



Proceedings of the 2021 conference on Big Data from Space (BiDS'21)

---From Insights to Foresight---

18-20 May 2021

Edited by P. Soille, S. Loekken, and S. Albani



EUROPEAN UNION
SATELLITE CENTRE
Analysis for decision making



This publication is a conference proceedings published by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information: Pierre.Soille at ec.europa.eu

EU Science Hub
<https://ec.europa.eu/jrc>

JRC125131

EUR 30697 EN

PDF ISBN 978-92-76-37661-3 ISSN 1831-9424 [doi:10.2760/125905](https://doi.org/10.2760/125905)

Luxembourg: Publications Office of the European Union, 2021

© European Union, 2021



The reuse policy of the European Commission is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content © European Union, 2021

How to cite this report: *Proceedings of the 2021 conference on Big Data from Space*, Soille, P., Loekken, S. and Albani, S., eds., EUR 30697 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-37661-3, [doi:10.2760/125905](https://doi.org/10.2760/125905), JRC125131.

Preface

While data have always been at the source of any scientific or technical endeavour, the term Big Data has gained momentum at the beginning of this millennium given the vast amounts of digital data generated from ever increasing sources ranging from individuals and the internet of things to sophisticated sensors in laboratories, on the ground, and in space. With an initial focus on addressing technical aspects such as handling data volume, velocity and variety, Big Data is now more about our capacity to extract meaningful information (insights) from multi-source data, and more recently also deriving the *foresight* to inform and support decisions that will shape our future.

In this context, *Big Data from Space* refers to the massive spatio-temporal Earth and Space observation data collected by a variety of ground-based & space-borne sensors and the synergy with data coming from other sources and communities. The continuous growth of Big Data from Space is motivated by the need for answering major societal challenges related to the impact of human activities on our planet in the case of Earth Observation, or fundamental questions related to the origin of our Universe in the case of space observation. While the growth of data volume, velocity, and variety is matched by technological advances related to sensors as well as information and communication technologies, the extraction of insights from the generated data is enabled by breakthroughs in data science and in particular thanks to recent progress in artificial intelligence. These developments are empowering new approaches and applications in various and diverse domains influencing life on Earth and societal aspects, from sensing cities, monitoring human settlements and urban areas to climate change, sustainable development goals, and security.

The main objectives of the 2021 Big Data from Space conference (BiDS'21) are:

- Bring to the scene new user needs and requirements related to the use of large amounts and varieties of data in different space domains such as Earth Observation (e.g., EU Copernicus programme), Space Science, Navigation and Telecommunications (e.g., EU Space programmes as Galileo and EU GovSatCom), mission operations and system engineering;
- Bring together major European actors in the fields of Space and data technologies, including research, industry, institutions, and users, to strengthen the communication and transfer of requirements, methods and technologies, and to reinforce an interdisciplinary approach;
- Explore and expand the ever increasing relevance of Big Data in European and global environmental policy initiatives and programmes, and the corresponding increasing complexity of applications and use cases;

- Discover and foster breakthrough data science processing and analysis techniques to extract insights and generate foresight, showing use cases wherever possible to facilitate future user uptake;
- Focus on new paradigms of data science addressing the entire value chain, i.e., building of reference training sets, data processing to extract information, information analysis to gather knowledge, and knowledge transformation in foresight;
- Maximise the uptake and impact of solutions exploiting multi-source spatio-temporal data linked with other data sources;
- Advance the upscale of new solutions from Research and Innovation (R&I) to operational use (e.g., for the security domain and informed policy making);
- Foster interoperability of platforms and services by promoting open standards, analysis ready data, and Application Programming Interfaces (APIs);
- Promote interdisciplinarity to respond to multi-sectorial challenges such as those put forward by the European Green Deal or the wide-ranging consequences of the Covid-19 pandemic;
- Promote cross-fertilisation with similar activities in other data intensive domains (e.g., high-energy physics, genomics, social media, internet of things, etc.).

