Photogrammetric Week '15 Dieter Fritsch (Ed.) Wichmann/VDE Verlag, Belin & Offenbach, 2015

Big Data Infrastructures for Processing Sentinel Data

Wolfgang Wagner, Vienna

ABSTRACT

Within the framework of the European earth observation programme Copernicus, a series of Sentinel satellites will be launched. These satellites provide operational sensing capabilities across the whole measurement spectrum, covering a broad range of applications in support to six thematic areas: land, marine, atmosphere, climate change, emergency management and security. Due to their advanced sensing concepts and outstanding spatio-temporal sampling characteristics, the Sentinel satellites will collect more data than any earth observation programme before. Every single day they will acquire several Terabyte of data, which means that over the years they will collect tens of Petabyte of environmental data. Clearly, if one would like to exploit the full wealth of information contained in this enormous data set, it must be possible to process the data over and over again with ever improving algorithms. But this is challenging, as storing, transferring and processing such big and diverse data sets is increasingly facing technological limitations (e.g. the relatively low I/O performance of today's computer platforms). Therefore, the current way of distributing the earth observation data to all interested users over the internet is not practrical, or affordable any longer. Instead, the task will be to "bring the software to data" rather than vice versa. This means that the processing of the Sentinel data will increasingly have to take place in large data centres or in the "cloud". Having recognised this problem early on, there are already several organisations and initiatives worldwide that prepare for the uptake of massive Sentinel data steams into their existing or planned Big Data infrastructures. The scope of this paper is to provide a brief overview of these initiatives, with a predominantly European perspective. The discussion highlights the need for increased cooperation and networking to maximise the tangible outputs and social benefits of the Sentinel programme.

1. INTRODUCTION

In the past, most scientific disciplines suffered from not having enough data for addressing many important research questions. It was quite usual that experiments or studies had to content with just a few dozens to a few hundreds measurements, which is why McFedries (2011) refers to this situation as data poor or small data. But thanks to powerful new instruments and simulators, many disciplines are now generating massive and diverse data sets that can be referred to as Big Data. Even though the availability of vast amounts of data is by itself not a guarantee for scientific quality (Lazer et al., 2014; Spiegelhalter, 2014), this data richness can nonetheless be expected to pave the way to exciting new research and applications. This fact has received wide recognition by scientists (Mattmann, 2013), industry (Mayer-Schönberger and Cukier, 2013) and funding agencies (Mervis, 2012) alike, particularly in the United States that currently assume a dominant role in the science and commercial exploitation of Big Data.

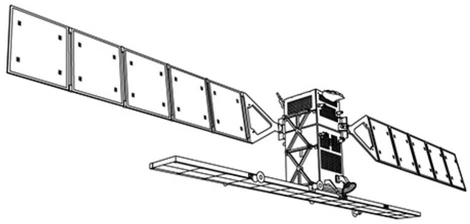
One of the disciplines that are on the brick of undergoing a transformative change into the Big Data era is earth observation (EO). One may wonder what is so exceptional about the current situation given that EO satellites have – from the very beginnings of the modern era of earth observation in the second half of the 20th century – acquired vast amounts of data? So far, progress in computer science and information technology has always allowed coping with the stepwise increases in data volume and diversity that happened with the start of each new generation of EO satellites and sensors. However, we have now arrived at a point where our capabilities to transfer, store, and process EO data increasingly fall behind our capability to generate new observational data. This is in fact one of the characteristics of the Big Data era that Schade (2015) describes as "a situation in where the volume, variety, velocity and veracity (3+1 Vs) in which data sets and streams become available challenges current management and processing capabilities". From a technological perspective, the fundamental problem is that growth rates in computing (and observational) capabilities have outstripped growth rates in storage and telecommunication capacities (Hilbert and

López, 2011). From a scientific perspective, the problem lies in the increasing complexity of computer models and algorithms needed to exploit diverse and abundant data.
With the start of the first Sentinel satellites, the challenges of the Big Data era have become imminent. Therefore, the scope of this paper is to discuss emerging Big Data infrastructures for processing of Sentinel data, whereas the thematic focus is on land applications. Please note that the term "Big Data infrastructure" is used here as a general expression encompassing all different kinds of infrastructures, from individual big data centres to distributed cloud environments. The structure of the paper is as follows: Section 2 gives a brief overview of the Sentinel programme. Section 3 discusses technical, scientific and organisational challenges that need to be addressed by Big Data infrastructures such as those mentioned in Section 4. Section 5 concludes by highlighting the need for increased cooperation and networking to maximise the social benefits of the Sentinel programme.

