


Concept Paper

A Transformative Concept: From Data being Passive Objects to Data being Active Subjects

Hans-Peter Plag^{1,2,3,*}  and Shelley-Ann Jules-Plag^{3,4} 

¹ Department of Ocean, Earth, and Atmospheric Science, Old Dominion University, Norfolk, VA 23529, USA

² Mitigation and Adaptation Research Institute, Old Dominion University, Norfolk, VA 23529, USA

³ Tiwah UG, 53547 Rossbach, Germany

⁴ Engineering, Management and System Engineering, Old Dominion University, Norfolk, VA 23529, USA; julesplag@gmail.com

* Correspondence: hpplag@tiwah.com

Received: 28 July 2019; Accepted: 29 September 2019; Published: 2 October 2019



Abstract: The exploitation of potential societal benefits of Earth observations is hampered by users having to engage in often tedious processes to discover data and extract information and knowledge. A concept is introduced for a transition from the current perception of data as passive objects (DPO) to a new perception of data as active subjects (DAS). This transition would greatly increase data usage and exploitation, and support the extraction of knowledge from data products. Enabling the data subjects to actively reach out to potential users would revolutionize data dissemination and sharing and facilitate collaboration in user communities. The three core elements of the transformative DAS concept are: (1) “intelligent semantic data agents” (ISDAs) that have the capabilities to communicate with their human and digital environment. Each ISDA provides a voice to the data product it represents. It has comprehensive knowledge of the represented product including quality, uncertainties, access conditions, previous uses, user feedbacks, etc., and it can engage in transactions with users. (2) A knowledge base that constructs extensive graphs presenting a comprehensive picture of communities of people, applications, models, tools, and resources and provides tools for the analysis of these graphs. (3) An interaction platform that links the ISDAs to the human environment and facilitates transaction including discovery of products, access to products and derived knowledge, modifications and use of products, and the exchange of feedback on the usage. This platform documents the transactions in a secure way maintaining full provenance.

Keywords: data discovery; metadata; knowledge base; graph data; intelligent semantic agents

1. Introduction

The current conceptual approach for discovery of Earth observation (EO) data and derived products is to a large extent based on a perception of data as passive objects. Extracting information and creating new knowledge from data often requires a high level of expertise. Users have to engage in often tedious search processes to discover data. Missing metadata reduce the chance to match data to requirements and determine applicability. Utilizing the data for research most often involves lengthy processes to access products and translate them into a format suitable for the purpose. For decision support, the high level of expertise required to extract information from data is a major obstacle. Feedback on the usability of data for different applications is mostly not collected and not available to users searching for data and knowledge. Semantic issues hamper discoverability and reduce usability of the data and products. Users who would benefit from collaborations often discover potential collaborators by chance. Linking of users with similar interests happens in social networks disconnected from data discovery and access tools. As a result, exploitation of Earth observations (EOs)

in Earth sciences is at a level much lower than desirable and feasible. The use of products and knowledge derived from Earth observations (EOs) for decision and policy making is also hampered by the level of expertise required to extract relevant information from data products and by the limited discoverability.

Currently, the challenges to the discovery, access and use of the increasingly comprehensive Earth observation (EO) data greatly limit the exploitation of the potential societal benefits of this global resource. In fact, the value of Earth observation (EO) as a ‘public good’ depends mainly of the conditions of access to that good [1]. At the same time, humanity is facing growing global threats, see, e.g., [2,3]. Humanity’s quest for sustainable development expressed in the United Nations’ Agenda 2030 [4] is hampered by a lack of information on the biosphere and humansphere, and much of this information could be extracted from Earth observations (EOs) [5]. Developing the interventions that can facilitate progress towards the seventeen Sustainable Development Goals (SDGs) set in the Agenda 2030 and monitoring progress toward the associated Targets requires comprehensive input from Earth observation (EO) communities, see, e.g., [6,7]. Sustainable development as defined in the Agenda 2030, as well as, developing sustainability in general requires a scientific paradigm shift toward systems thinking [8] and this transition has to be informed by comprehensive integrated Earth observation (EO) data. The current description of globally connected systemic and catastrophic risks captures poorly the role of human-environment interactions [9], and this creates a bias towards solutions that often ignore the new realities of the Anthropocene [10]. Understanding “Anthropocene risks”, i.e., risks that emerge from human-driven processes, interact with global social-ecological connectivity, and exhibit complex, cross-scale relationships [10], requires full and easy access to information that can be derived from Earth observation (EO) data and tools for the extraction. The large human-caused changes in the planetary physiology carry the risk of unexpected new phenomena with potentially global consequences and threats [9]. Examples are the emerging threats of sargassum blooms [11], the potential existence of a tipping points for a trajectory towards a “hothouse climate” [12], the possibility of ocean anoxic events [9], and the potential overload of the ocean with carbon [13].

Assessments of risks in general and “Anthropocene risks” in particular very often show a tendency to assume that the large risks are more likely in the far future [14]. For example, a potential state shift in the biosphere [15], reaching tipping points for a hothouse trajectory [12], or the overload of the ocean with carbon [13], etc., are all very often considered as a possibility in the far future, thus ignoring that there are potential hidden risks that could trigger such catastrophic events in the near future. Assessing risks, developing interventions to address the threats today and having early warnings concerning hidden risks also need full access to comprehensive Earth observations (EOs) to address the many knowledge gaps regarding catastrophic risks and to inform interdisciplinary and transdisciplinary mapping and tracking of the multitude of factors that could contribute to global catastrophic risks [16]. In the light of the challenges modern society is facing and the enormous value easy access to comprehensive and integrated Earth observation (EO) data and derived information would have for addressing these challenges, it seems imperative to transform the current relationship between data and users [5]. Thus, the goal of utilizing the societal benefits of Earth observations (EOs) has to be a major design criterion for systems that manage and provide access to such data.

1.1. Meeting Societal Data and Knowledge Needs

Over several decades, Earth observation (EO) communities have made efforts to increase the realization of the societal benefits of Earth observation (EO). The Integrated Global Observing Strategy (IGOS) initiated by the G7 in 1984 as a framework for Earth observations (EOs) was developed with the goal to identify what was essential to be observed in order to document comprehensively the changes that are happening on the planet [17]. The Integrated Global Observing Strategy Partnership (IGOS-P) was established in 1998 with the mandate to ensure that Earth observations (EOs) would respond to societal needs. This partnership brought together major organizations in the scientific and Earth observation (EO) fields and engaged in efforts to first identify what needs to be monitored and then to

facilitate the implementation of corresponding observing systems. IGOS-P used a well-defined theme approach to define the overall strategy, with the themes being motivated by real-world challenges [18]. The resulting IGOS-P theme reports documented very well the outcomes of the first step defining from observational needs for societally relevant themes, see, e.g., [19–23]. However, IGOS-P was less successful in the second step.

Already the Agenda 21 [24], which was a result of the World Summit in Rio in 1992, emphasized the need for coordinated Earth observations and for the creation of knowledge that would support decisions for sustainable development. The World Summit on Sustainable Development in Johannesburg in 2002 reconfirmed the need for coordinated Earth observations, and this led in 2003 to the initiation of the ad hoc Group on Earth Observations (GEO) with the task to develop in eighteen months an implementation plan for the Global Earth Observation System of Systems (GEOSS). The outcome of this activity resulted in 2005 in the establishment of Group on Earth Observations (GEO). The vision of GEO is a future where decisions can be informed by Earth observations. Considering the spectrum of challenges and threats to our global civilization, this is no longer a nice-to-achieve vision; it is a necessity for survival. For GEO, the tool for making progress towards this vision is Global Earth Observation System of Systems (GEOSS). Initially, GEOSS was intended to be integrated into an end-to-end feedback loop with GEOSS providing data and information in support of decision making and users providing feedback on information needs for the further development of GEOSS (Figure 1). Importantly, this initial concept included for GEOSS the task of integrating Earth observation (EO) data with other data and the use of Earth system models to generate the information and knowledge required by societal decision makers.

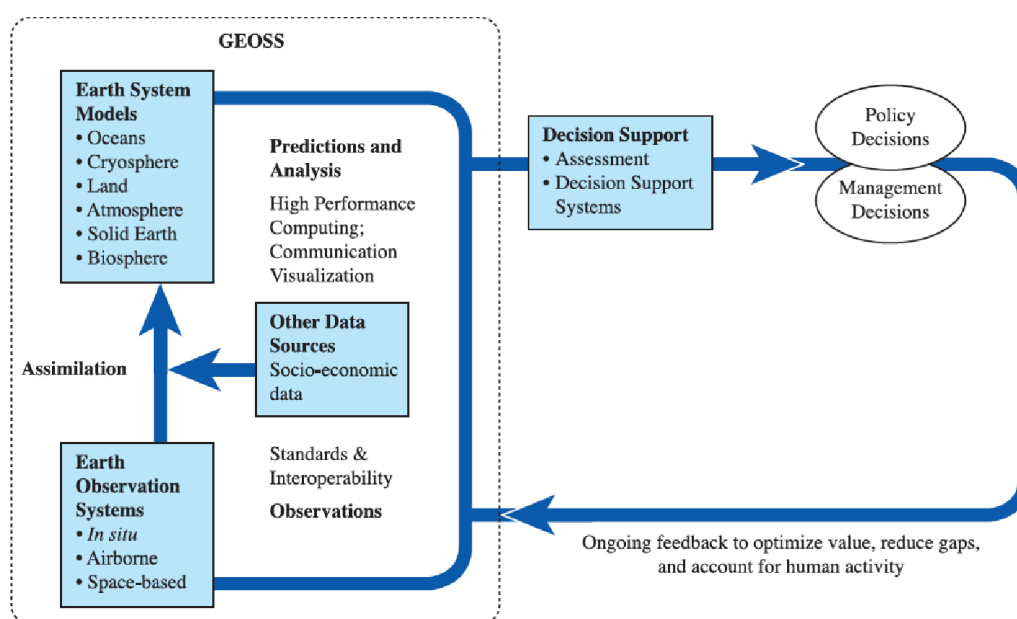


Figure 1. The initial concept for global Earth observation system of systems (GEOSS) emphasized its aim to inform decision making through an end-to-end feedback loop of data and knowledge supporting decision making and feedback from users informing the development of GEOSS. GEOSS was intended to integrate Earth observation (Earth observation (EO)) data with other data and Earth system models to provide the information needed for decision and policy making [25].

In the first ten years of GEO, considerable efforts were made on the feedback part of the loop to improve the knowledge of societal needs in support of defining EO priorities, both in communities of practice that mostly originated in IGOS-P themes, see, e.g., [26], and dedicated efforts to gain overviews of observational requirements derived from societal needs, see, e.g., [27–30]. For the development of GEOSS, the main effort was on improving data discoverability, availability, and accessibility, while the integration with other data and models had much lower priority. As a result, GEOSS

up to today serves best expert communities who have the capacity to search, access, and process the data. Efforts to combine data with a knowledge base remain at an early conceptual state. As recent as 2019, a new concept paper has been accepted by the GEO Executive Committee that proposes the development of a GEOSS Knowledge Hub mainly for expert communities as a framework for transforming Earth observation (EO) data to knowledge for decision making [31]. On the other hand, participatory workshops bringing Earth observation (EO) and science communities together with societal stakeholders again and again reveal that there is a lack of capacity outside relatively small expert communities for the extraction of information from Earth observations (EOs), see, e.g., [32].