The BiDS conference series is co-organised by the European Space Agency (ESA), the Joint Research Centre (JRC) of the European Commission, and the European Union Satellite Centre (SatCen). BiDS'21 emphasises not only on the insights that can be retrieved from Big Data from Space but also on the exploitation of these insights for foresight to improve our capacity to detect trends and model future evolution. This capacity is becoming increasingly important given the pace at which our World is changing. This is exemplified and reflected by the EU Destination Earth (DestinE) initiative and the related digital twin of the Earth. The objective of DestinE is to develop a very high precision digital model of the Earth to monitor and simulate natural and human activity, and to develop and test scenarios that would enable more sustainable development and support European environmental policies. The provision of more reliable scenarios of future evolution under different boundary conditions requires us to improve our understanding of Earth's dynamic systems besides their monitoring. Similarly to past editions of this conference, the 2021 edition provides a snapshot of the different research and innovation developments in the field of Big Data from Space including technical aspects and applications.

These proceedings contain the papers presented at the on-line BiDS'21 conference held on May 18-20 as an on-line conference. From a total of 63 submissions, reviewed in average by 3 programme committee members, 47 papers were accepted: 30 as oral presentations and the remaining 17 as poster presentations. Further to these presentations, the conference featured 5 keynote lectures from distinguished speakers that enlightened the audience with their experience in areas relevant to Big Data from Space:

1. *Digital Twins of the Earth System = Really Big Data*
by Peter Bauer
(European Centre for Medium-Range Weather Forecasts, UK)
2. *From Interactive Computing to Collaborative Science: Opportunities in the Cloud with Open Infrastructure*
by Fernando Pérez
(Berkeley Institute for Data Science, USA)
3. *Earth Observation + Machine Learning + System Modelling to Understand the Earth System*
by Markus Reichstein (Max Planck Institute for Biogeochemistry, Germany)
4. *Space: the Quantum Frontier*
by Radu Ionicioiu
(Horia Hulubei National Institute for R&D in Physics and Nuclear Engineering, Romania)
5. *Big Data Astronomy: Challenges and Opportunities*
by Leanne Guy
(AURA/Rubin Observatory, USA)

BiDS'21 was initially scheduled to be hosted at the Polytechnic University of Bucharest, Romania, with the support of the Romanian Space Agency (ROSA). This venue was announced during the closing session of BiDS'19 organised in November 2019. No one could have imagined the disruption that we would all have to face just 3 months later with the covid-19 pandemic. Accordingly, the BiDS'21 organising committee took the decision to move the conference to an on-line event during its first meeting in May 2020. This was not an easy decision given the enthusiasm from our colleagues from Romania to host the conference and promote Big Data from Space in Romania, particularly among students and young researchers. Retrospectively, this was the right decision as most of us are still limited in travelling not to speak of gathering at a conference that attracted 700 attendees in 2019! Each conference format has its own advantages and drawbacks. The creation of new contacts and sparkling new ideas and collaborations during coffee break, lunches, and dinners have for sure been very much missed. On the plus side, all on-line presentations including the recording of the keynote lectures not included in these proceedings can be accessed through the conference website: www.bigdatafromspace2021.org.

Great thanks go to all authors and presenters of BiDS'21 as well as the numerous participants (518 registrations from more than 54 different countries one week ahead of the conference start). Together, they have ensured the success of the 2021 conference on Big Data from Space. Special thanks go to the Programme Committee members who have guaranteed the quality of the conference programme and these proceedings.

