



Big Earth Data science: an information framework for a sustainable planet

Huadong Guo, Stefano Nativi, Dong Liang, Max Craglia, Lizhe Wang, Sven Schade, Christina Corban, Guojin He, Martino Pesaresi, Jianhui Li, Zeeshan Shirazi, Jie Liu & Alessandro Annoni

To cite this article: Huadong Guo, Stefano Nativi, Dong Liang, Max Craglia, Lizhe Wang, Sven Schade, Christina Corban, Guojin He, Martino Pesaresi, Jianhui Li, Zeeshan Shirazi, Jie Liu & Alessandro Annoni (2020): Big Earth Data science: an information framework for a sustainable planet, International Journal of Digital Earth, DOI: [10.1080/17538947.2020.1743785](https://doi.org/10.1080/17538947.2020.1743785)

To link to this article: <https://doi.org/10.1080/17538947.2020.1743785>



© European Union 2020



Published online: 23 Mar 2020.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Big Earth Data science: an information framework for a sustainable planet

Huadong Guo^{a,b}, Stefano Nativi^{lb c}, Dong Liang^{lb a,b}, Max Craglia^{lb c}, Lizhe Wang^d, Sven Schade^c, Christina Corban^{lb c}, Guojin He^a, Martino Pesaresi^{lb c}, Jianhui Li^e, Zeeshan Shirazi^a, Jie Liu^a and Alessandro Annoni^c

^aKey Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China; ^bUniversity of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing, China; ^cJoint Research Centre, European Commission, Ispra, Italy; ^dSchool of Computer Science, China University of Geosciences, Wuhan, China; ^eComputer Network Information Center, Chinese Academy of Sciences, Beijing, China

ABSTRACT

The digital transformation of our society coupled with the increasing exploitation of natural resources makes sustainability challenges more complex and dynamic than ever before. These changes will unlikely stop or even decelerate in the near future. There is an urgent need for a new scientific approach and an advanced form of evidence-based decision-making towards the benefit of society, the economy, and the environment. To understand the impacts and interrelationships between humans as a society and natural Earth system processes, we propose a new engineering discipline, Big Earth Data science. This science is called to provide the methodologies and tools to generate knowledge from diverse, numerous, and complex data sources necessary to ensure a sustainable human society essential for the preservation of planet Earth. Big Earth Data science aims at utilizing data from Earth observation and social sensing and develop theories for understanding the mechanisms of how such a social-physical system operates and evolves. The manuscript introduces the universe of discourse characterizing this new science, its foundational paradigms and methodologies, and a possible technological framework to be implemented by applying an ecosystem approach. CASEarth and GEOSS are presented as examples of international implementation attempts. Conclusions discuss important challenges and collaboration opportunities.

ARTICLE HISTORY

Received 30 November 2019
Accepted 13 March 2020

KEYWORDS

Big Earth Data; data science; sustainable development goals; digital transformation; Digital Earth; CASEarth; GEOSS

1. Introduction

Since the turn of the century, innovations in technology and greater affordability of digital devices have directed the ‘Industrial Revolution of Data,’ characterized by an explosion in quantity and diversity of real-time digital data in our lives. The amount of data being generated by people and machines alike is growing exponentially with development of smart devices and sensors, and our society is ‘entering an unprecedented period in terms of our ability to learn about human behavior’ (Leigh 2005; Onnela 2011; Mayer-Schönberger and Cukier 2014; Kessler 2017; Ghani et al. 2019). We now see a new movement that is concentrating on open access to knowledge in order to enable

CONTACT Stefano Nativi ✉ stefano.NATIVI@ec.europa.eu 📍 No 749 Via E. Fermi, 21027 Ispra, VA, Italy; Huadong Guo ✉ hdguo@radi.ac.cn 📍 No. 9 Dengzhuang South Road, Haidian District, Beijing 100094, China

© European Union 2020

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

re-use and value generation by third parties (Sampson et al. 2014; UN GGIM 2016; UN 2019; Maarooft 2019).

Big data (as a revolutionary innovation to see, explore, and understand world) has been seen as a 'strategic highland' in the new data-intensive era. It has gained attention from governments of nations worldwide (Sugimoto, Ekbia, and Mattioli 2016). In particular, the development of Earth observation technology and the hyper-connectivity of our Society is generating a huge amount of big data dealing with the natural and human phenomena of planet Earth. The term Big Earth Data was introduced bringing new impetus to Earth and social sciences (Guo et al. 2017; Guo 2017b). The heterogeneous Big Earth Data contains rich information (Fromm and Bloehdorn 2014; Nativi et al. 2015) to be mined and utilized in sophisticated frameworks.

Fuelled by big data, the emerging knowledge platforms and infrastructures require trustful relationships, with dependencies and shared responsibilities between all partners involved. Indications for the required shifts have already materialized in existing scientific and technical systems (Nativi et al. 2019). We have already witnessed a transition from an approach that principally favoured data sharing, to much more sophisticated improvements in data management and system architectural designs to generate information and knowledge (ITU-T 2017; Big Data Value Association 2019; Oliveira, de Fátima Barros Lima, and Farias Lóscio 2019). These application platforms utilize learning-based analytics for information generation. They are commonly conceived to engage both the public and the private sector. Today, innovative application sectors include smart city, health, and industry 4.0 (Song et al. 2017; Bohlen et al. 2018; Abidi and Abidi 2019; Wong, Zhou, and Zhang 2019).

To complete this digital transition, cover more complex application sectors, and meet the eminent challenges of urgently required international and cross-disciplinary collaborations, we envisage an additional step that improves transparency, reproducibility and knowledge co-creation. This new methodology is yet to be established in a way that connects globally shared and actionable knowledge from databases to local realities and activities. This approach requires novel insights, working methodologies, and sustainable systems that continuously evolve in order to meet the dynamic needs of modern society. It should aim at the best possible solutions for sustainable development and human wellbeing, including new methodologies for engagement across geographic scales, spearheading the independence of science from political agendas, and evidence-based proposals for democratizing data. At the same time, the application of this new approach has to be adopted to equally evolving constraints and boundary conditions, such as data ownership and control, data security requirements, feasibility of technical implementations and the increasing autonomy of machines.

Digital transformation coupled with increasing exploitation of natural resources (Guo et al. 2017) makes sustainability challenges more complex and dynamic than ever before. Given that we cannot expect these transitions to stop or even decelerate, there is an urgent need for a new scientific approach and an advanced form of evidence-based decision-making – towards the benefit of society, economy and the environment. It is but natural to look beyond the conventional, and explore these new avenues in support of global sustainable development. To generate the necessary knowledge on key interactions and connections existing between human society and the Earth system (e.g. natural phenomena), we argue the need for a new discipline of science, Big Earth Data science.

This science studies the huge bodies of information generated by the digital revolution and makes use of innovative technological frameworks, e.g. artificial intelligence, Internet of Things, and Digital Twins. Big Earth Data science is called upon to study societal changes, support human wellbeing, help the management of increasingly depleted natural resources, and equip us with the understanding to prepare for global changes manifesting with time (Guo et al. 2014). In summary, Big Earth Data science aims to provide the instrument to generate knowledge from the diverse, numerous, and complex data sources necessary to ensure a sustainable human society essential for the preservation of planet Earth.

1.1. Digital transformation

Throughout history, humans have constantly gone through a process of re-shaping the way society 'works' including patterns for information and communication (Norqvist 2018). We are living in the Digital Transformation era, where profound changes are taking place in the economy and society because of the uptake and integration of digital technologies in every aspect of human life. While digitalization disrupts society ever more profoundly, concern is growing about how it is affecting issues such as jobs, wages, inequality, health, resource efficiency and security (World Economic Forum 2019). On the other hand, the digitalization of our society is also creating unprecedented business and social opportunities in all sectors (Probst et al. 2018).

The digitalization of our society started many decades ago; recently, however, the development of technologies related to this process have seen a dramatic acceleration with the explosion of network connectivity and the trend of turning many aspects of our lives into data, which is subsequently transformed into valuable information (see datafication process). The advent of computing resources virtualization/democratization (see cloudification process), the introduction of data-driven artificial intelligence (see machine/deep learning), and the Internet-of-Things (see hyper-connectivity) significantly enabled this revolution. The digital transformation has changed our society, revolutionizing all the industrial, economic, and social sectors (Mayer-Schönberger and Cukier 2014); it is important to contextualize this transformation and understand the impact that the new technologies have been playing on the recognition of new sciences, such as Big Earth Data science.

1.2. Global sustainability in context

Unlike most life on Earth, humans are unique in their ability to alter their surroundings. Humans have built modern societies through progressive development towards the improvement of their lives. Unfortunately, this development has not been without consequences. For modern development, the economy is the prime concern, largely directed towards improving livelihood, to facilitate social progression. These oversights in realizing the consequences and delayed actions lead to counter-productive influences on our environment and our own wellbeing. On the 70th anniversary of the United Nations (UN) in September 2015, heads of state and delegates gathered at the UN headquarters in New York and announced 'Transforming our World: the 2030 Agenda for Sustainable Development' (United Nations 2015). This comprehensive sustainability framework is a product of the experiences of human society and presents a shared expectation for the future, a blueprint for countries to follow to move closer to global sustainable development in the next 15 years. The 17 Sustainable Development Goals (SDGs) incorporate various social, economic, environmental, and developmental targets and indicators, and have been endorsed by all countries with respective national implementation plans. The SDGs set an overall development context for other parallel global agendas for 2030 including the Paris Climate Agreement, Sendai Framework for Disaster Risk Reduction, and the New Urban Agenda, therefore representing the central mission of all toward 2030.

The 2030 Agenda for Sustainable Development is the manifestation of the international consensus to reduce risks and vulnerabilities of human populations, particularly ones globally underrepresented and weak. Collectively, the SDGs, the Sendai Framework for Disaster Risk Reduction and the Paris Climate Agreement are striving to reduce and reverse anthropogenic disturbances to ecology and climate and adopt a more sustainable approach to development. Progress towards the 2030 Agenda for Sustainable Development requires an understanding of global processes, the developmental challenges in a rapidly changing society and the impacts they have on the planet.

In the final report published by the Independent Expert Advisory Group (IEAG) on the Data Revolution for Sustainable Development (IEAG 2014), two global challenges for the current state of data were highlighted: the 'challenge of invisibility' and the 'challenge of inequality.' These challenges have generated important international initiatives debating the application of Big Data and AI for decision support in international frameworks such as the United Nations Global Pulse (see

<https://www.unglobalpulse.org/>), harnessing big data for development and humanitarian action, and the UN Big Data for Sustainable Development (see <http://www.un.org/en/sections/issues-depth/big-data-sustainable-development/index.html>), including the ‘AI for Good’ Global Summits (see <https://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx>).

1.3. Data science

Blei and Smyth data science

focuses on exploiting the modern deluge of data for prediction, exploration, understanding, and intervention. It emphasizes the value and necessity of approximation and simplification; it values effective communication of the results of a data analysis and of the understanding about the world and data that we glean from it; it prioritizes an understanding of the optimization algorithms and transparently managing the inevitable tradeoff between accuracy and speed; it promotes domain-specific analyses, where data scientists and domain experts work together to balance appropriate assumptions with computationally efficient methods. (Blei and Smyth 2017)

Therefore, like natural science, social science, or humanities, data science is best understood as a collection of disciplines with complementary foundations, perspectives, approaches, and aims. Data Science is multidisciplinary and employs mathematics, statistics and computer science, incorporating techniques like machine learning, cluster analysis, data mining and visualization. More than a decade ago, Jim Gray referred to data-intensive science as the ‘fourth paradigm’ of discovery, along with experimentation, theory, and numerical simulation (Hey, Tolle, and Tansley 2009). Today, Data Science explores where information comes from, what it represents and how it can be turned into a valuable resource in the creation of intelligence (Dhar 2013).

