China Data Cube (CDC) for Big Earth Observation Data: Practices and Lessons Learned

Qianqian Cao [1,2], Guoqing Li [1], Xiaochuang Yao [3,*] and Yue Ma [1,2]

1   Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China
2   School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
3   College of Land Science and Technology, China Agricultural University, Beijing 100083, China
*   Correspondence: yxc@cau.edu.cn; Tel.: +86-188-1135-6282

Abstract: In the face of tight natural resources and complex as well as volatile environments, and in order to meet the pressure brought by population growth, we need to overcome a series of challenges. As a new data management paradigm, the Earth Observation Data Cube simplifies the way that users manage and use earth observation data, and provides an analysis-ready form to access big spatiotemporal data, so as to realize the greater potential of earth observation data. Based on the Open Data Cube (ODC) framework, combined with analysis-ready data (ARD) generation technology, the design and implementation of CDC_DLTool, extending the support for data loading and the processing of international and Chinese imagery data covering China, this study eventually constructs the China Data Cube (CDC) framework. In the framework of this CDC grid, this study carried out case studies of water change monitoring based on international satellite imagery data of Landsat 8 in addition to vegetation change monitoring based on Chinese satellite imagery data of GF-1. The experimental results show that, compared with traditional scene-based data organization, the minimum management unit of this framework is a pixel, which makes the unified organization and management of multisource heterogeneous satellite imagery data more convenient and faster.

1. Introduction

In today's era, with the development of science and technology, our ability to obtain remote sensing data has gradually reached an unprecedented level [1]. Remote sensing data already embody the characteristics of big data, and remote sensing big data are receiving more and more attention from experts in academic fields and commercial application fields. Remote sensing big data have inherent characteristics, such as being dynamic, multiscale, and nonlinear, as well as external characteristics, such as being multisource, high-dimensional, and heterogeneous, which play an important role in multiple fields, such as atmospheric science, land use [2], vegetation and ecology [3,4], environmental science [5], and crustal evolution [6]. Traditional remote sensing data organization and management methods based on "scenes" have been unable to meet the management and application requirements of remote sensing big data due to the temporal and spatial fragmentation of data [7]. People have been exploring new data organization and management frameworks or technologies to make up for the shortcomings of existing data management methods, thereby improving the utilization efficiency of remote sensing big data.

Most of the traditional data management frameworks are created based on international standards such as those of the ISO (International Organization for Standardization) and OGC (Open Geospatial Consortium) [8] for representing and storing spatial information in data files (such as GeoTIFF and Esri Shapefile) as well as database systems (such as PostGIS and Oracle Spatial), and for serving spatial data and metadata through web

services. Most regional, national, and international remote sensing data management agencies still share and disseminate multisource Remote sensing data in the form of individual imagery data files through HTTP, FTP, and SSH protocols in web service portals [9]. In order to cope with the pressure faced by traditional data organization and management methods in storing, processing, and analyzing remote sensing big data, scientists and engineers have proposed and developed new frameworks based on new technologies, including cloud computing and distributed systems, such as the array database system (Array DBMS) [10], Google Earth Engine [11], data cubes [12–14], and cloud-based remote sensing data production systems [15,16]. It is known from the literature [6] that array database systems (such as RasDaMam [17], SciDB [18]) are centered on multidimensional arrays, and large arrays are split into index blocks, which are stored and shared among multiple computers to improve performance and efficiency. The Google Earth Engine was launched by Google in 2010 as a commercial cloud platform for the large-scale scientific analysis and visualization of geospatial datasets. The Open Data Cube (ODC), formerly known as the AGDC, is an earth observation data organization and analysis framework consisting of a series of open source data structures and tools. The ODC provides the data access mode of OGC web services (such as WCS, WM, and WMT), and also provides an open source Python API interface and Jupyter Notebook with application examples, so as to help users to use and share multisource remote sensing imagery data [19].

Based on the ODC basic framework, combined with actual needs and application scenarios, multiple research teams from all over the world have carried out a series of research studies. Australian researchers, based on archived satellite imagery data, such as that of Landsat and Sentinel, constructed Australian data cubes to monitor surface water changes [20–22] and land cover change maps [23], mangrove area expansion [24], etc. Scholars in Switzerland, combining the massive amounts of satellite imagery data covering [22] Switzerland, have carried out studies on snow cover [25], urbanization [26], vegetation [22], and water quality changes [27], hoping to improve their understanding of Swiss resources and further understand the environment [25]. African research teams, by building the ODC-based Digital Earth Africa, focused on application areas such as land transformation, urbanization, and water range changes that are consistent with the United Nations' Sustainable Development Goals and to meet national and regional decision-making needs [28–34]. In addition, research teams in countries such as Vietnam, Colombia, Brazil, Mexico, and China have used ODC for data management and application analyses in a variety of thematic areas, such as vegetation [34], hydrology [35,36], soils [37–39], and island ecology [40], using international and national satellite imagery data covering their countries.

Although data management frameworks such as the Google Earth Engine and Array DBMS facilitate the large-scale processing and application of satellite imagery data, users still need to put in a lot of effort and learn advanced techniques to be able to utilize these data management environments. Moreover, China's high-resolution imagery data require high security and are not suitable to be stored in a commercial cloud platform environment. Therefore, we choose the open source ODC framework to build a localized China Data Cube to meet the practical application requirements by extending data loading and application support for international imagery data and China imagery data. Specifically, based on the ODC open source framework, this study introduces ARD generation technology, designs and implements the CDC_DLTool tool, and extends data loading and application support for international imagery data, such as that of Landsat 8, and China imagery data, such as that of GF-1, covering the Chinese region. A localized China Data Cube is built to meet practical application requirements. Compared with the traditional scene-based data organization, the smallest management unit in this study is a pixel, which facilitates the dynamic analysis of long time series of multisource satellite imagery data.

The main objectives of this paper are to introduce the Earth Observation Data Cube, present designs of the China Data Cube applicable to Chinese satellite imagery data, design two research cases and perform related case studies, discuss the shortcomings of existing research work, and list future work.

2. Open Data Cube for Earth Observations (EODC)

The Earth Observation Open Data Cube (EODC) is a new paradigm shift from a "scene" approach to a pixel grid. This aims to realize the full potential of earth observation data by reducing the barriers posed by these big data and providing access to large-scale spatiotemporal data in the form of analysis-ready data (ARD) [36].

The conceptual architecture of the EODC is composed of four layers from bottom to top [35]: the data acquisition and input layer, which generates ARD through a series of preprocessing operations, such as radiometric and geometric corrections; the data cube infrastructure layer, where ARD are indexed and stored in the data cube through a Python API and related interfaces, providing N-dimensional matrix interfaces for tasks. The data and application platform layer provides users with services such as "virtual laboratory" and task management. The user interface and application layer provides users with applications in various research fields based on earth observation data by calling the underlying interface.