Considerable efforts have been made to measure the potential and actual societal benefits of Earth observations (EOs). For example, from 2009 to 2011, a community effort led by NASA aimed at an assessment of societal benefits of Earth observations (EOs) as a basis for the prioritization of Earth observation (EO) systems [27,33]. For several years, the GEO Work Programme included a Fundamental Task on Societal Benefits organizing a sequence of workshops addressing the assessment of societal benefits of Earth observations (EOs). NASA has set up the “VALUABLES” collaboration to measure how satellite information benefits people and the environment when it is used to make decisions [34]. However, very often the results of these assessments are published in reports and not easily available in digital format to link benefit-based knowledge needs to observational requirements.

New societal knowledge needs emerged in 2015 with the United Nations’ adoption of the 2030 Agenda for Sustainable Development [4], the adoption of the Sendai Framework for Disaster Risk Reduction 2015–2030 [35] by the United Nations, and the Paris Climate Agreement reached under the United Nations Framework Convention on Climate Change (UNFCCC). GEO has responded to the emergence of these agreements by including the support for the UN 2030 Agenda for Sustainable Development, the Paris Climate Agreement, and the Sendai Framework for Disaster Risk Reduction in the global priorities. Likewise, several United Nations agencies give the support of these frameworks high priority. Among others, the urgent need for a transformative digital ecosystem for the environment is emphasized by [5] to ensure that progress towards sustainability is informed by data.

Considering the example of the 2030 Agenda, the development and validation of interventions to reach the many targets associated with the seventeen Sustainable Development Goals (SDGs) pose wicked problems to society. Wicked problems are social or cultural problems that are difficult or impossible to solve because of incomplete and often contradictory knowledge, the large number of people and opinions involved, the heavy economic burden associated with progress towards a solution, and the interconnected nature of each problem with many other problems [36]. All of this applies to the Sustainable Development Goals (SDGs). In particular, knowledge on how to make progress towards the Sustainable Development Goals (SDGs) is incomplete and contradictory, reaching the SDGs even on a local level involves the whole of society, making progress requires a rethinking of economy [37], and the goals are strongly interconnected, see, e.g., [38–40]. Moreover, there are many interactions between the individual goals that are variable across different economic, social, and cultural settings [7].

Monitoring progress towards the targets associated with the Sustainable Development Goals (SDGs) requires metrics defined by a set of indicators, and developing indicators that provide useful quantitative metrics is a long process involving the scientific community, see, e.g., [41,42]. The United Nations Statistical Commission (UNSC) created the Inter-Agency and Expert Group on SDG Indicators (IAEG-SDGs) with the aim to develop a manageable indicator framework. Based on a proposal of the IAEG-SDGs, an initial framework with a total of 232 global indicators was adopted in 2017 by the United Nations General Assembly as a voluntary and country-led endeavor to monitor progress towards the SDG Targets. According to the level of data availability and methodological development, the SDG Indicators have been grouped in three different Tiers: From Tier I, for the ones having an established methodology and widely available data, to Tiers II and III, for those not having data available or no methodology established, respectively. As of 11 May 2018, the updated tier classification contains 93 Tier I indicators, 72 Tier II indicators, and 62 Tier III indicators [43]. However, actually

being able to quantify these indicators for individual countries poses an insurmountable challenge to small countries like the Small Island Developing States (SIDS) and those countries with very limited economic resources. Many of the indicators depend very much on Earth observations (EOs) and an integration of Earth observations (EOs) with other socioeconomic data and models [6,7,44,45].

Many efforts have focused on archiving and publishing datasets. An example is the World Data Center PANGAEA [46], which is a member of the ICSU World Data System. PANGAEA provides services for archiving, publishing, and re-usage of data [46]. Most of the datasets are open access, and a search engine provides a high level of discoverability. However, being a repository, the dataset are passive objects and extracting information from a dataset requires accessing the data and using expertise in the analyses of the data. The datasets are structured under a set of themes and sub-themes, which limits transdisciplinary approaches.

Efforts are also being made to utilize relationships between datasets and products to increase data discoverability and utilization. For example, the Linked Open Data Cloud (LODC) captures the relationships between an increasing number of datasets [47]. As of March 2019, the dataset contains 1239 datasets with 16,147 links. More datasets can be registered manually and links can be recorded. The LODC generates domain specific sub-clouds. Users can interactively explore the cloud to retrieve information of specific datasets or explore the relationships captured in the links. The full LODC is available for analyses. However, links to other objects such as applications, user types, processing tools, etc., are not comprehensively captured and feedback on the datasets is not solicited.

Recommender systems that would promote datasets and products to potential users are very limited in the Earth observation (EO) community. However, recommender systems are increasingly used for the promotion of commercial products. Commercial retailers increasingly use advanced algorithms including big data analyses, deep learning, deep search, and crowd-sourcing to bring their products to potential customers. In the early use of the Web, customers often had to carry out lengthy searches over limited domains to discover the products and services they were looking for, a conceptual approach that is denoted here as Customers Discover Products (CDP). The recent development in the commercial domain constitutes a transition to a conceptual approach where a framework enables products to discover potential customers, a concept denoted here as Products Discover Customers (PDC). Customers of, e.g., Amazon are informed when new books and other products appear on the market that might be of interest for them based on previous searches or purchases. Recommender systems have been developed and deployed in supermarkets to aid customers in decisions of what to choose from the large variety of products, see, e.g., [48]. Web advertisements are targeted to likely recipients based on social media behavior or Web searches. In Products Discover Customers (PDC), data from social media are increasingly collected and analyzed to explore connections among people and between people and products to propose and facilitate new connections. Extensive feedback on products and services is collected from customers and users and made available to inform decisions of other customers and users. In some cases, attempts are made to stimulate feedbacks with rewards, e.g., when hotels have very low numbers of reviews, Hotels.com offers coupons for special nights in return for reviews, and feedbackrewards.com manages for companies customer feedback programs using rewards for stimulating feedback [49].

Recent artificial intelligence (AI) developments have opened the door for intelligent software agents, see, e.g., [50,51]. Theoretical concepts have been developed to capture connections between societal agents, products, tools, activities, and transactions, and to construct graph data describing the chains and networks between these elements.

1.2. From Passive Data Objects to Active Data Subjects

The ability to design intelligent software agents that can represent a data product and provide comprehensive information derived from this product, combined with the ability to construct extensive graph data provides a basis for a transition in the Earth observation (EO) domain from the perception of Data as Passive Objects (DPO) to a perception of Data as Active Subjects (DAS). The DAS concept

has the overarching goal to greatly increase data usage and exploitation. It has the potential to revolutionize data discovery, sharing, dissemination and usage and by doing so greatly enhance the exploitation of Earth observations (EOs) for research and the realization of societal benefits. In contrast with the current DPO concept, in which datasets are passive and isolated in repositories, the DAS approach pairs datasets with intelligent software data agents that can connect and interact with other software and human agents. These software data agents are comparable to human agents who provide links between people (such as actors, musicians, etc.) and potential jobs. Similar to those human agents for people, the software data agents have full knowledge about the dataset(s) they represent, including among others comprehensive metadata as well as information on usability and applicability, and they have the ability to discover potential applications and users for their datasets(s).

The subject does the action. The object is the center of action. In the DPO perception, e.g., researcher X analyzed the global temperature data to quantify global warming. In the DAS perception, the global temperature dataset Y would inform that global heating has reached 0.1 °C per decade. In the first case, the temperature data is the object. In the second case, the data is the subject and this subject informs about knowledge it could extract from its data.

Another example would be a minister in a government who is in need to quantify one of the indicators for the SDGs. In the DPO world, the minister could have to engage a team of experts to discover and collect the relevant data, use appropriate processing tools, and, following a best practice, generate the quantitative indicator. In this case, all data used would be objects and even the indicator would be an object. However, in the DAS world, there would be a software agent representing this indicator, and this agent could inform the minister of the quantitative development of the indicator in the minister's country. This would be of great value particularly for the smaller and less resourceful countries such as the SIDS, see, e.g., [52].

Having active data-based subjects, these subjects also could have the capability to promote their data and knowledge to societal human agents who would benefit from this. Today, the dominating concept for data distribution is one of Users Discover Data (UDD). Within the Data as Active Subjects (DAS), a transition to a new concept of Data Discover Users (DDU) would be possible. This would be comparable to the ongoing transition in the commercial world mentioned above from Customers Discover Products (CDP) to Products Discover Customers (PDC).

1.3. Structure of The Paper

In the next section, the DAS concept is outlined in more detail. After an overview, three subsequent subsections discuss the three core elements of this concept, i.e., the Intelligent Semantic Data Agents (ISDAs) that are representing datasets, products and services (Section 2.2), the knowledge base that creates and provides access to extensive graph data (Section 2.3), and the interaction platform on which human users and ISDAs interact (Section 2.4). Section 3.1 explores the potential of DAS not only in terms of increased data exploitation but also in terms of capacity building, decision and policy making, and realization of societal benefits of Earth observations (EOs) and derived knowledge. Section 3.2 outlines a case study for the validation of the concept, and Section 3.3 provides thoughts on the implementation and identifies challenges for the implementation of DAS. Section 4 summarizes the main conclusions.

2. The DAS Concept

2.1. Overview

The overarching design criterion for the DAS concept (Figure 2) is the goal of enabling data products to actively respond to information and knowledge needs of societal users and to reach out to those who may benefit from knowing about a data product and having access to the product or information derived from the product. To some extent, this change in perception of data objects is comparable to the one from considering cars as passive objects that are driven by humans to cars as active subjects that provide

transportation to humans and other objects as needed. In the same way as autonomous cars may lead to a Gestalt shift [53] in how we perceive transportation, the transition to perceiving data as active subjects could lead to a Gestalt shift in how we perceive knowledge derived from data.

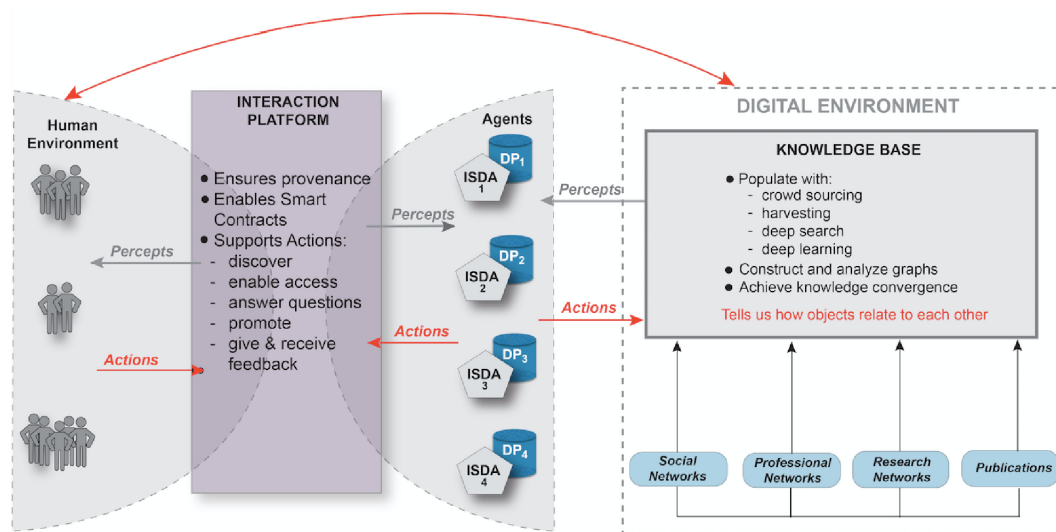


Figure 2. In the data as active subjects (DAS) concept, each intelligent semantic data agent (ISDA) represents a data product (DP). The ISDAs utilize the graph data in a knowledge base to discover applications and users that could benefit from their data products. They interact with those users, or users that contact them, to provide knowledge or manage access to data. All interactions that impact the data are recorded to ensure provenance. The knowledge base generates graph data based on information obtained through crowd sourcing or extracted from social and research networks and publications.

