Int J Appl Earth Obs Geoinformation 73 (2018) 245–252

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

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

Building a spatiotemporal index for Earth Observation Big Data

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:
GEOSS
Spatial data infrastructure
Data indexing
Information retrieval
Earth observation big data

ABSTRACT

With the rapid advancement of Earth Observation systems, Earth Observation data has been collected and accumulated at an unprecedented fast rate. Earth Observation Big Data emerged with new opportunities for human to better understand the Earth systems, but also pose a tremendous challenge for efficiently transforming Big Data into Earth Observation Big Value. Targeting on this challenge, a well-organized data index is a key to enhance the "Data-Value" transformation by accelerating the access to data. Although various data indexing approaches have been proposed with different optimization objectives, literature shows that there are still apparent limitations for Earth Observation data indexing. This paper aims to build a spatiotemporal indexing for Earth Observation Big Data. Specifically, a) to support various Earth Observation Data Infrastructures, we adopt an indexing framework to efficiently retrieve data with various textual, spatial and temporal requirements; b) a distributed indexing structure is designed to improve the index scalability; c) data access pattern is integrated to the indexing algorithm for both spatial and workload balancing. The results show that our indexing approach outperforms traditional indexing approaches and accelerates the access to Earth Observation data. We envision that data indexing will become a key technology that drives fundamental Earth Observation advancements in the Big Data era.

1. Introduction

With the rapid advancement of Earth Observation systems, satellite, airborne and ground-based remote sensing systems provide Earth Observation data with high spatial, temporal and radiometric resolutions (Desconnets et al., 2017; Lewis et al., 2015). According to the Committee on Earth Observation Satellites (CEOS) database, 174 Earth Observation satellites have been launched by over 20 agencies worldwide (CEOS, 2018). In the next 12 years, over 150 satellites will be launched for various Earth Observation missions including pollution monitoring, weather forecasting and surface deformation detection (Kar et al., 2015; Lei and Brown, 2017; Wang et al., 2015). Petabytes of Earth Observation data have been collected and accumulated at a global scale with an unprecedented fast rate (Guo et al., 2015). This explosively growing data is referred as the Big Data (Manyika et al., 2011). In addition to traditional Big Data characteristics, Earth Observation Big Data is highly multi-dimensional and exists in a "Variety" of spatial, temporal and spectral resolutions. Various observation condition (e.g., weather) often result in different uncertainty ranges ("Veracity"). Most importantly, the transformation of Earth Observation Big Data into Big "Value" needs to be fast supported ("Velocity").

Earth Observation is also one of the domains benefiting from the Big Data. By transforming Earth Observation Big Data into value-added products and information, we have a great opportunity to better understand the Earth systems. For example, the White House released the Climate Data Initiative to bring together Big Data and Climate Change (Ramachandran et al., 2016; Tilmes et al., 2015). To construct historical climate conditions and predict future climate changes, scientists need to collect, access and analyse Big Data from various Earth Observation systems such as atmosphere, hydrosphere, biosphere and lithosphere (Guo et al., 2015; Kang and Cressie, 2013). Meanwhile, the Big Data five "V" axes ("Volume", "Variety", "Veracity", "Velocity", "Value") also pose tremendous challenges for the management, access and analysis of Earth Observation data (Xie and Li, 2016; Nativi et al., 2015; Bargellini et al., 2013). Specifically, with the increase the data structure dimension (spatial to spatiotemporal), various access requirements (spatial, temporal and spatiotemporal) need to be supported while the computational complexity of data access is significantly increased. A

* 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.jag.2018.04.012
Received 18 January 2018; Received in revised form 27 April 2018; Accepted 29 April 2018
0303-2434/ © 2018 Elsevier B.V. All rights reserved.

mechanism to support fast Earth Observation Big Data access is urgently needed.

In order to fill the research gap, we propose a spatiotemporal indexing mechanism that optimizes the access to Earth Observation Big Data. In Section 2, a review of Earth Observation data indexing challenge and relevant literature is provided. Section 3 outlines our approach for Earth Observation Big Data indexing, including the indexing framework, indexing structure and optimization algorithm. Section 4 presents the experimental results across a range of data access scenarios. Section 5 discusses the indexing life-cycle, limitations and real-world examples. Section 6 provides conclusions and suggestions for future research.

