ARTICLE IN PRESS

Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Using spatiotemporal patterns to optimize Earth Observation Big Data access: Novel approaches of indexing, service modeling and cloud computing

Jizhe Xia[a,c], Chaowei Yang[b], Qingquan Li[a,c,*]

[a] Shenzhen Key Laboratory of Spatial Information Smart Sensing and Services, School of Architecture and Urban Planning & Research Institute for Smart Cities, Shenzhen University, Shenzhen, China

[b] NSF Spatiotemporal Innovation Center, George Mason University, Fairfax, VA 22030, United States

[c] Key Laboratory for Geo-Environmental Monitoring of Coastal Zone of the National Administration of Surveying, Mapping and GeoInformation, Shenzhen University, Shenzhen, China

ARTICLE INFO

Keywords:
Spatiotemporal optimization
Big Data
Cloud computing
Global operation
GEOSS

ABSTRACT

Based on our GEOSS Clearinghouse operating experience, we summarized the three Earth Observation (EO) Big Data access challenges, namely, fast access, accurate service estimation and global access, and two essential research questions: are there any spatiotemporal patterns when end users access EO data, and how can these spatiotemporal patterns be utilized to better facilitate EO Big Data access? To tackle these two research questions, we conducted a two-year pattern analysis with 2+ million user access records. The spatial pattern, temporal pattern and spatiotemporal pattern of user-data interactions were explored. For the second research question, we developed three spatiotemporal optimization strategies to respond to the three access challenges: a) spatiotemporal indexing to accelerate data access, b) spatiotemporal service modeling to improve data access accuracy and c) spatiotemporal cloud computing to enhance global access. This research is a pioneering framework for spatiotemporal optimization of EO Big Data access and valuable for other multidisciplinary geographic data and information research.

1. Earth Observation Big Data and GEOSS

With the rapid advancement of data acquisition technologies, terabytes to petabytes of data have been collected and accumulated at a global scale at an unprecedented rate. This explosively growing data is referred as the Big Data (Howe et al., 2008; Manyika et al., 2011). In general, the Big Data characteristics can be summarized by the "6 Vs", namely, Volume, Variety, Velocity, Veracity, Visualization and Value (Alexander, Hoisie, & Szalay, 2011; Khan, Uddin, & Gupta, 2014). Big Data characteristics are also genuine when Earth Observations data are considered (Schnase et al., 2014; Guo, Zhang, & Zhu, 2015, Kiemle, Molch, Schropp, Weiland, & Mikusch, 2016). A variety of EO systems (e.g., satellite) monitor the Earth on a daily basis and provide a tremendous amount ("Variety") of global observation data at a rapid rate ("Velocity"). These EO datasets exist in a "Variety" of spatiotemporal resolutions, distributions and formats with various data quality and uncertainty ranges ("Veracity"). Many typical EO applications require a fast "Velocity" to access EO Big Data and an effective presentation ("Visualization"). Most importantly, EO Big Data urgently need efficient analysis to produce "Value" for better scientific research, planning and decision making in various domains. EO Big Data not only provide a great opportunity to improve our understanding of the Earth system, but also pose tremendous challenges to discover, manage, access and analyze EO Big Data (Bargellini et al., 2013).

To respond to the EO Big Data challenges, the Group on Earth Observation (GEO) proposes a 10-year plan for building a Global Earth Observation System of Systems (GEOSS) to target nine societal benefit areas (Christian, 2005). GEOSS aims to coordinate various EO systems and provide EO data access for globally distributed GEO members and public end users (GEO, 2005). The GEOSS Common Infrastructure (GCI) is an official implementation of the GEOSS concept, and the GCI generally consists of four principal components (Fig. 1; Desconnets et al., 2017; Bai & Di, 2011; Xia, Yang, Gui, Liu, & Li, 2014; Nativi et al., 2015; Santoro, Nativi, & Mazzetti, 2016):

- The GEOSS Portal provides a single Internet access point (http://

* Corresponding author at: Shenzhen Key Laboratory of Spatial Information Smart Sensing and Services, School of Architecture and Urban Planning & Research Institute for Smart Cities, Shenzhen University, Shenzhen, China.

E-mail address: liqq@szu.edu.cn (Q. Li).

https://doi.org/10.1016/j.compenvurbsys.2018.06.010
Received 31 January 2018; Received in revised form 28 June 2018; Accepted 29 June 2018