The DAS concept introduced here hinges on three core elements (Figure 2):

1. Intelligent Semantic Data Agents (ISDAs) that are software agents that represent data products. They have the goal to serve potential users and to increase the exploitation of the societal benefits of the data product they represent. To achieve this, an ISDA has comprehensive knowledge about the data product it represents including quality, uncertainties, access conditions, previous uses, user feedbacks, etc. These non-human software agents have the semantic capabilities to communicate with potential users in the human environment and comprehensive graph data in the knowledge base. The ISDAs also have semantic and pragmatic descriptors that allow them to meaningfully interconnected with software agents of other datasets through complex and dynamic relations. These relations are continuously updated as users interact with the data agents and provide feedback on the data.
2. A knowledge base that can construct and analyze extensive graphs presenting a comprehensive picture of the elements in a community of people, applications, models, tools, and resources. Earth observation (EO) data is mostly polyglot spatial data representing properties at points, lines, or polygons in space and their changes over time (Figure 3). Graph data captures the connections between objects and can consist, e.g., of property graphs linking persons, network graphs linking locations, semantic graphs linking language elements in ontologies, and more generalized graphs linking diverse objects such as data sets, information needs, and societal agents. Polyglot data are helpful in answering questions such as “how did land cover change over time at this point?” Graph data can answer questions such as “which researcher could benefit from land cover data?” The knowledge base will focus on graph data providing links between, e.g., knowledge needs and data types, user types and applications, publications and datasets, processing tools and datasets. None of the objects linked in the graph data resides in the knowledge base.

3. An interaction platform to negotiate and execute “contracts” under which users gain access to knowledge extracted from data, access data, modify data, use data and provide feedback on their usage, and to document these interactions in a secure and reliable way maintaining full provenance.

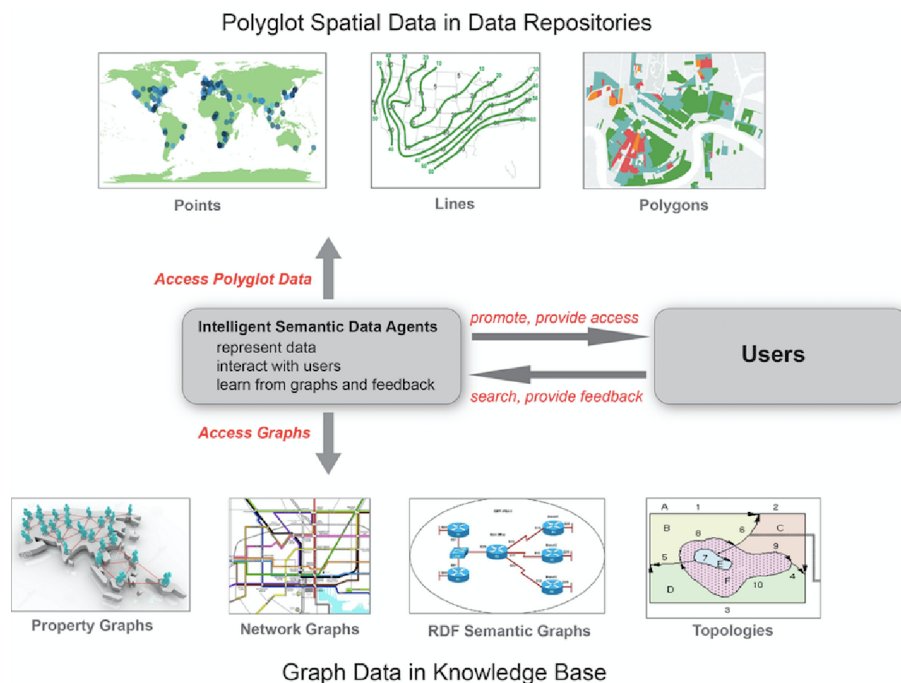


Figure 3. In the DAS concept, graph data capturing the properties and connections in diverse networks (people, applications, models, datasets) are used by data agents representing data to match users and data both on request (searches) and through promotion. The data agents “learn” from user feedback and dynamically adjust to changes in the graphs.

In the DAS concept, datasets and products derived from Earth observations (EOs) are associated with the Intelligent Semantic Data Agents (ISDAs) that can communicate semantic information in response to queries including access conditions, derived knowledge, quality, uncertainties, guidance on applicability, and user feedback. Conceptually, these ISDAs utilize the graph data in the knowledge base to explore the user landscape in search for users that might have interest in the data (Figure 2). They can interact with users as well as other ISDAs. An ISDA will also have knowledge about tools that can make use of the data or derive other products from the data. The sharing of this knowledge with users facilitates rapid capacity building in the use of the data and broadens the range of scientific applications of the data represented by the ISDA. Thus, the DAS concept provides remedies to many of the current issues associated with a perception of passive data objects paired with passive metadata that often are maintained separately from the actual data. All interactions with a data agent are either integrated into the agent as an innate part or recorded in the provenance system.

The knowledge base in the DAS concept uses deep searches, big data analyses and crowd sourcing to map for specific use cases the user landscape in the communities engaged in research and applications and to identify their knowledge and information needs. Based on deep searches and deep learning, graphs of user types, what they do, their tools, and their potential needs are constructed from publications, social networks, social media communications, and observation inventories. The graph data are analyzed to enable the ISDAs to promote their data products to users with potentially matching interests and needs.

The ISDAs utilize the interaction platform for communication and interactions with users. This platform provides a system that tracks interactions with users, ensures provenance and increase

reproducibility of research that is based on the represented data. The matching of users and data products takes place on this interaction platform, which will ensure provenance. The interactions are handled with an approach similar to smart contracts. Searches and feedbacks are analyzed by the knowledge base to update graphs and by the ISDAs to add intelligence to the ISDAs and to enable them to identify new potential use cases for the data they represent.

2.2. Intelligent Semantic Data Agents

The introduction of the software Intelligent Semantic Data Agents (ISDAs) (Figure 2) is a concept that has the potential to revolutionize the interaction of users and data. The principle idea is comparable to the human agent of, e.g., a movie star, who has the task to promote the actor and to negotiate new engagements for the actor. Ideally, the human agent has all relevant information about the actor, including past engagements, preferred partners, limitations, and preferences, and fully understands the capabilities of the actor. Similarly, an ISDA has all relevant information about a dataset, including comprehensive provenance, related datasets, models and applications to be used by users, user types that might be interested, applicability and limitations, quality and uncertainties, and more. The ISDA has the task to promote the dataset actively to potential users (thus making progress toward the Data Discover Users (DDU) concept), to respond to queries, to inform about the dataset, to provide derived information (e.g., selected statistics, subsets, etc.), receive feedback from users, and to learn from user interactions to be better prepared for future users.

From a semantic point of view, the knowledge base will formulate the semantics of the domain, such that each data product has a meaning attached to it. However, it will go beyond the semantics of datasets to a pragmatic approach, in which a data product is represented by an agent that is aware of the data product's meaning and is capable of learning potential use cases of the data product. Thus, data products will be represented by agents (the ISDAs) that can act on knowledge within the knowledge base and generate new knowledge.

Data products present in the graphs of the knowledge base will be represented by ISDAs that act on their behalf. The ISDAs are purposive software agents whose aim is to facilitate the interaction between users and the data product. In particular, an ISDA will be able to respond to questions about its data product, provide access to parts or all of the data product, and solicit feedback on the data product. Initially, the ISDAs will be goal-based agents [50,51] but they will have to evolve into learning agents. The ISDAs can request specific analytics from the knowledge base to discover potential users and to enter into communication with them. In particular, it can find users with the skills and interest to use the data or who might need these data to corroborate a published study, even if these potential users did not know of the existence of the data. The ISDAs will be able to use the social media and contact information of users in the knowledge base to enter in communication with them. A core research question on the path to implementation is how rich the data description will have to be to enable these capabilities.

The ISDAs are capable of executing complex transaction patterns with users, such as granting access, executing custom queries to aggregate, truncate, convert, randomly sample data, and provide references or meta-data. For that, the agents will adopt a transaction processing framework to manage its interactions with other agents and users [54]. The concept of rough set [55–58] can be considered as a capability of the ISDAs.

The ISDAs will be able to grow from initial “seeds” with very limited capabilities into fully developed “adult” agents that have access to all the information related to the dataset, including all uses, experiences, feedbacks. Thus, the agents gain in knowledge as the knowledge base becomes more complete. A deep-learning algorithm will be used to further enrich the information available to an ISDA about the represented dataset so that it can link to users with potentially matching interests and needs and inform users about products of potential interest to them, including the data sharing and access conditions.

The ISDAs will also benefit from a generalization of the concept of digital object identifier which comprehensively identifies a dataset including the relevant metadata, the ISDA, and derived datasets in a consistent identification scheme. Having the main identifier pointing to the ISDA instead of the dataset itself will ensure that a user who aims to access the data always will have access to the full history of transformations and applications of the data.

2.3. The Knowledge Base

The knowledge base is envisioned as an extended version of the existing Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB), which has the main function to construct and analyze graph data capturing the connections between datasets, products, applications, user types, and other elements in scientific communities and society at large. To the extent permissible under privacy and personal data protection regulations (such as the European General Data Protection Regulation (GDPR)), individual persons can be integrated into graph data. This knowledge base provides the graph data and analytical tools to connect users and facilitate collaborations.

Graph data consists of two basic elements: The nodes (or vertices), and the links (or edges) between these nodes. Both the nodes and links are objects that are characterized by a set of properties. Each link is associated with two nodes. Links can be directional with head and tail nodes or bidirectional. In the Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB), the nodes are not limited in terms of what objects can constitute a node. For example, nodes can be as diverse as a specific person, a group or type of humans (e.g., a user type), a dataset, an information need, a societal goal, a modeling software, or a specific observation sensor. The set of properties for each class of nodes and links is dynamic and can be extended as more information about an object becomes available. Importantly, each node and link has a unique identifier.