The EODC framework is based on open source software (datacube-core, datacube-dataset-config, datacube-explorer, datacube-notebooks, datacube-docker, odc-tools, etc.) [12] and an API interface, using GDAL, Xarray, Numpy, Matplotlib, and other library files to achieve the loading and analysis of satellite imagery data, the construction of multidimensional arrays, the analysis of calculations, the analysis of the results of graphical plotting, etc. The EODC supports reading satellite imagery data in various data formats (e.g., GeoTIFF, NetCDF, HDF, etc.) [37] and using two data structures, DataSet and DataArray, to represent, in memory, the EODC corresponding to multidimensional satellite imagery data, and to build an EODC model based on multidimensional arrays for easy calculation and analysis. Finally, EODCs use the PostgreSQL [38] database to manage the EO data stored in the file system [35].

The EODC can be deployed in environments such as local file systems, cloud platforms, and high-performance computing, providing users with pixel-level data computation and processing capabilities, allowing them to flexibly implement data analysis algorithms and applications for specific application scenarios with the help of existing analysis tools, thus solving the difficulties encountered when managing and analyzing data based on traditional scenic satellite imagery [41]. However, the EODC also has some shortcomings. Although the EODC open sources the source code of the ODC core technology framework and some example code based on the Jupyter Notebook, it is still not possible to use the application and data directly in the new environment. Users must follow the relevant documentation to manually install, configure, and, if necessary, write interface plug-ins to reproduce the results of an instance. This requires EODC users to have a relevant professional background and programming skills. It can be seen that building a Chinese data cube based on the EODC is both necessary and technically difficult.

3. Development of the China Data Cube (CDC)

Data cubes, a new type of multisource satellite imagery data management framework, were first well-developed and promoted in countries such as Australia and Switzerland. The National Data Centre for Earth Observation Science (NODA), as the only national scientific data center in the field of earth observation science in China officially recognized by the Ministry of Science and Technology and the Ministry of Finance, plays a crucial role in the field of application services for matters of national importance. The NODA has also been following the technical development of the ODC, and proposed the creation of the CDC initiative in 2018 [39]. The goal of the CDC is to efficiently manage the massive amount of multisource international and China satellite imagery data covering the Chinese region based on the ODC, taking into account the characteristics and practical application needs of Chinese satellite imagery data, to provide a reference for relevant researchers to make quick decisions and formulate policies. In this paper, the details and processes of data loading and organizational management in the CDC are described using Landsat 8 and GF-1 data as examples, respectively, as shown in Figure 1.
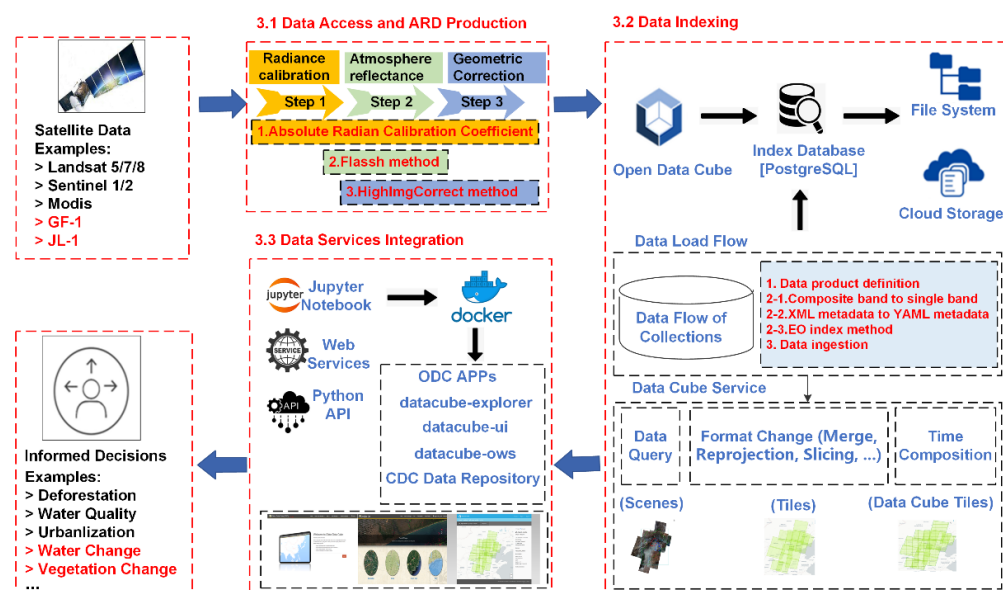
Figure 1. China Data Cube (CDC) architecture diagram.

3.1. Data Access and ARD Production

The CDC is built based on the ODC software suite. The ODC is open source and was initiated in 2016 by organizations and research institutions such as GA, CSIRO, Australian National University, NASA, CEOS, and USGS [40]. The ODC aims to provide for rapid access, storage management, and the analysis of large amounts of gridded satellite earth observation data in a management framework. In detail, the ODC is capable of cataloging large volumes of satellite EO data; the ODC provides Python-based application programming interfaces (APIs) for data analysis; and the ODC can also track data sources such that quality control and updates can be performed.

The systematic and regular delivery of analysis-ready data (ARD) is essential to facilitate the generation of useful information products and support the development of end user applications [42]. CEOS defines ARD as "satellite data that has been processed to minimum requirements and organized into a form that allows immediate analysis with minimal additional user effort, in the shortest possible time, and interoperability with other data sets" [41]. ARD reduce the burden of fully utilizing satellite data by providing specifications that limit data preparation to produce relevant, consistent, normalized, and interoperable data. These specifications save time and effort and minimize the cost of preprocessing data, while leveraging the knowledge and expertise of users, allowing them to spend more time analyzing data rather than searching and preprocessing them. These requirements involve parameters such as radiometric and geometric corrections, atmospheric corrections, and metadata descriptions. In optical imageries the ARD level corresponds to the surface reflectance product, while in radar images it corresponds to the radiation-normalized backward-scattering product [41].

Since the Landsat 8 data provided by the USGS have been processed by the EPSA program to be generated to the LaSRC level, they can provide a surface reflectivity level (L2A) data file, and no further preprocessing is needed to reach the ARD standard. For domestic satellite imageries, the data distribution unit provides users with data at the L1A level, which require further preprocessing, such as radiometric correction and geometric correction, to reach the ARD level. In this paper, absolute radiometric calibration coefficients are used to achieve the radiometric calibration of GF-1 imagery data, and the FLASSH method [43] is used to achieve the atmospheric correction of GF-1 imagery data. The geometric normalization of the GF-1 imagery data is achieved by using the HighImgCorrect method [44]. Finally, the quality of the GF-1 imagery data products is improved to the L4

level, which can meet the requirements of quantitative remote sensing analyses, such as water body change, vegetation change, urbanization, and coastline change.

