data

MDPI

Editorial

Earth Observation Open Science: Enhancing Reproducible Science Using Data Cubes

Gregory Giuliani [1,2,*], Gilberto Camara [3], Brian Killough [4] and Stuart Minchin [5]

1 Institute for Environmental Sciences, University of Geneva, Bd Carl-Vogt 66, CH-1205 Geneva, Switzerland
2 United Nation Environment Programme, Science Division, GRID-Geneva, 11 Chemin des Anémones, CH-1219 Châtelaine, Switzerland
3 Group on Earth Observations, 7bis Avenue de la Paix, Case Postale 2300, 1211 Geneva, Switzerland; gcamara@geosec.org
4 National Aeronautics and Space Administration, Langley Research Center, MS 457, Hampton, VA 23681, USA; brian.d.killough@nasa.gov
5 Geoscience Australia, GPO Box 378, Canberra, ACT 2601, Australia; stuart.minchin@ga.gov.au
* Correspondence: gregory.giuliani@unige.ch; Tel.: +41-(0)22-379-07-09

Received: 21 November 2019; Accepted: 21 November 2019; Published: 25 November 2019

Abstract: Earth Observation Data Cubes (EODC) have emerged as a promising solution to efficiently and effectively handle Big Earth Observation (EO) Data generated by satellites and made freely and openly available from different data repositories. The aim of this Special Issue, "Earth Observation Data Cube", in *Data*, is to present the latest advances in EODC development and implementation, including innovative approaches for the exploitation of satellite EO data using multi-dimensional (e.g., spatial, temporal, spectral) approaches. This Special Issue contains 14 articles covering a wide range of topics such as Synthetic Aperture Radar (SAR), Analysis Ready Data (ARD), interoperability, thematic applications (e.g., land cover, snow cover mapping), capacity development, semantics, processing techniques, as well as national implementations and best practices. These papers made significant contributions to the advancement of a more Open and Reproducible Earth Observation Science, reducing the gap between users' expectations for decision-ready products and current Big Data analytical capabilities, and ultimately unlocking the information power of EO data by transforming them into actionable knowledge.

Keywords: open science; reproducibility; earth observations; data cube; analysis ready data; remote sensing; satellite imagery

---

Planet Earth is currently on an unsustainable pathway. Increasing pressures on natural resources induced by human activities are globally affecting the environment. Regular and continuous monitoring is necessary to assess, understand, and mitigate these environmental changes [1–3]. Consequently, timely and reliable access to data describing physical, chemical, biological, and socio-economic conditions can provide the basis for reliable and accountable scientific understanding and knowledge about the limits of our planet. This access to data can support informed decisions and evidence-based policies for the efficient use of our planet's resources [4,5].

To facilitate environmental monitoring, our planet has been under continuous observations from satellites since 1972 [6,7]. Today, remotely sensed Earth Observations (EO) data have already exceeded the petabyte-scale and increasingly are freely and openly available from different data repositories [8]. This poses a number of issues in terms of Volume (e.g., data volumes have increased by 10 in the last 5 years); Velocity (e.g., Sentinel-2 is capturing a new image of a given place every 5 days); and Variety (e.g., different type of sensors, spatial/spectral resolutions). Traditional approaches to the acquisition, management, distribution, and analysis of satellite EO data have limitations (e.g., data size, heterogeneity and complexity) that impede the massive use and analysis of Big Earth Data.

The fact that the full information potential of EO data has not yet been realized and therefore remains still underutilized is explained by various reasons: (1) it requires scientific knowledge to understand what data is needed—optical (which resolution?)—radar (which type?); (2) it is difficult to access and download the increasing volumes of data generated by satellites; (3) there is a lack of expertise and computing resources to efficiently prepare and utilize EO data; (4) the particular structure of EO data and (5) the significant effort and cost required to store and process data limit its effective use.