The knowledge base uses big data analysis techniques to map the user landscape in the communities engaged in research and applications and identify their knowledge and information needs. It generates graph data that describe user types and their potential needs based on publications and social media communications and links them to tools and datasets. In utilizing published information on persons, such as paper authorship and owners of data and processing tools, it will be important to ensure compliance to privacy and personal data protection regulations, such as the General Data Protection Regulation (GDPR). Individual persons can be integrated as nodes into the graphs. During the development of the Global Earth Observation System of Systems (GEOSS) User Requirements Registry (URR), which initially only captured user types, users of the User Requirements Registry (URR) repeatedly requested the possibility to link themselves to user types and establish a social network of users within the User Requirements Registry (URR) [30]. It is expected that similar requests are made for the knowledge base. The knowledge base also maps the Earth observation (EO) landscape in terms of available datasets, products, and processing tools. The research communities are being mapped in terms of research topics, needs, and challenges, as well as the tools available to process and analyze data and to use data for modeling and simulation. An important source for mapping research communities is the comprehensive publication and citation data compiled in rapidly expanding research knowledge hubs. Increasingly, journals require information on data and tools used for the research published in a paper, see, e.g., [59]. This information can be exploited to inform the construction of graph data and to increase the knowledge and skills of the ISDAs. The development of the graph data also is based on deep searches and deep learning from scholarly and other publications, social networks, etc. In particular, the knowledge base will employ parallel crawlers to inform the construction of graph data.

The knowledge base requires the capability to provide the information needed to bring data and products to potential users. This capability has to be based on the full spectrum of graph theory. This includes the detection of components and communities applying, e.g., the deep search algorithms depth-first search (DFS) [60] and Kosaraju, see, e.g., [61], and the concept of weakly connected components, label propagation, and spacification [62]. Evaluating community structures

can focus on conductance, modularity, and clustering coefficients [63], and this provides a basis to identify collaboration potentials between research groups and individuals. Ranking and walking along graphs provides a basis for prioritization as well as discovery of relevant nodes in support of data promotion and can be based on algorithms applying pageranks and different centralities, see, e.g., [64], random walking and sampling. Path-finding facilitates the identification of users who's requirements could be a match for a dataset, applying, e.g., Dijkstra's [65] and Bellman-Ford's [61] algorithms. Importantly, detection of unreliable or fake information [66,67] has to be integrated into the graph development processes.

The Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB) provides extensive search and feedback utilities and the analysis of both searches and feedbacks with deep learning methods can further improve the capability to add intelligence to the Intelligent Semantic Data Agents (ISDAs). Crowd-sourcing opportunities can be used to gather both primary graph data and feedback on data and the performance of the ISDAs. The lexicon (ontology) contained in the Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB) as the primary source for all semantic aspects will grow based on deep learning from other registries and from user interactions. The Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB) provides access to a large set of user needs (originally collected in the Global Earth Observation System of Systems (GEOSS) User Requirements Registry (URR) [68,69]) and observational requirements (partly harvested from OSCAR, see <http://www.wmo-sat.info/oscar>). The Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB) explores existing and new data repository in an effort to link Earth observations (EOs) and the global community of potential users.

Big data analytics on the graph data in an extended version of the Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB) is at the core of the DAS concept. In the current DPO concept datasets are passive and isolated in repositories. In contrast to this, the DAS approach will create the graph data of a "Web of things" where each dataset will be represented by a node with semantic and pragmatic descriptors, and meaningfully interconnected with the other entities (other datasets, users, models, instruments, etc.) through complex and dynamic relations, which will be updated as users and ISDAs interact with the graph data and provide feedback.

The graph data requires a generic model for metadata (referred to below as metamodel) that enables the networked representation of a population of entities and their mutual relations. Since the system is open-ended, and the final extent of all datasets that may be added is not known at inception, it would be illusive to attempt to create a fixed and comprehensive ontology that would encompass every future addition of datasets in the knowledge base.

A dataset provides a partial, biased, and time-bounded description of an object of interest in the real world. This means that the dataset expresses a reference in a semiotic relationship that involves the real world object as a referent, and the specific form of the data as symbol. The data provider and data users relate with the dataset both at a semantic level to uncover the meaning expressed in it, and also at a pragmatic level to achieve some practical ends, communicative or otherwise. In this sense, datasets seem to be more complex objects to manipulate and recommend automatically than products on Amazon or videos of Youtube. Even the individuation of the real-world object to which the data is pointing is subject to the researcher's interests and underlying theories or a user's preconceptions and world view. Similarly, the characteristics of the object represented by the dataset depend on the technical means of observation, on the methodology adopted, and on the level of fidelity decided by the data provider.

Other aspects to be covered in the DAS approach involve the origin of the data (what actors made it available), how it was obtained, for instance, whether the measurement is punctual or longitudinal, whether the data originated from a model (and what kind of model), a survey, observations (what kind of sensor), and what use-cases the data can support. The Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB) will also have to enforce integrity rules through mechanisms like reputation management, voting, and read/copy/write access rules, to make sure that

datasets are not tampered with, and that single source of truth principles are maintained for every given data entity.

An important step towards the implementation of the DAS concept is the introduction of an extensible metamodel that covers these aspects of the graph data, so that the Intelligent Semantic Data Agents (ISDAs) initiated by data providers may represent their associated datasets as precisely as possible, that advanced search capabilities may be implemented, and that the big data algorithms have a rich basis upon which to analyze a continuously growing knowledge base, and ultimately bring the data to those data users who need it.

Besides the graph-data metamodel, an important ingredient for the DAS concept is the introduction of advanced machine learning algorithms to bring the data to potential users. Broadly speaking, machine learning refers to capability of a computer program to learn a knowledge-intensive task while improving its performance on the task as it gains more experience [70]. The task at hand is the suggestion of datasets and potential collaborators to a set of users. The performance corresponds to the practical value of the suggested datasets to the users, while the experience is derived from the feedback obtained from users regarding the quality of the suggestions. The machine learning algorithms will take advantage of the underlying structure of the graph data, the similarity between datasets, and the similarity between users as obtained from social media and scholarly publications. The machine learning techniques that can be used to achieve this include clustering, collaborative filtering, case-based reasoning, and deep learning.

Clustering is a computing task in which a set of objects is segmented in subsets such that the objects in one cluster are more similar to each other than the objects out of the cluster [70]. Clustering can be used to create categories of datasets on the one hand, and categories of users and applications on the other hand. The clustering of datasets can be performed by applying the highly connected subgraph algorithm [71] on the graph data. Datasets will be found in the same cluster if they are highly connected in the graph data, which would mean that the datasets within one cluster will share relevant variables and methodological features. The similarity metric of the clustering algorithm will be continuously adapted based on the feedback received from users. Thus, as the algorithm gains in experience, the clustering of the datasets will result in groups more and more homogeneous, thereby enabling more customized suggestions. Since the graph links have different semantics, the same dataset element will potentially belong to multiple clusters, for instance geographic clusters, data fidelity clusters, topical clusters, etc.

Using social network data (such as Facebook posts or Twitter hashtags), parsed publications, research knowledge hubs with citation data, newspaper articles (particularly those discussing science-related topics), co-citation analysis, as well as past patterns of dataset search and use, it will be possible to similarly cluster the users into multiple groups based on their scientific disciplines, their application domains of interest, their geographic area of focus, etc. Here again, as the algorithm learns more about the relevant properties that users share, they will be placed in clusters that become more and more specific, so that the recommendations will become more accurate.

Collaborative filtering uses the ratings and feedback provided by users of a product to recommend the same product to users with a high level of similarity. A commonly used similarity metric is the Pearson correlation [72] or the vector cosine-based similarity [73]. In this approach, crowd-sourced user feedback is exploited to provide better suggestions. This method may be inadequate at the beginning when user feedback data is sparse, but improves exponentially as user data becomes more widespread [74]. Collaborative filtering works well in combination with the clustering method described before, since, initially, recommendations may be forwarded to users in the same cluster, as they share some similarity.

In case-based reasoning, properties of datasets and of users entities are utilized to match users and products. The cases encode knowledge such as “users sufficiently similar with user u and who accessed dataset with property x also used dataset with property y .” As such, case-based reasoning will exploit the results of the clustering algorithms. Case-based reasoning algorithms are often based on decision trees [75] and have some major benefits: They are suitable for non-formalized knowledge

domains, they are robust and easy to maintain, and they allow for incremental improvement. However, just as with collaborative filtering, the approach becomes computationally inefficient when the domain is too dynamic and when the number of cases becomes very large [76].

To remedy these shortcomings, deep learning, based on restricted Boltzmann machines [77] are emerging as very promising techniques for data intensive learning tasks, owing to the availability of parallelized computational resources. These techniques use successive layers of neural networks and perform computations of increasing levels of abstraction to discover a hierarchy of features, from low-level features to higher level ones [78], i.e., a bottom-up approach. Deep-learning algorithms have been successfully applied to computer vision and language processing and have only recently begun to be used in commercial recommender systems [79]. As shown in [80], deep-learning algorithm can be used to learn about the attitudes of a user toward a dataset from the review text of dataset posted by users and the features of the product itself, and thereby match datasets with types of users to maximize the utility of a dataset for a certain type of user.

2.4. Interaction Platform

The interaction platform is the space in which users and ISDAs interact with each other (Figure 2) and where a track record of these interactions is being kept. Users of the platform can take on the role of data provider, who want to make datasets available to a community of users, and data users who may be scientists who need some data in the context of their research or other social agents (individuals, governmental bodies, NGOs) who may have interest in knowledge derived from the data to answer practical questions relevant to their problems.

Experience and events should be captured in schemes that provide a complete history of a given dataset. While such a scheme for the recording of the transactions could be based on blockchains, there are concerns that this would be far too demanding in terms of energy, see, e.g., [81]. Blockchain is an emerging interaction paradigm for transmission and storage of information without centralized control. It is a secure and distributed database that is hosted locally by the human or software agents engaged in a transaction. It contains the history of all transactions performed by these agents, without a centralized intermediary, thereby allowing each participant to independently verify validity of a chain of interactions. Furthermore, blockchains can be made public or limit access to only users with specified credentials.

The first blockchain was introduced by Bitcoin [82], but its use as an architectural model for secure user interaction has now expanded beyond the domain of digital currencies [83]. User transactions are structured in blocks. Each block is validated by an algorithmic key or “proof-of-work.” Once a block is validated, it is timestamped and added to the chain of blocks and becomes publicly visible to the members of the network. The decentralized, transparent and robust nature of blockchain makes it particularly well adapted for a distributed and intelligent data search system. However, the choice of whether to use one of the existing blockchains (for a discussion of potential candidates, see, e.g., [84]) or to develop a new blockchain dedicated to data and knowledge-related transactions would be a difficult one. In addition, there are concerns that the trust in blockchains is not fully justified [85]. An important application of blockchains is to provide provenance particularly with respect to transfers of ownership in something. This comes with a very high use of resources. In fact, a white paper developed by the World Economic Forum states that the energy consumed in the blockchain network is unsustainable [81]. Energy consumption can be reduced significantly depending on the consensus algorithms used [86], and replacing the “proof-of-work” algorithms by “proof-of-stake” or “proof-of-authority” results in drastically reduce energy consumption decoupled from the number of users engaged in a blockchain [87]. For the access to data, tools to process the data, information derived from data, and knowledge created using the data, the ownership in general remains with the originator, and only the rights to access, processing, use and further distribution are points of negotiation. For this purpose, provenance may be achieved without blockchains. However,

a distributed ledger that validates and records transactions between several ISDAs as well as between ISDAs and human agents seems to be mandatory for the interaction platform.