The ODC provides definition files and organization scripts for earth observation imagery data products such as Landsat and MODIS, which can automatically generate YAML [45] profiles and imagery metadata files from imagery files, but lacks relevant access methods and tools for Chinese satellite imagery data, such as those of GF-1. To this end, based on GDAL, YamlDotNet, and other components, this paper designs and implements the CDC_DLTool middleware, which realizes the acquisition of spectral information, band information, spatial location, and other contents from Chinese satellite imagery files and generates YAML metadata documents conforming to ODC standards; the specific operation flow is shown in Figure 2.
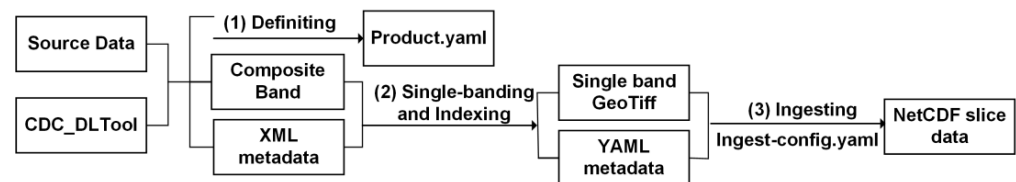
Figure 2. Workflow of loading Chinese satellite imagery data based on CDC_DLTool.

3.2. Data Indexing Based on the CDC Grid

As a new data management method, the CDC grid realizes the paradigm shift from "scene" to pixels, and it can efficiently store satellite imagery data with multitime, multispatial, multispectrum, and multiattribute characteristics. At the same time, the CDC grid also takes into account the temporal and spatial correlation of satellite imagery data, avoiding the temporal and spatial fragmentation of the original "scene" management approach, and making it easy and efficient to analyze the imagery data for long time series applications. Figure 3 shows the data query and retrieval process based on the CDC grid, from which the flow of pixel data in the memory can be clearly seen.
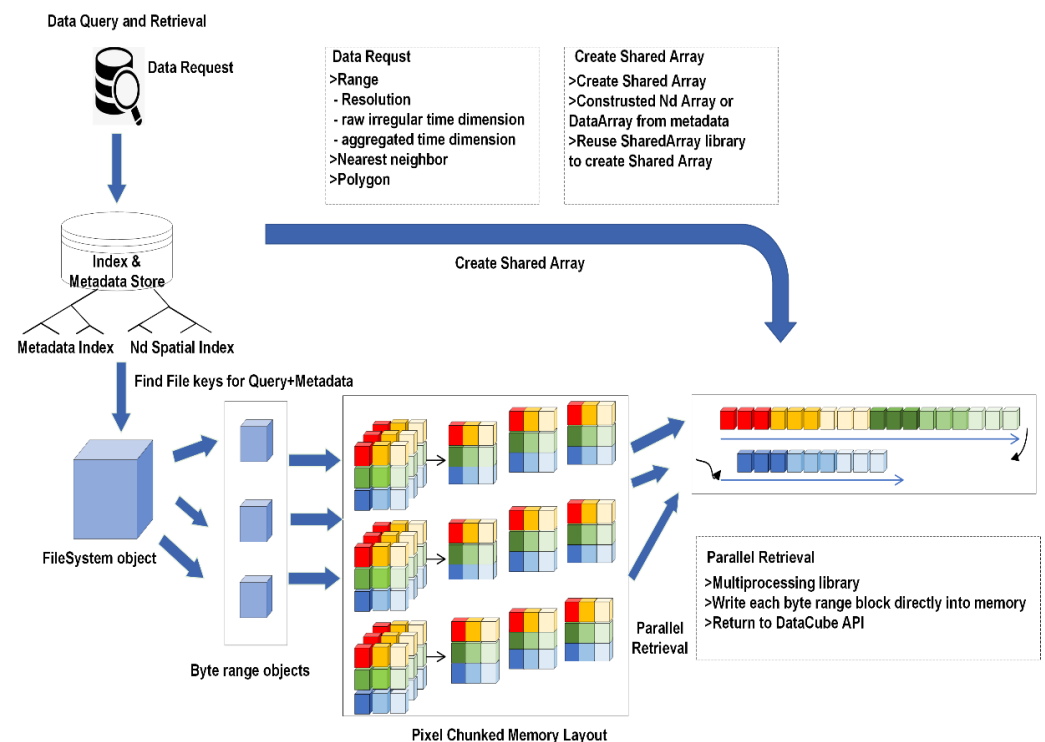
**Figure 3.** Pixel-based grid data query and retrieval.

3.3. Data Storage Strategy and Services

In the process of constructing the CDC, a data storage strategy is an important step in the management of large amounts of multisource heterogeneous satellite imagery data. The optimal resampling scheme needs to be determined based on the spatial resolution and magnitude of the managed satellite imageries to ensure that all of the observed values (i.e., pixels) have the same characteristics, such as spatial resolution. The AGDC resamples Landsat data with a spatial resolution of 30 m and MODIS data with a spatial resolution of 250 m into a grid with a spatial resolution of 25 m [23]. In the CDC grid framework, Landsat 8 data and GF-1 data are stored in two different sets, which retain the original spatial resolution. At the same time, the CDC keeps the original data of Landsat and GF-1 data to ensure that users can decide whether to use panchromatic bands according to the actual needs.

4. Case Study and Results

4.1. Water Body Change Monitoring of the Baiyangdian Lake

The Baiyangdian Lake, the largest freshwater lake in the urban agglomeration of Beijing, Tianjin, and Hebei Provinces [46], is located in An'xin County, Hebei Province, with a water area of 366 km^2, as shown in Figure 4a. Due to the complexity of Baiyangdian, the lake wetland and its surrounding water play an important role in several processes that maintain the normal function of the local ecosystem, including supplying water for the growth of vegetation such as reeds, increasing groundwater supply, improving the local climate system, and protecting biodiversity, as shown in Figure 4b [47]. In recent years, the water bodies of the Baiyangdian Lake have suffered serious impacts due to people's excessive interventions, causing many ecological problems, such as the eutrophication of the water bodies and a reduction in the water body area. Therefore, there is an urgent need to study the changes in the water body area of the Baiyangdian Lake to prevent the further deterioration of the current ecological problems.
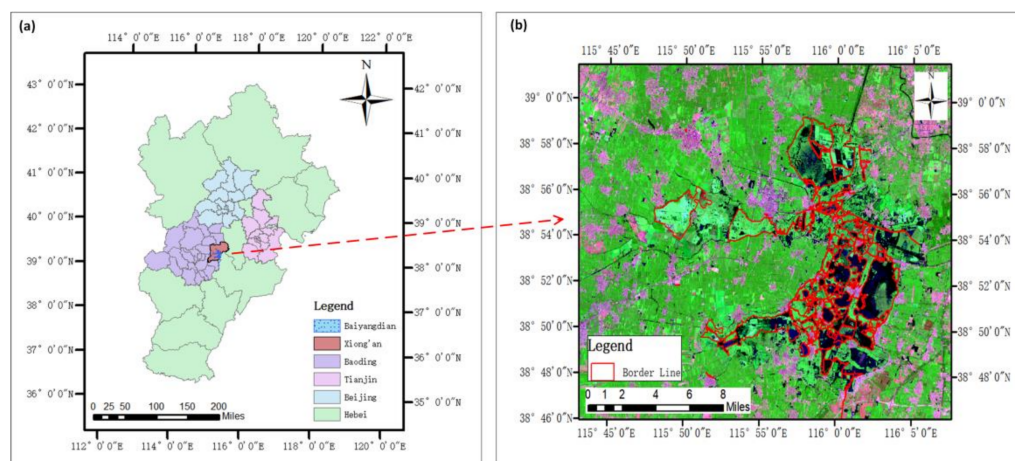