2. THE SENTINEL PROGRAMME

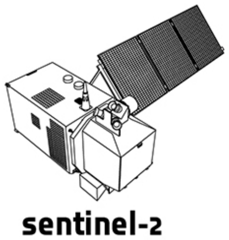
2.1. Overview

The Sentinel programme is the most comprehensive and ambitious civil EO programme that Europe has ever engaged in. It is carried out within the framework of the Copernicus programme, which is one of the two pillars of the European Union's space ambitions. It is chiefly implemented by the European Space Agency (ESA) and the European Commission in cooperation with their member states. Likely, the Sentinel programme is currently the most ambitious EO programme worldwide, complementing – and in many aspects surpassing – the environmental monitoring capabilities of the United States and other leading space nations. This unique position of the Sentinel programme stems from the fact that it consists of a series of Sentinel satellites that offer operational sensing capabilities across the whole measurement spectrum, covering a broad range of applications. In the following a short overview will be given of those Sentinel missions that are of particular relevance for land applications, namely Sentinel-1 (radar imaging), Sentinel-2 (high-resolution optical imaging), and Sentinel-3 (medium-resolution optical monitoring). For all these missions it is planned to operate two satellites in parallel to increase coverage and operational data availability.



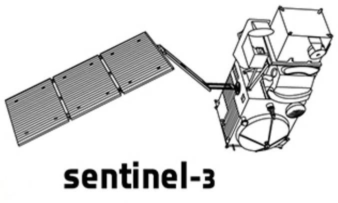
sentinel-1

Sentinel-1 is a Synthetic Aperture Radar (SAR) mission for ocean and land monitoring. Sentinel-1 is the continuity mission to the SAR instruments flown on board of ERS and ENVISAT. Key applications over land include the monitoring of topographic movements (land subsidence, glacier flow, etc.) and hydrologic processes (soil moisture, water bodies, etc.). Sentinel-1A was launched on 3 April 2014; Sentinel-1B launch is planned for spring 2016.



sentinel-2

Sentinel-2 is a high-resolution optical imaging mission, continuing and substantially adding to the EO capabilities of the Landsat and SPOT programmes. From all Sentinel satellites, Sentinel-2 is expected to attract the highest interest from the land applications community. Key applications over land include the mapping of vegetation, soil, water, and urban areas. Sentinel-2A was launched on 23 June 2015; Sentinel-2B launch is foreseen for mid-2016.



Sentinel-3 is a multi-instrument mission to monitor the sea-surface topography, sea- and land-surface temperature and ocean- and land-surface colour with two medium-resolution multi-spectral optical sensors and a dual-frequency synthetic aperture radar altimeter. The launches for Sentinel-3A and 3B are planned for fall 2015 and spring 2017 respectively.

2.2. Sentinel-1

To illustrate the Big Data challenges posed by the Sentinel programme, some of the characteristics of the Sentinel-1 mission are discussed in the following. The Sentinel-1 mission is implemented through a constellation of two satellites (units A and B), each carrying a Synthetic Aperture Radar operating at a frequency of 5.405 GHz (C-band). The main novelty of the Sentinel-1 mission is that, in contrast to its predecessor missions ERS and ENVISAT, it was designed to provide systematic acquisitions over land and oceans. Over land, the Sentinel-1 SAR instruments is predomintaly operated in the so-called Interferometric Wide Swath (IW) mode that covers a swath of 250 km at a ground resolution of 5×20 m^2. For each 98,5 minutes orbit, this mode can be operated up to 25 minutes if the other two high-date-rate modes (StripMap and Extra Wide Swath modes) are not switched on. As a result, using the two satellite units, complete global coverage can be achieved within 6 to 8 days. Over Europe and Canada the temporal revisit time will be about 3 days. The main advantage of such a systematic data acquisition approach is that it allows the build-up of long data time series at equidistant time intervals (Geudtner et al., 2014). This will especially support applications such as soil moisture monitoring (Hornacek et al., 2012) and SAR interferometry time series analysis (Hooper et al., 2012). As a result, Sentinel-1 will provide backscatter measurement at an unprecedented spatio-temporal sampling, which is expected to be of high value for a broad range of land applications (Wagner et al., 2012).