2. Big Earth Data as a new engineering science

One of the most popular definitions of science is: ‘the intellectual and practical activity encompassing the systematic study of the structure and behaviour of the physical and natural world through observation and experiment’ (Oxford Dictionary of English 1998). Overall, according to the Encyclopædia Britannica, ‘a science involves a pursuit of knowledge covering general truths or the operations of fundamental laws.’

In the history of science there have been important evolutions, such as the distinction of natural and social sciences (studying the natural and physical worlds, respectively) and the introduction of engineering sciences, empirical sciences studying the artificial rather than the natural or physical worlds – so in a category not allowed for in the traditional division. A valuable example of engineering empirical science is the Computer Science, which studies ‘the phenomena surrounding computers’ (Newell and Simon 1976). In general, empirical engineering sciences conduct experiments and construct theories to explain the results of them (Burkholder 2017).

Science and technology have always had a close relation; this has become particular evident in the age of the last industrial revolutions. While technology is considered as an application of scientific knowledge to systems and sub systems, technology advancements have also enabled the emergence of new sciences. Valuable examples are Galileo’s telescope, which helped creating the modern astronomy, and the microscope that helped biologists introducing micro-biology and enabled the disciplines studying micro-organisms. This is the case for several other technological developments that enhanced scientists’ capabilities of observing and experimenting, making them discover new phenomena and insights that require innovative and systematic studies, i.e. a new science. Recently, Data Science was introduced to indicate the ‘systematic study of the organization, properties, and analysis of data and its role in inference, including our confidence in the inference’ (Dhar 2013).

As to science, another philosophical debate has concerned reductionism versus holism. While reductionism theorizes that the nature of complex phenomena is reduced to the nature of sums of simpler or more fundamental phenomena, holism argues that all the properties of a given system cannot be determined or explained by the sum of its component parts alone – on the contrary, the system as a whole contributes to determine the parts behaviour. This sounds particular important to

study the complex Earth system (including the social and economic human components) and its global changes.

The data-driven science approach and the importance of scientists' collaboration to study trans and multi-disciplinary phenomena promise to change the scientific method and result in a practice called 'science 2.0' (named after web 2.0) (Kobro-Flatmoen et al. 2012). Enabled by the disrupting digital transformation of our Society, for first time in human history, there exist platforms able to collect world-wide observations and measurements (i.e. global big data) to monitor planetary phenomena, almost in real-time. These observations span the natural and physical disciplines and their combination results in such a great number of variables (observed an incredible number of times) that only the recent advancement of artificial intelligence (like a modern telescope or microscope) allows us to investigate generating insights about the observed phenomena.

In our view, the introduction of platform ecosystems for the management and analytics of Big Earth Data can be seen as the introduction of microprocessor computers in the early seventies. Therefore, we argue about the need to introduce a new empirical engineering science for a systematic study of the phenomena around the Big Earth Data analytics ecosystems – the combined use of Big Data, AI instruments, and online platforms to observe and test planet Earth as a whole – encompassing natural and physical worlds as well as data realm. This scientific field can be called Big Earth Data science and, being a data-driven science, can be considered as a sub-field of Data Science. Big Earth Data science is important to study the design and architecture of big earth data ecosystem and its application in the field of digital transformation of our society and Global sustainability of planet Earth.

We argue that Big Earth Data science must apply a holistic approach building on natural and social sciences as well as the engineering sciences dealing with big data and artificial intelligence. To generate actionable and trustable knowledge, Big Earth Data science is required to study the Big Earth Data platform ecosystems as programmed and living organisms conceived to understand planet Earth phenomena. The holistic universe of discourse characterizing Big Earth Data ecosystems is depicted in Figure 1.

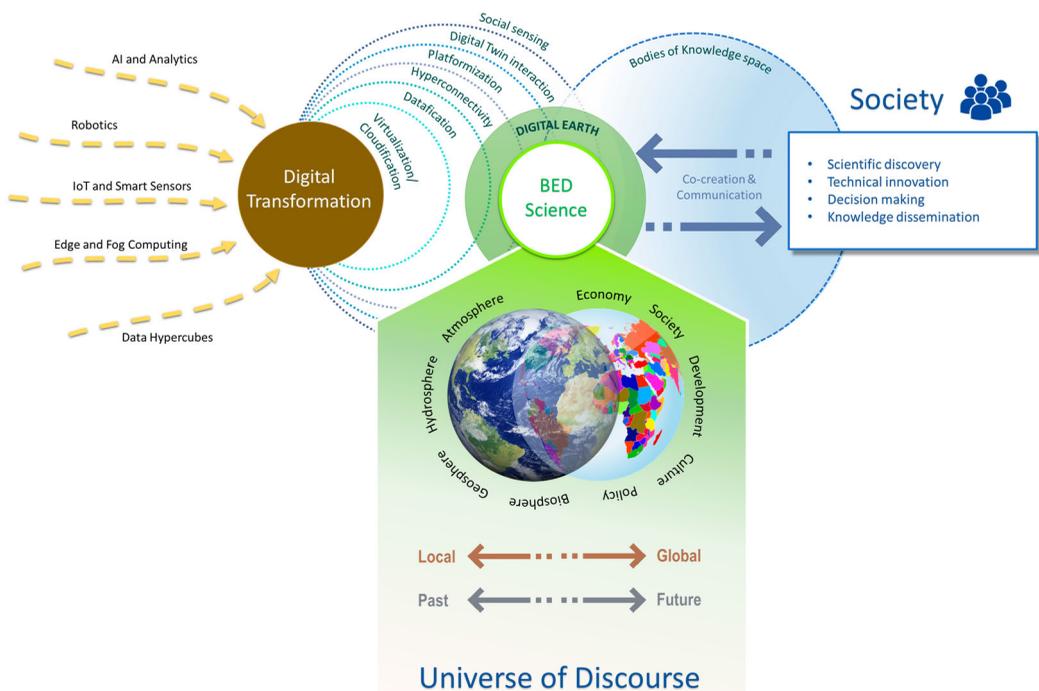


Figure 1. Big Earth Data science context, universe of discourse, enabling processes, and paradigms.

Being an engineering science, an important aspect of Big Earth Data science is the development of new basic understanding by empirical inquiry – i.e. this is how Big Earth Data analytics ecosystems solve problems and generate knowledge. The availability of high volume of (cross-disciplinary) observations and heuristic search (i.e. generating and testing potential solutions) are commonly utilized by Big Earth Data analytics ecosystems to generate actionable information. Therefore, Big Earth Data science is called to study what is known about heuristic search and review the empirical findings that show how the ecosystem enables action to be intelligent. In other words, to understand how data is collected and organized as well as how information is processed and intelligence is achieved to address the grand questions on our planet sustainability.

All sciences want to characterize the essential nature of the systems they study. They commonly do that recurring to a (simply) stated theory about the structural units out of which the entities (and related phenomena) of a domain can be composed (Newell and Simon 1976) – e.g. the cell doctrine in biology, plate tectonics in geology and the germ theory in diseases. For the domain of Big Earth Data ecosystems, we argue that big data, learning-driven artificial intelligence, and network-based platform are the fundamental units to be investigated. These units set the terms within which more detailed knowledge can be developed. This manuscript tries to characterize the Big Earth Data analytics ecosystem in the ‘Foundational paradigms and methodologies’ section – see also Figure 2.

2.1. Digital Earth and the Big Earth Data science

Big Earth Data science encompasses methodological and technological activities to study Big Earth Data analytical ecosystem as an organism that supports systemic discovery of information from data that deals with the planet Earth. This includes the development and deployment of a large variety of methodologies and technologies that enable collection, storage, retrieval and accessibility of data on different natural and social domains within an efficient analytical environment. A Big Earth Data ecosystem must be capable of integrating diverse inputs within a geographic context, resulting accessible to different communities, ensuring democratization of data and information – e.g. by making

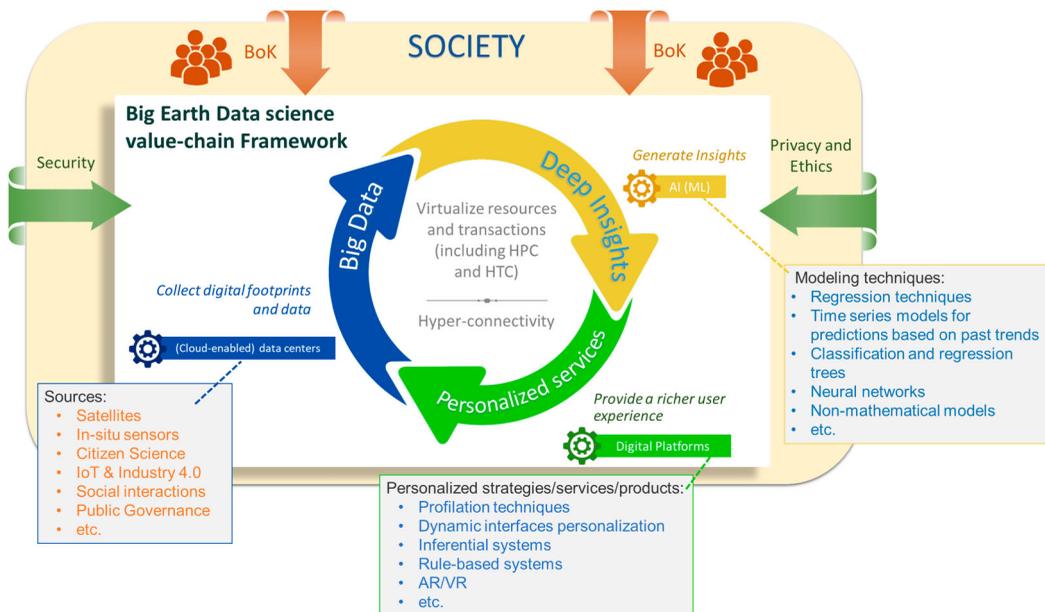


Figure 2. The Big Earth Data science high-level value-chain framework, building on a simplified model of the datafication paradigm (i.e. the three-step cycle).

hardware limitations redundant. Big Earth Data ecosystem enables and empowers communities to take action based on verifiable facts and develop their capacity irrespective of their developmental status.

An important objective of Big Earth Data science is the scientific comprehension, modelling and application of the processes to generate information from data and provide the knowledge required by the society to address global changes and enable a truly sustainable development. Big Earth Data science research is important to contribute addressing grand societal questions, including:

- How do we observe, describe and understand the global changes in a transformed society?
- How do we identify and understand the complexity of natural and social (including economic) phenomena, under the framework of Digital Earth?
- How do we envision, support and implement a more sustainable development, prevent degradation of the natural environment and depletion of natural resources, and preserve values of equality and poverty reduction?