Figure 4. (**a**) The selected study area of case study one, the Baiyangdian Lake, located in the urban agglomeration of Beijing, Tianjin, and Hebei Provinces; (**b**) The Landsat 8 remote sensing imageries covering the Baiyangdian Lake, displayed by false color imageries using bands 7, 5, and 3 (Date: 18 September 2019).

With the rapid development of science and technology, it has become increasingly important to map and detect changes in lake waters through satellite imageries, especially because satellites capture and provide data in the visible and infrared spectral bands, where it is relatively easy to distinguish between land and water [48–50]. This makes optical satellite imageries suitable for monitoring changes in the area of lake waters. In this study, Landsat 8 satellite imagery data were used to extract the area of water bodies in the Baiyangdian Lake. Landsat 8 imagery data were downloaded for the period of 2013 to

2021 with less than 20% cloud cover, which were selected from June to September. Taking advantage of the Landsat 8 data, monitoring studies of water body coverage information can provide scientific data for the development of effective measures to improve the ecological environment of the reservoir basin in the future.

In order to test the usability of the CDC grid, based on the loaded Landsat 8 imagery data and the water body extraction algorithm WOfS [51], this study calculated the changes in the spatial distribution of water bodies in the Baiyangdian Lake, as shown in Figure 5. By comparing and analyzing the annual changes in the water bodies of the Baiyangdian Lake, a theoretical basis for dynamic changes in domestic and industrial generation can be provided. For example, from the second graph in the first row, it can be seen that there was a relatively significant increase in the volume of water in the Baiyangdian Lake in 2014, due to the introduction of external water from the South–North Water Transfer Project. As can also be seen in the third graph in the second row, there was also an incremental change in the quantity of the Baiyangdian Lake in 2018, thanks to a series of protection policies for the Baiyangdian Lake issued after the establishment of the Xiong'an New Area.
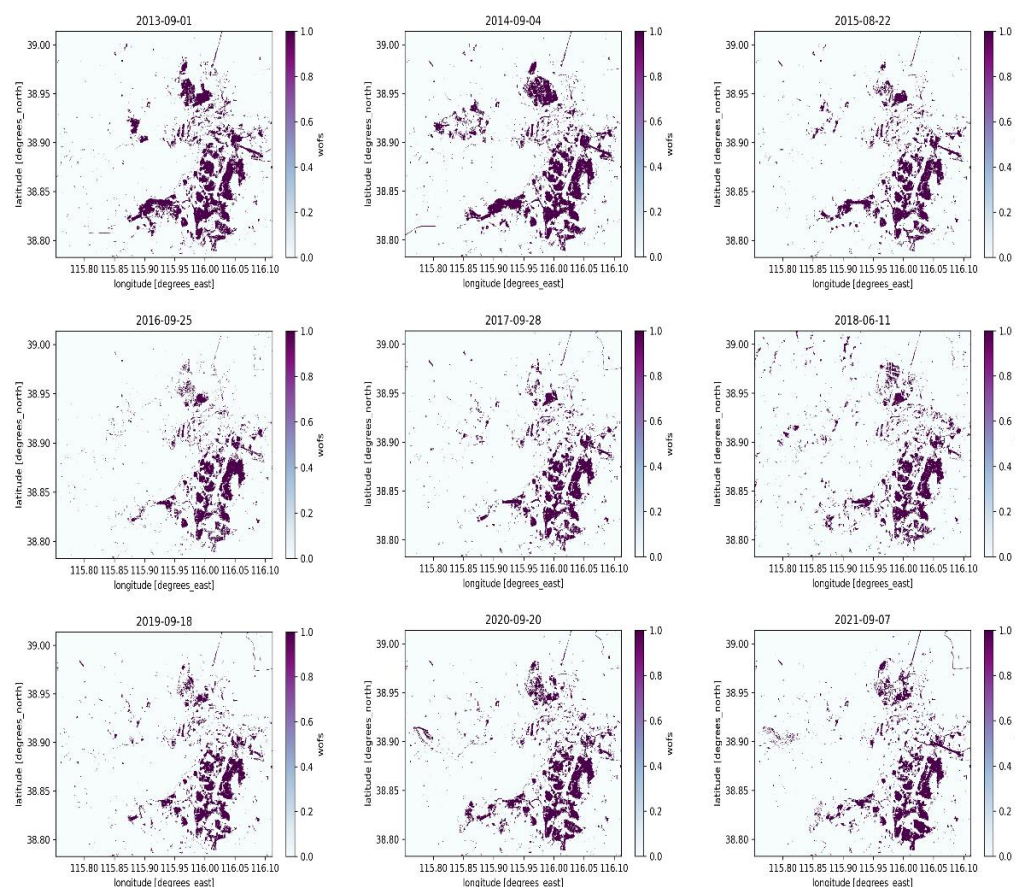
Figure 5. Spatial distribution of water body changes in the Baiyangdian Lake from 2013 to 2021.

Figure 6 shows the total number of observable water bodies in the Baiyangdian Lake based on the CDC grid and the WOfS water body extraction algorithm for the 9 years between 2013 and 2021. The values of the total number of observable water bodies are from large to small, in line with the colored bars (dark blue, blue, green, yellow, red) on the right. The darker the colored bars are, the larger are the total number of observable water bodies in the study area. From this, it can be assumed that the change of water body in Baiyangdian Lake is smaller, and its ecological environment is more balanced. The dark-blue area in the purple rectangle at the bottom right of Figure 6 indicates that water has been present in this area for nine consecutive years. In contrast, it can be seen that the light-green areas in the red rectangles presented in the left part, top-right part, and