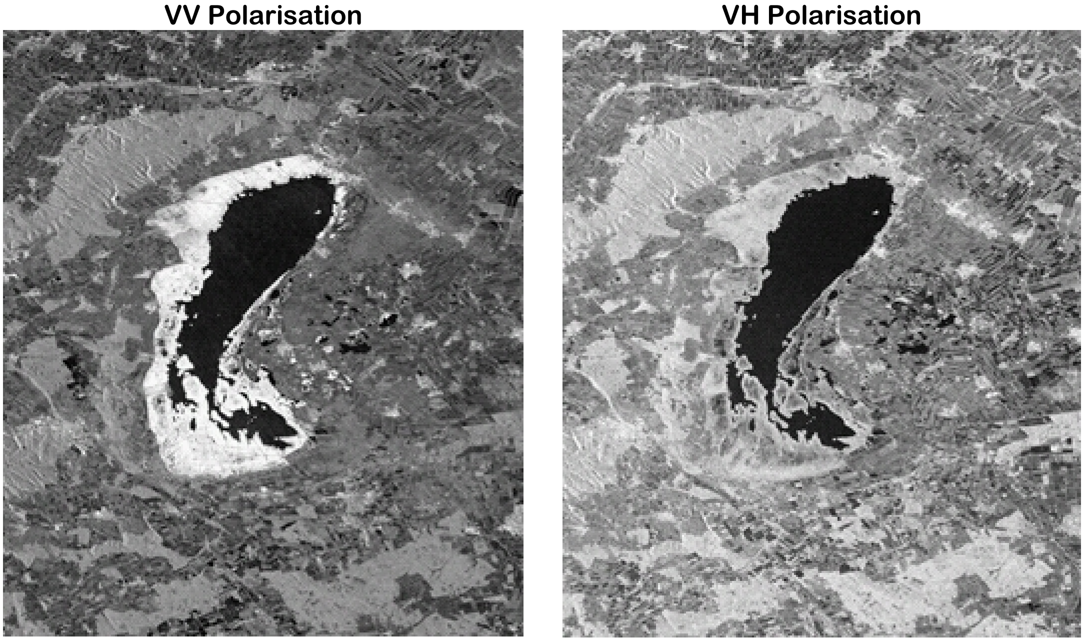


Figure 1: Like polarised (VV, left image) and cross polarised (VH, right image) Sentinel-1 images of the Lake Neusiedl acquired on 2 May.

In addition to the systematic acquisition capability of Sentinel-1, its SAR instrument offers some advanced measurement capabilities, such as the possibility to acquire in dual polarisation. As an example of this capability, Figure 1 shows Sentinel-1 images acquired in IW dual polarisation mode over the Lake Neusiedl, which is located in the east of Austria at the border to Hungary. While in the like-polarised (VV) image a belt of very high backscatter surrounding the lake is clearly visible, this is not the case in the cross-polarised (VH) image. The reason for this very different appearance of the images lies in the different physical interaction mechanisms taking place at the two different polarisations. The backscatter values at VV polarisation is caused by double bounces of the microwave pulses over the lake's reed belt; at VH polarisation volume scattering by the reed dominates.

The high spatio-temporal coverage and the additional measurement capabilities lead to a situation where Sentinel-1 will acquire much more data every year than its predecessor instrument, the Advanced Synthetic Aperture Radar (ASAR), did over the complete 10 years lifetime of the ENVISAT satellite (Figure 2). Once the two satellite units are operational, about 1,8 Terabyte of raw data will be acquired each day. As a result, after the nominal mission life time of 7 years, over 1 Petabyte of raw data will be available. 1 Petabyte is 1 000 000 000 000 000 Bytes, a very large number even by today's standards. If we would like to exploit the wealth of information contained in this enormous data set, it is clear that it must be possible to process it over and over again with ever improving algorithms.

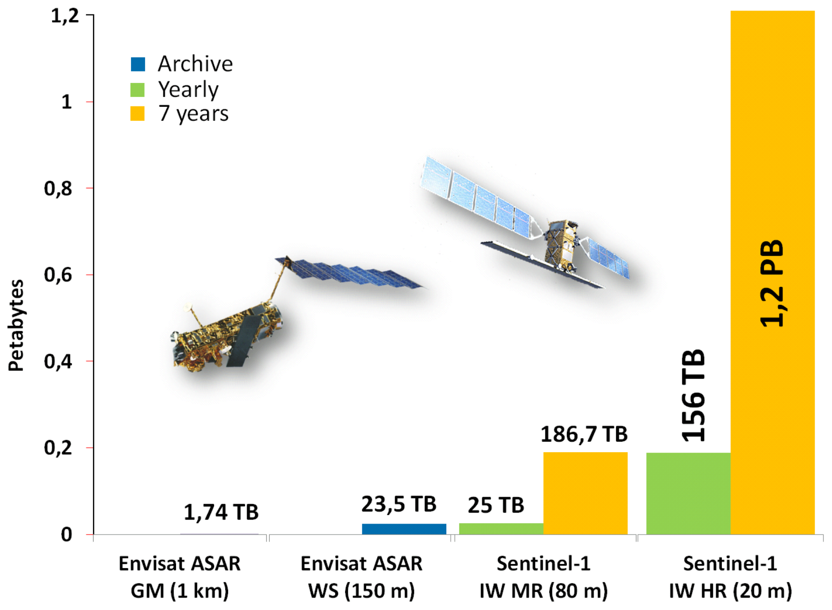


Figure 2: Growth of data volume from ENVISAT ASAR to Sentinel-1.

3. CHALLENGES

3.1. Technical Challenges

As already mentioned in the introduction, storage and telecommunication capabilities are the two technical bottlenecks one has to deal with when working with very large data volumes (Hilbert and López, 2011). As pointed out by Szalay and Blakeley (2009), the analysis of observational datasets is severely limited by the relatively low I/O performance of most of today's computing platforms. Even high-performance numerical simulations are increasingly feeling the "I/O bottleneck." Therefore, to avoid the situation where the performance or capacity of the entire system is limited by a single or small number of components, the hardware components (storage, network, processors, etc.) of Big Data infrastructures have to be chosen in such as way as to balance the system in an optimum way.