Digital Earth concept (Gore 1998) was introduced as a global strategic contributor to scientific and technological developments, and a catalyst in finding solutions to international scientific and societal issues (ISDE 2009). In 2019, the International Society on Digital Earth (ISDE) met in Florence (Italy) and recognized that ‘Digital Earth is a multidisciplinary collaborative effort that engages different societies to monitor, address and forecast natural and human phenomena’ (ISDE 2019). ISDE is working to see Digital Earth expanding its role in ‘accelerating information transfer from science to applications in support to the implementation of UN Sustainable Development Goals and in support of a sustainable management of global environmental commons’. In this framework, Digital Earth calls on Big Earth Data science to study big data analytics ecosystems and platforms to comprehend the complex Earth system coupled with social systems – utilizing data from Earth observation and social sensing. Big Earth Data science is asked to develop theories for understanding the mechanisms of how such a social-physical system develops, operates, and evolves (Guo, Goodchild, and Annoni 2020). Big Earth Data science will become a milestone of Digital Earth in a new stage:

- It will extend and redefine the virtual domain for implementing a digital Earth system, such as complexity science, advanced artificial intelligence techniques, digital twins, augmented reality, ubiquitous sensing techniques, and advanced data engineering techniques.
- It will serve as the science studying the infrastructures, platforms, and tools for the ecosystem of social sciences and Earth sciences (for example) implementing a human-centred regional development system, developing an Urban Brain, and implementing a sustainable infrastructure for knowledge dissemination.
- It will contribute to a number of missions that help a sustainable planet, e.g. identify, understand and implement feasible government actions for climate change, implement sustainable cities and human society, and develop sustainable infrastructure that provides clean, accessible water for all.

The next section will introduce the Big Earth Data science universe of discourse, framing the domain context of this science. Then we will introduce a set of foundational paradigms and processes for Big Earth Data science, stemming from data science, GIScience, and digital transformation methodologies, followed by a discussion of the main Big Earth Data science technology and engineering frameworks. Two on-going international programmes aiming at implementing such a framework will serve as examples. Finally, the last section will discuss Big Earth Data science challenges and opportunities.

2.2. The universe of discourse

Big Earth Data science is about the study of the introduction and impact of big data analytical platform (like an artificial system or organism) in addressing the phenomena and the questions of a complex universe (domain) of discourse, which encompasses the set of natural, social and economic events characterizing our planet (Figure 1). This class of entities encloses the local and global changes affecting natural cycles and processes as well as the tight interconnections with human society, i.e. our social and economic systems. Over these entities, certain elements (variables) of interest are utilized to model or describe the relevant changes of the Earth system (e.g. oceans, land surface, atmosphere, cryosphere, and biosphere) and others to formalize societal changes. Traditionally, they have been managed and presented separately within different frameworks and tools. Big Earth Data science aims at overcoming these cultural, disciplinary and technological barriers, in a multi-scale and multi-temporal framework, from local to global and vice versa, in a variety of aspects from change detection to sustainable development planning.

The geo-spatial and temporal context provides a robust framework to integrate data on inter-linked natural systems and also to associate them with the social, economic and cultural phenomena (Goodchild 2004) which are essential to generating knowledge about the complex systems and processes characterizing our planet (Goodchild 2009). Collectively, this multifaceted data on Earth provides the opportunity to formulate scientific inquiry, contemplating anthropogenic drivers and vulnerabilities, towards understanding dynamic natural systems and contextualizing the information in terms of potential for sustainable societal development, their vulnerabilities, and risks. Big Earth Data science, therefore, requests the application of (Geographic) Data Science methods, processes, and systems to methodologically study the composite phenomena of planet Earth and their changes – with the challenges of multi- and trans-disciplinary collaborations.

As present social challenges are inherently complex, this new approach to science is not only characterized by collaborations between a large amount of diverse scientific disciplines (belonging to natural sciences as well as social sciences and humanities) but also by the need of trans-disciplinary knowledge sharing and co-creation. Scientists must collaborate closely with actors from civil society, industry, and public institutions in order to address socially relevant and sound research questions allowing rich sets of perspectives to be incorporated into the solution and decision-making process.

Since these interconnected situations appear across all geographic and temporal scales, Big Earth Data is inherently multi- and trans-disciplinary, as shown in Figure 1.

Big Earth Data science presents an opportunity to observe the Earth system in a new holistic, multi-disciplinary and trans-disciplinary way. The digital transformation of our society presents remarkable opportunities to understand and simulate Earth as a number of interrelated digital systems (Guo, Wang, and Liang 2016). In this area, important challenges that need to be considered by Big Earth Data analytical platforms include:

- What are the most promising roles of digital twins to address social challenges?
- How will people interact in the interplay of the physical world, digital twins and augmented reality?
- How can we advance and connect semantic reference systems to enable multi- and trans-disciplinary research in global changes and sustainable development domains?
- How can we best utilize the Earth system data from multiple sources for scientific value creation?

By adopting a paradigm of scientific discovery driven by data, Big Earth Data science establishes strong relationships with domain applications (i.e. society), the current digital transformation processes, their interaction patterns, and technological drivers (Figure 1). For this reason, Big Earth Data science can be seen as a collection of disciplines (i.e. Earth System, Data, Social and Economic sciences) with a set of complementary foundations, perspectives, approaches and

aims – all sharing a grand mission: our planet’s sustainability. Finally, Big Earth Data science methodologies and investigation approaches make use of some foundational paradigms and processes, which stem from data science, the digital transformation of our society and the democratization of science. In turn, these are enabled by recent and upcoming technology revolutions, as depicted in [Figure 1](#).

3. Foundational paradigms and methodologies

3.1. Datafication and Big Earth Data ecosystems

Over the past decade, data-intensive logic and practices have come to affect domains of modern life ranging from marketing and policy making to entertainment and education; at every turn, there is evidence of ‘datafication’ (Mayer-Schönberger and Cukier 2014) or the conversion of qualitative aspects of life into quantified data (Ruckenstein and Schüll 2017). Datafication became a novel paradigm characterizing the digital transformation of our society. This paradigm mainly builds on three digitalization processes (JRC 2018), as depicted in [Figure 2](#):

- (a) Big data sensing and collection: the collection, aggregation and contextualization of digital artefacts and ‘footprints’ generated by humans, machines and things. This pervasive process produces what is often referred to as big data. Virtual (e.g. cloud-enabled) data centres play an important enabling role. Social networks, public government digitalization, the Internet-of-Things (IoT) and the new generation of remote sensing instruments are among the best sources for big data.
- (b) Deep insight generation: the recognition of valuable insights by analysing the collected big data, i.e. big data analytics. Today, this largely makes use of learning-based artificial intelligence (AI) technologies, i.e. machine and deep learning models.
- (c) Service personalization: making use of the generated intelligence, specialized online platforms interact with users and stakeholders to provide personalized services and offer a richer user experience, leading the platform economy to thrive.

Growing digital connectivity of humans and machines, or hyper-connectivity (Wellman 2001), supports the implementation of the datafication paradigm, empowered by the significant advancements and virtualization of high-performance and high-throughput computing infrastructures, i.e. HPC and HTC.

The datafication paradigm has caused a shift from data scarcity to a data overload, consequently affecting the scientific process, which has been constrained by the need to create data (by experiments and observations) and mathematical models to fill data gaps. The process has now been transformed and enriched, and large amounts of data are available before the formation of research questions. Therefore, unlike before, we see the emergence of methods that create hypotheses from (big) data analysis and almost exclusively operate on virtual representations of reality.

Datafication has also facilitated and advanced the replicability of data science experiments and the reproducibility of results (Ostermann and Granell 2017) to an unprecedented level. Easier access to data (using online platforms) as well as algorithm and code re-usability (using public interfaces) provide ground to follow and understand the scientific process. On the other side, AI algorithms still raise challenges for transparency and reproducibility (Craglia et al. 2018). In summary, datafication and the online platform era have benefited science and the scientific processes by attracting a large user base, diversifying perspectives, and providing opportunities for innovation, collectively broadening the scope of development and improving contributions to scientific knowledge.

3.2. WoT and the data twin interaction pattern

The Web first connected digital documents, allowing the discovery of information. Then the Web 2.0 revolution connected people and communities, enabling the disruptive business of social networks. The present Web generation (sometimes referred to as Web 3.0) enables the externalization (i.e. virtualization) of practically any digital resource and capability (e.g. auto-scaling infrastructures, service platforms, and complex domain-specific software applications) moving most of society's transactions and processes onto the network. Datafication, hyper-connectivity, and cloud computing have played a fundamental role in this digital transformation. IoT and 5G (the fifth-generation architecture of cellular networks) are promising to yet again revolutionize (or rather to further expand) the Web, by connecting 'everything' and generating new business intelligence. Simple devices (e.g. sensors), complex real-time systems (e.g. moving vehicles), sophisticated monitoring, and forecasting models (e.g. real-time imagery recognition) will be online exchanging information. Foundational to this revolution are digital 'things' (systems) interacting with the physical world. For example, sensors get the information from the physical world, while actuators act upon it. According to several international standardization organizations (e.g. ISO, ITU and IEC), this innovative global infrastructure is conceived for the information society to enable advanced services through the interconnection of objects that belong to the physical and virtual worlds (ISO/IEC 2016).

From an engineering point of view, such an infrastructure that supports an ecosystem of diverse platforms and domain applications and calls upon the Web to interact and interoperate, can be implemented. This introduces the Web-of-Things (WoT), a technological framework defined by W3C (W3C 2019). WoT aims to connect real-world things/systems to the Web, creating a decentralized IoT, where things are linkable, discoverable and usable. In this framework, a really promising interaction pattern is named digital twin: a virtual representation (i.e. digital model) of a connected real thing or a set of things representing a complex domain environment. According to its complexity, the digital representation (i.e. the twin) may reside on a virtual platform (e.g. a cloud) or on an edge system. A digital twin can be used to represent real-world things/systems that may not be continuously online, or to run simulations of new constructions, applications and services, before they are deployed to the real world (W3C 2019). Valuable examples of digital twins spawn from complex systems like aeroplanes to multifaceted infrastructures like metro rail, or human settlements like cities (Ponti et al. 2019) and even countries.

In the near future, it might be possible to connect (in the virtual world) the diverse digital twins representing extremely complex and vast domains. These twins would represent those infrastructures and systems monitoring the reality as well as those acting in (and contributing to shaping) the real world – e.g. natural phenomena and social processes domains. In such a framework, it is possible to define a digital twin of Earth.

Digital Earth concept can serve as the virtual representation (i.e. the digital twin) of Earth. To achieve that, it is necessary to have a profound digital transformation of our planet to provide access to vast amounts of scientific, social and cultural information that enable the understanding of the Earth system, human activities and their interconnections. The digital twin of Earth can be a metaphor, i.e. a way of conveying the complexity of Earth system processes in simple to grasp, intuitive ways, highlighting certain aspects and hiding those that are less relevant. It can be the basis for powerful narratives and storytelling about the choices facing humanity and how each of us as individuals, organisations, businesses, and governments can make a difference. Combined with modelling techniques and visual representations of different scenarios, the digital twin of Earth could become the missing tool to develop collective consciousness and action. As an example, the recent report from the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (UN SDG 2019) makes a strong case and provides considerable evidence about the threat to biodiversity and the unprecedented rate at which species are disappearing. However, if this data and the

scenarios for action could be modelled and visualised in a Digital Earth framework, it would make a much greater impact in engaging society and policy. Notably, although physical interactions might happen at limited geographic scales, global challenges and the interplay with local actions can be incorporated using digital representations of the Earth system. This allows for the integration of global knowledge into decision making on the ground (“Think global act local”) and at the same time enables the integration and extrapolation of tacit knowledge into an Earth-spanning knowledge base. With all inputs sensed and collected from Earth and human society via datafication, hyper-connection, platformization and digital twin methods, a well-organized Big Earth Data science system emerges.