Pierre Soille, Sveinung Loekken, and Sergio Albani

ARTIFICIAL INTELLIGENCE AND BIG DATA TECHNOLOGIES FOR COPERNICUS DATA: THE EXTREMEEARTH PROJECT

Manolis Koubarakis¹, George Stamoulis¹, Dimitris Bilidas¹, Theofilos Ioannidis¹, George Mandilaras¹, Despina-Athanasia Pantazi¹, George Papadakis¹, Vladimir Vlassov², Amir H. Payberah², Tianze Wang², Sina Sheikholeslami², Desta Haileselassie Hagos², Lorenzo Bruzzone³, Claudia Paris³, Giulio Weikmann³, Daniele Marinelli³, Torbjørn Eltoft⁴, Andrea Marinoni⁴, Thomas Kræmer⁴, Salman Khaleghian⁴, Habib Ullah⁴, Antonis Troumpoukis⁵, Nefeli Prokopaki Kostopoulou⁵, Stasinou Konstantopoulos⁵, Vangelis Karkaletsis⁵, Jim Dowling^{2,6}, Theofilos Kakantousis⁶, Mihai Datcu⁷, Wei Yao⁷, Corneliu Octavian Dumitru⁷, Florian Appel⁸, Silke Migdall⁸, Markus Muerth⁸, Heike Bach⁸, Nick Hughes⁹, Alistair Everett⁹, Åshild Kierbech⁹, Joakim Lillehaug Pedersen⁹, David Arthurs¹⁰, Andrew Fleming¹¹, Andreas Cziferszky¹¹

¹ National and Kapodistrian University of Athens ² KTH Royal Institute of Technology, Stockholm ³ University of Trento

⁴ UiT The Arctic University of Norway ⁵ National Center for Scientific Research - Demokritos ⁶ Logical Clocks AB

⁷ German Aerospace Center (DLR) ⁸ VISTA Remote Sensing in Geosciences GmbH ⁹ Norwegian Meteorological Institute

¹⁰ Polar View ¹¹ British Antarctic Survey

ABSTRACT

ExtremeEarth is a three-year H2020 ICT research and innovation project. Its main objective is to develop Artificial Intelligence and big data technologies that scale to the large volumes of big Copernicus data, information and knowledge, and apply these technologies in two of the European Space Agency (ESA) Thematic Exploitation Platforms (TEP): Food Security and Polar.

Index Terms— ExtremeEarth, Earth Observation, Linked Geospatial Data, Artificial Intelligence, Deep Learning, Copernicus, Food Security, Polar Regions

1. INTRODUCTION

Copernicus data is a paradigmatic case of big data giving rise to all relevant challenges, the so-called 5-Vs: volume, velocity, variety, veracity, and value, as it is documented in recent reports, such as the 2019 Copernicus Sentinel Data Access Report and the Copernicus Market Report of the same year. Copernicus data today is freely available not only through the Copernicus Open Access Hub but also through the five Data and Information Access Services (DIAS), where computing power is also available close to the data. Some related facilities of the Earth Observation (EO) ecosystem in Europe are the Thematic Exploitation Platforms (TEPs) of the European Space Agency (ESA), which enable user communities to collaborate using a virtual workspace where EO data, non-EO data, tools, and computing power are available. Today most of the TEPs run on a DIAS (e.g., the Food Security and Polar TEPs run on CREODIAS).

This work is supported by the ExtremeEarth project funded by European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 825258.

ExtremeEarth¹ is positioned in this prosperous European EO ecosystem and has three objectives: (i) extracting information and knowledge from big Copernicus data using scalable algorithms, (ii) managing this information and knowledge efficiently, and (iii) integrating it with other data sources to develop demo applications of economic, environmental and societal value.

ExtremeEarth is currently in its final year. Its main achievements so far are the following: (i) two implemented use cases focusing on Food Security and the Polar Regions, (ii) new deep learning architectures for crop type mapping in the context of the Food Security use case, (iii) new deep learning architectures for sea ice mapping in the context of the Polar use case, (iv) the development and open publication of very large datasets for training the deep architectures, (v) scalable semantic technologies for managing, as big linked geospatial data, the information and knowledge extracted from Copernicus data, and (vi) the ExtremeEarth platform that brings all the above technologies together and is used to implement the two use cases.

The rest of the paper presents the above contributions.

2. THE FOOD SECURITY USE CASE

Food Security is a very challenging issue of this century, especially given the changing Earth environment. Irrigation is an important dimension of it requiring reliable water resources either from ground water or from surface water. A large portion of fresh water is linked to snowfall, snow/ice storage and seasonal release. Therefore, water availability maps are an important EO-based product that can support farmers in decision making and irrigation information management.