But setting up Big Data infrastructures for processing of EO data is not a pure matter of a balanced and cost-efficient design of the hardware, it is also a question of designing the entire EO data management and processing system in such as way as to (i) minimise the transfer of data from/to external remote data centres, (ii) avoid the unnecessary duplication of data while at the same time making sure that there is an easily recoverable backup of "crucial" data sets, and (iii) minimise I/O operations within the entire computer network. Optimising Big Data infrastructures in such a way means that it is essentially a continuous process, where optimisation will often have to be traded with generalisation.

To illustrate why all this is important for the Sentinel data, let us consider the following example: After several years in orbit, the Sentinel-1 "raw" data volume will be several Petabytes large. Let's assume that for our task at hand (e.g. to analyse climate chance impacts on soil moisture and inland water), we need to retrieve 1 Petabyte of Sentinel-1 data stored in ESA's data archive and process these data on our own computers. Assuming a download speed of 100 Mbit/s, the download of this data set over the internet takes about 2,5 years, which is clearly impractical. Alternative solutions are to use dedicated (and expensive) high-speed internet connections or to upload the data on external hardware drives and send these per surface mail. This shows that transferring such large data volumes may be prohibitively expensive and time consuming; hence it must be avoided as far as possible. In any case, let's assume that we have succeeded of having uploaded the 1 Petabyte of Sentinel-1 raw data on our own computer. Using a powerful personal computer with multiple cores, preprocessing (geocoding, radiometric correction, resampling, etc.) of the Sentinel-1 data can currently be done with 1-10 Mbit/s. Assuming a processing speed of 4 Mbit/s the processing of the entire data set would take about 63 years when using just one personal computer! This illustrates that regular reprocessing of the Sentinel-1 data archive is only possible with supercomputers or in a highly scalable cloud computing environment.

3.2. Scientific Challenges

The analysis of Big Data does not just imply the application of existing scientific methods to much bigger data volumes; it usually requires much more. This is because Big Data are normally much more diverse and capture more phenomena than comparable smaller data sets. Consider again the example of Sentinel-1: Due its much improved spatio-temporal sampling characteristics compared to its predecessor SAR mission ERS and ENVISAT, it will observe more physical phenomena taking place at the Earth's surface. This is of course a highly describe feat, but means that our algorithms for analysing and interpreting the data need to be significantly advanced beyond the currently available models. Curiously, it implies that the models will become even more "data thirsty" to unambiguously describe the growing number of phenomena captured by the data. Consequently, it can be expected that the joint use of multiple satellites (e.g. Sentinel-1 and Sentinel-2) will become more and more important.

Advancing our EO data models will be particularly challenging for process-oriented physical models, as it is often quite difficult to introduce new sub-processes without revising the basic structure of the models. This is less of a problem for statistical approaches such as artificial neural networks (ANNs) or support vector regression (SVR), as for these methods it is straight forward to use multiple data inputs, and new data sources can be easily integrated. Possibly, this is also the reason behind the focus of the Big Data community on mere statistical correlations between data sets (Mayer-Schönberger and Cukier, 2013). But Big Data analytics that aim just for finding correlations in the data do not shed light on the inner workings of observed phenomena; they just serve to identify useful proxy data. Also, statistical data models are in general ill suited for capturing “extremes” that are, almost per definition, not well represented by the data. Therefore, pursuing physical approaches that allow gaining insights into the processes observed by the EO satellites continues to be of high importance also in the Big Data era.

3.3. Organisational Challenges

As data volume and diversity increases, and algorithms become more complex, cooperation and sharing of data and source code becomes increasingly important for the EO community. This is the first main driver for rethinking the organisational model in earth observation. The other main driver is that the current way of distributing the EO data to all interested users over the internet is not practical, or affordable any longer. This is why one of the central paradigms of Big Data is to “*bring the software to the data*” rather than vice versa. When this paradigm is properly implemented, it has a series of implications that all together will completely change the way of how the EO community is organised. The first implication is that we will see the emergence of an increasing number of Big Data infrastructure in the coming years, encompassing different infrastructures such as individual big data centres to distributed cloud computing environments. These infrastructures may grow out of existing capacities, or may be newly founded. In any case, they must be able to attract a large EO data user community, making them “decentralised nodes” in an increasingly complex EO infrastructure network (Figure 3).