bottom-left part of Figure 6 indicate that the values identified as water bodies in these three parts range from one to three, indicating that the Baiyangdian Lake is dynamically changing. More in-depth studies will be carried out later, to determine whether the sizes of the water bodies are increasing or decreasing.
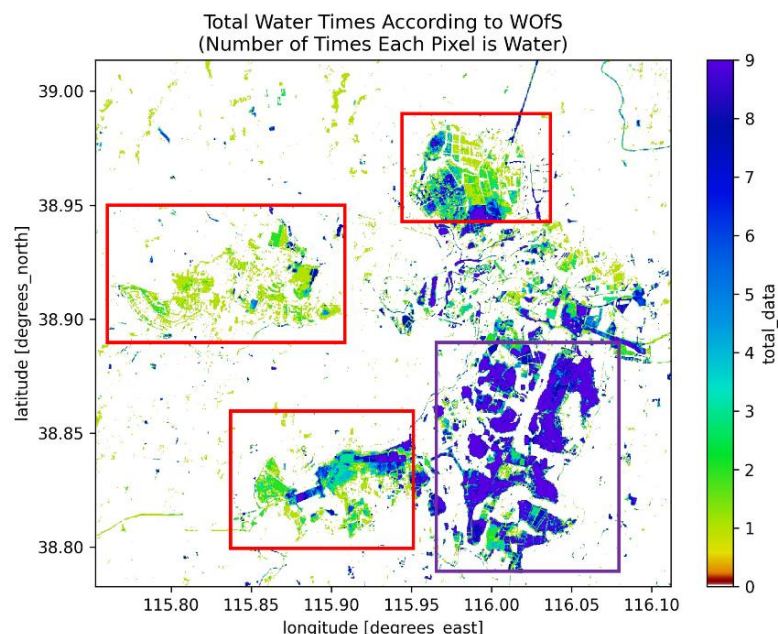
Figure 6. Total water times according to WOfS in the Baiyangdian Lake from 2013 to 2021. The light-green areas in the red rectangles indicated the dynamic changing in the Baiyangdian Lake, while the dark blue area in the purple rectangle indicated water of the Baiyangidan Lake is persistent.

4.2. Vegetation Change Detection in the Beijing Suburbs

The forest area of Huairou District, Beijing, reaches 164,242 hectares, with a forest coverage rate of 77.38%. Forests play an important role in maintaining water and soil, regulating the climate, purifying the atmosphere, preventing noise, and maintaining the ecological balance of nature [52]. In order to verify the availability of the China GF-1 imagery data stored in the CDC framework, this study selected a small area (40.8050° N, 40.8574° N, 116.5207° E, and 116.653° E) in Huairou District, Beijing, for vegetation change monitoring, as shown in Figure 7.
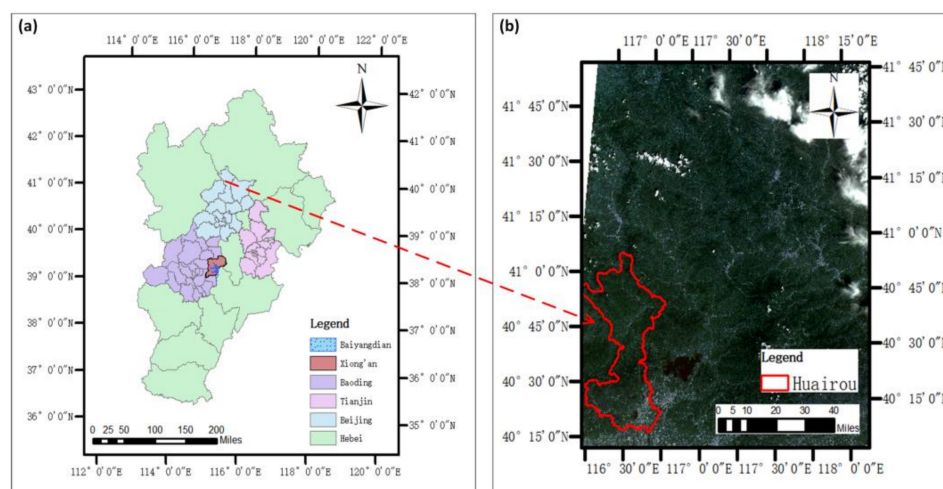

Figure 7. Location of case study area two (**a**) map projection; (**b**) satellite photo (the GF-1 imagery of Huairou, Beijing; data date: 21 August 2021).

Since the GF-1 imagery data downloaded from the official data website are of an L1A level, further data preprocessing steps, such as radiometric correction and geometric correction, are required to generate ARD products. Figure 8a,b show the changes in the vegetation spectral curve in the process of radiometric and geometric correction processing of the GF-1 imagery data used in this study. Figure 9b,c show the geometric deformation and corresponding pixel position shift of the GF-1 imagery data covering the study area before and after geometric correction. The error in the geometric correction process was 1.508 pixels. On the metric scale of meters, the error in the geometric correction of this 16-m spatial resolution GF-1 imagery data was 24.126 m.
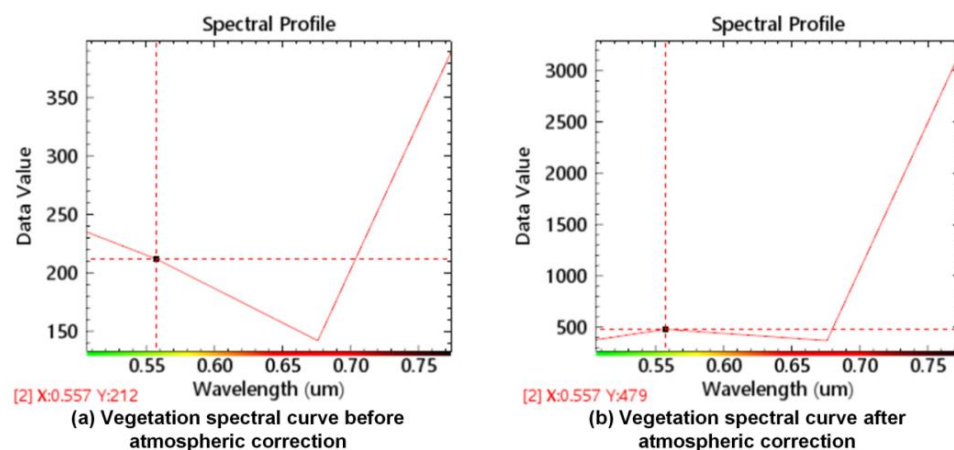
Figure 8. Comparison of vegetation spectral curves before and after atmospheric correction in the study area.
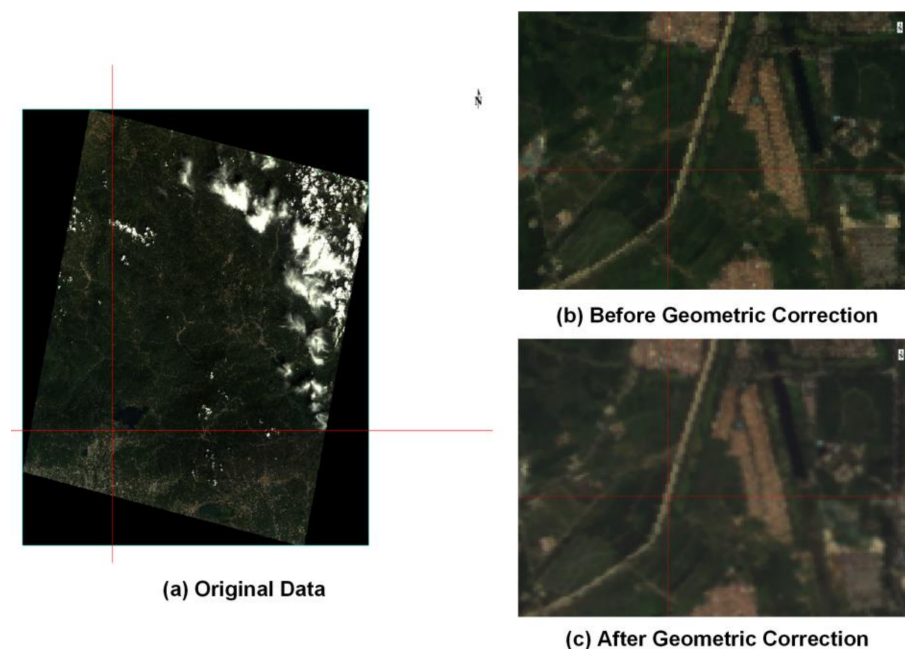
Figure 9. Comparison of the study area before and after geometric correction.