Addressing Big Data challenges such as Volume, Velocity and Variety, requires a change of paradigm and a move away from traditional data-centric approaches (e.g., local processing and data distribution methods to lower the barriers caused by data size and related complications in data management [9,10]. In particular, data volume and velocity will continue to grow as the demands increase for decision-support information derived from these data [11]. Using the cloud, it is now possible to move algorithms and tools to data, making large volumes of EO data available to a wide range of users, enabling them to handle and visualize data they are interested in without having to download them and consequently avoiding large-scale data transfers that can impede the efficient and effective use of EO data [12,13].

To tackle these issues and bridge the gap between users' expectations and current Big Data analytical capabilities, EO Data Cubes (EODC) have emerged as a new paradigm revolutionizing the way users can interact with EO data and providing a promising solution for the storage, organization, management, and analysis of Big EO data [14]. The main objective of EODC is to facilitate EO data usage by addressing Volume, Velocity, Variety challenges and providing access to large spatio-temporal data in an analysis-ready format [15].

Different EODC implementations are currently operational, such as Digital Earth Australia [16], the Swiss Data Cube [17], the EarthServer [18], the E-sensing platform [19] the Copernicus Data and Information Access Services (DIAS) [20] or the Google Earth Engine [21]. These initiatives are paving the way for broadening the use of EO data to larger communities of users, supporting decision-makers with timely and actionable information converted into meaningful geophysical variables and ultimately unlocking the information power of EO data.

All these developments would not have been possible without Free and Open Data policies to facilitate access to data and Open Source code to efficiently develop software solutions [22]. Open Science is a new approach to research and educational processes, which seeks to make scientific research more collaborative and transparent and to make knowledge accessible by using digital technologies and new collaborative tools [23]. Achieving reproducible knowledge requires exposing all parts of an application (e.g., code, data, executable) [24]. Therefore, Open Science is considered as an umbrella term encompassing all practices that aim to remove barriers to sharing any type of output (e.g., research data), resources (e.g., scientific publications), methods (e.g., lab notes) or tools (e.g., software). This is a practice of science to achieve more openness and to enable others to collaborate and contribute under terms that enable the reuse, redistribution and reproduction of research and its underlying data and methods [25]. In particular, with the advent of cloud computing, knowledge is easier to share [11]. Open Science is fundamental in a 21st Century where Science is embedded in societal decision-making. Increased openness and transparency are effective means to fight fake news and post-truth [26].

Despite the fact that in the EO domain various open science practices are already adopted, such as the Open Standards provided by the Open Geospatial Consortium (OGC) [27], Open Source software [28], Open Code Library (e.g., Open remote sensing http://openremotesensing.net) or the IEEE Remote Sensing Code Library (http://www.grss-ieee.org/publication-category/rscl/), data sets and algorithm evaluation standards (http://dase.grss-ieee.org), or Open Data licenses for Landsat and Sentinel data [29], EO Open Science remains underestimated and various socio-cultural, technological, political, organizational, economic and legal challenges (e.g., lack of recognition and rewards, overload for opening data, changing working procedure, missing political endorsements–strategies–policies, unclear legal frameworks) need to be addressed to adequately realize its full potential.

This Special Issue (https://www.mdpi.com/journal/data/special_issues/EODC) presents some of the most recent advancements in the use and implementation of EODC. They significantly contribute to the advancement of a more Open and Reproducible EO Science and help to reduce the gap between data and knowledge.

Most of the Open Science facets (https://www.fosteropenscience.eu/content/what-open-science-introduction) are covered by the contributions of this Special Issue. First of all, the 14 papers are accessible in *Open Access* under the terms and conditions of the Creative Commons Attribution (CCBY) license. This means that these research outputs are distributed online and freely available, removing the barriers to copying or reuse by applying an open license copyright. Together with FAIR guiding principles, this allows sharing of findings and streamlining of the creation of new data products by making them Findable, Accessible, Interoperable and Reusable [30,31].