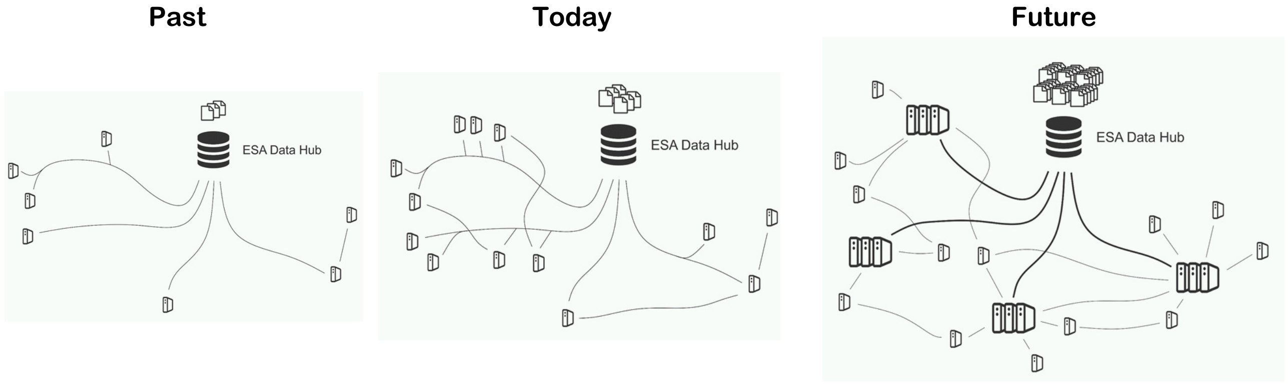


Figure 3: Distributing of earth observation data from a centralised data hub to the users through an increasingly complex infrastructure network.

Once these decentralised nodes have large user communities, individual users can easily start cooperating with other users who work on the same platform. Such cooperation may involve sharing their experiences when working with certain types of data and software, or to directly share data and code (for free, or on a licence- or pay-per-use basis). This will lead to further specialisation of the EO experts, and ultimately to more professional algorithms and data processing chains. Last but not least, it is also clear that there will be benefits from cooperation between the different infrastructures, which is why the development of a federated infrastructure for processing Sentinel and other EO data can be foreseen. All of this will be quite challenging as the underlying Big Data technologies are still evolving, and their landscape remains dynamic (Schade, 2015).

4. BIG DATA INFRASTRUCTURES

The need to move geospatial data analysis, and more specifically earth observation data processing, into the "cloud" has been recognised by many organisations worldwide. Consequently, several organisations and initiatives worldwide have already started, or are preparing for the uptake of Sentinel data into their existing or planned Big Data infrastructures. For the following discussion, I have selected just a few of these initiatives (with an emphasis on the situation in Europe) to give a first impression of the diversity of this field/sector. For further examples and more in-depth technical discussions I would like to refer the readers e.g. to Yue et al. (2013) and Sugumaran et al. (2015).

4.1. Private Sector Offers

Earth observation is one of the disciplines that are often cited as examples for scientific domains that face a data "deluge" (McFedries, 2011). Other commonly cited examples are high-energy physics, genomics, and earth sciences in general. Therefore, it is not surprising that many commercial IT companies have started to offer cloud services for processing of EO data.

The most advanced – and for scientific purposes free – offer comes from Google through its Earth Engine (https://earthengine.google.org/). As described on their website "*Google Earth Engine* brings together the world's satellite imagery and makes it available online with tools for scientists, independent researchers, and nations to mine this massive warehouse of data to detect changes, map trends and quantify differences on the Earth's surface." Google Earth Engine is attracting an increasing number of scientists worldwide, with the first landmark study published by Hansen et al. (2013) on the global mapping of forest cover change at a spatial resolution of 30 m for the years 2000 to 2010. Google plans to host all Sentinel data, putting the Earth Engine in a good position to become a premier platform for the scientific exploitation of Sentinel data.

Another commercial IT giant that has recently moved into the earth observation market is Amazon. Amazon offers through its *Amazon Web Service* (AWS) access to Landsat 8 data from 2015 onwards (http://aws.amazon.com/public-data-sets/landsat/). The data themselves are free, but users need to pay for Amazons cloud computing services, whereas costs are composted of the following three components: compute, storage, and data transfer out. Huang et al. (2010) describe their first experiences when implementing a geospatial data and service catalogue on AWS.

Also European based commercial IT providers have identified earth observation as a potential business case for their cloud service offers. Within the framework of the *Helix Nebula* (HN) partnership (http://www.helix-nebula.eu/), also often referred to as the European Science Cloud, they have engaged in a cooperation with three of Europe's leading research organisations – including ESA – in order to provide computing capacity and services that elastically meet big science's growing demand for computing power. So far, the initiative has deployed and validated three high-profile flagship experiments in high energy physics, life sciences and earth science, on commercial cloud services hosted by multiple suppliers. Despite many unsolved problems related to the standardisation and federation of cloud services, HN has credibly demonstrated the high potential of cloud computing technology for addressing the problem of big data in science.

All of the above examples involve commercial cloud providers that have moved into the field of earth observation, but the other way is also of course possible. Just as one example for such a venture, I mention here the Munich-based company *CloudEO*. Their business model is nicely summarised in Figure 4, which has been taken from their website (www.cloudeo-ag.com).