This case study is based on the NDVI (normalized difference vegetation index) values to determine the vegetation change in the same area at two different time periods. Different baseline NDVI threshold ranges can be set for different vegetation types, e.g., 0.6 to 0.9 for dense vegetation and 0.2 to 0.6 for grassland. Figure 10a,b represent the change in dense vegetation in the study area between the two time periods of 2015 and 2017, based on the GF-1 imagery data stored and managed in the CDC grid framework. This is an NDVI threshold plot showing the change in green pixels within the threshold range. Based on

this map, shown in Figure 10c, it can be seen that dense vegetation changes significantly elsewhere except the ridges, which can provide a data reference for forest vegetation change studies [3].
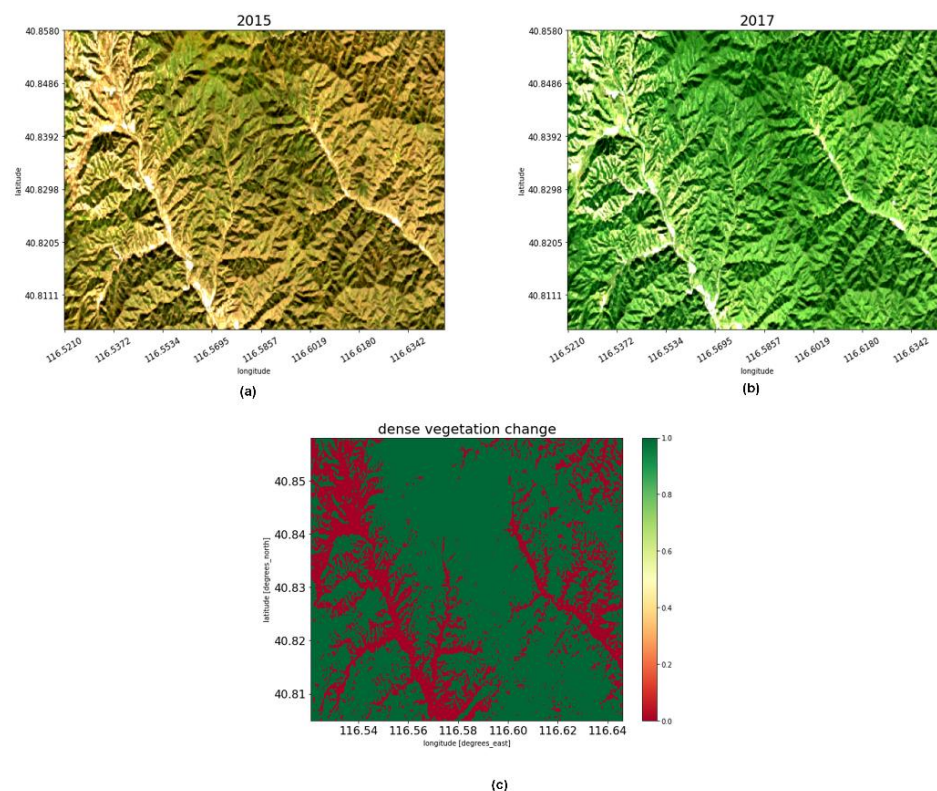
Figure 10. Result of the NDVI grass change detection between 2015 and 2017. (**a**) Partially enlarged RGB imagery view of the selected area in 2015, (**b**) partially enlarged RGB imagery view of the selected area in 2017, and (**c**) NDVI change of the selected area.

5. Discussion and Conclusions

Of particular concern to data managers is the question of how to efficiently manage the massive amounts of EO data generated every day and overcome the limitations as well as problems encountered in the downloading of data and the transmission of data providers and high-performance computing infrastructures. The freely accessible EO Data Cube, capable of bringing global ARD into local infrastructure, makes the EODC one of the most popular EO data management tools [13]. The main research purpose of this paper was to introduce the EODC, propose the design of the CDC, suitable for Chinese satellite imagery data, and then design two research cases as well as conduct related case studies.

The China Data Cube data management framework constructed in this paper has the following main features: (1) Based on the existing framework and open source code of the Open Data Cube, the CDC_DLTool middleware is designed and implemented to extend support for the data loading and processing of international and Chinese satellite imagery data covering the Chinese region. (2) In order to verify the reliability of this CDC framework, this study conducted case studies on the spatial variation in water bodies and vegetation based on Landsat 8 and GF-1 imagery data, respectively. The advantages of this research framework are mainly focused on two aspects. First, the minimum data management unit in our research framework is pixels, which realizes a paradigm shift from the traditional scene organization mode to the pixel organization mode. This facilitates the organization and management of heterogeneous satellite imagery data from multiple sources in a consistent way. Second, compared with the traditional scene-based organization, this research framework is much more efficient in data management in terms of imageries storage, imageries retrieval, and imageries processing [53]

The current CDC data management framework project is in its infancy. In the process of remote sensing big data generation and practice, there are many sources of satellite imagery data, more complex data formats, and more diverse project needs. Combining the construction of the CDC grid project with the actual needs of using and managing multisource satellite imagery data is an important challenge to be considered and solved in the future. In addition, the current phase of research has not yet considered working on a cloud platform, and future research work plans have considered improving the CDC framework by building a localized private cloud. Future research will also study and learn from existing geospatial cloud platforms such as Sentinel Hub, Google Earth Engine and Microsoft Azure, and conduct quantitative comparative analyses in terms of data storage and computation.