With the different innovative solutions that are available to implement EODC, one of the major challenges is to prevent them from becoming silos of information. Interoperability is consequently an important aspect to consider. Giuliani et al. [32] demonstrated how widely adopted geospatial standards can be used to enhance the interoperability of EODC and can help in delivering and leveraging the power of EO data building, efficient discovery, access and processing services. However, to harness the information potential of satellite EO data, syntactic interoperability is not sufficient. As numerical sensory data have no semantic meaning, EO data lack semantics. Augustin et al. [33] clarify and share their definition of semantic EODC, demonstrating how they enable different possibilities for data retrieval, semantic queries based on EO data content, and semantically enabled analysis. Semantic EODC are the foundation of the EO data expert system and can facilitate deriving knowledge, as presented by Plag and Jules Plag [34].

Regarding *Open Data*, one of the main topics concerns the development of Analysis Ready Data (ARD) for Synthetic Aperture Radar (SAR) imagery. Indeed, if the provision of optical ARD is becoming common, the complexity of SAR data makes them challenging to developed procedures for the regular provision of SAR ARD. Truckenbrodt et al. [35] and Ticehurst et al. [36] assessed the feasibility of automatically producing analysis-ready radiometrically terrain-corrected (RTC) Synthetic Aperture Radar (SAR) gamma nought backscatter data from Sentinel-1. Both studies concluded that the European Space Agency (ESA) Sentinel Application Platform (SNAP) toolbox (https://step.esa.int/main/toolboxes/snap/) is a valid solution for producing Sentinel-1 ARD products. One important reward in publishing open data is the possibility to be cited. Providing a reference to data similar to scientific journal articles or conference papers is increasingly recognized as an essential practice leading to the recognition of data as important research outputs. Data citation supports (1) attribution and credit; (2) collaboration and reuse of data; (3) enables reproducibility of findings; (4) faster and efficient research progress, and (5) provides means to share data with (future) researchers. Schubert et al. [37] presented a solution for an operational service on dynamic data citation to enable the effective reuse of EO data in a collaborative and reproducible manner.

To benefit from the large volume of EO data made available with Data Cubes, recent *Open Source* developments were allowed to implement solutions in open source geoinformation and statistical software. Gebbert et al. [38] have developed spatio-temporal topological operators in the GRASS GIS software to enable the effective use of heterogenous (e.g., extent, granularity) spatio-temporal EO data. Similarly, Appel et al. [39] introduced an open source C++ library and R package for the construction and processing of on-demand data cubes from satellite image collections, and showed how it supports interactive method development workflows where data users can initially try methods on small subsamples before running analyses on high resolution and/or large areas. Finally, *Open Standards* such as the OGC Web Map Service (WMS), together with modern web browser capabilities, has enabled time-series analysis directly within a web-based application [40].

To reach the objective of facilitated and reproducible analysis of EO data, as well as empowering a large community of users to benefit from satellite EO data, *Open Notebooks* appear as promising solutions. They help to document research developments as reproducible experiments and facilitate the sharing of scientific data analysis. Electronic Lab Notebooks (ELN), such as Jupyter Notebooks, are replacing paper lab notebooks with digital versions. Kopp et al. [41] showed that such notebooks simplify access and use for end-users, enabling a wide variety of web and desktop applications. Poussin et al. [42] demonstrated the benefits of Open and Reproducible Science using a snow detection algorithm, developed in Switzerland and shared as an open notebook, to monitor snow cover evolution for the last three decades in the Gran Paradiso National Park in Italy. Furthermore, Lucas et al. [43] developed a conceptual framework to implement a Land Cover Change model, providing Australia and other countries using the Open Data Cube (ODC) environment with the opportunity to routinely generate land cover maps from Landsat or Sentinel-1/2 data, at least annually and using a consistent and internationally recognized taxonomy.

Finally, an important aspect related to any new technology lies in developing new capacities to reach large adoption, acceptance and commitment. Asmaryan et al. [44] presented how effective knowledge transfer, using *Open Educational resources*, has been achieved between Switzerland and Armenia for developing and implementing the first version of an Armenian Data Cube. This ultimately can support National Open Data Cubes to contribute to country-level development policies and practices [45].

To conclude, we believe that EODC have the potential to achieve the vision of transforming data into actionable knowledge by lowering the entry barrier to massive-use Big Earth Data analysis and therefore act as an information technology enabler. Ultimately, it can provide an effective mean to build socially robust, replicable, and reusable knowledge, to generate decision-ready products based on Open Science.