3.3. Digital Earth and the co-creation and communication approach

3.3.1. Communities of practice and bodies of knowledge

Parallel to the datafication and platformization of society (Van Dijck, Poell, and de Waal 2018), we are also witnessing an increased polarisation in the political debate in many countries due to various political, economic and development challenges in the world. Fuelled by the growing accessibility of news and the emergence of social communication platforms, the political narrative (as a product of social, economic, cultural and religious identities as well as interest and beliefs) is strongly shaping the scientific narrative, rather than the converse, leading to disinterest in the scientific process. Some socio-political classes have started questioning the credibility of experts, the role of science in policy support and, more broadly, the authority of science in society.

With the ambition to study the complex interaction between physical and social phenomena, we need to recognize the limits of the model of a value-free science, which is typical of the physical sciences. Science is not value free, and scientists make value judgements every time they create models or select which phenomena merit study along with which variables best represent them. It is crucial to develop a more open and participative model of science by exposing such values and assumptions and recognizing that there is not a single problem space in which to look for answers, but rather multiple spaces with competing values and views. Funtowicz and Ravetz (1993, 2018) call this ‘post-normal science’ that is an approach designed to apply ‘where facts are uncertain, values in dispute, stakes high, and decision urgent’.

The acknowledgement that may meet various legitimate or problematic spaces and perspectives allows for greater humility and reflexivity in science, facilitating openness and participation towards finding a shared ‘framing’ of the problem. For example, first, a collective understanding of what are the important questions should be established and only then should the methods, data, and tools to address them be developed. Multi-disciplinarily, trans-disciplinarily (i.e. involving not just other disciplines but also non-academic stakeholders, see Vaccari et al. 2012), public participation and citizen science are crucial but only if already involved from the beginning, at the initial stage of framing the problem space (i.e. *frame first, compute later*).

In addition to a new, more open approach to defining problems and building evidence, we also need new forms of communication. Researchers are increasingly recognizing that emotions have an important role to play in policy-making, which is not just rational and evidence-based as generally portrayed. The greater the complexity of issues, uncertainty of process, and risk of outcomes, the greater the role of emotions in navigating the complex decision environment. For this reason, greater attention has now been given to the use of metaphors and narratives to communicate scientific evidence – see storytelling, for instance. In this respect, the notion of Digital Earth (Gore 1998; Craglia et al. 2008; Goodchild et al. 2012; Ehlers et al. 2014) has a key role to play.

To communicate and collaborate, social and professional communities use bodies of knowledge, shared among them. A body of knowledge (BoK) can be defined as the systematic collection of activities and outcomes in terms of their values, constructs, models, principles and instantiations (Romme 2016). Once a community of people identifies relevant issues of concern, a set of BoK is introduced to deliver robust and actionable solutions – where ‘relevance’ and ‘robustness’ adhere to the scientific as

well as social dimension. The multi- and trans-disciplinary settings required by Big Earth Data science imply the collaboration of people with different concerns, needs, understandings, theories and models of reality. Diversities in these Bodies of Knowledge stem from different scientific disciplines (and schools), social hierarchies of membership, priorities and policy objectives. They differentiate perspectives and opinions on shared goals and are essential for viable and sustainable solutions. However, productive outcomes of debate are threatened by the limited strength of communication, the understanding of problems and the agreement on a set of clear objectives. These shortcomings can be mitigated by accepting plurality in opinions with the motivation to negotiate, starting with the challenge of identifying the actual problem to deriving the relevant research questions and methodologies that should ultimately have positive societal impacts.

The high degree of diversity and collaboration of social and professional communities demands an advanced communication approach and ways to express different BoK with community-specific vocabularies and values, assumptions, possible constraints, etc. In addition, to avoid misunderstandings or a lack of shared understanding, we need approaches to inter-compare and translate between diverse BoK as a prerequisite for fruitful negotiations between such different actors. In other words (and in addition to reference systems that express and translate spatial-temporal information), we need a new generation of semantic reference systems (Kuhn 2003; Ortman et al. 2014) that allow us to formalize our BoK and their interrelationships to the full extent.

A semantic reference system constitutes two parts (Janowicz and Scheider 2010). First, a 'semantic reference frame' (e.g. an ontology) formally defines the intended meaning of terms that are used within a given BoK. These terms might refer to entities (e.g. 'lakes' or 'water bodies'), processes (e.g. 'rainfall' or 'evaporation') and physical properties (essential variables such as 'temperature' and 'humidity'). Second, a 'semantic datum' grounds the meaning of the terms used in a particular semantic reference frame. When dealing with models of physical phenomena, semantic datum can define the very basic terms that are used within the reference frame by connecting the terms to observable structures in the real world. For example, repeatable physical processes are used in order to define the zero values and units of measure in the International System of Units (Bureau of Weights and Measures 2006). Recently, several Earth observation communities tried to introduce semantic reference systems by defining shared ontologies and essential variables (Lehmann et al. 2019; Siricharoen and Pakdeetrakulwong 2014; Li, Song, and Tian 2019).

Having such reference systems in place, including projections from one system to another, will ultimately unambiguously describe the intended meaning of terms used by different communities and different BoK. It will also offer translations between them and thereby serve as an essential element for multi- and trans-disciplinary collaborations.

4. Big Earth Data science technological framework

Benefiting from the new paradigm introduced by the digital transformation of our society, Big Earth Data science is concerned with the generation of data-driven knowledge to help address challenges to the planet's sustainability. From a technological point of view, this requires the implementation of high-level value-chain frameworks (Figure 2). The Big Earth Data science technology framework consists of three wide-ranging tasks: (a) collecting and aggregating big data, (b) producing valuable insights from it (by processing and analysing data) and (c) interpreting these insights in the context of real world problems to work out intelligence to support societal needs. The main possible sources of big data as well as popular analytics techniques and knowledge generation tools are shown in Figure 2.

The Big Earth Data science value-chain framework is embedded in society and applies the collaborative and co-creation paradigms discussed previously. For this reason, in addition to data provision, society also contributes to the framework by specifying a set of constraints (i.e. security, privacy and ethics) and BoK stemming from stakeholders and users.

4.1. Implementing a Big Earth Data science framework using a collaborative approach

The Big Earth Data science framework aims at realizing a value-chain system for big data. The big data paradigm is commonly defined as the distribution of data systems across horizontally coupled, independent resources to achieve the scalability needed for the efficient processing of extensive data-sets (NIST 2018). Actually, scalability and distribution needs originated some of the framework requirements previously introduced, i.e. cloud computing model, application distribution and data locality. Therefore, for the framework implementation, it is recommended to adopt an ecosystemic approach (Jansen and Cusumano 2012), that is, the realization of an international digital ecosystem consisting of different and distributed components, services and resources. Ecosystem modularization allows multi-organizational and heterogeneous systems to interoperate at different levels, all contributing to the framework's implementation. The stakeholders engaged in realizing this ecosystem may compete in other sectors, but they will collaborate for the ecosystem's success – i.e. the mission of the Big Earth Data science framework.

In keeping with the ecosystemic approach, the Big Earth Data science value-chain framework is built in a network environment (e.g. the Web) where components interact via effective and flexible interfaces (e.g. Web APIs). Usually, a Big Earth Data science high-level framework is engineered through a big data analytics platform (i.e. the Big Earth Data Science Analytics Platform) and each of the wide-ranging tasks of the framework is commonly split into independent and cohesive sub-tasks exposed as (network) services, as depicted in Figure 3 (Craglia et al. 2018). In the figure, each high-level task is composed of two implementation service components. Naturally, different sub-tasks and granularity levels are possible. The implementation components are generally managed by different organizations, all contributing to the ecosystem. These services generate valuable business opportunities and contribute to the final value offered by the ecosystem to society – e.g. policy makers, citizens and the communities working in the diverse domains of society dealing with global changes and sustainable development. The platform must also implement a set of ethical, privacy and security constraints characterizing the Big Earth Data science high-level framework as agreed by its stakeholders and society.

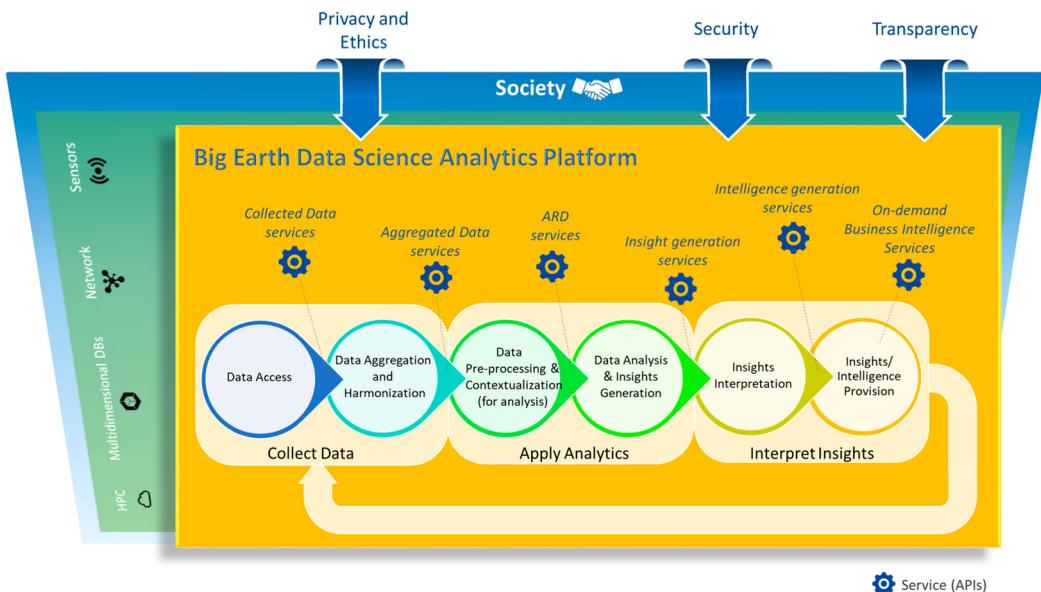


Figure 3. Ecosystem of the Big Earth Data Science Analytics Platform.

4.2. Collect data components

4.2.1. Data access

To be ingested into the Big Earth Data science analytics platform, data may be collected through numerous and various sources (see the datafication and hyperconnectivity processes of our society) using different techniques and methods. Therefore, this task aims at recognizing and connecting to the most appropriate sources, including: maps, official statistics, IoT, video data, citizen science, crowdsourcing and volunteer monitoring, social media, online analytics and app tracking, transactional data.

4.2.2. Data aggregation and harmonization

Since collected data can be very heterogeneous (e.g. unstructured or structured/semi-structured, time sensitive or almost invariant, one large file or a set of many tiny files), they cannot be processed straightforwardly by standard databases and requires some elaboration and homogenization. This is the work of the data aggregation and harmonization task, involving a variety of techniques and methods to make highly heterogeneous and multi-sourced data usable as a logical whole. In the Big Earth Data domain, for data harmonization and aggregation, a common solution consists of first defining a unified data abstraction model and then mapping or extending heterogeneous data models against the unified data abstraction (Nativi, Craglia, and Pearlman 2013; Nativi and Bigagli 2009; Nativi et al. 2015). In addition to a unified data abstraction, a set of 'information integrity' rules must be applied to ensure the integrity of data sources and the correctness of the final results.