For the management of interactions between agents (data agents, models, persons, repositories, etc.), a concept similar to that of “smart contracts” could be developed. These “smart contracts” would automatically perform delegated terms of a contract without user intervention. The traceability of blockchains or a similar distributed ledger would allow the capture of events and user experiences as blockchain-based schemes to provide a complete history of datasets addition, access, purchase, updates, etc. To the extent possible, protocols would facilitate, verify, or enforce the negotiation or performance of a “contract” between a user and the ISDA representing the data product. With this concept, many aspects of the transactions could be made partially or fully self-executing, self-enforcing, or both. Conceptually, this approach provides security superior to traditional more open transactions. The “smart contract” concept seamlessly interfaces with a distributed ledger.

However, as noted above, blockchains are very demanding in terms of computational resources and energy, and a careful assessment of the trade-off between the amount of resources needed and the level of security, perseverance, and documentation achieved needs to be carried out to inform the design of the interaction platform.

3. Discussion

3.1. Current Status and New Contributions

Many Earth observation (EO) communities have made considerable efforts to improve data discoverability and accessibility. In particular, Group on Earth Observations (GEO) has made a significant contribution serving users of data with means to discover data, see, e.g., [88]. In many scientific communities, efforts have been made towards the integration of data and modeling tools. A particular focus has been on the development of data models that support interdisciplinary and cross-disciplinary data integration, see, e.g., [89]. Harmonization of metadata across thematic areas and beyond poses a major challenge, see, e.g., [90]. Brokering of data and metadata for a large number of datasets is often at the core of efforts to overcome this challenge, see, e.g., [88,91]. The need for new transformative approaches is acknowledged, see, e.g., [5,92].

For the development of Earth observation (EO) systems with high scientific and societal benefits, comprehensive knowledge of information needs is mandatory. Over the last few decades, there have been abundant efforts at national and international levels to assess user needs that constitute requirements for Earth observations (EOs). Examples are the reports produced by IGOS-P themes, see, e.g., [19–21,23,93], and the reports that resulted from the GEO task US-09-01a, see, e.g., [94]. In most cases, mapping of user landscapes was based on limited surveys, user forums, or literature reviews by experts with emphasis mostly on one or another methodology. Surveys of users often resulted in limited responses, and the main input was provided by expert groups and communities (see, e.g., [33] and the references therein). The output of most of these efforts consists mainly in written reports with no functionality for further machine and algorithm-based analyses. While these reports have a high value, exploitation is low. Repositories of observational requirements (such as OSCAR, see <http://www.wmo-sat.info/oscar>) are mostly limited to relational databases and in most cases lack a linkage of the observational requirements to societal users and their decision and policy making processes. In most cases, feedback capabilities are limited or absent and users have limited opportunities to comment on and augment the information in the repositories. The construction and analysis of graphs is not supported in these approaches. However, implementing DAS can build on these initiatives and utilize the resulting reports and repositories in the construction of graphs.

The Global Earth Observation System of Systems (GEOSS) User Requirements Registry (URR) aimed to construct graphs that represented user types, applications, observational requirements, and needs in terms of research, technology, infrastructure, and capacity [69,95,96]. These graphs captured the connectivity between instances in one group as well as cross-group interdependencies.

The experience with the URR shows that users wanted the graphs to be extended to include far more groups, such as models, tools, people, data, knowledge, decision and policy making, etc. [69]. This user-based request was one of the main motivations for the transition of the URR to the Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB).

Identifying and describing comprehensively all characteristics of data relevant for the matching of potential users still presents a major challenge to the provision of sufficient metadata. Different communities have different views and understandings of a given characteristics, and this severely hampers harmonization (see, e.g., the discussion on data quality in [97]).

Abundant efforts have been made to map the user landscapes for Earth observations (EOs). For example, numerous efforts have been made to characterize those users engaged in water sustainability that depend on information derived from Earth observations (EOs), see, e.g., [94]. However, the full picture of the user landscape has not been captured, at least not in a form that could easily be analyzed by algorithms to discover unexploited linkages and unmatched needs. In the past, focus has been too much on writing reports and articles and not on getting the information on needs and requirements in a form available for machine-based analyses. The reports (e.g., the set of reports produced by US-09-01; see [33]) often disappear in shelves and are not really used in guiding the development of observing systems and knowledge services or in linking users and data.

Recent developments in unstructured databases allow for a far more flexible approach to data that represents a system of graphs. Advances in big data analysis and the availability of abundant information in social media, research networks, social communication channels, governmental and non-governmental Web sites, and online publications enables the machine-based construction of complex graphs that include, among others, also the decision and policy making processes and agents that depend on evidence and knowledge derived from Earth observations (EOs). Likewise, improvements in the presentation and analysis of graph data open new avenues for comprehensive user assessments and the detailed mapping of user landscapes. Importantly, the theory for the analysis of graph data is fully developed (see Section 2.3) and provides a powerful tool for those who need to explore the landscape in order to identify and engage with users, discover gaps, and improve the services they provide to better meet the needs of the users. Utilizing these recent developments, efforts have been made to utilize large Web-based knowledge sources to develop new avenues for access to data sources. For example, knowledge has been extracted from Wikipedia to link this knowledge to data by [98]. Other efforts aim at unifying the access to knowledge, see, e.g., [99]. The Linked Open Data Cloud (LODC) provides an opportunity to publish data and integrate it into a graph connecting data across many domains [47].

Despite the many efforts to improve access and usability of Earth observations (EOs), to increase knowledge of information needs, and to link users better to available data and knowledge resources, the current techniques available to Earth scientists and other users to discover and access data are still at a very low level with respect to comprehensive discovery, easy access, options for feedback, etc. The separation of passive metadata from the actual data often leads to incomplete metadata with crucial information missing. This has major impacts on provenance and reproducibility of research. Data citation is also impacted by incomplete metadata, see, e.g., [100]. What appears necessary is a fundamental transformation, a “Gestalt shift”, in the view of how data and users should interact [5]. The DAS concept could provide for this transformation.

The overall DAS concept is fully developed (see Figure 2). The knowledge base builds on the Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB). The SEE-IN KB is being developed as a knowledge base to construct, store, present, and analyze complex systems of graphs. It is populated with graphs fully capturing the stakeholder landscapes for societally relevant themes. It provides the means to explore the graphs to discover connectivity and to identify gaps in terms of unmatched linkages. The current version of the Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB) is at a prototype level with respect to storing graph-data. In most approaches to graph data, the concept of triples is used, where a triple

consists of two nodes (a subject and an object) and a link or predicate connecting these two nodes (e.g., the Resource Description Framework (RDF), see [101]). In a number of approaches, the nodes carry information on the links (in and out links) they are attached to (an example is the “Oracle Big Data Spatial and Graph” package; see, e.g., [102]). The Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB) does not include information on links with the nodes. Moreover, links are not necessarily directional. This generalized graph data model provides on the one hand for more flexibility and on the other hand requires more analytical skills to extract relationships from the graph data.

The artificial intelligence (AI)-based construction of graphs applying deep search and deep learning methodology is at the beginning. The currently available knowledge base design specification and architecture description needs to be further detailed and discussed with relevant communities, including the providers of GEOSS infrastructure. Comments from experts communities, including the private sector, will be crucial for the full conceptual development.

The concept for ISDAs is developed in terms of functionality and desired capabilities. Conceptually, the ISDAs are fully software agents that represent specific datasets and have the authority to answer queries from potential users and to negotiate with interested users conditions of data access and use. The ISDAs could be designed similar to Web servers giving access to information about extended metadata and contents of datasets, as well as derived attributes. Among others, an ISDA can provide the full dataset it represents or subsets of it in a user-requested format, can give access to tools used to process the data, and answer questions that require certain processing of the data. The ISDAs can access and query the graph data in the knowledge base to discover potential users and contact these users with promotional information about their dataset. They also can collect feedback from users of their datasets and provide this feedback to the knowledge base. The machinery the ISDAs could initially work on is the Web. For example, a dedicated main domain could ensure easily recognizable URLs and enhanced browsers could facilitate the communication between ISDAs and humans. In a later stage, a new framework for the world of the ISDAs could be created. The ecosystem of the ISDAs would be a core part of the digital ecosystem for the environment and the planet envisioned by [5,92].

The main advantage of the DAS concept is the fact that ISDAs are local agents associated with the data products where these data products exist. Thus, the need to publish data in archives or repositories, to develop large catalogs of datasets, etc., would be much reduced or disappear.

The interaction platform is conceptually developed in terms of the software and human actors and the documentation of interactions to ensure provenance. It will provide a matching and recording framework, where users and ISDAs can interact in promotion of datasets and in transactions that can lead to data modifications (e.g., for data providers) and use of data (for users). The platform could utilize the blockchain concept to ensure provenance, but a key question to be researched is the trade-off between the amount of resources needed and the level of security, perseverance, and documentation achieved needs to be carried out to inform the design of the interaction platform (see Section 2.4).

3.2. Validation through Case Studies

Detailed case studies addressing societal problems in a transdisciplinary approach could provide validation of the DAS concept. Initially, focus should be on broad scientific communities that depend heavily on Earth observations (EOs) and are researching societally relevant problems. Most of the problems related to sustainable development or developing sustainability are wicked problems [36] or super-wicked problems [103], and for most of these problems transdisciplinary collaborative approaches are most suited to address the problem [104].

Problems that appear to be ideal candidates for such case studies are within the Food-Water-Energy Nexus (FWEN). The FWEN provides an excellent example of interactions in a complex system of systems [7,38,44] with many potentially severe societal consequences [105]. In particular, a water crisis has been identified as a global catastrophic risk, see, e.g., [106]. Earth observations (EOs) are crucial to address the FWEN comprehensively, see, e.g., [107,108], and to

make progress towards the SDGs. The FWEN links sustainability of water use to almost all of humanity's activities. Achieving the 2030 Agenda for Sustainable Development [4] is conditioned by addressing the FWEN and making progress towards global food, water, and energy sustainability. The Sustainable Development Goals (SDGs) 2 (no hunger), 6 (clean water), and 7 (affordable and clean energy) are directly interdependent, while almost all other SDGs are impacted or are impacting the sustainability within these three domains. This makes the landscape of users depending on knowledge of the state and trends in the planetary physiology including the water, nitrogen and phosphorus cycles a very complex one. Diagnosing the time and spatial patterns of problems and co-developing and validating solutions for food, water and energy-related problems constitutes a suite of wicked problems. Addressing these issues requires access to comprehensive data, and the need for increased cross-domain data sharing has been emphasized within the relevant domains e.g., [109]. Likewise, building capacity to use the available cross-domain knowledge for decision and policy making and management of the relevant cycles in the planetary physiology is a complex task that needs to use many different avenues to engage with users in their activities. Comprehensive knowledge of the landscape of stakeholder, decision makers, and knowledge providers engaged in sustainability in a form that supports matchmaking, collaboration, and participatory activities is a prerequisite for identifying problems as well as providing evidence and knowledge to those who need this, and to build capacity.

The goal of such case studies would be to improve the understanding of the relationship between the FWEN and modern global change, including modern climate change, changes in the nitrogen and phosphorus cycles, and loss of biodiversity, and to develop transformative interventions that could change the trajectory of the underlying system towards desirable futures. The knowledge base would be used to construct the graph data relevant for research and user communities related to these challenges and to construct a data Web of relevant datasets. ISDAs for these datasets would be trained and would interact with researchers in the participating communities to discover and access data products. The ISDAs would also promote data products to potential users. Feedback collected from those participating in the use cases would provide a basis to validate and improve the DAS components. The communities that ideally should be involved in this validation include, among others, the Group on Earth Observations (GEO) Initiative "Earth Observations in Service of the 2030 Agenda for Sustainable Development" (<http://eo4sdg.org/>), the GEO Water Cycle Community of Practice (http://www.earthobservations.org/wa_igwco.shtml), the Future Earth Sustainable Water Future Programme (<https://water-future.org/>), and the Sustainable Water-Energy-Food Nexus Working Group (http://water-future.org/working_groups/sustainable-w-e-f-nexus-working-group/).

