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:

Bargellini, 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
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.

Particularly, the Ground Segment Coordination Body (http://earth.esa.int/gscb/) participated as technical advisor. 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 (http://www.congrexprojects.com/13c10):

- 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, as per table below.

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’s J Kaye opened the event together with European Commission’s K Glinos, R Schulte-Braucks, G Ollier, and ESA.

<table>
<thead>
<tr>
<th>Day 1, Wednesday 5th June</th>
<th>Day 2, Thursday 6th June</th>
<th>Day 3, Friday 7th June</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:00-09:00 On-site Registration</td>
<td>09:00-11:00 Session 2a</td>
<td>09:00-11:00 Session 4</td>
</tr>
<tr>
<td>09:00-11:00 Opening</td>
<td>(BIG) BEST PRACTICES</td>
<td>SOME VIEWS IN PROSPECT</td>
</tr>
<tr>
<td>Chair: K Glinos, EC DG Connect</td>
<td>Chair: K Glinos, EC DG Connect</td>
<td>Chair: H Zunker, EC DG Enterprise and Industry</td>
</tr>
<tr>
<td>Rapporteur: M Dittr, European Space Agency</td>
<td>Rapporteur: V Guidetti, European Space Agency</td>
<td>Rapporteur: P Bergellini, European Space Agency</td>
</tr>
<tr>
<td>11:00-11:30 Coffee Break</td>
<td>11:00-11:30 Coffee Break + Posters</td>
<td>11:20-11:45 Coffee Break</td>
</tr>
<tr>
<td>11:30-13:00 Session 1a</td>
<td>11:30-13:30 Session 2b</td>
<td>11:45-12:00 Summary of main points covered. Discussion, recommendations and planning. Close-out of the event.</td>
</tr>
<tr>
<td>Chair: R Schulte-Braucks, EC DG Enterprise and Industry</td>
<td>(BIG) BEST PRACTICES</td>
<td></td>
</tr>
<tr>
<td>Rapporteur: B Haersch, European Space Agency</td>
<td>Chair: K Casey, NOAA National Oceanographic Data Center</td>
<td></td>
</tr>
<tr>
<td>13:30-14:30 Lunch Break</td>
<td>13:30-14:30 Lunch Break</td>
<td>13:00 The End</td>
</tr>
<tr>
<td>14:30-16:30 Session 1b</td>
<td>14:30-16:30 Session 3a</td>
<td></td>
</tr>
<tr>
<td>Chair: P Kjeld, European Environment Agency</td>
<td>Chair: G Ollier, EC DG Research and Innovation</td>
<td></td>
</tr>
<tr>
<td>Rapporteur: V Guidetti, European Space Agency</td>
<td>Rapporteur: B Greco, European Space Agency</td>
<td></td>
</tr>
<tr>
<td>16:30-17:00 Coffee Break</td>
<td>16:30-17:00 Coffee Break + Posters</td>
<td></td>
</tr>
<tr>
<td>17:00-18:30 Session 1c</td>
<td>17:00-19:30 Session 3b</td>
<td></td>
</tr>
<tr>
<td>Chair: K Casey, NOAA National Oceanographic Data Center</td>
<td>Chair: G Pierchval, Open Geospatial Consortium</td>
<td></td>
</tr>
<tr>
<td>Rapporteur: H Laur, European Space Agency</td>
<td>Rapporteur: J Farres, European Space Agency</td>
<td></td>
</tr>
<tr>
<td>18:30-20:00 Welcome Cocktail and Posters</td>
<td>20:00 Buffet Dinner</td>
<td></td>
</tr>
</tbody>
</table>

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

Big data from an EEA perspective
Big data initiatives in Ireland

Big Data Challenges in the Maritime Domain

A global, near real-time map of ocean activity – the challenges and opportunities of space based AIS
APPs4GMES - development of operational services for GMES using cloud computing technology