¹<http://earthanalytics.eu/>

The goal of the Food Security use case of ExtremeEarth is to *develop high resolution water availability maps* for agricultural areas, allowing a new level of detail for wide-scale irrigation support for farmers [11]. The Danube river basin is the area where the results of the use case have been demonstrated so far. This area was selected for the following reasons: (i) variability in water supply due to changing precipitation patterns leading to extremes events (floods and droughts), (ii) significant portion of irrigated agriculture, (iii) significant water supply from water storage by snow in the Alps, (iv) large interest of demo users, and (v) strong economic, environmental and societal value.

The first stage of this use case was the collection of user requirements during a workshop which was organized by VISTA in Munich in March 2019. The user requirements drove the design and implementation of the Food Security use case which is shown graphically in Figure 1.

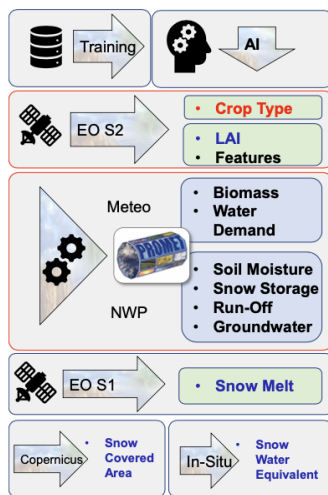


Fig. 1: The Food Security use case

The implementation of the use case draws on the following information: (i) crop type and leaf area index computed using Sentinel-2 images, (ii) biomass, water demand, soil moisture, snow storage, snow run off and groundwater computed using the proprietary land surface modelling software PROMET of VISTA, (iii) snowmelt from Sentinel-1 data, (iv) snow cover products from the Copernicus CryoLand service, and (v) snow water equivalent from in-situ sensors.

The outputs of the use case are field specific irrigation recommendations for specific demo applications in Austria, Hungary and Romania. These consist of recommendations regarding when and how much to irrigate, and yield forecasts with and without optimized irrigation plans.

The implementation of the processing chain of the Food Security use case has been done in the Food Security TEP using the ExtremeEarth platform (see Section 7 and [3]). The deep learning algorithms used for crop type mapping are discussed in Section 4. The semantic technologies that are used are discussed in Section 6.

3. THE POLAR USE CASE

The anticipated economic development of the Arctic, partially driven by reductions in sea ice cover, will see an increase in maritime shipping activity. High quality, timely and reliable information about sea ice and iceberg conditions is vital to ensure that vessels can navigate efficiently and safely with minimal risk to the environment. This information is required by vessels in many sectors, including cargo transport, fisheries, tourism, research vessels, resource exploration and extraction, destination shipping and national coast guard vessels.

The goal of the ExtremeEarth Polar use case is to *produce high resolution ice charts* from massive volumes of heterogeneous Copernicus data. The first stage of the use case was the collection of user requirements during the user workshop of March 2019. Two key technical requirements that resulted from this workshop were: (i) SAR data (Sentinel-1 and other third party missions) were considered the most reliable source of information for the use case, since they are already used widely for operational sea ice charting, and (ii) automatic products to be produced by ExtremeEarth had to maintain the high resolution of this data and the ice charts derived from it (300 meters or better). The technical requirements drove the design and implementation of the Polar use case which is shown graphically in Figure 2.

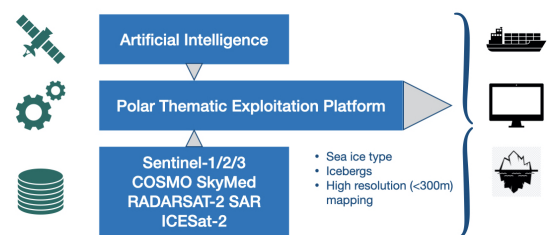


Fig. 2: The Polar use case

The implementation of the use case draws on the following information: (i) Level-1 Sentinel-1 images, (ii) training data compiled manually by expert ice analysts from a variety of sources including other satellite data such as Sentinel-2 and -3 visible and infrared optical, COSMO SkyMed and RADARSAT-2 SAR, and ICESat-2 sea ice freeboard, and in addition shipboard observations from Ice Watch².