4.3. Apply analytics components

4.3.1. Data pre-processing and contextualization

This task aims to pre-process aggregated and harmonized data to generate ready-to-use datasets for feeding the analysis models – Analysis-Ready-Data (ARD) (Holmes 2018) and Data Cubes (Nativi, Mazzetti, and Craglia 2017; Giuliani et al. 2019). Data pre-processing removes data of poor quality in the dataset (e.g. satellite data with high cloud cover and sensor data affected by noise) and data that is not relevant for the analysis to be done. This task generally implements the following functionalities: data cleaning, data contextualization, data re-structuring and data indexing. The basic principle of data pre-processing is to maintain as much of the original amount of data as possible, avoiding adding other information. However, data contextualization may significantly reduce the amount of data to be processed, according to the analysis to be carried out – generating an ARD body suitable for a specific application area.

Data cleaning is used to transform raw data into an understandable format, as real world data can be incomplete (lacking attribute values, lacking certain attributes of interest, or containing only aggregated data), noisy (containing errors or outliers) and inconsistent (containing discrepancies in codes or names).

Data contextualization is the process of identifying the data relevant to an entity of interest (e.g. a natural/social phenomenon or a place) based on the contextual information characterizing the entity itself.

Data restructuring faces the case where a data (physical and logical) storage structure must be changed in order to facilitate its processing.

Data indexing aims to achieve fast access to the platform data. For example, data indexing may include the principle of space–time neighbours – i.e. data tiles that are temporally continuous and spatially close to each other form an index tree.

4.3.2. Data analysis and insights generation

This task aims at generating useful insights (or information), i.e. an understanding of the content of the datasets collected and pre-processed by the Big Earth Data science platform. This task makes use

of big data processing and analysis solutions, which commonly include techniques and methods dealing with data mining, machine learning, collective intelligence and data visualization.

4.4. Insight interpretation components

4.4.1. Insight interpretation and intelligence generation

The aim of this task is to generate actionable intelligence (or knowledge) by interpreting the information produced from analysis. Knowledge discovery refers to the overall process of discovering actionable knowledge from available information. It involves the evaluation and possibly interpretation of the patterns revealed through the process of data analysis in order to support decision-making. Typically, knowledge discovery exists within data mining and often the data mining methods apply knowledge discovery processes, too. There exist several instruments to convert information into actionable knowledge, including expert systems, rule-based systems, decision support systems, and automatic inferential systems.

4.4.2. Insights/intelligence provision

This task intends to manage requests from platform users and provide specific intelligence requested by them within their business domain. The insights generated from the analytics components of the Big Earth Data science platform are of great value to understanding global changes and planning a more sustainable development, e.g. better resource management and utilization as well as environmental protection. Therefore, this component should be capable of facilitating the platform's interaction with users with the aim to generate greater application value. To this end, the ecosystem partners must carefully consider the service provision policy, including privacy or ethical aspects. The body of user requests and provided answers are collected by the platform to be ingested by the platform itself, becoming one of the data sources.

5. Examples implementing a Big Earth Data science framework

5.1. GEOSS

Another valuable attempt to implement a Big Earth Data science framework is represented by GEOSS (Global Earth Observation System of Systems). Established in 2005, the Group on Earth Observation (GEO) (see <https://www.earthobservations.org>) is a voluntary partnership of governments and organizations; presently, GEO Member governments include 108 nations and the European Commission, more than 130 Participating Organizations, and 11 Observers. A central part of GEO's Mission is to build GEOSS. This system-of-systems is a set of coordinated, independent Earth observation, information and processing systems that interact and provide access to diverse information for a broad range of users in both public and private sectors (GEO 2016). GEOSS links these systems to strengthen the monitoring of the state of the Earth. It facilitates the sharing of environmental data and information collected from the large array of observing systems contributed by countries and organizations within GEO. Further, GEOSS wants to ensure that these data are accessible, of identified quality and provenance, and interoperable to support the development of tools and the delivery of information and knowledge services. In doing that, GEOSS aims at increasing our understanding of Earth processes and enhances predictive capabilities that underpin sound decision-making (GEO 2016).

Due to the evolution of information and computing technologies, GEOSS recently evolved into a multi-organizational and global ecosystem transforming its common infrastructure into a Web-based platform: the 'GEOSS Platform' that exposed a set of high level APIs (GEOSS Platform Operation Team 2017, 2018; GEO DAB Team 2019). The GEOSS ecosystem and the role played by the GEOSS Platform are depicted in Figure 4. The GEOSS Platform implements a set of brokering (i.e. mediation, aggregation, and harmonization) services exposing a set of high level APIs, for machine-

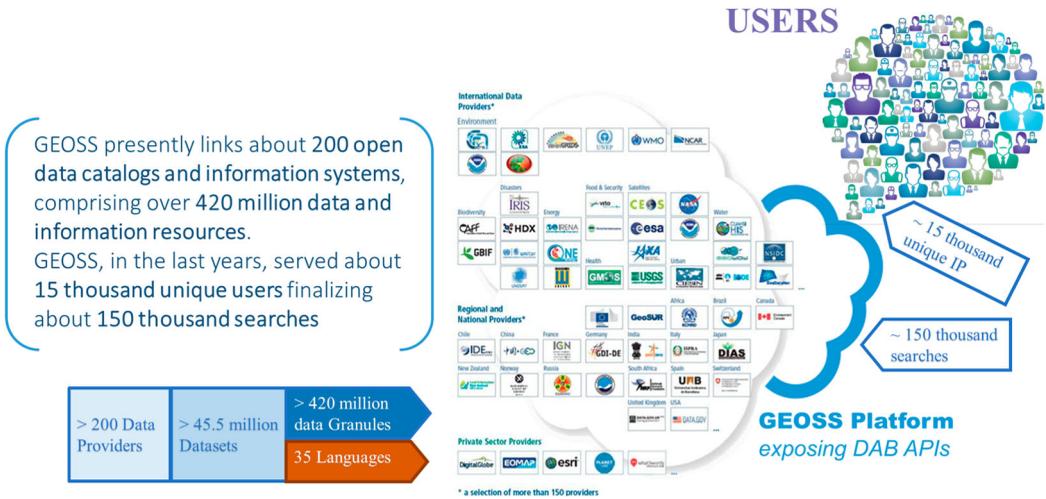


Figure 4. The GEOSS Ecosystem and the GEOSS Platform role [numbers refer to the year 2017].

to-machine interaction. Three kinds of interfaces are implemented: (a) standard Web service interfaces, based on de-jure and de-facto geospatial standard specification; (b) RESTful API, an application programming interface enabling the interaction with GEOSS through the exchange of HTTP requests/responses transferring JSON-encoded messages; and (c) Web API, a Javascript library for rapid development of mobile and desktop applications to serve a wide audience of mobile/desktop app developers. GEOSS contents, structure and relationships across institutions and disciplines were explored by (Craglia et al. 2017).

To continue leveraging these successes through 2025, GEO will evolve GEOSS and its infrastructures to meet current and emerging high-level needs (GEO 2016). The GEOSS Platform will augment the present EO data discovery and access functionalities by introducing data analytics services to generate the actionable information required by decision makers. This is carried out by collecting and using the necessary knowledge generated by the many activities and communities of practice contributing to GEO (Nativi, Craglia, and DeLoatch 2018). The envisioned GEOSS framework is

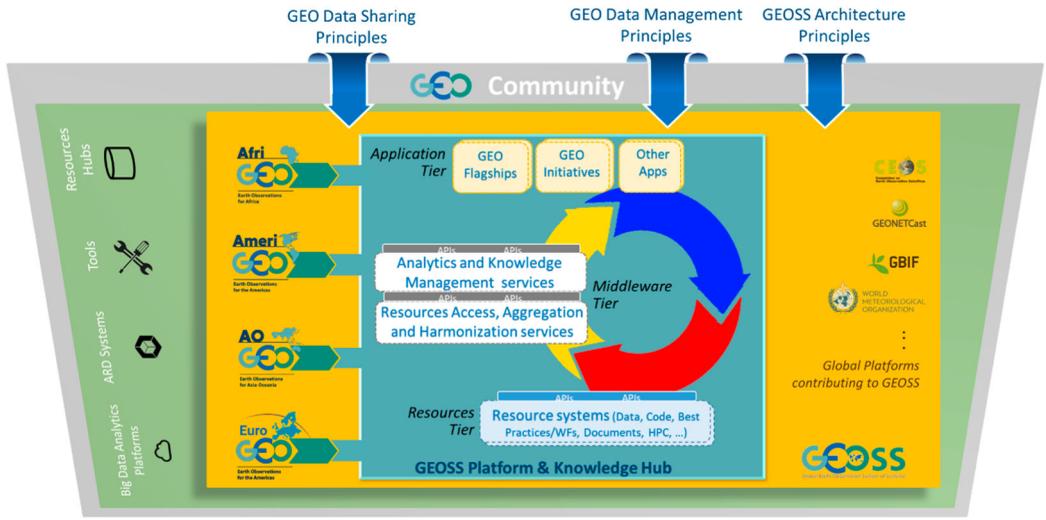


Figure 5. The advanced GEOSS framework to generate knowledge from Earth observations in a collaborative way.

shown in [Figure 5](#). It builds on the capacities provided by the most recent ICT development (e.g. online data/programme code hubs, Analysis-Ready Data systems, Big Data Analytics platforms, AI Tools, etc.) and must implement (by design) the principles advocated by the GEO Community – i.e. GEO Data Sharing and Management principles (GEO 2014 2015) as well as the GEOSS Architectural principles. To this advanced GEOSS framework, several international and GEO programmes/initiatives are going to contribute, as depicted in [Figure 5](#). In particular, the GEO Regional initiatives (i.e. AfriGEO, AmeriGEO, AOGE0, and EuroGEO) will play a key role to enhance the resource management functionalities of the GEOSS Platform and develop a set of new services characterizing a knowledge hub. The high-level architecture of the planned GEOSS Platform & Knowledge Hub is depicted in [Figure 5](#). This is the architecture of a collaborative digital ecosystem, consisting of three tiers: (1) the resources tier, containing the necessary data, algorithms/code, workflows, best practices/publications, and HPC infrastructures shared via online distributed systems; (2) the middleware tier, containing the services layer to access, aggregate, and harmonize the required resources, as well as the layer with the analytics services useful to generate information and knowledge from data – for example, by executing an AI model either referenced in a formalized workflow or in a best practice document; and (3) the applications tier, which contains the applications making use of the generated insights to deliver actionable information to decision makers. In turn, these applications can produce new resources (e.g. data, algorithms, documents) contributing to the resource tier.