Author Contributions: G.G.: conceptualization, writing—original draft preparation; G.C.: writing—review and editing; B.K.: writing—review and editing; S.M.: writing—review and editing.

Acknowledgments: The authors gratefully acknowledge the administrative and technical support of the Data journal team. We want to thank the authors who contributed towards this Special Issue on "Earth Observations Data Cube", as well as the reviewers who provided the authors with comments and very constructive feedback.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rockström, J.; Bai, X.; DeVries, B. Global sustainability: the challenge ahead. *Glob. Sustain.* **2018**, *1*, 1–3. [CrossRef]
2. Steffen, W.; Richardson, K.; Rockström, J.; Cornell, S.E.; Fetzer, I.; Bennett, E.M.; Biggs, R.; Carpenter, S.R.; De Vries, W.; De Wit, C.A.; et al. Planetary boundaries: Guiding human development on a changing planet. *Science* **2015**, *347*, 1259855. [CrossRef] [PubMed]
3. Biermann, F.; Bai, X.; Bondre, N.; Broadgate, W.; Chen, C.-T.A.; Dube, O.P.; Erisman, J.W.; Glaser, M.; van der Hel, S.; Lemos, M.C.; et al. Down to Earth: Contextualizing the Anthropocene. *Glob. Environ. Chang.* **2016**, *39*, 341–350. [CrossRef]
4. Giuliani, G.; Nativi, S.; Obregon, A.; Beniston, M.; Lehmann, A. Spatially enabling the Global Framework for Climate Services: Reviewing geospatial solutions to efficiently share and integrate climate data & information. *Clim. Serv.* **2017**, *8*, 44–58.
5. Lehmann, A.; Chaplin-Kramer, R.; Lacayo, M.; Giuliani, G.; Thau, D.; Koy, K.; Goldberg, G.; Sharp, R., Jr. Lifting the Information Barriers to Address Sustainability Challenges with Data from Physical Geography and Earth Observation. *Sustainability* **2017**, *9*, 858. [CrossRef]
6. Zhu, Z. Science of Landsat Analysis Ready Data. *Remote Sens.* **2019**, *11*, 2166. [CrossRef]
7. Wulder, M.A.; Loveland, T.R.; Roy, D.P.; Crawford, C.J.; Masek, J.G.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Belward, A.S.; Cohen, W.B.; et al. Current status of Landsat program, science, and applications. *Remote Sens. Environ.* **2019**, *225*, 127–147.