The outputs of the use case are sea ice concentration and type maps, displaying stages of development (in accordance with the World Meteorological Organization Sea Ice Nomenclature), including fraction of leads and ridges, over the Polar Regions, at a resolution of 300 meters or better.

The implementation of the processing chain of the Polar use case has been done in the Polar TEP [2] using the ExtremeEarth platform (see Section 7 and [3]). The deep learning algorithms used for sea ice classification are discussed in Section 5. The semantic technologies that are used are discussed in Section 6.

²<https://icewatch.met.no>

4. DEEP LEARNING FOR CROP TYPE MAPPING

The determination of crops using satellite images is an important component of the pipeline of the Food Security use case discussed in Section 2. For this task, University of Trento developed a deep neural network architecture for crop type mapping using Sentinel-2 image time series [13]. This classification task presents many challenges: (i) the considered time series are noisy, due to the presence of clouds that corrupts the multi-temporal spectral signature, thus affecting the classification results, (ii) time series of different tiles are made up of images acquired in different dates (different temporal sampling), and (iii) a large training dataset of labeled samples is needed to train the deep model.

To address these challenges, the methodology of [13] consists of three main steps: (i) a preprocessing step that generates temporally homogeneous time series of images across tiles that accurately represent the phenological behavior of the crops, (ii) an extraction step that automatically establishes a large training dataset leveraging publicly available crop type maps based on farmer declarations in a large area of Austria, and (iii) a multi-temporal deep learning classification algorithm based on a Long Short Term Memory neural network. The proposed approach achieves more balanced classification results compared to existing state-of-the-art methods obtaining a mean F1 score of 78.32% and an overall accuracy of 85.86%. The approach of [13] has recently been implemented in Hopsworks (see Section 7) and has been deployed in the Food Security TEP.

An important contribution of ExtremeEarth in this context is the development of the training dataset mentioned above which consists of around 1 million pixels of 16 Sentinel-2 images located in Austria, where each pixel is labelled with one of 13 crop types. The dataset will soon be available in the web site of the project.

5. DEEP LEARNING FOR SEA ICE CHARTING

The core of the Polar use case of ExtremeEarth is sea ice classification. For this task, UiT, KTH and DLR have developed multiple deep neural network architectures (LDA, CNNs, variational auto-encoders, GANs, etc.) described in more detail in [5, 6, 7]. Some of these architectures have been implemented in Hopsworks (see Section 7) and have been deployed in the Polar TEP.

An important contribution of ExtremeEarth in this context is the development of three training datasets for sea ice classification: (i) A training dataset consisting of 63,048 patches of 30 Sentinel-1 images located in the European Arctic where each patch is labelled with one of 6 ice types. This dataset was developed by expert photo-interpretation and it was used to train three of the CNNs. (ii) A training dataset consisting of around 62 million patches of 24 Sentinel-1 images located in the Belgica Bank of the Greenland Sea, where each patch is labelled with one of 11 ice types. This dataset was developed using active

learning and it was used to train the LDA model and one of the CNNs. (iii) A training dataset consisting of 18,000 patches of 12 Sentinel-1 images located in the Danmarkshavn (East coast of Greenland), where each patch is labelled with one of 2 classes (ice or water). This dataset was developed by expert photo-interpretation and it was used to train one of the CNNs.

The first and the third of the above datasets are publicly available on the web site of the project³ and the same will be true for the second one very soon.

To advance the international state of the art in this area, ExtremeEarth also organized a workshop on “Machine learning for operational sea ice charting” during ESA’s Φ -week 2020.

6. BIG DATA TECHNOLOGIES

The previous sections presented the two use cases of ExtremeEarth and the deep learning algorithms deployed in these use cases. The other technical dimension of the project, which is important in the development of the two use cases, is the utilization of linked data technologies that scale to large volumes of heterogeneous geospatial data available in geographically dispersed data sources. To tackle this important challenge, University of Athens and Demokritos have developed the following big data systems:

- GeoTriples-Spark, a scalable implementation of GeoTriples [8] on top of Apache Spark for transforming geospatial data from their legacy formats (e.g., shapefiles) into RDF.
- JedAI-spatial, a scalable system for interlinking RDF data sources by discovering topological relations among geographic features present in these sources [12].
- Strabo 2, a scalable geospatial RDF store developed using Apache Spark and Apache Sedona.
- A scalable extension of the system SemaGrow [1] for federating geospatial data sources.