Author Contributions: Conceptualization, Q.C.; G.L. and X.Y.; Data curation, Q.C.; Formal analysis, Q.C.; Software, Q.C. Writing—review and editing, Y.M. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Yao, X.; Li, G.; Xia, J.; Ben, J.; Cao, Q.; Zhao, L.; Ma, Y.; Zhang, L.; Zhu, D. Enabling the Big Earth Observation Data via Cloud Computing and DGGS: Opportunities and Challenges. *Remote Sens.* **2020**, *12*, 62. [CrossRef]
2. Song, S.; Wang, S.; Ye, H.; Guan, Y. Exploratory Analysis on the Spatial Distribution and Influencing Factors of Beitang Landscape in the Shangzhuang Basin. *Land* **2022**, *11*, 418. [CrossRef]
3. Xie, J.; Hüsler, F.; Jong, R.; Chimani, B.; Asam, S.; Sun, Y.; Schaepman, M.; Kneubuehler, M. Spring Temperature and Snow Cover Climatology Drive the Advanced Springtime Phenology (1991–2014) in the European Alps. *J. Geophys. Res. Biogeosci.* **2021**, *126*. [CrossRef]
4. Han, F.; Fu, G.; Yu, C.; Wang, S. Modeling Nutrition Quality and Storage of Forage Using Climate Data and Normalized-Difference Vegetation Index in Alpine Grasslands. *Remote Sens.* **2022**, *14*, 3410. [CrossRef]
5. Xie, J.; Sun, Y.; Liu, X.; Ding, Z.; Lu, M. Human Activities Introduced Degenerations of Wetlands (1975–2013) across the Sanjiang Plain North of the Wandashan Mountain, China. *Land* **2021**, *10*, 1361. [CrossRef]
6. Liu, P. A survey of remote-sensing big data. *Front Env Sci-Switz* **2015**, *3*. [CrossRef]
7. Giuliani, G.; Camara, G.; Killough, B.; Minchin, S. Earth Observation Open Science: Enhancing Reproducible Science Using Data Cubes. *Data* **2019**, *4*, 147. [CrossRef]
8. OGC. OGC Standards and Supporting Documents. Available online: http://www.opengeospatial.org/standards/ (accessed on 22 June 2022).
9. Müller, M.S. Service-oriented Geoprocessing in Spatial Data Infrastructures. Master's Thesis, Technische Universität Dresden, Dresden, Germany, 2016.
10. Merticariu, G.; Misev, D.; Baumann, P. *Towards a General Array Database Benchmark: Measuring Storage Access*; Springer International Publishing: Toronto, ON, Canada, 2015; pp. 40–67.
11. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–22. [CrossRef]
12. Open Data Cube. Available online: https://www.opendatacube.org/ (accessed on 22 June 2022).
13. Sudmanns, M.; Augustin, H.; Killough, B.; Giuliani, G.; Tiede, D.; Leith, A.; Yuan, F.; Lewis, A. Think global, cube local: An Earth Observation Data Cube's contribution to the Digital Earth vision. *Big Earth Data* **2022**, 1–29. [CrossRef]
14. Xu, C.; Du, X.; Jian, H.; Dong, Y.; Qin, W.; Mu, H.; Yan, Z.; Zhu, J.; Fan, X. Analyzing large-scale Data Cubes with user-defined algorithms: A cloud-native approach. *Int. J. Appl. Earth Obs.* **2022**, *109*, 102784. [CrossRef]
15. Yan, J.; Liu, Y.; Wang, L.; Wang, Z.; Huang, X.; Liu, H. An Efficient Organization Method for Large-Scale and Long Time-Series Remote Sens. Data in a Cloud Computing Environment. *IEEE J.-Stars* **2021**, *14*, 9350–9363. [CrossRef]
16. Yan, J.; Ma, Y.; Wang, L.; Choo, K.-K.R.; Jie, W. A cloud-based Remote Sens. data production system. *Future Gener. Comput. Syst.* **2018**, *86*, 1154–1166. [CrossRef]

17. Baumann, P.; Dehmel, A.; Furtado, P.; Ritsch, R.; Widmann, N. The Multidimensional Database System RasDaMan. *Acm. Sigmod. Record* **1998**, *27*, 575–577. [CrossRef]

18. Stonebraker, M.; Rogers, J.; Battle, L.; Papaemmanouil, O. SciDB DBMS Research at MIT. *IEEE Data Eng. Bull.* **2013**, *36*, 21–30.

19. Dhu, T.; Dunn, B.; Lewis, B.; Lymburner, L.; Phillips, C. Digital earth Australia—Unlocking new value from earth observation data. *Big Earth Data* **2017**, *1*, 64–74. [CrossRef]

20. Krause, C.E.; Newey, V.; Alger, M.J.; Lymburner, L. Mapping and Monitoring the Multi-Decadal Dynamics of Australia's Open Waterbodies Using Landsat. *Remote Sens.* **2021**, *13*, 1437. [CrossRef]

21. Malthus, T.J.; Lehmann, E.; Ho, X.; Botha, E.; Anstee, J. Implementation of a Satellite Based Inland Water Algal Bloom Alerting System Using Analysis Ready Data. *Remote Sens.* **2019**, *11*, 2954. [CrossRef]

22. Lucas, R.; Mueller, N.; Siggins, A.; Owers, C.; Clewley, D.; Bunting, P.; Kooymans, C.; Tissott, B.; Lewis, B.; Lymburner, L. Land cover mapping using digital earth Australia. *Data* **2019**, *4*, 143. [CrossRef]

23. Lewis, A.; Lymburner, L.; Purss, M.B.J.; Brooke, B.; Evans, B.; Ip, A.; Dekker, A.G.; Irons, J.R.; Minchin, S.; Mueller, N.; et al. Rapid, high-resolution detection of environmental change over continental scales from satellite data—The Earth Observation Data Cube. *Int. J. Digit. Earth* **2016**, *9*, 106–111. [CrossRef]

24. Brooke, B.; Lymburner, L.; Lewis, A. Coastal dynamics of Northern Australia–Insights from the Landsat Data Cube. *Remote Sens. Appl.* **2017**, *8*, 94–98. [CrossRef]

25. Chatenoux, B.; Richard, J.P.; Small, D.; Roeoesli, C.; Wingate, V.; Poussin, C.; Rodila, D.D.; Peduzzi, P.; Steinmeier, C.; Ginzler, C. The Swiss data cube, analysis ready data archive using earth observations of Switzerland. *Sci. Data* **2021**, *8*, 295. [CrossRef] [PubMed]

26. Honeck, E.; Castello, R.; Chatenoux, B.; Richard, J.-P.; Lehmann, A.; Giuliani, G. From a Vegetation Index to a Sustainable Development Goal Indicator: Forest Trend Monitoring Using Three Decades of Earth Observations across Switzerland. *ISPRS Int. J. Geo.-Inf.* **2018**, *7*, 455. [CrossRef]

27. Giuliani, G.; Chatenoux, B.; Piller, T.; Moser, F.; Lacroix, P. Data Cube on Demand (DCoD): Generating an earth observation Data Cube anywhere in the world. *Int. J. Appl. Earth Obs.* **2020**, *87*, 102035. [CrossRef]

28. Giuliani, G.; Chatenoux, B.; Bono, A.D.; Rodila, D.; Richard, J.P.; Allenbach, K.; Dao, H.; Peduzzi, P. Building an Earth Observations Data Cube: Lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD). *Big Earth Data* **2017**, *1*, 18. [CrossRef]

29. Killough, B. The impact of analysis ready data in the Africa regional data cube. In Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5646–5649.

