Earth observing data operators and Space Agencies are highly interested in following evolution on these topics. Particularly, ESA’s own Programmes have to be steadily adapted to technological evolution.
- **Taking the environment to the user, bringing processing to the data** is a concept unanimously supported within the public present at the event. It is perceived as complementary to web services based approaches (service oriented architectures remain widely used and promoted). User platforms are conceived primarily to allow users focussing on productive work, e.g. development of new algorithms and services, automating as much as possible time consuming and less productive tasks like data discovery, aggregation, and processing.

Following the excellent feedback received, ESA has confirmed the availability to convene a follow-on event in the near term. Global developments on big Earth observing data have successfully taken-off and will continue steadily, including the very next appointment at ESA ‘s Living Planet Symposium 2013, in Edinburgh, UK.