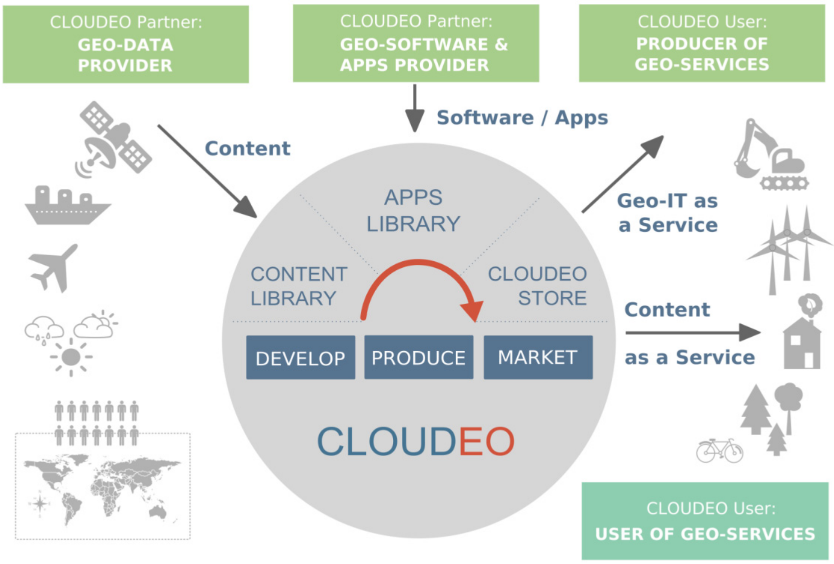


Figure 4: Business model of CloudEO. Image taken from http://www.cloudeo-ag.com/how-it-works.

4.2. Public Sector Initatives

Public sector initiatives with the aim to establish Big Data infrastructures capable of processing Sentinel data have mostly sprung up from national and international space agencies already involved in setting up the Sentinel ground segment. On a technical level, the challenges caused by the big data volumes became clear at an early stage of the development of the Sentinel programme. However, translating these technical requirements into a plan for action with corresponding policies regulating the responsibilities of ESA, EC and their member states has taken time. In fact, this process is still ongoing.

On a national level, progress towards the establishment of collaborative infrastructures for processing Sentinel data has e.g. been made by some of the bigger European countries (Germany, France, United Kingdom, …), but also some of the smaller countries have launched successful initiatives (e.g. Norway, Austria, etc.). Each country has followed its own approach, with different levels of involvement of the private industry. For example, the French *THEIA Land Data Center* (http://www.theia-land.fr/) is a national inter-agency organization designed to foster the use of images coming from the space observation of land surfaces. The Centre organizes the French skills in this field and relies on a distributed data centre infrastructure in order to provide data to more than 400 laboratories and 100 graduate schools on the national territory. In the United Kingdom the *Climate, Environment and Monitoring from Space* (CEMS) facility (also referred to as Satellite Applications Catapult, https://sa.catapult.org.uk/cems#) was established. It is a purpose-built facility offering users access to CEMS data and various platform services.

On an international level, ESA has recently selected so-called *Thematic Exploitation Platforms* for five thematic application areas: coastal environment, forestry, hydrology, polar, and urban. The TEPs correspond to virtual workspaces providing a user community interested in a common Earth science topic with very fast access to: (i) large volumes of EO data, (ii) computing resources, (iii) processing software, and (iv) general platform capabilities. One ambition of the TEP programme is to keep the platform architecture open, scalable and independent of any specific IT infrastructure. This shall allow an easy expansion of the platform's capabilities, and encourage the development of a wider network of other thematic exploitation platforms. As an example for a TEP, see e.g. the website of the Polar TEP (http://p-tep.polarview.org/).

4.3. Public-Private Partnerships

Many of the above discussed Big Data infrastructures involve collaboration between public and private partners, at least to some degree (one may even say this for Google and Amazon as they benefit – like anybody else – from open data policies). However, true public-private partnerships where both sides invest and have comparable influence in the decision making process are rare. One such example is the Earth Observation Data Centre for Water Resources Monitoring (EODC), which was founded as a public-private partnership in 2015 (https://www.eodc.eu/). As described by Wagner et al. (2014), it follows the example of the successful cooperation model of the 52°North Initiative for Geospatial Open Source Software (52north.org/), adopted to its own goals and mission statement (Figure 5). In this cooperation model, partners from the private and public sectors work together in so called communities, which have clearly specified goals, e.g. the federation of infrastructure resources or joint software developments. The cooperation is facilitated and promoted by the EODC GmbH. Principal Cooperation Partners have particular responsibilities in this framework, i.e. they lead communities and participate in the Advisory Board, thereby being able to influence the strategic orientation of the EODC.

Figure 5: EODC cooperation model.