To evaluate Strabo 2 and SemaGrow, the same partners have developed and published two benchmarks: Geographica 2 [4] and GeoFedBench [14].

All of the above systems are deployed in the two use cases. In both use cases, information and knowledge extracted from satellite images (e.g., crop type maps) together with data from auxiliary data sources are encoded in RDF using the ontology of the relevant use case. Then, the use case is implemented using the above big data systems. For example, in the Food Security use case, we use an ontology to model data sources such as water availability, crop conditions and irrigation information (see Section 2). The ontology also integrates these data sources with the results of the deep learning algorithms and the PROMET model, so that we can provide irrigation recommendations for specific crop fields in an area of interest.

Another example of the use of the above linked geospatial data technologies in ExtremeEarth is [10], where we show how

³<http://earthanalytics.eu/datasets.html>

to use geospatial interlinking algorithms, such as the ones implemented in JedAI-spatial, to produce automatic workflows for combining in-situ observational data with satellite images. For the Polar use case, this has been done using observations from the Ice Watch system of MET Norway, which collects data from ships performing visual sea ice observations while navigating the Arctic. This in-situ observational data record the time, point locations, and other important properties of sea ice. Interlinking these observations with satellite images has enabled MET Norway to validate and improve the interpretation of satellite images, improve routine ice charts, and assist in building deep learning algorithm training datasets.

7. THE EXTREMEEARTH PLATFORM

The ExtremeEarth platform brings together the deep learning architectures and the big data technologies presented above and applies them to the development of the two use cases.

The platform is based on Hopsworks, a data intensive AI platform from Logical Clocks. Hopsworks⁴ is an open-source framework for the development and operation of machine learning models, available as a managed platform on AWS and Azure and self-managed (open-source or Enterprise version). It has certain unique features that makes it appropriate for the development of deep learning algorithms for EO data: it provides tools to build end-to-end machine learning pipelines, a feature store, management of machine learning artifacts and assets such as experiments and models, first-class support for popular open-source machine learning frameworks such as TensorFlow, PyTorch, Keras and Scikit-Learn, integration with data science tools such as Jupyter notebooks, and infrastructure monitoring functionalities. Hopsworks provides a horizontally scalable platform for deep learning with GPUs and SDKs for hyper-parameter tuning and elastic model serving.

ExtremeEarth has demonstrated that Hopsworks is an excellent platform for developing the two use cases using the big linked geospatial data systems presented above, as it offers a convenient collaborative environment for building data pipelines. For example, a user can import a specific dataset in a project, transform it into RDF and securely share the results with specific other users or projects, who then can perform further processing, such as interlinking or querying. Hopsworks supports dynamic roles for users accessing and processing such datasets, which enables data owners to securely give access to datasets in a project, knowing the data cannot be exported outside the project or cross-linked with other data sources outside the project. This security model is built on TLS certificates and enables Hops⁵ to operate as the only multi-tenant Hadoop platform. In order to perform these tasks, users and developers only need to interact through the human-usable interface of the platform, that offers ready-to-use

⁴<https://www.logicalclocks.com/>

⁵<https://hopsworks.readthedocs.io/en/stable/overview/introduction/what-hops.html>

deployments of popular cloud data storage and processing tools like Apache Hive, Apache Spark and Apache Kafka. Also, using this interface the users can collaborate in order to specify and execute their data pipeline in a Jupyter Notebook and effortlessly monitor the execution progress and inspect the results.

Finally, we have shown that by implementing the big data systems of Section 6 using Hopsworks, we can outperform competitor systems and scale to TBs of geospatial data [9].