3.3. Considerations for Implementation

To ensure broad acceptance and support for the transition from the DPO perception to a DAS perception, the design and implementation of the DAS concept should be further developed in a participatory modeling. The planning of a versatile, secure, efficient, and active system linking observations and users for the benefit of society constitutes a wicked problem, and participatory modeling could be the first step in a collaborative approach to this problem. Group on Earth Observations (GEO) could utilize its convening power to bring a wide range of stakeholders together for such a participatory modeling. Again, the FWEN and related SDGs could be the societal challenge for this participatory modeling effort to focus on.

As a result of this effort, the design specifications for the DAS concept would be further detailed, including a detailed description of the functionality. The architecture will have to consider distributed cloud-based elements and will most likely require modifications of the current graph data model. The current graph data model separates the graph information from the objects. In many other graph software implementation, objects carry part of the graph information, and it will have to be researched whether a complete separation of objects and links is desirable and feasible within legal constraints. The specification will include the description of the methodology for the construction, presentation and analysis of graph data as well as the functionality for user feedback collections. For the latter,

potential legal constraints will have to be assessed to ensure that the collection of user information is conform with legal requirements.

The concept for the ISDAs as representatives of datasets and products has to ensure that the ISDAs have semantic capabilities. A core research question to address is how rich the data description available to the ISDAs will have to be to enable these capabilities. The development of a genuine knowledge model that enables AI to reason and search is a necessity for the implementation of the DAS concept.

The specification of a communication protocols for the ISDAs is an important step towards implementation. The methodology for self-learning ISDAs can be based on deep learning methodology to increase their knowledge relevant to the data they represent as well as the potential and actual applications and users of the data. To some extent, the ISDAs could utilize crawlers to collect relevant information. It is anticipated that ISDAs will be initiated as minimal seeds and then grow into more adult ISDAs. A research question relates to the minimum capability of the seeds necessary for them to grow. Among others, the ISDAs will need limited data processing capabilities to extract rough datasets or statistical or average properties, and they will have “magnifying glasses” to allow users to zoom into large datasets. They also should have the capability to provide data in a format requested by a user. Thus, a user would not have to know anything about the details of how the data are actually stored in the original data archive.

The generic design specification and architecture of the virtual interaction platform for ISDAs and users requires careful considerations. In terms of interactions, the platform will support the capabilities of the ISDAs to respond to user queries, identify users and needs and to promote data accordingly or to suggest collaborations between users to users. For this, the ISDAs will need to utilize and analyze the graph data available in the knowledge base to assess where their data would be beneficial. The ISDAs will be able to provide access to data in various ways. Actual transactions could be recorded in a scheme derived from blockchains to ensure provenance of both the original data and derived products,

The knowledge base is currently implemented as an extension of the already existing Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB). The SEE-IN KB contains considerable graph data for several research areas including water cycle, geohazards, health, and air quality. The data model of the Socio-Economic and Environmental Information Needs Knowledge Base (SEE-IN KB) is specifically designed for graph data, and a methodology for graph construction based on deep search and deep learning approaches is being implemented.

The GEOSS Common Infrastructure provides access to a large number of datasets. It will be important to ensure that the three core elements of DAS can communicate with the GEOSS Common Infrastructure (GCI) to train ISDAs for relevant datasets and to allow access to the knowledge base, ISDAs and interaction platform through the GCI.

4. Conclusions

The amount, quality, and diversity of Earth observations (EOs) is rapidly increasing but exploitation of this extremely valuable resource is hampered by limited discoverability, lack of information on applicability, and insufficient capacity in extracting relevant information from this resource for knowledge creation. Most efforts to improve in all these aspects are incremental improvements of existing concepts. At the same time, as outlined in Section 1, humanity in the Anthropocene is challenged increasingly with global catastrophic risks while aiming for more sustainability. Assessing and addressing these risks requires comprehensive information on the biosphere, the humansphere and the impacts of the humansphere and technosphere on the biosphere.

In this situation, a transformational paradigm shift in the relationship between data and users is required. The transition from the DPO to a DAS perception could facilitate this “Gestalt shift” and would have far reaching transformational consequences. In particular, it is expected that this transition would provide novel ways of integrating data into transdisciplinary approaches to wicked

problems discussed, e.g., by [110]. Implementing the United Nations' 2030 Agenda for Sustainable Development [4] poses many wicked problems to society, and most of the seventeen Sustainable Development Goals (SDGs) detailed in the agenda have all the additional properties of super-wicked problems identified by [103]. In particular, for most of the Sustainable Development Goals (SDGs), there is no central authority for the implementation, time is running out, and those who are causing the challenge are now attempting to solve the problem. For the validation of the DAS concept, use cases can be built around selected wicked problems associated with the implementation of the Sustainable Development Goals (SDGs).

The implementation of the DAS concept requires a major community effort and GEO could use its convening power to bring together selected communities for pilot projects aiming at the further development and validation of the DAS concept. A specific use cases of interest would be the Food-Water-Energy Nexus (FWEN) and the related Sustainable Development Goals (SDGs) 2 (no hunger), 6 (clean water), and 7 (clean energy). A DAS-related use case would aim at understanding the relationship between the FWEN and modern global change, including modern climate change, changes in the nitrogen and phosphorus cycles, and loss of biodiversity.

Author Contributions: The authors contributed equally to all sections.

Funding: The authors would like to acknowledge the European Union "Horizon 2020 Program" that funded the ConnectinGEO (Grant Agreement no. 641538) projects. Part of the work for one author (Plag) was conducted under NASA grant 80NSSC17K0241.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	artificial intelligence
CDP	Customers Discover Products
DAS	Data as Active Subjects
DDU	Data Discover Users
DFS	Depth-first search
DPO	Data as Passive Objects
EO	Earth observation
FWEN	Food-Water-Energy Nexus
GCI	GEOSS Common Infrastructure
GDPR	General Data Protection Regulation
GEO	Group on Earth Observations
GEOSS	Global Earth Observation System of Systems
IGOS	Integrated Global Observing Strategy
IGOS-P	Integrated Global Observing Strategy Partnership
IAEG-SDGs	Inter-Agency and Expert Group on SDG Indicators
ISDA	Intelligent Semantic Data Agent
LODC	Linked Open Data Cloud
PDC	Products Discover Customers
RDF	Resource Description Framework
SDG	Sustainable Development Goal
SEE-IN KB	Socio-Economic and Environmental Information Needs Knowledge Base
SIDS	Small Island Developing States
UDD	Users Discover Data
UNFCCC	United Nations Framework Convention on Climate Change
UNSC	United Nations Statistical Commission
URR	User Requirements Registry

References

- Harris, R.; Miller, L. Earth observation and the public good. *Space Policy* **2011**, *27*, 194–201. [CrossRef]
- Cotton-Barratt, O.; Farquhar, S.; Halstead, J.; Schubert, S.; Snyder-Beattie, A. *Global Catastrophic Risks 2016*; Technical Report; Global Challenge Foundation, Global Priorities Project: Stockholm, Sweden; Oxford, UK, 2016.
- World Economic Forum. *Global Risks 2019*, 14th ed.; Technical Report; World Economic Forum: Geneva, Switzerland, 2019.
- United Nations. *Transforming our World: The 2030 Agenda for Sustainable Development*; Technical Report A/RES/70/1; United Nations: New York, NY, USA, 2015.
- Campbell, J.; Jensen, D.E. *The Promise and Peril of a Digital Ecosystem for the Planet*; Technical Report; United Nations Environment Programme: Nairobi, Kenya, 2019. Available online: https://medium.com/@davidedjensen_99356/building-a-digital-ecosystem-for-the-planet-557c41225dc2 (accessed on 25 September 2019).
- Ryan, B. Open data for Sustainable Development. *Geospatial World*, 14 August 2016.
- Jules-Plag, S.; Plag, H.P. Supporting Agenda 2030's Sustainable Development Goals—Agend-Based Models and GeoDesign. *ApoGeoSpatial* **2016**, *31*, 24–30.
- Taylor, G. *Evolution's Edge—The Coming Collapse and Transformation of our World*; New Society Publishers: Gabriola Island, BC, Canada, 2008.
- Baum, S.D.; Handoh, I.C. Integrating the planetary boundaries and global catastrophic risk paradigms. *Ecol. Econ.* **2014**, *107*, 13–21. [CrossRef]
- Keys, P.W.; Galaz, V.; Dyer, M.; Matthews, N.; Folke, C.; Nyström, M.; Cornell, S.E. Anthropocene risk. *Nat. Sustain.* **2019**, *2*, 667–673. [CrossRef]
- Wang, M.; Hu, C.; Barnes, B.B.; Mitchum, G.; Lapointe, B.; Montoya, J.P. The great Atlantic Sargassum belt. *Science* **2019**, *365*, 83–87. [CrossRef]
- Steffen, W.; Rockström, J.; Richardson, K.; Lenton, T.M.; Folke, C.; Liverman, D.; Summerhayes, C.P.; Barnosky, A.D.; Cornell, S.E.; Crucifix, M.; et al. Trajectories of the Earth System in the Anthropocene. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 8252–8259. [CrossRef]
- Rothman, D.H. Thresholds of catastrophe in the Earth system. *Sci. Adv.* **2017**, *3*, e1700906. [CrossRef]
- Baum, S.D. The far future argument for confronting catastrophic threats to humanity: Practical significance and alternatives. *Futures* **2015**, *72*, 86–96. [CrossRef]
- Barnosky, A.D.; Hadly, E.A.; Bascompte, J.; Berlow, E.L.; Brown, J.H.; Fortelius, M.; Getz, W.M.; Harte, J.; Hastings, A.; Marquet, P.A.; et al. Approaching a state shift in Earth's biosphere. *Nature* **2012**, *486*, 52–58. [CrossRef]
- Avin, S.; Wintle, B.C.; Weitzdörfer, J.; hÉigeartaigh, S.S.Ó.; Sutherland, W.J.; Rees, M.J. Classifying global catastrophic risks. *Futures* **2018**, *102*, 20–26. [CrossRef]
- Dahl, A.L. IGOS from the perspective of the global observing systems and their sponsors. In Proceedings of the 27-th International Symposium on Remote Sensing of Environment: Information for Sustainability, Tromsø, Norway, 8–12 June 1998; Norwegian Space Centre: Oslo, Norway, 1998; pp. 92–94.
- IGOS-P. *The Integrated Global Observing Strategy (IGOS) Partnership Process*; Technical Report, IGOS Partnership, 2003; IGOS Process Paper, Version of 19 March 2003; World Meteorological Organization: Geneva, Switzerland, 2003.
- IGOS-P Ocean Theme Team. *An Ocean Theme for the IGOS Partnership*; Technical Report, IGOS Integrated Global Observing Strategy; NASA: Washington, DC, USA, 2001.
- Lawford, R.; The Water Theme Team. *A Global Water Cycle Theme for the IGOS Partnership*; Technical Report, IGOS Integrated Global Observing Strategy, 2004; Report of the Global Water Cycle Theme Team, April 2004; ESA Publications Division: Noordwijk, The Netherlands, 2004.
- Marsh, S.; The Geohazards Theme Team. *Geohazards Theme Report*; Technical Report, IGOS Integrated Global Observing Strategy; BRGM: Orleans, France, 2004.
- Townshend, J.R.; The IGOL Writing Team. *Integrated Global Observations of the Land: A Proposed Theme to the IGOS Partnership—Version 2*; Technical Report, IGOS Integrated Global Observing Strategy, 2004; Proposal Prepared by the IGOL Proposal Team, May 2004; FAO: Rome, Italy, 2004.
- IGOS. *A Coastal Theme for the IGOS Partnership—For the Monitoring of our Environment from Space and from Earth*; IOC Information Document No. 1220; UNESCO: Paris, France 2006; 60p.