One of the central goals of the EODC is to provide collaborative IT infrastructure for archiving, processing, and distributing EO data. The approach chosen by EODC is to work together with scientific organisations, research centres and cloud providers, and to manage and coordinate – as far as necessary – the utilisation of their infrastructures. Within this framework, EODC has worked together with the Technische Universität Wien (TU Wien) and the Austrian meteorological service

ZAMG, to establish a Big Data infrastructure for processing of Sentinel data in Vienna (Figure 6). Key components of this infrastructure, which went operational in spring 2015, are (i) a platform for 24/7 processing, (ii) a virtual environment for collaborative algorithm development, (iii) Petabyte storage, and (iv) access to the Vienna Scientific Cluster 3 (VSC-3), which was ranked 85th in the November 2014 worldwide TOP500 supercomputer list.

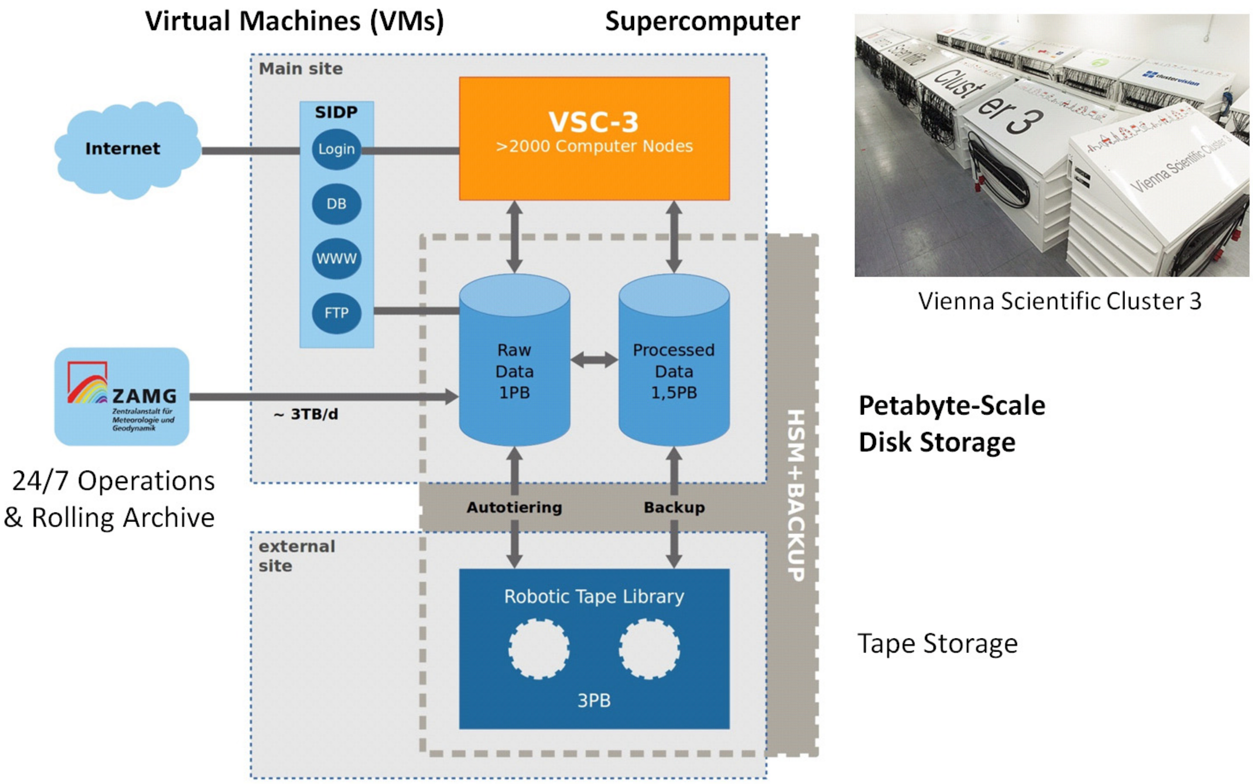


Figure 6: EODC infrastructure in Vienna hosted by the Technische Universität Wien (TU Wien) and the Austrian meteorological service ZAMG.

5. CONCLUSIONS

The Sentinel satellites represent a step change in our capability to collect environmental data. This makes them very attractive to an increasing number of applications, but raises at the same time the question of how to cope with such massive amounts and diversity of data? The only plausible way appears to be a complete overhaul of how the processing and analysis of earth observation (EO) data is currently organised. Instead of distributing the data to hundreds or thousands of users over the internet, it is necessary to follow one of the central paradigms of the Big Data era, namely to bring the “software to the data”. Several organisations and initiatives have recognised this need at an early stage, and are now working towards the establishment of data centres and cloud computing environments capable of hosting and processing Petabyte of Sentinel data. Despite their starting points were in general very different, one can nonetheless discern two (inverse) directions of how this task is approached. The first approach is to start offering EO specific services (e.g. access to EO data and expert software) on existing general-purpose cloud computing environments. This is the path typically followed by commercial cloud providers and big IT companies. The second approach is to build new, or expand existing, dedicated EO data processing infrastructures, a path typically followed by space agencies and EO expert organisations. As most of the discussed initiatives are at an early stage of development, the level of cooperation between them is yet limited. However, considering the significant technical, scientific and organisational challenges as discussed in Section

3, it appears indispensable to strengthen the level of cooperation between them, e.g. by forming public-private partnerships. Only this will allow building up the necessary scientific and technical capacities required to fully exploit the wealth of information provided by the Sentinel satellites.