2. Data indexing challenges and related works

In response to the Earth Observation Big Data challenges, the intergovernmental Group on Earth Observation (GEO) was launched at the World Summit on Sustainable Development by the Group of Eight (G8). In 2005, the GEO proposed a 10-years' plan: Global Earth Observation System of Systems (GEOSS) for coordinating globally distributed Earth Observation systems (Nativi and Bigagli, 2010). The GEOSS Common Infrastructure (GCI) is the official implementation of the GEOSS concept and the GCI consists four principal components (Fig. 1 Nativi et al., 2015; Bai and Di, 2011; Xia et al., 2014; Christian, 2008):

- **The GEO Portal** (http://www.geoportal.org/) provides a Graphical User Interface (GUI) for GEOSS end users to access Earth Observation Big Data.
- **The GEOSS Component and Service Registry** enables the resource registration capability for data providers to nominate their Earth Observation resources to the entire GEOSS community.
- **The GEOSS Clearinghouse** provides the engine for the GEOSS by managing and coordinating Earth Observation Big Data information (metadata) for other GEOSS core components.
- **The Discovery and Access Broker** adopts several brokering approaches to discover and register Earth Observation resources from multidisciplinary data providers. In 2011, the GEO planned a significant change to GCI framework by integrating other three GCI components to the Discovery and Access Broker.

The GEOSS Clearinghouse program was launched in 2008 and collaboratively designed by the GEO, Joint Center for Intelligent Spatial Computing for Water/Energy Science at George Mason University (CISC), Federal Geographic Data Committee (FGDC) and U.S. Geology Survey (USGS) (Huang et al., 2010). By 2017, the GCI prototype has been published for the access to over 80 million Earth Observation data (Nativi et al., 2015). Learned from the GEOSS Clearinghouse operating experience in the past few years, how to efficiently index Earth Observation Big Data has become one of the most critical issues to the success of GEOSS. Due to the Big Data "Volume", "Velocity" and "Variety", the cost to retrieve required information (e.g., global hurricane data from 2011 to 2015) from the entire Earth Observation dataset has been significantly increased. On the other hand, the fast

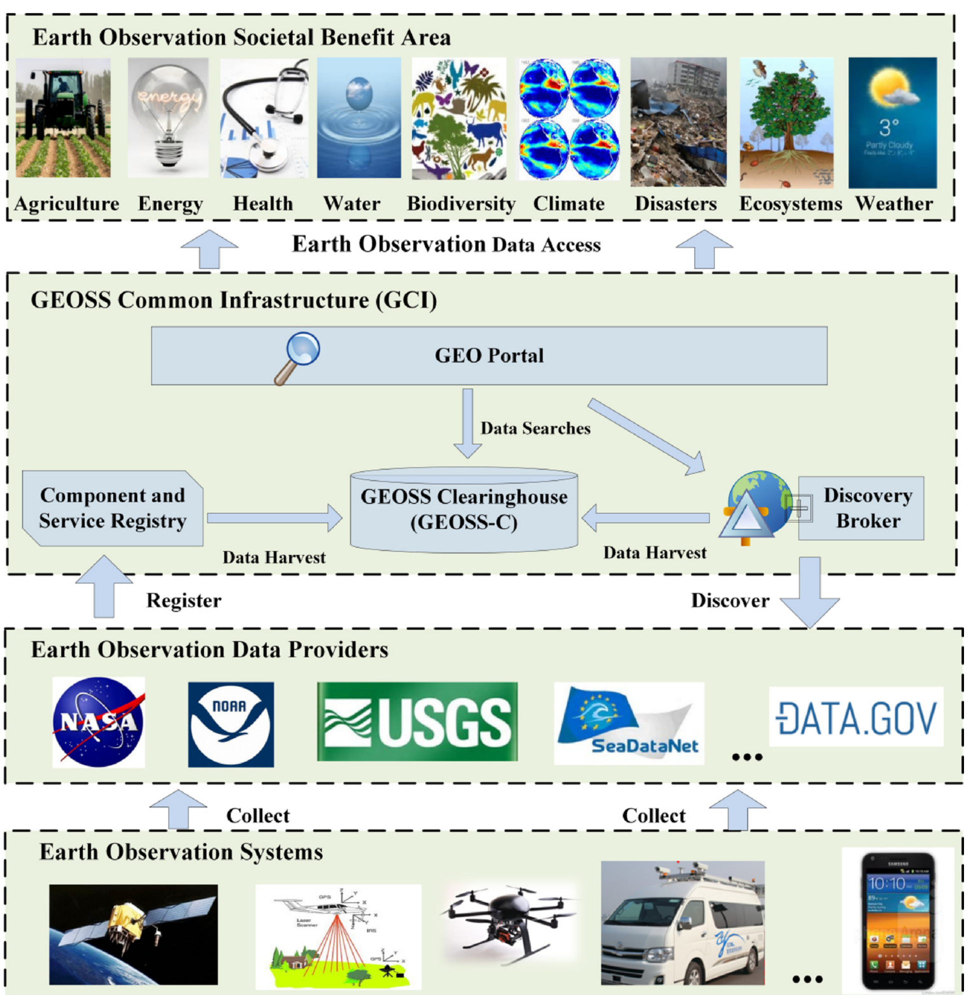

Fig. 1. A brief view of GEOSS Infrastructure.

Download English Version:

https://daneshyari.com/en/article/8867752

Download Persian Version:

https://daneshyari.com/article/8867752

Daneshyari.com