24. United Nations Sustainable Development. In Proceedings of the AGENDA 21, United Nations Conference on Environment & Development, Rio de Janeiro, Brazil, 3–14 June 1992; Technical Report; United Nations: New York, NY, USA, 1992. Available online: <http://sustainabledevelopment.un.org/content/documents/Agenda21.pdf> (accessed on 15 August 2019).
25. GEO. *Global Earth Observing System of Systems GEOSS—10-Year Implementation Plan Reference Document*; Technical Report GEO 1000R, Group on Earth Observations; ESA Publications Division: Noordwijk, The Netherlands, 2005. Available online: <http://earthobservations.org> (accessed on 10 August 2019).
26. LeCozannet, G.; Salichon, J. *Geohazards Earth Observation Requirements*; Technical Report BRGM/RP-55719-FR; BRGM: Orléans, France, 2007.
27. Zell, E.; Huff, A.K.; Carpenter, A.T.; Friedl, L. A user-driven approach to determining critical earth observation priorities for societal benefit. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1594–1602. [[CrossRef](#)]
28. Plag, H.P.; Rizos, C.; Rothacher, M.; Neilan, R. The global geodetic observing system (GGOS): Detecting the fingerprints of global change in geodetic quantities. In *Advances in Earth Observation of Global Change*; Springer: Berlin, Germany, 2010.
29. Plag, H.P.; Ondich, G.; Kaufman, J.; Foley, G.; Pignatelli, F. The GEOSS User Requirement Registry: A Versatile Tool for the Dialog Between Users and Providers. In Proceedings of the 34th International Symposium on Remote Sensing of the Environment, Sydney, Australia, 10–15 April 2011.
30. Plag, H.P.; Foley, G.; Jules-Plag, S.; Kaufman, J.; Ondich, G. The GEOSS user requirement registry (URR): Linking users of GEOSS across disciplines and societal benefit areas. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium IEEE, Munich, Germany, 23–27 July 2012.
31. EAG. *Results-Oriented GEOSS: A Framework for Transforming Earth Observation Data to Knowledge for Decision Making*; Technical Report, Group on Earth Observation, Executive Committee; Report Prepared by the Expert Advisory Group for the 48th Meeting of the Executive Committee; Group on Earth Observation: Geneva, Switzerland 2019.
32. Plag, H.; The Workshop Participants. *Implementing and Monitoring the Sustainable Development Goals in the Caribbean: The Role of the Ocean, 2018, Saint Vincent, Saint Vincent and the Grenadines, 17–19 January 2018*; Technical Report; GEOSS Science and Technology Stakeholder Network (GSTSN): Rossbach, Germany, 2018. Available online: http://www.gstss.org/2018_Ocean_SDGs (accessed on 21 September 2019).
33. Group on Earth Observations. *Task US-09-01a: Critical Earth Observation Priorities*, 2nd ed.; Technical Report; Group on Earth Observations: Geneva, Switzerland, 2012. Available online: <http://sbageotask.larc.nasa.gov> (accessed on 15 July 2019).
34. Valuables. Resources for the Future. Available online: <https://www.rff.org/valuables/> (accessed on 15 August 2019).
35. UNISDR. *Sendai Framework for Disaster Risk Reduction 2015–2030*, 1st ed.; Technical Report UNISDR/GE/2015-ICLUX EN5000; UNISDR: Geneva, Switzerland, 2015. Available online: http://www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf (accessed on 15 July 2019).
36. Rittel, H.W.J.; Webber, M.W. Dilemmas in a general theory of planning. *Policy Sci.* **1973**, *4*, 155–169. [[CrossRef](#)]
37. UNRISD. *Policy Innovations for Transformative Change—Implementing the 2030 Agenda for Sustainable Development*; Unrisd Flagship Report 2016; United Nations Research Institute for Social Development: Geneva, Switzerland, 2016.
38. Nilsson, M.; Griggs, D.; Visbeck, M. Policy: Map the interactions between Sustainable Development Goals. *Nature* **2016**, *534*, 320–322. [[CrossRef](#)] [[PubMed](#)]
39. Griggs, D.J.; Nilsson, M.; Stevance, A.; McCollum, D. (Eds.) *A Guide to SDG Interactions: From Science to Implementation*; Technical Report; International Council for Science: Paris, France, 2017; doi:10.24948/2017.01. [[CrossRef](#)]
40. Singh, G.G.; Cisneros-Montemayor, A.M.; Swartz, W.; Cheung, W.; Guy, J.A.; Kenny, T.A.; McOwen, C.J.; Asch, R.; Geffert, J.L.; Wabnitz, C.C.; et al. A rapid assessment of co-benefits and trade-offs among Sustainable Development Goals. *Mar. Policy* **2018**, *93*, 223–231. [[CrossRef](#)]
41. Alcamo, J.; Chenje, M.; Ghai, A.; Keita-Ouane, F.; Leonard, S.A.; Niamir-Fuller, M.; Nobbe, C. *Embedding the Environment in Sustainable Development Goals*; UNEP Post-2015 Discussion Paper 1, Version 2; UNEP: Nairobi, Kenya, 2013.

42. Leadership Council of the Sustainable Development Solutions Network. *Indicators for Sustainable Development Goals*; Technical Report, Draft Report for Public Hearing; Sustainable Development Solutions Network of the United Nations: New York, NY, USA, 2014.
43. IAEG-SDGs. *Tier Classification for Global SDG Indicators—11 May 2018*; Technical Report; Intern-Agency Expert Group for SDG Indicators, United Nations: New York, NY, USA, 2018.
44. Jules-Plag, S.; Plag, H.P. Supporting the Implementation of SDGs. *Geospatial World*, 15 August 2016. Available online: <http://www.geospatialworld.net/article/supporting--implementation--sdgs/> (accessed on 10 July 2019).
45. Plag, H.P.; Jules-Plag, S.A. A Goal-Based Approach to the Identification of Essential Transformation Variables in Support of the Implementation of the 2030 Agenda for Sustainable Development. *Int. J. Digit. Earth* **2019**. doi:10.1080/17538947.2018.1561761. [CrossRef]
46. PANGAEA Team. PANGAEA. Data Publisher for Earth & Environmental Science. Available online: <https://pangaea.de> (accessed on 28 August 2019).
47. McCrae, J.P.; Abele, A.; Buitelaar, P.; Cyganiak, R.; Jentzsch, A.; Andryushechkin, V.; Debattista, J. The Linked Open Data Cloud. Available online: <https://www.lod-cloud.net/> (accessed on 27 August 2019).
48. Christodoulou, P.; Christodoulou, K.; Andreou, A.S. A real-time targeted recommender system for supermarkets. In *Proceedings of the 19th International Conference on Enterprise Information Systems— Volume 2, Porto, Portugal, 26–29 April 2017*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 703–712. [CrossRef]
49. The Performance Edge, Inc. Feedback Rewards—Guest Feedback and Rewards Program. Available online: <http://www.feedbackrewards.com/> (accessed on 27 August 2019).
50. Russell, S.J.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2003.
51. Weiss, G. *Multiagent Systems*, 2nd ed.; The MIT Press: Cambridge, MA, USA, 2013.
52. Plag, H.P. Implementing and Monitoring the Sustainable Development Goals in the Caribbean: The Role of the Ocean. Presented at the Meeting of the Steering Committee of the GEO Initiative “Ocean and Society: Blue Planet”, Saint Vincent, Saint Vincent and the Grenadines, 15 March 2018.
53. Stevenson, H. Emergence: The Gestalt Approach to Change. Available online: <http://www.clevelandconsultinggroup.com/articles/emergence-gestalt-approach-to-change.php> (accessed on 15 August 2019).
54. Dietz, J. *Enterprise Ontology - Theory and Methodology*; Springer: Berlin/Heidelberg, Germany, 2006.
55. Pawlak, Z. Rough sets. *Int. J. Parallel Program.* **1982**, *11*, 341–356. [CrossRef]
56. Bazan, J.; Szczuka, M.; Wojna, A.; Wojnarski, M. On the evolution of rough set exploration system. In *Proceedings of the RSCTC 2004, LNAI 3066, Uppsala, Sweden, 1–5 June 2004*; Tsumoto, S., Ed.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 592–601, doi:10.1007/978-3-540-25929-9_73. [CrossRef]
57. Ziarko, W. Rough sets as a methodology for data mining. In *Rough Sets in Knowledge Discovery 1: Methodology and Applications*; Polkowski, L., Skowron, A., Eds.; Physica-Verlag: Heidelberg, Germany, 1998; pp. 554–576.
58. Chen, H.; Li, T.; Luo, C.; Horng, S.J.; Wang, G. A decision-theoretic rough set approach for dynamic data mining. *IEEE Trans. Fuzzy Syst.* **2015**, *23*, 1958–1970. [CrossRef]
59. Neukom, R.; Barboza, L.A.; Erb, M.P.; Shi, F.; Emile-Geay, J.; Evans, M.N.; Franke, J.; Kaufman, D.S.; Lücke, L.; Rehfeld, K. Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era. *Nat. Geosci.* **2019**. [CrossRef]
60. Tarjan, R.E. Depth-first search and linear graph algorithms. *SIAM J. Comput.* **1972**, *1*, 146–160. [CrossRef]
61. Cormen, T.H.; Leiserson, C.E.; Rivest, R.L.; Stein, C. *Introduction to Algorithms*, 3rd ed.; The MIT Press: Cambridge, MA, USA, 2009.
62. Soman, J.; Narang, A. Fast community detection algorithm with GPUs and multicore architectures. In *Proceedings of the 2011 IEEE International Parallel and Distributed Processing Symposium*, Anchorage, AK, USA, 16–20 May 2011; IEEE Computer Society: Washington, DC, USA, 2011; doi:10.1109/IPDPS.2011.61. [CrossRef]
63. Adamic, L.A.; Adar, E. Friends and neighbors on the web. *Soc. Netw.* **2003**, *25*, 211–230. [CrossRef]
64. Newman, M.E.J. *Networks: An Introduction*; Oxford University Press: Oxford, UK, 2010.
65. Sniedovich, M. Dijkstra’s algorithm revisited: The dynamic programming connexion. *J. Control Cybern.* **2006**, *35*, 599–620.
66. Cook, J.; Lewandowsky, S. *The Debunking Handbook*; University of Queensland: St. Lucia, Australia, 2011.
67. Pennycook, G.; Cheyne, J.A.; Barr, N.; Koehler, D.J.; Fugelsang, J.A. On the reception and detection of pseudo-profound bullshit. *Judgm. Decis. Mak.* **2015**, *10*, 549–563.