Under GEO, several initiatives have demonstrated how Big Earth Data plays an insightful role in monitoring targets, tracking progress and contributes toward achieving SDGs agenda implementation. A valuable example is the GEO Initiative ‘Earth Observations in Service of the 2030 Agenda’, which issued a report presenting a number of case studies provided by governments and agencies to demonstrate specific applications of Big Earth Data to SDGs and national ambitions (e.g. indicators 15.2 on forest monitoring and indicator 14.1 on water quality monitoring) (Anderson et al. 2017). The GEO Human Planet Initiative supports UN-Habitat in the provision of global baseline data on built-up areas (Corbane et al. 2019) and population dynamics (Freire et al. 2020) for reporting on SDG 11, related to sustainable cities and communities. The baseline datasets were derived by combining different sources of remote sensing, census surveys, and other socio-economic variables. The increasing capability and diversity of operational satellites observing our planet provide significant opportunity for Big Earth Data science to support a large number of SDG indicators, bringing more accurate, spatially explicit, and frequently updated evidences. For example, as the main EU contribution to GEO, Copernicus programme (with its free data and thematic services information, see <https://www.copernicus.eu/en>) along with its data and information data analytical environments (DIAS, see <https://www.copernicus.eu/et/juurdepaas-andmetelete/dias>) provide a valuable help to monitor the status of achievement of several SDG indicators.

5.2. CASEarth

As a strategic priority research programme of the Chinese Academy of Sciences (CAS), the Big Earth Data Science Engineering Project (CASEarth) has been designed to develop an effective data-to-knowledge platform. The implementation of CASEarth could be expected to achieve substantial breakthroughs for Big Earth Data science services.

CASEarth ([Figures 6 and 7](#)) has been developing a big data and cloud service platform as well as a big data-driven Digital Earth Science Platform, integrating data, models, and services and promoting multidisciplinary integration. It has the potential to break through the bottleneck of data sharing (Guo 2017a).

The objectives of CASEarth include the following:

- (1) Developing data collection standards in the fields of resources, environment, ecology and biology, and developing an open, multi-terminal-oriented data integration management system that can be used in all fields;

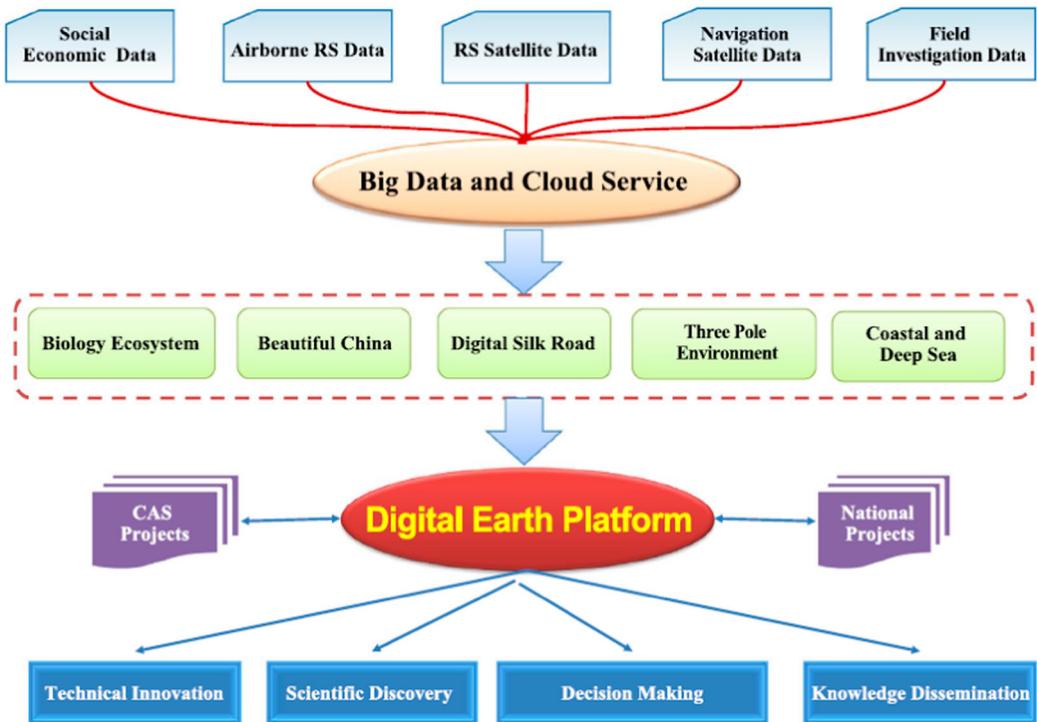


Figure 6. The framework of CASEarth (Guo 2017b).

- (2) Formulating an open sharing mechanism for data (environmental, ecological and biological) and forming a multi-source data sharing cloud service platform connecting various fields and departments;
- (3) Developing integrated data mining, visualization and decision support solutions for resources, environment, ecology and biology, and developing a dynamic and scalable model integration system based on cloud computing and big data technology.

The overall technical architecture is divided into three parts: resource management and cloud services, the CASEarth engine, and public services and applications (Guo 2017b).

- Resource management and cloud service layer

Based on the Big Earth Data Cloud Service Infrastructures, the resource management and cloud service layer aggregates CASEarth satellite data, Earth observation data, ground monitoring data, and socioeconomic data by building cloud services such as high-performance computing, elastic big data processing, and cloud computing. Thereby, a big data resource library and special database are built, ultimately forming the capability for data services.

- CASEarth engine

The CASEarth engine layer provides computational engines, data access engines, data analysis engines, model algorithm libraries, and visualization engines for specific technological needs.

- Public service and application layer

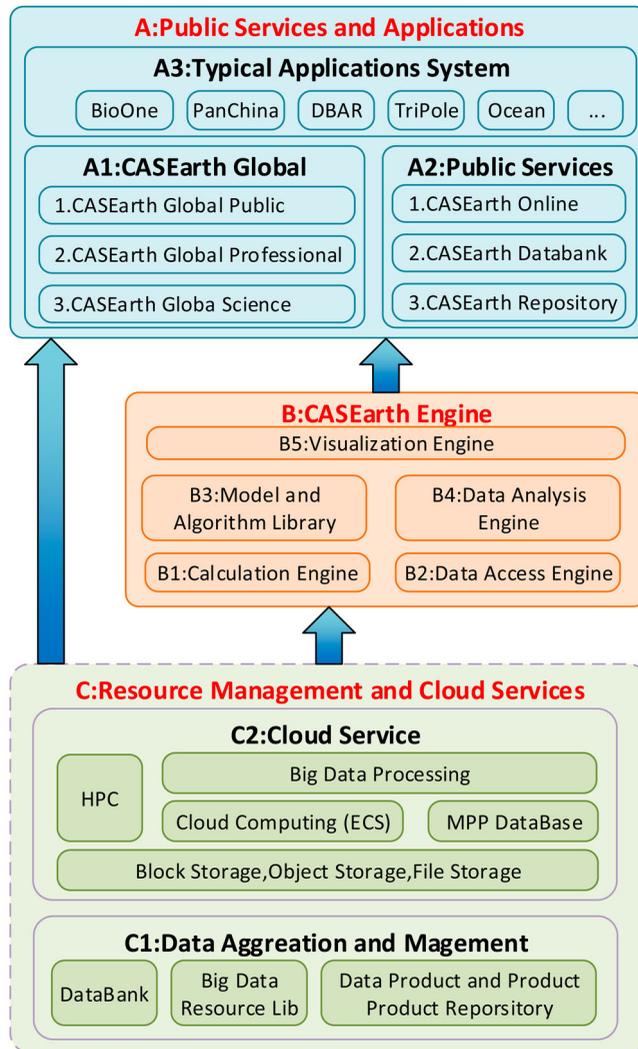


Figure 7. Overall technical architecture of CASEarth.

The public service and application layer mainly includes Digital Earth and application systems in five major fields, and is a platform for cross-disciplinary research and exporting and displaying results.

Currently, the Big Earth Data Sharing Service Platform developed by CASEarth has been used for multiple services. The platform provides global users with systematic, diversified, dynamic, continuous, and standardized Big Earth Data with a Global Unique Identifier (GUID), and promotes the formation of new data sharing models by constructing a data sharing system that integrates data, computing, and services. At present, the total amount of shared data is about 5 PB, and about 3 PB will be updated each year.

A sub-system, Big Earth Data for Sustainable Development Goals (SDGs) in the Belt and Road region has also been established. It is designed to: (1) handle vast amounts of satellite imagery and socioeconomic data; (2) provide web-based interactive data exploration; (3) offer distributed scientific computing algorithms; (4) provide a cloud computing environment for Big Earth Data science analytics; (5) offer specific APIs for various users; and (6) provide a new data model for sustainable development decision-making (Guo 2018).

The CASEarth system has been successfully utilized for a large variety of sustainability-oriented studies. Internationally, CASEarth has provided the design and data framework for the Drought-Watch system developed for Mongolia under a technology transfer agreement, which was completed and officially handed over to the Mongolian National Remote Sensing Center of China in September 2018. Similar systems are also being developed for implementation in Cambodia and Sri Lanka under similar agreements. CASEarth had success in developing a particulate matter observation system using the CASEarth Big Earth Data science platform based on remote sensing data. Similarly, an indicator and modelling system for eutrophication assessment in coastal regions has been successfully developed. It has been utilized for assessing progress towards a sustainable cropping system in China. Rapid developments towards cloud-based systems are being made and a Big Earth Data-powered forest management system is under development. To facilitate the decision-making process, CASEarth is utilizing Big Earth Data science for a ‘River basin sustainable development decision support system (RisDSS)’, which is currently being implemented in the Heihe River Basin. Aside from SDGs, CASEarth also closely interacts with two other international frameworks to include several links to spatial information on Earth: the United Nations Framework Convention on Climate Change (see <https://unfccc.int/>), and the Sendai Framework for Disaster Risk Reduction 2015–2030 (see <https://www.unisdr.org/we/coordinate/sendai-framework>).

6. International collaboration for implementing a Big Earth Data science framework

Existing and future international programmes that engage in international collaboration on sustainable development and global challenges can significantly contribute to the implementation of the value-chain framework, depicted in [Figure 2](#). Driven by Big Earth Data science technology, the transnational digital framework can encompass diverse governance, social, business and technological aspects. The framework must be characterized by a set of system requirements that affect all of the mentioned aspects and are shared at the international level. In addition, a set of high-level architectural requirements are necessary to ensure the interoperability and effectiveness of the components and resources contributing to the framework. The system and interoperability requirements that characterize the Big Earth Data science value-chain framework include:

- FAIRness of the framework components, services and resources: to apply the FAIR principles (Wilkinson et al. 2016) to the framework components, services and data must be Findable, Accessible, Interoperable and Reusable. This includes data systems, analysis models, algorithm implementations and high-performance infrastructures. In particular, each component, service and resource should be described to allow its usability by humans and machines – see for example APIs.
- Openness of the framework components, services and resources: for an effective stewardship of the framework components, services and resources (and for interoperability) all these elements must be open. For example, computing infrastructures must be transparent to support the application of big data analysis methods by third parties. Similarly, infrastructure openness is necessary to allow different contributors and consumers having strong technical support for resource sharing and reuse.
- Virtualization model (or cloud computing model): this model entails ubiquitous, convenient and on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. Computing virtualization enables elasticity and flexibility in the allocation of physical resources.
- Apply the Distributed Application paradigm: the framework must support the development of applications in a virtual and distributed environment. For example, this is achieved by applying the Web-as-a-Platform (WaaS) pattern, avoiding traditional enterprise level applications and bringing comprehensive components (working collaboratively) based on high-performance data and computing infrastructures – i.e. cloud infrastructures.