8. Woodcock, C.E.; Allen, R.; Anderson, M.; Belward, A.; Bindschadler, R.; Cohen, W.; Gao, F.; Goward, S.N.; Helder, D.; Helmer, E.; et al. Free Access to Landsat Imagery. *Science* **2008**, *320*, 1011. [CrossRef]
9. Nativi, S.; Mazzetti, P.; Craglia, M. A view-based model of data-cube to support big earth data systems interoperability. *Big Earth Data* **2017**, *1*, 75–99. [CrossRef]
10. Nativi, S.; Mazzetti, P.; Santoro, M.; Papeschi, F.; Craglia, M.; Ochiai, O. Big Data challenges in building the Global Earth Observation System of Systems. *Environ. Model. Softw.* **2015**, *68*, 1–26. [CrossRef]
11. Nativi, S.; Santoro, M.; Giuliani, G.; Mazzetti, P. Towards a knowledge base to support global change policy goals. *Int. J. Digit. Earth* **2019**, 1–29. [CrossRef]
12. Boulton, G. The challenges of a Big Data Earth. *Big Earth Data* **2018**, *4471*, 1–7. [CrossRef]
13. Guo, H. Big Earth data: A new frontier in Earth and information sciences. *Big Earth Data* **2017**, *1*, 4–20. [CrossRef]
14. Baumann, P.; Misev, D.; Merticariu, V.; Huu, B.P. Datacubes: Towards Space/Time Analysis-Ready Data. In *Service-Oriented Mapping*; Springer: Cham, Switzerland, 2019; pp. 269–299.
15. Dwyer, J.; Roy, D.; Sauer, B.; Jenkerson, C.; Zhang, H.; Lymburner, L. Analysis Ready Data: Enabling Analysis of the Landsat Archive. *Remote Sens.* **2018**, *10*, 1363.
16. Dhu, T.; Dunn, B.; Lewis, B.; Lymburner, L.; Mueller, N.; Telfer, E.; Lewis, A.; McIntyre, A.; Minchin, S.; Phillips, C. Digital earth Australia—Unlocking new value from earth observation data. *Big Earth Data* **2017**, *1*, 64–74. [CrossRef]
17. Giuliani, G.; Chatenoux, B.; De Bono, A.; Rodila, D.; Richard, J.-P.; Allenbach, K.; Dao, H.; Peduzzi, P. Building an Earth Observations Data Cube:Lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD). *Big Earth Data* **2017**, *1*, 100–117. [CrossRef]
18. Baumann, P.; Mazzetti, P.; Ungar, J.; Barbera, R.; Barboni, D.; Beccati, A.; Bigagli, L.; Boldrini, E.; Bruno, R.; Calanducci, A.; et al. Big Data Analytics for Earth Sciences: The EarthServer approach. *Int. J. Digit. Earth* **2016**, *9*, 3–29. [CrossRef]
19. Camara, G.; Ribeiro, G.; Vinhas, L.; Ferreira, K.R.; Cartaxo, R.; Simões, R.; Llapa, E.; Assis, L.F.; Sanchez, A. The e-Sensing architecture for big Earth observation data analysis. In Proceedings of the 2017 Conference on Big Data from Space (BiDS'17), Toulouse, France, 28–30 November 2017; pp. 1–4.
20. European Commission. The DIAS: User-friendly Access to Copernicus Data and Information. 2018. Available online: https://www.copernicus.eu/sites/default/files/Copernicus_DIAS_Factsheet_June2018.pdf (accessed on 25 November 2019).
21. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [CrossRef]
22. Ferrari, T.; Scardaci, D.; Andreozzi, S. *The Open Science Commons for the European Research Area. Earth Observation Open Science and Innovation*; Springer: Cham, Switzerland, 2018; pp. 43–67.
23. European Commission. *Open Innovation, Open Science, Open to the World — A Vision for Europe*; Directorate-General for Research and Innovation: Brussels, Belgium, 2016.
24. Peng, R.D. Reproducible Research in Computational Science. *Science* **2011**, *334*, 1226–1227. [CrossRef]
25. McKiernan, E.; Bourne, P.; Brown, C.; Buck, S.; Kenall, A.; Lin, J.; McDougall, D.; Nosek, B.A.; Ram, K.; Soderberg, C.K.; et al. How open science helps researchers succeed. *eLife* **2016**, *5*, e16800. [CrossRef]
26. Cornell, S.; Berkhout, F.; Tuinstra, W.; Tàbara, J.D.; Jäger, J.; Chabay, I.; de Wit, B.; Langlais, R.; Mills, D.; Moll, P.; et al. Opening up knowledge systems for better responses to global environmental change. *Environ. Sci. Policy* **2013**, *28*, 60–70. [CrossRef]
27. Maso, J.; Zabala, A.; Serral, I.; Pons, X. Remote Sensing Analytical Geospatial Operations Directly in the Web Browser. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Delft, The Netherlands, 1–5 October 2018; Volume XLII–4, pp. 403–410.
28. Grizonnet, M.; Michel, J.; Poughon, V.; Inglada, J.; Savinaud, M.; Cresson, R. Orfeo Toolbox: Open Source Processing of Remote Sensing Images. Open Geospat. *Data Softw. Stand.* **2017**, *2*, 15.
29. Ryan, B. The benefits from open data are immense. *Geospat. World* **2016**, 72–73.
30. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; Santos, L.B.D.; Bourne, P.E.; et al. Comment: The fair guiding principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
31. Stall, S.; Yarmey, L.; Cutcher-Gershenfeld, J.; Hanson, B.; Lehnert, K.; Nosek, B.; Parsons, M.; Robinson, E.; Wyborn, L. Make scientific data FAIR. *Nature* **2019**, *570*, 27. [CrossRef] [PubMed]