68. Plag, H.P.; Adegoke, J.; Bruno, M.; Christian, R.; Digiaco, P.; McManus, L.; Nicholls, R.; van de Wal, R. Observations as decision support for coastal management in response to local sea level changes. In *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Volume 2)*, Venice, Italy, 21–25 September 2009; Hall, J., Harrison, D.E., Stammer, D., Eds.; ESA: Paris, France, 2010; doi:10.5270/OceanObs09.cwp.69. [CrossRef]
69. Plag, H.P.; McCallum, I.; Fritz, S.; Jules-Plag, S.; Nyenhuis, M.; Nativi, S. The GEOSS Science and Technology Service Suite: Linking S&T Communities and GEOSS. *E3S Web Conf.* **2013**, *1*, 28003. doi:10.1051/e3sconf/20130128003. [CrossRef]
70. Michalski, R.S.; Carbonell, J.G.; Mitchell, T.M., Eds. *Machine Learning: An Artificial Intelligence Approach*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.
71. Hartuv, E.; Schmitt, A.O.; Lange, J.; Meier-Ewert, S.; Lehrach, H.; Shamir, R. An algorithm for clustering cDNA fingerprints. *Genomics* **2000**, *66*, 249–256. [CrossRef] [PubMed]
72. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing, Springer Topics in Signal Processing 2*; Springer: Berlin/Heidelberg, Germany, 2009; doi:10.1007/978-3-642-00296-0_5. [CrossRef]
73. Hameed, M.A.; Al Jadaan, O.; Ramachandram, S. Collaborative filtering based recommendation system: A survey. *Int. J. Comput. Sci. Eng.* **2012**, *4*, 859–876.
74. Linden, G.; Smith, B.; York, J. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet Comput.* **2003**, *7*, 76–80. [CrossRef]
75. Houeland, T.G. An efficient random decision tree algorithm for case-based reasoning systems. In *Proceedings of the FLAIRS 24th International Florida Artificial Intelligence Research Society Conference*, Palm Beach, FL, USA, 18–20 May 2011; AAAI Press: Menlo Park, CA, USA, 2011.
76. Dalal, S.; Athavale, D.V.; Jindal, K. Case retrieval optimization of case-based reasoning through knowledge-intensive similarity measures. *Int. J. Comput. Appl.* **2011**, *34*, 12–18.
77. Larochelle, H.; Bengio, Y. Classification using discriminative Restricted Boltzmann Machines. In *Proceedings of the 25th International Conference on Machine Learning*, Helsinki, Finland, 5–9 July 2008; ACM: New York, NY, USA, 2008; pp. 536–543. [CrossRef]
78. Lee, H.; Grosse, R.; Ranganath, R.; Ng, A.Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *Proceedings of the 26th International Conference on Machine Learning*, Montreal, QC, Canada, 14–18 June 2009; ACM: New York, NY, USA, 2009; pp. 609–616.
79. Wang, H.; Wang, N.; Yeung, D.Y. Collaborative Deep Learning for Recommender Systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Sydney, NSW, Australia, 10–13 August 2015; ACM: New York, NY, USA, 2015; pp. 1235–1244. [CrossRef]
80. Van den Oord, A.; Dieleman, S.; Schrauwen, B. Deep content-based music recommendation. In *Advances in Neural Information Processing Systems 26*; Burges, C.J.C., Bottou, L., Welling, M., Ghahramani, Z., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2013; pp. 2643–2651.
81. Forum, W.E. *Realizing the Potential of Blockchain—A Multistakeholder Approach to the Stewardship of Blockchain and Cryptocurrencies*; Technical Report; World Economic Forum: Davos, Switzerland, 2017. Available online: http://www.weforum.org/docs/WEF_Realizing_Potential_Blockchain.pdf (accessed on 13 September 2019).
82. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System, 2009. Available online: metzdowd.com (accessed on 10 February 2018).
83. Swan, M. *Blockchain: Blueprint for a New Economy*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2015.
84. Van Rijmenam, M. The Top 11 Blockchains for Enterprise Organisations, and Why. Available online: <https://vanrijmenam.nl/11-blockchains-enterprise-organisations-why/> (accessed on 13 September 2019).
85. Schneier, B. There's No Good Reason to Trust Blockchain Technology, 2019. *Wired*. Available online: <https://www.wired.com/story/theres-no-good-reason-to-trust-blockchain-technology/> (accessed on 13 September 2019).
86. Hijgenaar, S. Not All Blockchains are Created Equal When It Comes to Energy Consumption. Available online: <https://www.cgi.com/canada/en/blog/utilities/not-all-blockchains-are-equal-when-it-comes-to-energy-consumption> (accessed on 13 September 2019).
87. Matthews, K. 4 Ways to Counter Blockchain's Energy Consumption Pitfall. <https://www.greenbiz.com/article/4-ways-counter-blockchains-energy-consumption-pitfall> (accessed on 16 August 2019).

88. Boldrini, E.; Craglia, M.; Mazzetti, P.; Nativi, S. The brokering approach for enabling collaborative scientific research. In *Collaborative Knowledge in Scientific Research Networks*; Diviacco, P., Fox, P., Pshenichny, C., Leadbetter, A., Eds.; IGI Global: Hershey, PA, USA, 2015; pp. 283–304. [CrossRef]
89. Hsu, L.; Mayorga, E.; Horsburgh, J.; Carter, M.; Lehnert, K.; Brantley, S. Enhancing Interoperability and Capabilities of Earth Science Data using the Observations Data Model 2 (ODM2). *Data Sci. J.* **2017**, *16*. [CrossRef]
90. Hu, Y.; Janowicz, K.; Prasad, S.; Gao, S. Metadata Topic Harmonization and Semantic Search for Linked-Data-Driven Geoportals: A Case Study Using ArcGIS. *Trans. GIS* **2015**, *19*, 398–416. [CrossRef]
91. Khalsa, S.J.S. Data and Metadata Brokering—Theory and Practice from the BCube Project. *Data Sci. J.* **2017**, *16*. [CrossRef]
92. Campbell, J.; Jensen, D.E. *Could a Digital Ecosystem for the Environment Have the Potential to Save the Planet?*; Technical Report; National Council for Science and the Environment: Washington, DC, USA, 2019. Available online: <https://science.nasa.gov/national-council-science-and-environment-ncse-2019> (accessed on 25 September 2019).
93. Barrie, L.A.; The IGACO Writing Team. *An integrated Global Atmospheric Chemistry Observation Theme for the IGOS Partnership*; Technical Report, IGOS Integrated Global Observing Strategy; WMO: Geneva, Switzerland, 2004.
94. Unninayar, S.; Task Team. *GEO Task US-09-01a: Critical Earth Observations Priorities—Water Societal Benefit Area*; Technical Report; Group on Earth Observations—User Interface Committee: Geneva, Switzerland, 2016.
95. Plag, H.P.; Ondich, G.; Kaufman, J.; Foley, G. The GEOSS User Requirement Registry—Supporting a User-Driven Global Earth Observation System of Systems. *Imaging Notes* **2010**, *25*, 28–33.
96. Plag, H.P.; Jules-Plag, S.; Callaghan, C.; McCallum, I. Linking science and technology communities to GEOSS. In *Towards a Sustainable GEOSS (Global Earth Observation System of Systems)—Some Results of the EGIDA Project*; Nativi, S., Mazzetti, P., Plag, H.P., Eds.; Aíon: Florence, Italy, 2013; pp. 13–34, ISBN 978-88-98262-05-2.
97. Yang, X.; Blower, J.D.; Bastin, L.; Lush, V.; Zabala, A.; Masó, J.; Cornford, D.; Díaz, P.; Lumsden, J. An integrated view of data quality in Earth observation. *Philos. Trans. A Math. Phys. Eng. Sci.* **2013**, *371*, 20120072. [CrossRef] [PubMed]
98. Lehmann, J.; Isele, R.; Jakob, M.; Jentzsch, A.; Kontokostas, D.; Mendes, P.N.; Hellmann, S.; Morsey, M.; van Kleef, P.; Auer, S.; et al. DBpedia—A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semant. Web* **2012**, *6*, 167–195.
99. DBpedia Team. DBpedia—Global and Unified Access to Knowledge. Available online: <https://wiki.dbpedia.org/> (accessed on 21 September 2019).
100. McCallum, I.; P. Plag, H.; Fritz, S. Data Citation Standard: A Means to Support Data Sharing, Attribution, and Traceability. *E3S Web Conf.* **2013**, *1*, 28002, doi:10.1051/e3sconf/20130128002. [CrossRef]
101. W3C. *RDF 1.1 Concepts and Abstract Syntax*; Technical Report; W3C: Keio, Japan, 2014. Available online: <https://www.w3.org/TR/rdf11-concepts/> (accessed on 6 June 2019).
102. Oracle. *Oracle Big Data Spatial and Graph—Property Graph: Features and Performance*; Technical Report, ORACLE Technical Whitepaper; Oracle: Redwood City, CA, USA, 2017.
103. Levin, K.; Cashore, B.; Bernstein, S.; Auld, G. Overcoming the tragedy of super wicked problems: Constraining our future selves to ameliorate global climate change. *Policy Sci.* **2012**, *45*, 123–152. [CrossRef]
104. Roberts, N. Wicked Problems and Network Approaches to Resolution. *Int. Public Manag. Rev.* **2000**, *1*, 1–19.
105. Obersteiner, M.; Walsh, B.; Frank, S.; Havlík, P.; Cantele, M.; Liu, J.; Palazzo, A.; Herrero, M.; Lu, Y.; Mosnier, A.; et al. Assessing the land resource–food price nexus of the Sustainable Development Goals. *Sci. Adv.* **2016**, *2*, 10.1126/sciadv.1501499. [CrossRef]
106. World Economic Forum. *Global Risks 2016*, 11th ed.; Technical Report; World Economic Forum: Geneva, Switzerland, 2016.
107. García, L.E.; Rodríguez, D.J.; Wijnen, M.; Pakulski, I., Eds. *Earth Observation for Water Resources Management: Current Use and Future Opportunities for the Water Sector*; World Bank Group: Washington, DC, USA, 2016; doi:10.1596/978-1-4648-0475-5. [CrossRef]
108. Keskinen, M.; Someth, P.; Salmivaara, A.; Kumm, M. Water-Energy-Food Nexus in a Transboundary River Basin: The Case of Tonle Sap Lake, Mekong River Basin. *Water* **2015**, *7*, 5416–5436. [CrossRef]

109. Lehmann, A.; Giuliani, G.; Ray, N.; Rahman, K.; Abbaspour, K.C.; Nativi, S.; Craglia, M.; Cripe, D.; Quevauviller, P.; Beniston, M. Reviewing innovative Earth observation solutions for filling science-policy gaps in hydrology. *J. Hydrol.* **2014**, *518*, 267–277. [[CrossRef](#)]
110. Brown, V.A.; Harris, J.A.; Russell, J.Y. (Eds.) *Tackling Wicked Problems—Through the Transdisciplinary Imagination*; Earthscan: New York, NY, USA, 2010.



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