- Towards zero data egress (or data locality): due to the variety of data types and the volume of present data series, data egress (i.e. data leaving a network in transit to an external location) must be avoided as much as possible. Data storage and analysis procedures should reside on the same high performing infrastructure (e.g. cloud infrastructure). The resources in an infrastructure should be dynamically configured to meet the needs of data analysis. Interoperability mechanisms among diverse infrastructures must be supported to accomplish complex sharing operations – e.g. multi-cloud interoperability solutions. This approach is also known as data locality, i.e. data processing occurring at the location of the data storage.
- Ethics and security by design: the ethical and security (including privacy and transparency) requirements and constraints stemming from society must be considered in the design of the framework, engaging stakeholders and users at the early stage of the framework conception.
- Framework evolution: With the constant development of new scientific theories and technological devices, the Big Earth Data science framework must be able to update/evolve and add new components and resources (e.g. data sources, data analytics models and high-performance infrastructures) dynamically. Furthermore, the system should constantly evolve the path of solving complex issues, so that the application value of data is constantly advanced.

In addition to FAIRness and openness, other valuable requirements for the data framework include:

- Data uniqueness: to ensure data discoverability, findability and usability, each framework resource should be assigned with a globally unique persistent identifier (PID) during the publishing and archiving process (LIBER Europe 2017). PIDs help users to quickly locate and index unique datasets or metadata that correspond to them.
- Data standardization: To maximize data value and facilitate wide circulation of data among scientific and social communities, standardization should cover the whole process of the data life cycle, including trust and ethical aspects.
- Data comprehensibility: To enhance re-usability, data publishers should carefully describe each dataset with metadata as complete as possible. In particular, metadata elements should describe the generation process, including the behaviour of data producers, editors, publishers, etc., as well as the terms of use and reuse of data. Metadata and data sets can be managed separately, linked via the data PID.

To generate intelligence, the data analytics functionalities of the value-chain framework must support:

- Interactivity: There should be robust algorithms for exploring scientific innovation that guarantee the reproducibility and trustworthiness of results. The framework needs to enable the interaction among different research approaches to make sure results are reliable. This requires a real-time interactive environment for experts during data analysis to ensure data interpretability.
- Visualization: Information visualization (e.g. change trends, score size and object relationships) can provide a better understanding of data and information content and support quick decision-making.

7. Discussion and conclusions

7.1. Big Earth Data science challenges and collaboration opportunities

Studying our planet's sustainability using data science must face several challenges belonging to the different domains and technological areas that constitute the context of Big Earth Data science – see Figure 1. They deal with diverse research and innovation areas, including:

- (a) Holistic view – the models representing the social, economic and natural spheres of our planet as well as their interactions and integration to form a holistic view of Earth.

- (b) Socio-technological view – ethical and privacy challenges related to data analytics and artificial intelligence;
- (c) Economic-technological (or sustainability) view – there is a clear need for innovative, collaborative governance models for digital ecosystems;
- (d) Interoperability view – the advent of IoT is requiring the use of new computing architectures to move intelligence from the centre to the edge of the network (fog and edge computing); another valuable example is inter-cloud interoperability.

The transition from Earth data systems and data sharing to Earth data ecosystems and intelligence generation has started but it is not yet complete. To facilitate and accelerate this process, there is a need to introduce a new science: Big Earth Data science. One of its main objectives is to define a reference framework for this new paradigm. The CASEarth and GEOSS initiatives are two examples of this transition and re-engineering process.

We have argued that Big Earth Data science is becoming an important new scientific discipline to meet the current and future challenges that are a product of societal evolution, digital transformation, and resource scarcity. This new discipline is tasked with integrating the consolidated knowledge of the world and making it accessible to individuals at different levels of the decision and policy formulation process. Acknowledging challenges such as invisibility and inequality, it thereby pushes the limits of sustainable development by leveraging digital and social transformations, and helps us to prepare for a better future. Two examples, CASEarth and GEOSS, provided evidence of the ongoing transition from a data management approach to knowledge sharing and value co-creation. These two examples also illustrated that we still face major research questions and unmatched requirements for systems engineering.

We thus see a need to open a new phase of basic and applied research, as well as systems engineering. This will require sharing and common access to global knowledge, so that every citizen of the world has equal opportunities – without any bias by geographic region or restrictions imposed by demographic or cultural diversity. This globally shared transversal knowledgebase will have to be equipped with socially accepted methods and tools that allow for contextualisation and use in many diverse settings, including cross-border scenarios, national contexts, and local applications.

Earth observation data, research, and IT infrastructures do provide essential ingredients to this newly requested approach to address global challenges. However, this is far from being enough, and new emerging possibilities are highly desirable – for example, integration of insights about human behaviour and sophisticated mechanisms to advance dialogues between scientific disciplines, different stakeholders, and decision-makers at various geographic scales (Onnela 2011). The proposed Big Earth Data science provides the opportunity to research and deliver required scientific and technical components towards the realization of this globally united solution. It will improve the well-being of individuals and communities around the world.

Acknowledgements

The author is grateful for helpful comments from many researchers and colleagues. All views and errors are the responsibility of the author.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The research was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant numbers XDA19030000 and XDA19090000) and by the DG Research and Innovation of the European Commission (H2020 grant number 34538).

ORCID

Stefano Nativi  <http://orcid.org/0000-0003-3185-8539>
 Dong Liang  <http://orcid.org/0000-0001-9147-7792>
 Max Craglia  <http://orcid.org/0000-0001-6244-6428>
 Christina Corban  <http://orcid.org/0000-0002-2670-1302>
 Martino Pesaresi  <http://orcid.org/0000-0003-0620-439X>

References

- Abidi, S. S., and S. R. Abidi. 2019. "Intelligent Health Data Analytics: A Convergence of Artificial Intelligence and Big Data." *Healthcare Management Forum* 32: 178–182.
- Anderson, K., B. Ryan, W. Sonntag, A. Kavvada, and L. Friedl. 2017. "Earth Observation in Service of the 2030 Agenda for Sustainable Development." *Geo-spatial Information Science* 20 (2): 77–96.
- Big Data Value Association. 2019. "Towards a European Data Sharing Space: Enabling Data Exchange and Unlocking AI Potential." BDVA. <http://www.bdva.eu/node/1277>.
- Blei, D. M., and P. Smyth. 2017. "Science and Data Science." *PNAS* 114 (33): 8689–8692.
- Bohlen, V., L. Bruns, N. Menz, F. Kirstein, and S. Schimmler. 2018. Open Data Spaces: Towards the IDS Open Data Ecosystem. Fraunhofer Focus. <http://publica.fraunhofer.de/dokumente/N-555772.html>.
- Bureau of Weights and Measures. 2006. *The International System of Units (SI)*. 8th ed. ISBN 92-822-2213-6.
- Burkholder, L. 2017. "Computing." In *A Companion to the Philosophy of Science*, edited by W. H. Newton-Smith, 44–52. Bodmin: Blackwell Publishers.
- Corbane, C., M. Pesaresi, T. Kemper, P. Politis, A. J. Florczyk, V. Syrris, M. Melchiorri, and P. Soille. 2019. "Automated Global Delineation of Human Settlements from 40 Years of Landsat Satellite Data Archives." *Big Earth Data* 3 (2): 140–169.
- Craglia, M., A. Annoni, P. Benczur, P. Bertoldi, B. Delipetrev, G. De Prato, C. Feijoo, E. Fernandez Macias, et al. 2018. "Artificial Intelligence: A European Perspective." *JRC113826*. doi:10.2760/936974.
- Craglia, M., M. F. Goodchild, A. Annoni, G. Camara, M. Gould, W. Kuhn, D. Mark, et al. 2008. "Next-Generation Digital Earth: A Position Paper from the Vespucci Initiative for the Advancement of Geographic Information Science." *International Journal of Spatial Data Infrastructures Research* 3: 146–167.
- Craglia, M., J. Hradec, S. Nativi, and M. Santoro. 2017. "Exploring the Depths of the Global Earth Observation System of Systems." *Big Earth Data Journal* 1 (1–2): 21–46.
- Dhar, V. 2013. "Data Science and Prediction." *Communications of the ACM* 56 (12): 64–73.
- Ehlers, M., P. Woodgate, A. Annoni, and S. Schade. 2014. "Advancing Digital Earth: Beyond the Next Generation." *International Journal of Digital Earth* 7 (1): 3–16. doi:10.1080/17538947.2013.814449.
- Freire, S., M. Schiavina, A. J. Florczyk, K. MacManus, M. Pesaresi, C. Corbane, O. Borkovska, et al. 2020. "Enhanced Data and Methods for Improving Open and Free Global Population Grids: Putting 'Leaving No One Behind' into Practice." *International Journal of Digital Earth* 13 (1): 61–77.
- Fromm, H., and S. Bloehdorn. 2014. "Big Data – Technologies and Potential." In *Enterprise Integration*, edited by G. Schuh, and V. V. Stich, 107–124. Berlin: Springer.
- Funtowicz, S., and J. Ravetz. 1993. "Science for the Post-Normal Age." *Futures* 25: 739–755. Accessed May 4, 2019. http://www.uu.nl/wetfilos/wetfil10/sprekers/Funtowicz_Ravetz_Futures_1993.pdf.
- Funtowicz, S., and J. K. Ravetz. 2018. "Post-Normal Science." *Companion to Environmental Studies* 443 (447): 443–447.
- GEO. 2014. *GEOSS Data Sharing Principles Post-2015*. file:///C:/Users/Stefano/Downloads/ExCom31_14_GEOSS%20Data%20Sharing%20Principles%20Post-2015.pdf.
- GEO. 2015. *GEO Data Management Principles*. https://www.earthobservations.org/documents/dswg/201504_data_management_principles_long_final.pdf.
- GEO. 2016. *GEO Strategic Plan 2016-2025: Implementing GEOSS*. Geneva. https://earthobservations.org/documents/GEO_Strategic_Plan_2016_2025_Implementing_GEOSS.pdf.
- GEO DAB Team. 2019. *GEO DAB APIs Documentation*. <https://www.geodab.net/apis>.
- GEOSS Platform Operations team. 2018. *Executive Report of the 3rd GEO Data*. Geneva. https://www.earthobservations.org/documents/me_201805_dpw/me_201805_dpw_summary_report.pdf.
- GEOSS Platform Operation Team. 2017. *The GEOSS Platform: All You Need to Know to Become a GEO Data Provider*. Geneva. http://www.earthobservations.org/documents/gci/201711_gci_manual_01.pdf.
- Ghani, N. A., S. Hamid, A. T. Ibrahim Hashem, and E. Ahmed. 2019. "Social Media big Data Analytics: A Survey." *Computers in Human Behavior* 101: 417–428.
- Giuliani, G., J. Masó, P. Mazzetti, S. Nativi, and A. Zabala. 2019. "Paving the Way to Increased Interoperability of Earth Observations Data Cubes." *Data Journal* 4 (3): 1–23.
- Goodchild, M. F. 2004. "GIScience, Geography, Form, and Process." *Annals of the Association of American Geographers* 94 (4): 709–714.