32. Giuliani, G.; Masó, J.; Mazzetti, P.; Nativi, S.; Zabala, A. Paving the Way to Increased Interoperability of Earth Observations Data Cubes. *Data* **2019**, *4*, 113. [CrossRef]
33. Augustin, H.; Sudmanns, M.; Tiede, D.; Lang, S.; Baraldi, A. Semantic Earth Observation Data Cubes. *Data* **2019**, *4*, 102. [CrossRef]
34. Plag, H.-P.; Jules-Plag, S.-A. A Transformative Concept: From Data Being Passive Objects to Data Being Active Subjects. *Data* **2019**, *4*, 135. [CrossRef]
35. Truckenbrodt, J.; Freemantle, T.; Williams, C.; Jones, T.; Small, D.; Dubois, C.; Thiel, C.; Rossi, C.; Syriou, A.; Giuliani, G. Towards Sentinel-1 SAR Analysis-Ready Data: A Best Practices Assessment on Preparing Backscatter Data for the Cube. *Data* **2019**, *4*, 93. [CrossRef]
36. Ticehurst, C.; Zhou, Z.-S.; Lehmann, E.; Yuan, F.; Thankappan, M.; Rosenqvist, A.; Lewis, B.; Paget, M. Building a SAR-Enabled Data Cube Capability in Australia Using SAR Analysis Ready Data. *Data* **2019**, *4*, 100. [CrossRef]
37. Schubert, C.; Seyerl, G.; Sack, K. Dynamic Data Citation Service—Subset Tool for Operational Data Management. *Data* **2019**, *4*, 115. [CrossRef]
38. S Gebbert, S.; Leppelt, T.; Pebesma, E. A Topology Based Spatio-Temporal Map Algebra for Big Data Analysis. *Data* **2019**, *4*, 86. [CrossRef]
39. Appel, M.; Pebesma, E. On-Demand Processing of Data Cubes from Satellite Image Collections with the gdalcubes Library. *Data* **2019**, *4*, 92. [CrossRef]
40. Maso, J.; Zabala, A.; Serral, I.; Pons, X. A Portal Offering Standard Visualization and Analysis on top of an Open Data Cube for Sub-National Regions: The Catalan Data Cube Example. *Data* **2019**, *4*, 96. [CrossRef]
41. Kopp, S.; Becker, P.; Doshi, A.; Wright, D.J.; Zhang, K.; Xu, H. Achieving the Full Vision of Earth Observation Data Cubes. *Data* **2019**, *4*, 94. [CrossRef]
42. Poussin, C.; Guigoz, Y.; Palazzi, E.; Terzago, S.; Chatenoux, B.; Giuliani, G. Snow Cover Evolution in the Gran Paradiso National Park, Italian Alps, Using the Earth Observation Data Cube. *Data* **2019**, *4*, 138. [CrossRef]
43. Lucas, R.; Mueller, N.; Siggins, A.; Owers, C.; Clewley, D.; Bunting, P.; Kooymans, C.; Tissott, B.; Lewis, B.; Lymburner, L.; et al. Land Cover Mapping using Digital Earth Australia. *Data* **2019**, *4*, 143. [CrossRef]
44. Asmaryan, S.; Muradyan, V.; Tepanosyan, G.; Hovsepyan, A.; Saghatelyan, A.; Astsatryan, H.; Grigoryan, H.; Abrahamyan, R.; Guigoz, Y.; Giuliani, G. Paving the Way towards an Armenian Data Cube. *Data* **2019**, *4*, 117. [CrossRef]
45. Dhu, T.; Guiliani, G.; Juárez, J.; Kavvada, A.; Killough, B.; Merodio, P.; Minchin, S.; Ramage, S. National Open Data Cubes and Their Contribution to Country-Level Development Policies and Practices. *Data* **2019**, *4*, 144. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).