Integrating Satellite EO with UAV and In-Situ for Real-Time Visual Situational Awareness
Applications of large-scale Floating Car Data collection from consumer navigation systems

Enabling EO Data Exploitation – the GeoFarm approach of DLR
Exploring of Heterogeneous Earth-Observation Data Archives using Enriched Metadata and Data Mining Tools
Near-real time operational processing of soil moisture indicators for Africa
Ten years of EO data analyses on active volcanoes: lesson learnt

Global Deforestation through Time: Big Data meets Scalable Visualizations
Google Earth Engine: A Global-Scale Geospatial Analysis Platform
The US Big Data Initiative

Using Earth Observation data for finding impact craters - the Kamil crater as example
Pertinence of Processing Heterogeneous Legacy Earth Observation Data

Developing EO Service Markets (the importance of data Policy)

(BIG) BEST PRACTICES

Big Data for a Big Ocean – Big Data Efforts at the US National Oceanographic Data Center
Experience with managing a multi-petabyte meteorological archive

Nephelae: a platform for data intensive science on oceans

Experience of processing of ESA data

First Results from CEMS (A)ATSR Hosted Processing Pilot Projects

The Fourth Paradigm: Data-intensive Scientific Discovery
Climate Data: Challenges & Opportunities, with focus on the ESA Climate Change Initiative (CCI)
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

EXAMPLES OF INFRASTRUCTURE TOOLS

Earth Observation User Exploitation Platforms
Processing and making available pan-European very high resolution Earth observation data

P Kjeld, European Environment Agency
B Fennell, International Technology Programmes Enterprise Ireland
K Bryan, NATO Center for Maritime Research and Experimentation
J Allan, exactEarth Europe Ltd
G Ruecker, ZEBRIS GbR
H Skinnemoen, Ansur
PA Krootjes, TomTom
E Diedrich, German Aerospace Center
D Espinoza Molina, German Aerospace Center
EM Haas, GeoVille
M Silvestri, Istituto Nazionale di Geofisica e Vulcanologia
J de la Torre, Vizzuality
T Erickson, Google Inc.
G Strawn, Federal Networking and Information Technology Research and Development, US National Coordination Office
E Perozzi, Deimos Space
G Sawyer, European Association of Remote Sensing Companies

K Casey, NOAA National Oceanographic Data Center
M Fuentes, European Centre for Medium-Range Weather Forecasts
JF Piolle, Institut français de recherche pour l'exploitation de la mer
N Houghton, European Space Agency

P Kershaw, Science and Technology Facilities Council
T Hey, Microsoft
PP Mathieu, European Space Agency

P Walker, Plymouth Marine Laboratory
J Styles, Assimila Ltd
I Spence, Spacemetric
M Defilippi, SARMAP

L Jordan, ESRI

S Loekken, European Space Agency
D De Marchi, European Commission Joint Research
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. They are grouped per lines of action, into Challenges, User perception, Technology curve, A way forward.

Challenges

- 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 and phenomena otherwise not accessible. 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 and use data of relevance to a given issue. Earth observing satellite data are much bigger than before, mass processing is more and more required, interdisciplinary competences are increasingly required. The above requires innovative solutions, more suitable and flexible than the traditional data download at user’s end element. They shall naturally lead to setting up downstream services for mass processing or for running sophisticated ad-hoc algorithms.

User perception

- 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 available, it is perceived outreach and users educated to use data facilities are needed.
- 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).

Technology curve

- 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.

A way forward

- The wide scope and response given by the many received contributions highlights a high potential to address complex issues via new public-private partnerships. To this end, further characterisation of partnership models leading to successful sustained businesses is needed.
- It is common perception that future big data initiatives will trigger a number of activities with favourable economic impact, freeing resources untapped so far.
- Relying on decades of experience, consolidated processes and a high degree of coordination among its key players, Earth observing data assets certainly represent a major case to pursue further in the big space data domain.
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.