- Goodchild, M. F. 2009. "Geographic Information Systems and Science: Today and Tomorrow." *Annals of GIS* 15: 3–9.
- Goodchild, M. F., H. Guo, A. Annoni, L. Bian, K. de Bie, F. Campbell, M. Craglia, et al. 2012. "Next-generation Digital Earth." *Proceedings of the National Academy of Sciences* 109 (28): 11088–11094. doi:10.1073/pnas.1202383109.
- Gore, A. 1998. "The Digital Earth: Understanding our Planet in the 21st Century." Speech Given at the California Science Center, Los Angeles, California, on January 31, 1998. ESRI. Accessed May 6, 2019. http://portal.opengeospatial.org/files/?artifact_id=6210.
- Guo, Huadong. 2017a. "Big Data Drives the Development of Earth Science." *Big Earth Data* 1 (1–2): 1–3. doi:10.1080/20964471.2017.1405925.
- Guo, Huadong. 2017b. "Big Earth Data: A New Frontier in Earth and Information Sciences." *Big Earth Data* 1 (1–2): 4–20. doi:10.1080/20964471.2017.1403062.
- Guo, Huadong. 2018. "Steps to the Digital Silk Road." *Nature* 554: 25–27.
- Guo, Huadong, Michael F. Goodchild, and Alessandro Annoni. 2020. *Manual of Digital Earth*. Singapore: Springer.
- Guo, H., Z. Liu, H. Jiang, C. Wang, J. Liu, and D. Liang. 2017. "Big Earth Data: A New Challenge and Opportunity for Digital Earth's Development." *International Journal of Digital Earth* 10 (1): 1–12.
- Guo, H. D., L. Z. Wang, F. Chen, and D. Liang. 2014. "Scientific Big Data and Digital Earth." *Chinese Science Bulletin* 59 (35): 5066–5073.
- Guo, H. D., L. Z. Wang, and D. Liang. 2016. "Big Earth Data from Space: A New Engine for Earth Science." *Chinese Science Bulletin* 61 (7): 505–513.
- Hey, T., K. M. Tolle, and S. Tansley, eds. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmont: Microsoft Research. ISBN: 978-0-9825442-0-4. 252 pages.
- Holmes, C. 2018. "Analysis Ready Data Defined. Medium: Planet Stories." Online document. <https://medium.com/planet-stories/analysis-ready-data-defined-5694f6f48815>.
- ISDE. 2009. "Beijing Declaration on Digital Earth-September 12, 2009." The 6th International Symposium on Digital Earth. <https://web.archive.org/web/20120227195943/http://159.226.224.4/isde6en/hyxx11.html>.
- ISDE. 2019. "Florence DECLARATION, ISDE11." 11th International Symposium on Digital Earth, Florence, September 12. <http://www.digitalearth-isde.org/news/851>.
- ISO/IEC. 2016. "Information Technology – Internet of Things Reference Architecture (IoT RA)." ISO/IEC CD 30141:20160910(E) 67 pages.
- ITU-T. 2017. "AI for Good: Global Summit Report." ITU-T. https://www.itu.int/en/ITU-T/AI/Documents/Report/AI_for_Good_Global_Summit_Report_2017.pdf.
- Janowicz, K., and S. Scheider. 2010. "Semantic Reference Systems." In *Encyclopedia of Geography*, edited by B. Warf, 2530–2531. Thousand Oaks, CA: Sage. doi:10.4135/9781412939591.n1020.
- Jansen, S., and M. Cusumano. 2012. "Defining Software Ecosystems: A Survey of Software Platforms and Business Network Governance." Proceedings of IWSECO 2012. <http://ceur-ws.org/Vol-879/paper4.pdf>.
- JRC. 2018. "Artificial Intelligence: A European Perspective." JRC report N.113826. ISBN: 978-92-79-97217-1. <https://publications.jrc.ec.europa.eu/repository/bitstream/JRC113826/ai-flagship-report-online.pdf>.
- Kessler, B. 2017. "Internet of Humans – How We Would Like the Internet of the Future to Be." EU Digital Single Market by European Commission. <https://ec.europa.eu/digital-single-market/en/news/internet-humans-how-we-would-internet-future-be>.
- Kobro-Flatmoen, A., G. Langdon, C. Wright, J. Block, L. J. Gilarranz, J. J. Lever, R. P. Rohr, et al. 2012. "NextGenVoices – Results." *Science* 335 (6064): 36–38.
- Kuhn, W. 2003. "Semantic Reference Systems." *International Journal of Geographical Information Science* 17 (5): 405–409. doi:10.1080/1365881031000114116.
- Lehmann, A., J. Maso', S. Nativi, and G. Giuliani. 2019. "Towards Integrated Essential Variables for Sustainability." Under publication on the International of Digital Earth.
- Leigh, P. 2005. "The Ecological Crisis, the Human Condition, and Community-Based Restoration as an Instrument for Its Cure." *Ethics in Science and Environmental Politics* 9: 3–15.
- Li, W., M. Song, and Y. Tian. 2019. "An Ontology-Driven Cyberinfrastructure for Intelligent Spatiotemporal Question Answering and Open Knowledge Discovery." *ISPRS International Journal of Geo-Information* 8 (11): 496. doi:10.3390/ijgi8110496.
- LIBER Europe. 2017. "Implementing FAIR Data Principles: The Role of Libraries." Online document. <https://libereurope.eu/blog/2017/12/08/implementing-fair-data-principles-role-libraries/>.
- Maarof, A. 2019. "Big Data and the 2030 Agenda for Sustainable Development." UN-ESCAP. https://www.unescap.org/sites/default/files/Final%20Draft_%20stock-taking%20report_For%20Comment_301115.pdf.
- Mayer-Schönberger, V., and K. Cukier. 2014. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray. 242 pages.
- Nativi, S., and L. Bigagli. 2009. "Discovery, Mediation, and Access Services for Earth Observation Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2 (4): 233–240.
- Nativi, S., M. Craglia, and I. DeLoatch. 2018. "GEOSS Evolve Initiative." CEOS online document. http://ceos.org/document_management/Working_Groups/WGISS/Meetings/WGISS-46/3.%20Wednesday/2018.10.24_15.10_GEOSS%20Evolve-6.pdf.

- Nativi, S., M. Craglia, and J. Pearlman. 2013. "Earth Science Infrastructures Interoperability: The Brokering Approach." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (3): 1118–1129.
- Nativi, S., P. Mazzetti, and M. Craglia. 2017. "A View-Based Model of Data-Cube to Support Big Earth Data Systems Interoperability." *Big Earth Data* 1 (1–2): 75–99.
- Nativi, S., P. Mazzetti, M. Santoro, F. Papeschi, M. Craglia, and O. Ochiai. 2015. "Big Data Challenges in Building the Global Earth Observation System of Systems." *Environmental Modelling & Software* 68: 1–26.
- Nativi, S., M. Santoro, G. Giuliani, and P. Mazzetti. 2019. "Towards a Knowledge Base to Support Global Change Policy Goals." *International Journal of Digital Earth*. doi:10.1080/17538947.2018.1559367.
- Newell, A., and H. A. Simon. 1976. "Computer Science as Empirical Inquiry: Symbols and Search." *Communications of the ACM* 19 (3): 113–126.
- NIST Big Data Public Working Group. 2018. NIST Big Data Interoperability Framework: Volume 1, Definitions. NIST Special Publication 1500-1r1.
- Norqvist, L. 2018. Analysis of the Digital Transformation of Society and its Impact on Young People's Lives', published at the Council of Europe website.
- Oliveira, M. I., G. de Fátima Barros Lima, and B. Farias Lóscio. 2019. "Investigations into Data Ecosystems: A Systematic Mapping Study." *Knowledge and Information Systems* 61: 589–630.
- Onnela, J.-P. 2011. "Social Networks and Collective Human Behavior." UN Global Pulse. <https://www.unglobalpulse.org/2011/11/social-networks-and-collective-human-behavior/?highlight=Social%20Networks%20and%20Collective%20Human%20Behavior>.
- Ortmann, J., G. De Felice, D. Wang, and D. Daniel. 2014. "An Egocentric Semantic Reference System for Affordances." *Semantic Web* 5 (6): 449–472.
- Ostermann, F. O., and C. Granell. 2017. "Advancing Science with VGI: Reproducibility and Replicability of Recent Studies Using VGI." *Transactions in GIS* 21 (2): 224–237.
- Oxford Dictionary of English. 1998. London: Oxford University Press.
- Ponti, M., M. Micheli, H. Scholten, and M. Craglia. 2019. *Internet of Things: Implications for Governance*. Brussels: JRC of the European Commission.
- Probst, L., V. Lefebvre, C. Martinez-Diaz, N. U. Bohn, D. Klitou, and J. Conrads. 2018. *Digital Transformation Scoreboard 2018 – EU Businesses Go Digital: Opportunities, Outcomes and Uptake*. Luxembourg: Publications Office of the European Union.
- Romme, G. 2016. *The Quest for Professionalism: The Case of Management and Entrepreneurship*. Oxford: Oxford University Press.
- Ruckenstein, M., and N. Dow Schüll. 2017. "The Datafication of Health." *Annual Review of Anthropology* 46: 261–278.
- Sampson, D., D. Ifenthaler, J. Spector, and P. Isaias. 2014. *Digital Systems for Open Access to Formal and Informal Learning*. Cham: Springer International Publishing.
- Siricharoen, W. V., and U. Pakdeetrakulwong. 2014. "A Survey on Ontology-Driven Geographic Information Systems." Fourth international conference on Digital Information and Communication Technology and its Applications (DICTAP), 180–185, Bangkok. IEEE.
- Song, H., R. Srinivasan, T. Sookoor, and S. Jeschke. 2017. *Smart Cities: Foundations, Principles, and Applications*. Hoboken, NJ: John Wiley & Sons.
- Sugimoto, C. R., H. R. Ekbja, and M. Mattioli. 2016. *Big Data Is Not a Monolith*. Boston, MA: MIT Press.
- UN. 2019. "Big Data and Artificial Intelligence." UN Global Pulse. <https://www.unglobalpulse.org/>.
- UN GGIM. 2016. "Transforming Our World' with Geospatial Information, Prospects and Challenges for Application of High Resolution Earth Observation Data in Sectors Supporting SDG Agenda, Country Profiles & Intelligence." UN GGIM. <http://ggim.un.org/meetings/GGIM-committee/documents/GGIM6/TRANSFORMING%20OUR%20WORLD.pdf>.
- United Nations. 2015. *Transforming Our World: The 2030 Agenda for Sustainable Development*. New York: United Nations.
- UN SDG. 2019. "UN Report: Nature's Dangerous Decline 'Unprecedented'; Species Extinction Rates 'Accelerating.'" Online document. <https://www.un.org/sustainabledevelopment/blog/2019/05/nature-decline-unprecedented-report/>.
- Vaccari, L., M. Craglia, C. Fugazza, S. Nativi, and M. Santoro. 2012. "Integrative Research: The EuroGEOSS Experience." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (6): 1603–1611.
- Van Dijk, J. A., T. Poell, and M. de Waal. 2018. *The Platform Society*. New York: Oxford University Press.
- W3C. 2019. Web of Things (WoT) Architecture. <https://www.w3.org/TR/wot-architecture/>.
- Wellman, B. 2001. "Physical Place and Cyberplace: The Rise of Personalized Networking." *International Journal of Urban and Regional Research* 25 (2): 227–252. <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.169.5891>.
- Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 160018. doi:10.1038/sdata.2016.18.
- Wong, Z. S., J. Zhou, and Q. Zhang. 2019. "Artificial Intelligence for Infectious Disease Big Data Analytics." *Infection, Disease & Health* 24: 44–48.
- World Economic Forum. 2019. "Understanding the Impact of Digitalization on Society." Online document. <http://reports.weforum.org/digital-transformation/understanding-the-impact-of-digitalization-on-society/>.