6. ACKNOWLEDGEMENTS

I would like to thank my colleagues from TU Wien and EODC, and all our partners from the EODC network for the many fruitful discussions that were the basis for writing this paper. Vahid Naeimi, Bernhard Bauer-Marschallinger, Thomas Mistelbauer, Hans Thüminger and Duy Nguyen Ba from TU Wien helped me with the preparation of the figures. Furthermore, the support by the Austrian Space Application Programme (Projects 844350 "Prepare4EODC" and 88001 "WetMon") and the European Space Agency (Contract No. 4000107319/12/I-BG "EODC Water Study") is gratefully acknowledged.

7. REFERENCES

Geudtner, D., Torres, R., Snoeij, P., Davidson, M. and Rommen, B., 2014: Sentinel-1 System capabilities and applications, IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1457-1460.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O. and Townshend, J. R. G., 2013: High-resolution global maps of 21st-century forest cover change. Science 342 (6160), pp. 850-853.

Hilbert, M. and López, P., 2011: The World's Technological Capacity to Store, Communicate, and Compute Information. Science 332 (6025), pp. 60-65.

Hooper, A., Bekaert, D., Spaans, K. and Arikan, M., 2012: Recent advances in SAR interferometry time series analysis for measuring crustal deformation. Tectonophysics 514, pp. 1-13.

Hornacek, M., Wagner, W., Sabel, D., Truong, H. L., Snoeij, P., Hahmann, T., Diedrich, E. and Doubkova, M., 2012: Potential for High Resolution Systematic Global Surface Soil Moisture Retrieval via Change Detection Using Sentinel-1. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 5 (4), pp. 1303-1311.

Huang, Q., Yang, C., Nebert, D., Liu, K. and Wu, H., 2010: Cloud computing for geosciences: deployment of GEOSS clearinghouse on Amazon's EC2. In: Proceedings of the ACM SIGSPATIAL international workshop on high performance and distributed geographic information systems, pp. 35-38.

Lazer, D.M., Kennedy, R., King, G. and Vespignani, A., 2014: The parable of Google Flu: Traps in big data analysis. Science 343, pp. 1203-1205.

Mattmann, C. A., 2013: Computing: A vision for data science. Nature 493 (7433), pp. 473-475.

Mayer-Schönberger, V. and Cukier, K., 2013: Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt Publishing Company, New York.

McFedries, P., 2011: The coming data deluge. IEEE Spectrum 48 (2), pp. 19-19.

Mervis, J., 2012: Agencies Rally to Tackle Big Data. Science 336 (6077), pp. 22-22.

Schade, S., 2015: Big Data breaking barriers – first steps on a long trail. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XL-7/W3, pp. 691-697.

Spiegelhalter, D., 2014: The future lies in uncertainty. Science 345 (6194), pp. 264-265.

Sugumaran, R., Hegeman, J. W., Sardeshmukh, V. B. and Armstrong, M. P., 2015: Processing remote-sensing data in cloud computing environments. In: P. S. Thenkabail (Editor), Remote Sensing Handbook. CRC Press, pp. 549-558.

Szalay, A. S. and Blakeley, J. A., 2009: Gray's laws: database-centric computing in science, The Fourth Paradigm - Data-intensive scientific discovery, pp. 5-11.

Wagner, W., Fröhlich, J., Wotawa, G., Stowasser, R., Staudinger, M., Hoffmann, C., Walli, A., Federspiel, C., Aspetsberger, M., Atzberger, C., Briese, C., Notarnicola, C., Zebisch, M., Boresch, A., Enenkel, M., Kidd, R., von Beringe, A., Hasenauer, S., Naeimi, V. and Mücke, W., 2014: Addressing grand challenges in earth observation science: The Earth Observation Data Centre for Water Resources Monitoring. In: ISPRS Technical Commission VII Mid-term Symposium 2014, Volume II-7, Istanbul, Turkey, pp. 81-88.

Wagner, W., Sabel, D., Doubkova, M., Hornacek, M., Schlaffer, S. and Bartsch, A., 2012: Prospects of Sentinel-1 for land applications. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 22-27 July 2012, pp. 1741-1744.

Yue, P., Zhou, H., Gong, J. and Hu, L., 2013: Geoprocessing in Cloud Computing platforms–a comparative analysis. International Journal of Digital Earth 6 (4), pp. 404-425.