30. Yuan, F.; Repse, M.; Leith, A.; Rosenqvist, A.; Milcinski, G.; Moghaddam, N.F.; Dhar, T.; Burton, C.; Hall, L.; Jorand, C.; et al. An Operational Analysis Ready Radar Backscatter Dataset for the African Continent. *Remote Sens.* **2022**, *14*, 351. [CrossRef]

31. Yuan, F.; Lewis, A.; Leith, A.; Dhar, T.; Gavin, D. Analysis Ready Data for Africa. In Proceedings of the 2021 IEEE International Geoscience and Remote Sens. Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 1789–1791.

32. Mubea, K.; Mfundisi, K.; Yuan, F.; Burton, C.; Boamah, E. Analysing Effects of Drought on Inundation Extent and Vegetation Cover Dynamics in the Okavango Delta. In Proceedings of the AGU Fall Meeting Abstracts, New Orleans, LA, USA, 1 December 2021; p. 0652.

33. Halabisky, A.M.; Mubea, K.; Mar, F.; Yuan, F.; Burton, C.; Birchall, E.; Moghaddam, N.F.; Adimou, G.; Mamane, B.; Ongo, D.; et al. Water Observations from Space: Accurate maps of surface water through time for the continent of Africa. *ESSOAr* **2021**, *9*. [CrossRef]

34. Burton, C.; Yuan, F.; Chong, E.-F.; Halabisky, M.; Ongo, D.; Mar, F.; Addabor, V.; Mamane, B.; Adimou, S. Co-Production of a 10 m Cropland Extent Map for Continental Africa using Sentinel-2, Cloud Computing, and the Open Data Cube. *J AGU Fall Meeting Abstracts* **2021**, 0924. [CrossRef]

35. Lewis, A.; Oliver, S.; Lymburner, L.; Evans, B.; Wyborn, L.; Mueller, N.; Raevksi, G.; Hooke, J.; Woodcock, R.; Sixsmith, J. The Australian Geoscience Data Cube—Foundations and lessons learned. *Remote Sens. Environ.* **2017**, 276–292. [CrossRef]

36. Xu, D. *Research on the Key Techniques of Multi-source Remote Sens. Big Data Management under the Cloud Computing Environment*; University of Chinese Academy of Sciences: Beijing, China, 2018.

37. Unidata | NetCDF. Available online: https://www.unidata.ucar.edu/software/netcdf/ (accessed on 1 March 2022).

38. PostgreSQL: The world's most advanced open source database. Available online: https://www.postgresql.org/ (accessed on 1 June 2022).

39. Yao, X.; Liu, Y.; Cao, Q.; Li, J.; Huang, R.; Woodcock, R.; Paget, M.; Wang, J.; Li, G. China Data Cube (CDC) for Big Earth Observation Data: Lessons Learned from the Design and Implementation. In Proceedings of the 2018 International Workshop on Big Geospatial Data and Data Science (BGDDS), Wuhan, China, 22–23 September 2018; pp. 1–3.

40. Ross, J.; Killough, B.; Dhu, T.; Paget, M. *Open Data Cube and the Committee on Earth Observation Satellites Data Cube Initiative*; IAC: Adelaide, Australia, 2017; Volume 17, p. 6.

41. Lewis, A.; Lacey, J.; Mecklenburg, S.; Ross, J.; Hosford, S. CEOS Analysis Ready Data for Land (CARD4L) Overview. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 7407–7410.

42. Dwyer, J.L.; Roy, D.P.; Sauer, B.; Jenkerson, C.B.; Lymburner, L. Analysis ready data: Enabling analysis of the landsat archive. *Remote Sens.* **2018**, *10*, 1363. [CrossRef]

43. San A, B. Evaluation of different Atmospheric Correction Algorithms for EO-1 Hyperion Imagery. 2010. Available online: https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.222.1799 (accessed on 22 June 2022).
44. Yongquan, Z.; Xiaojun, S.; Ping, T. Spatial Consistency Analysis and Relative Geometric Correction of Low Spatial Resolution Multi\|source Remote Sens. Data. *Remote Sens. Technol. Appl.* **2014**, *29*, 155–163.
45. The Official YAML Web Site. Available online: https://yaml.org/ (accessed on 1 June 2022).
46. Yinghu, L.I.; Baoshan, C.; Zhifeng, Y. Influence of hydrological characteristic change of Baiyangdian on the ecological environment in wetland. *J. Nat. Resour.* **2004**, *19*, 62–68. [CrossRef]
47. Zhuo, L.A.; Wja, B.; Ww, C.; Zheng, C.C.; Zl, B.; Jl, A. Ecological risk assessment of the wetlands in Beijing-Tianjin-Hebei urban agglomeration—ScienceDirect. *Ecol. Indic.* **2020**, *117*. [CrossRef]
48. Louati, M.; Saidi, H.; Zargouni, F. Shoreline change assessment using Remote Sens. and GIS techniques: A case study of the Medjerda delta coast, Tunisia. *Arab. J. Geosci.* **2015**, *8*, 4239–4255. [CrossRef]
49. Alesheikh, A.A.; Ghorbanali, A.; Nouri, N. Coastline change detection using Remote Sensing. *Int. J. Environ. Sci. Technol.* **2007**, *4*, 61–66. [CrossRef]
50. Durduran, S.S. Coastline change assessment on water reservoirs located in the Konya Basin Area, Turkey, using multitemporal landsat imagery. *Environ. Monit Assess* **2010**, *164*, 453–461. [CrossRef] [PubMed]
51. Mueller, N.; Lewis, A.; Roberts, D.; Ring, S.; Melrose, R.; Sixsmith, J.; Lymburner, L.; McIntyre, A.; Tan, P.; Curnow, S.; et al. Water observations from space: Mapping surface water from 25years of Landsat imagery across Australia. *Remote Sens. Environ.* **2016**, *174*, 341–352. [CrossRef]
52. Available online: http://www.bjhr.gov.cn/ywdt/mtgz/202106/t20210603_2404698.html (accessed on 22 June 2022).
53. Cao, Q.; Li, G.; Yao, X.; Jia, T.; Yu, G.; Zhang, L.; Xu, D.; Zhang, H.; Shan, X. GF-1 Satellite Imagery Data Service and Application Based on Open Data Cube. *Appl. Sci.* **2022**, *12*, 7816. [CrossRef]