8. SUMMARY

We gave an overview of ExtremeEarth and its main contributions up to today. As the project reaches its conclusion, the ExtremeEarth team is working on the following problems: validation of the deep learning models, detailed experimental evaluation of the implemented big linked geospatial data systems using Geographica 2 and GeoFedBench, and integrating all available technologies to build demos of the two use cases in the Food Security and Polar TEPs.

REFERENCES

- [1] A. Charalambidis, A. Troumpoukis, and S. Konstantopoulos. Semagrow: optimizing federated SPARQL queries. In *SEMANTICS*, 2015.
- [2] A. Everett, A. Marinoni, A. Cziferszky, D. Arthurs, D.-A. Pantazi, G. Stamoulis, G. Mandilaras, J. L. Pedersen, M. Datcu, N. Hughes, P. Wagner, S. Khaleghian, T. Kræmer, and Å. Kierbech. Implementation and evaluation of the Polar use case-v1. ExtremeEarth Deliverable D5.3, Available from <http://earthanalytics.eu/deliverables.html>, 2020.
- [3] D. H. Hagos, T. Kakantousis, V. Vlassov, S. Sheikholsami, T. Wang, J. Dowling, A. Fleming, A. Cziferszky, M. Muerth, F. Appel, D.-A. Pantazi, D. Bilidas, G. Papadakis, G. Mandilaras, G. Stamoulis, M. Koubarakis, A. Troumpoukis, and S. Konstantopoulos. Software architecture for Copernicus Earth observation data. In *BiDS*, 2021.
- [4] T. Ioannidis, G. Garbis, K. Kyzirakos, K. Bereta, and M. Koubarakis. Evaluating geospatial RDF stores using the benchmark Geographica 2. *Journal on Data Semantics*, 2021.
- [5] C. Karmakar, C. O. Dumitru, G. Schwarz, and M. Datcu. Feature-free explainable data mining in SAR images using latent Dirichlet allocation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:676–689, 2021.
- [6] S. Khaleghian, T. Kræmer, A. Everett, Åshild Kierbech, N. Hughes, T. Eltoft, and A. Marinoni. Synthetic aperture radar data analysis by deep learning for automatic sea ice classification. In *The European Conference on Synthetic Aperture Radar*, 2021.
- [7] T. Kræmer, S. Khaleghian, A. Marinoni, C. Dumitru, M. Datcu, and T. Eltoft. Deep architectures implementation for the Polar use case-v1. ExtremeEarth Deliverable D2.3.
- [8] K. Kyzirakos, D. Savva, I. Vlachopoulos, A. Vasileiou, N. Karalis, M. Koubarakis, and S. Manegold. GeoTriples: Transforming geospatial data into RDF graphs using R2RML and RML mappings. *J. Web Sem.*, 2018.
- [9] G. Mandilaras and M. Koubarakis. Transforming big geospatial data into linked data. In *Forthcoming*, 2021.
- [10] G. Mandilaras, D.-A. Pantazi, M. Koubarakis, N. Hughes, A. Everett, and Å. Kierbech. Ice monitoring with ExtremeEarth. In *2nd Workshop on Large Scale RDF Analytics*, 2020.
- [11] S. Migdall, S. Dotzler, C. Miesgang, F. Appel, M. Muerth, H. Bach, G. Weikmann, C. Paris, D. Marinelli, and L. Bruzzone. Water stress assessment in Austria based on deep learning and crop growth modelling. In *Submitted to BiDS*, 2021.
- [12] G. Papadakis, G. Mandilaras, N. Mamoulis, and M. Koubarakis. Progressive, holistic geospatial interlinking. In *The Web Conference*, 2021.
- [13] C. Paris, G. Weikmann, and L. Bruzzone. Monitoring of agricultural areas by using Sentinel 2 image time series and deep learning techniques. In *SPIE Remote Sensing Conference*, 2020.
- [14] A. Troumpoukis, S. Konstantopoulos, G. Mouchakis, N. Prokopaki-Kostopoulou, C. Paris, L. Bruzzone, D.-A. Pantazi, and M. Koubarakis. GeoFedBench: A benchmark for federated GeoSPARQL query processors. In *ISWC*, 2020.