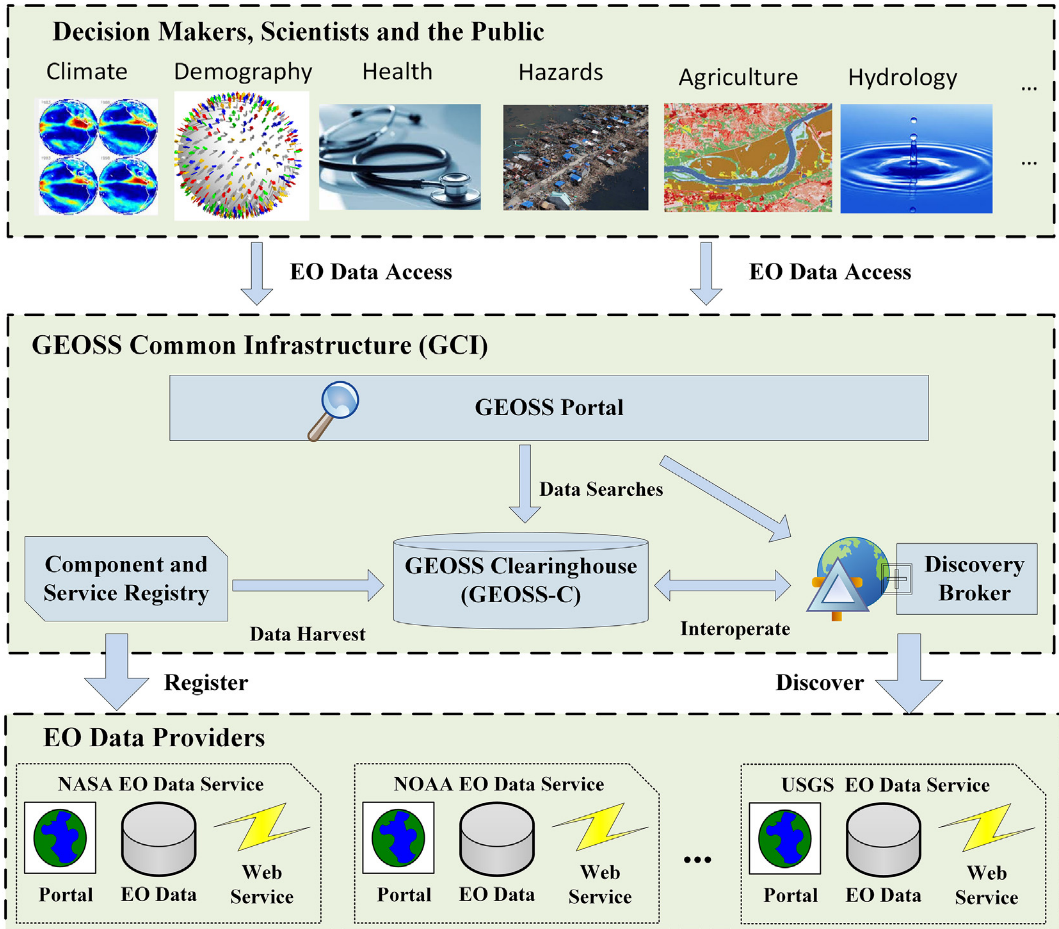

Fig. 1. A brief view of GEOSS infrastructure.

www.geoportal.org/) for EO Big Data provided by GEOSS members. End users can discover and access EO Big Data though the GEOSS portal with spatial, temporal and thematic criteria.

- The GEOSS Component and Service Registry (GEOSS-CSR) collects EO data that are self-nominated by data providers through a resource registration process.
- The GEOSS Clearinghouse (GEOSS-C) is a driver for the GEOSS by managing and providing access to and descriptions of EO data (EO metadata) for other GEOSS core components.
- The Discovery and Access Broker (GEO-DAB) employs a series of brokering approach, interoperability standards and cutting-edge Big Data technologies to discover EO data from multidisciplinary data providers, such as NASA, NOAA, and USGS.

This GCI offers a global, flexible and interoperable EO Big Data framework allowing global end users to access an extraordinary range of EO data (Nativi et al., 2015). However, it is still a considerable challenge for GCI to provide effective EO Big Data access (Section 2) and two critical research questions should be answered:

- Are there any spatiotemporal patterns when end users access EO Big Data?
- If yes, how can these spatiotemporal patterns be utilized to better facilitate EO Big Data access?

To respond two research questions, we first explore spatiotemporal patterns during the EO data access process (Section 3). Then, we introduce how spatiotemporal patterns can be utilized to offer potential solutions to address the relevant EO Big Data challenges (Sections 4–6). Finally, Section 7 provides conclusions and suggestions for future research.

Our major contributions include: 1) we explored the spatiotemporal patterns of user–data interactions based on 2+ million user EO data access records in GEOSS-C; based on these patterns, we 2) proposed a spatiotemporal index mechanism to accelerate EO data access, 3) designed a spatiotemporal service model to improve EO service evaluation accuracy, and 4) implemented a cloud computing framework to better support global EO data access.

2. Challenges and related works

By 2015, the GEOSS prototype has been developed to enable the entire GEOSS community and general public to have access to over 80 million heterogeneous EO datasets (Nativi et al., 2015). Members at the Joint Center for Intelligent Spatial Computing for Water/Energy Science (CISC) at George Mason University (GMU) have collaborated with officials from the Federal Geographic Data Committee (FGDC), GEO and the U.S. Geology Survey (USGS) in designing, developing and operating the GEOSS-C prototype. Currently, the GEOSS-C prototype is deployed and operated by GMU at the Fairfax, Virginia, U.S. By 2015, the GEOSS-C has provided over 2 million access request to other GEOSS components and public users from 140+ countries. Learning from the GEOSS-C operating experience over the past few years, we summarized three challenges stemming from the GEOSS-C providing effective EO Big Data access: