Computers & Geosciences 92 (2016) 1–8

Contents lists available at ScienceDirect

Computers & Geosciences

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

Case study

Investigating metrics of geospatial web services: The case of a CEOS federated catalog service for earth observation data

Weiguo Han [a,*], Liping Di [b], Genong Yu [b], Yuanzheng Shao [c], Lingjun Kang [b]

[a] Visiting Scientist Programs, University Corporation for Atmospheric Research, 3090 Center Green Drive, Boulder, CO 80301, USA
[b] Center for Spatial Information Science and Systems, George Mason University, 4400 University Drive MS 6E1, Fairfax, VA 22030, USA
[c] Collaborative Innovation Center of Geospatial Technology, Wuhan University, 129 Luoyu Road, Wuhan, Hubei 430079, China

A R T I C L E I N F O

Article history:
Received 25 November 2014
Received in revised form
17 February 2016
Accepted 8 April 2016
Available online 9 April 2016

Keywords:
Federated catalog service
Catalog service for the web
Geospatial web services
Web service metrics
Earth observation

A B S T R A C T

Geospatial Web Services (GWS) make geospatial information and computing resources discoverable and accessible over the Web. Among them, Open Geospatial Consortium (OGC) standards-compliant data, catalog and processing services are most popular, and have been widely adopted and leveraged in geospatial research and applications. The GWS metrics, such as visit count, average processing time, and user distribution, are important to evaluate their overall performance and impacts. However, these metrics, especially of federated catalog service, have not been systematically evaluated and reported to relevant stakeholders from the point of view of service providers. Taking an integrated catalog service for earth observation data as an example, this paper describes metrics information retrieval, organization, and representation of a catalog service federation. An extensible and efficient log file analyzer is implemented to retrieve a variety of service metrics from the log file and store analysis results in an easily programmable format. An Ajax powered Web portal is built to provide stakeholders, sponsors, developers, partners, and other types of users with specific and relevant insights into metrics information in an interactive and informative form. The deployed system has provided useful information for periodical reports, service delivery, and decision support. The proposed measurement strategy and analytics framework can be a guidance to help GWS providers evaluate their services.

1. Introduction

Open Geospatial Consortium (OGC) has made a series of specifications for geospatial Web services (GWS) covering discovery, access, portrayal, and processing of geospatial data. The specifications, like Web Map Service (WMS), Web Coverage Service (WCS), Web Feature Service (WFS), Web Processing Service (WPS), and Catalog Service for the Web (CSW), have been extensively adopted and implemented in the industry and academic community. The OGC standard-compliant services contributed to make geospatial data and computing resources discoverable and accessible over the Web, and greatly facilitated sharing and interoperability of geospatial information from distributed sources. The wide adoption of OGC services raised the need to monitor and evaluate performance of GWS from both service providers and consumers.

As an international interagency organization, Committee on Earth Observation Satellite (CEOS) coordinates satellite Earth Observation (EO) programs between space agencies of its member countries. The goal of CEOS Working Group on Information Systems and Services (WGISS) is to provide EO data management systems and services to worldwide users. Initiated by National Aeronautics and Space Administration (NASA) and National Oceanic and Atmospheric Administration (NOAA) in 2010, CEOS WGISS Integrated Catalog (CWIC) project aims to build a federated catalog system which provides inventory level search results from EO data catalog systems of CEOS members through a standard unified interface (Enloe and Yapur, 2011). Currently, the operational CWIC service (http://cwic.wgiss.ceos.org) has integrated EO data catalog systems from NASA Earth Observing System (EOS) Clearing House (ECHO), NOAA Group for High Resolution Sea Surface Temperature (GHRSST), United States Geological Survey (USGS) Land Surface Imaging (LSI), Brazil National Institute for Space Research (Instituto Nacional de Pesquisas Espaciais, INPE), and Canada Centre for Mapping and Earth Observation (CCMEO). It has been served as one of data sources for the Societal Benefit Area (SBA) of Agriculture and Disasters activities in the Global Earth Observation System of Systems (GEOSS) Architecture Implementation Pilot (AIP) projects (Percivall et al., 2013).

* Corresponding author.
E-mail address: hanwg.bj@gmail.com (W. Han).

http://dx.doi.org/10.1016/j.cageo.2016.04.005
0098-3004/

CWIC project stakeholders, sponsors, developers, and partners are most concerned with the questions about metrics of this integrated catalog service, such as "*How many users have accessed CWIC service in the past month?*", "*How long does it take CWIC service to handle GetRecordById request?*", "*Which data collections or datasets of USGS LSI are requested most often?*", "*Which countries are CWIC users from?*", and so on. Such metrics information should be collected, organized, and presented for evaluation of overall performance and impacts of CWIC service. This paper will address these and other related requirements. To make the process smoother and easier, a log file is necessary to track request processing for metrics information collection, a log file parser is needed to retrieve metrics information from the log file and organize them in an easy-to-display format, and a web dashboard is required to represent metrics information in informative tables and charts. And this mechanism can also be extended and applied to other similar Web services.

The reminder of this paper is organized as the following. Section 2 reviews the progresses in evaluation of GWS metrics and introduces the system requirements. Section 3 presents technical approaches in the implementation of CWIC metrics information collection, extraction, and representation. In Section 4, system functions are demonstrated to provide detailed and actionable insights into CWIC metrics. Section 5 discusses the experiences from metrics monitoring and analysis of CWIC. Finally, Section 6 summarizes conclusions and directions for future work.

2. Related works

2.1. Geospatial web services

Many organizations have followed OGC standards to publish their geospatial data, information, and services in an open and interoperable way. These standards are broadly grouped into three categories, i.e. data, processing, and catalog.

Geospatial data services support geospatial data customization and retrieval according to input parameters. For example, WCS services provide access to geographical coverages through standard operations (Whiteside and Evans, 2008), WFS services offer vector data manipulation and retrieval (Vretanos, 2002), and WMS services handle geospatial data rendering and portrayal (de la Beaujardière, 2006). These data services are well supported in both commercial and open source GIS software. The issues related to quality of this kind of service have been addressed by many researchers. From the perspective of service consumers, Zhang et al. (2010) used metrics of precision and recall to evaluate WFS query results; Horák et al. (2011) measured response time, error occurrence, availability, and performance of WMS services by repeating same requests; Wu et al. (2011) presented a new approach to monitor and assess quality of WMS services and developed a mechanism to choose better map layers for decision making support; Gui et al. (2013) leveraged Geospatial Cyber-infrastructure (GCI) components to build a search engine framework for geospatial resources discovery and registry, and developed a quality monitoring and evaluation module to assess accessibility and performance of registered OGC data services. In addition, Giuliani et al. (2013) proposed a new approach to evaluate performance of WFS and WCS services on the server side and provided service providers with guidance on service quality improvement.

Geospatial processing services offer operations for geospatial data transformation and processing derived from geospatial models, algorithms, and applications. The WPS specification defines standard interfaces for discovery of, publishing of, and binding to geospatial process (Schut, 2007), so WPS services can be composed in scientific workflows to perform complex tasks over distributed geospatial resources (Cepicky and Becchi, 2007; Kiehle et al., 2007). Measurements of data transfer fluency and processing control in the workflow can be used to evaluate quality and performance of interoperability (Gorgan et al., 2012). Scholten et al. (2006) analyzed four performance-related factors for geoprocessing services, including caching, network adaptation, data granularity, and communication mode. Sun et al. (2012) developed a prototype system called GeoPWTManager to chain geo-processing services and monitor and visualize performance of these services.

Geospatial catalog services provide geospatial information registry, description, discovery, and access. OGC CSW specification defines standard interfaces to register, publish and search geospatial data, information, and services in the metadata catalogs (Voges and Senkler, 2005). The general query criteria contain spatial extent, temporal range, and dataset identifier, etc. This specification has been adopted and implemented in many applications, like GeoBrain Catalog Federation service (Bai et al., 2007), GEOSS Component and Service Registry (Bai et al., 2012), Group on Earth Observations (GEO) Discovery and Access Broker (DAB) (Nativi et al., 2013), *GeoNetwork* (http://geonetwork-opensource.org), and *deegree* (http://www.deegree.org). However, in comparison with the other two categories of geospatial service, less effort has been devoted to measure quality of catalog services which is critical for both service providers and end users, and there have been fewer publications on this topic so far.

2.2. Web analytics

Web analytics tools collect and display metrics of a website or Web application, and give powerful indicators on its performance, capacity, and availability. In these tools, analytical statistics are performed on the measurements of each aspect of a website or web application to provide information like visits, ranking, and processing time on traffic history. The metrics results are presented in a detailed web traffic dashboard with interactive tables and colorful graphs for decision making support.

Server log file analysis and page tagging are two common technical solutions on visit information collection. In the former method, the log file or database collecting web activities is parsed and analyzed through self-hosted web analytics software. AWStats (http://awstats.sourceforge.net) is an open source web analytics application for processing visit information from the server log file and presenting them visually within static HTML reports. Other open source alternatives to AWStats are Analog, Webalizer, and W3Perl. Google Analytics (http://www.google.com/analytics), belonging to the latter approach, is one of the most popular web analytics programs in the world today. Users only need to embed a snippet of JavaScript tracking code in their web pages, Google Analytics will help them track visitors along with their activities from browser cookies and learn full pictures of their websites, such as where visitors are from and where web traffic comes from. Other free web analytics services from different vendors include Yahoo! Web Analytics (http://web.analytics.yahoo.com/), Bing Webmaster (http://www.bing.com/toolbox/webmaster), Quantcast Measure (https://www.quantcast.com), etc.

Miller et al. (2002) analyzed service metrics using a Quality of Service (QoS) model with three dimensions (i.e. time, cost, and quality). Giuliani et al. (2013) pointed out that quality of downloading services like geospatial data services should be evaluated using three criteria: 1) performance, 2) capacity, and 3) availability, and these criteria can be measured on either server side or client side. Khaled et al. (2010) proposed to enhance metadata information on quality using ISO 19119 standard, and suggested that the GWS quality should be evaluated based on spatial data quality (currency, freshness, timeless, etc), web service quality (execution time, latency, throughput), and general quality (consultation, rating, citation, etc). In addition to related studies in Section 2.1, some free or commercial web applications have been built to provide health and quality information on GWS. For example, Service Status Checker (SSC, http://registry.fgdc.gov/statuschecker/index.php), provided by Federal Geographic Data Committee (FGDC), performs health tests on multiple kinds of GWS like OGC WCS, WMS, WFS, CSW, and Sensor Observation Service (SOS), and data services from ArcGIS Servers and ArcIMS servers. Spatineo (http://www.spatineo.com) is another great Web application for monitoring and reporting quality and performance of registered geospatial data services. Strictly speaking, these GWS monitoring and reporting applications are limited to evaluating service performance and availability through sending test requests and checking responses at a regular interval (e.g. 10 min), so these applications actually act as consumers of the evaluated service and burden them with unnecessary test requests, so a GWS monitoring system from the perspective of service provider would be better to analyze service quality.

For a federated catalog system like CWIC, the catalog provider requires specific metrics information that can be aggregated and displayed visually by different categories (like data collection, operation, or date) on its service requested by the system users. With new catalogs or data collections integrated, it is desired that the defined metrics for them can be presented easily and directly. The above mentioned Web analytics systems can't meet these requirements fully, so it is necessary to build an applicable metrics application for the federated catalog service.

2.3. CWIC – federated catalog service

With development and popularization of EO technologies, many national and international organizations have built their own catalog systems for EO satellite data collection, registry, discovery and access. In many Earth Science research and applications, EO data of interest must be queried and obtained from these heterogeneous systems across organizations. A catalog federation system providing a standard-based and consistent interface can allow users to search and access EO data across distributed sources. Promoted by CEOS, CWIC service is such a system, which aims to federate major EO data catalogs from CEOS member agencies and provide community portals (or clients) a single point of entry to the participant catalogs (Enloe and Yapur, 2011).

The mediator-wrapper architecture is leveraged in the design and implementation of CWIC. OGC CSW serves as the standard of query interface, which stipulates the protocol between CWIC client and CWIC server. Detailed description of CWIC architecture and technical implementation can be found in the previous publication (Shao et al., 2013).

CWIC service has been consumed by project partners and other users in their applications. A CSW-complaint user interface named CWIC Start (https://api.echo.nasa.gov/cwic-start) has been developed from NASA ECHO Reverb web client to search and access EO data for a variety of Earth Science disciplines and domains through CWIC (Farley et al., 2011). Other two operational CWIC clients are USGS LSI Explorer (http://lsiexplorer.cr.usgs.gov/) and GeoBrain Online Analysis System (GeOnAS) (Han et al., 2011). CWIC service also contributed to the GEOSS AIP activities as a mediated access enabler (Percivall et al., 2013).

2.4. CWIC metrics requirements

With more integrated catalogs/datasets and more client requests, it is desired by CWIC stakeholders (participant agencies, like NASA and NOAA) to build a mechanism to evaluate reliability and quality of CWIC service and its integrated catalog services. They also would like to know who are requesting data from CWIC and how many users are using these data, so do the catalog providers. It is required that CWIC requests can be aggregated and ranked by different categories (e.g. catalog or dataset). Moreover, these metrics are expected to be presented interactively and responsively through a Web application. These requirements are collected directly from the monthly CWIC team telecons and the annual CWIC Developer's meetings, and discussed and refined intensively at the weekly CWIC development team tag-up telecons.

Previously, a Java program was developed to analyze historical CWIC service requests, and monthly statistics were manually extracted from the program output and filled in a monthly report that was finally submitted to project stakeholders and sponsors for review. Obviously, this non-automated process is time consuming and laborious, so an efficient and effective approach is proposed in next section to monitor client accesses in a fully automatic way and present metrics information to various types of users in a more user-friendly web presentation.

3. Technical approach

Monitoring metrics associated with various requesting features is an important measure of GWS's facility. A common web metrics system for CWIC is required by both stakeholders and catalog providers.

General architecture of CWIC is shown in the lower part of Fig. 1, *CWIC Mediator* acts as a broker between CWIC service and integrated catalog services, each wrapper (or connector) in it is developed to mediate communication with the corresponding catalog service. More specific information about wrapper implementation can be perused in the paper presented by Shao et al. (2013). Because the integrated catalog services are heterogeneous and don't only serve CWIC project, it is not easy for catalog providers to extract CWIC request information from their log files with different format and submit them to CWIC develop team for summarization. And it is desired that CWIC development team can provide catalog providers with monitoring information about their running catalog services for CWIC. Therefore it makes sense to build a comprehensive web analysis system with CWIC which can collect and present metrics seamlessly.

CWIC Mediator is implemented as a *Servlet*, so the method of page tagging is not applicable and the other method, log file analysis, is adopted. Thereby, *CWIC Mediator* also serves as a log generator for metrics information collection. The processing flow is shown in the upper part of Fig. 1.

3.1. Metrics definition

To address the requirements, the following metrics are tracked and analyzed:

1) The number of requests, which means that how many requests are processed (successfully or not) during the specified period. The requests can be organized and aggregated by various attributes, like catalog, dataset, or day of week;
2) The processing time of request, which indicates that how much time the service takes to process individual request. The maximum, minimum, and average values can be calculated;
3) The country of user, which represents where the user requesting the service is from, so the user distribution can be presented in a map.

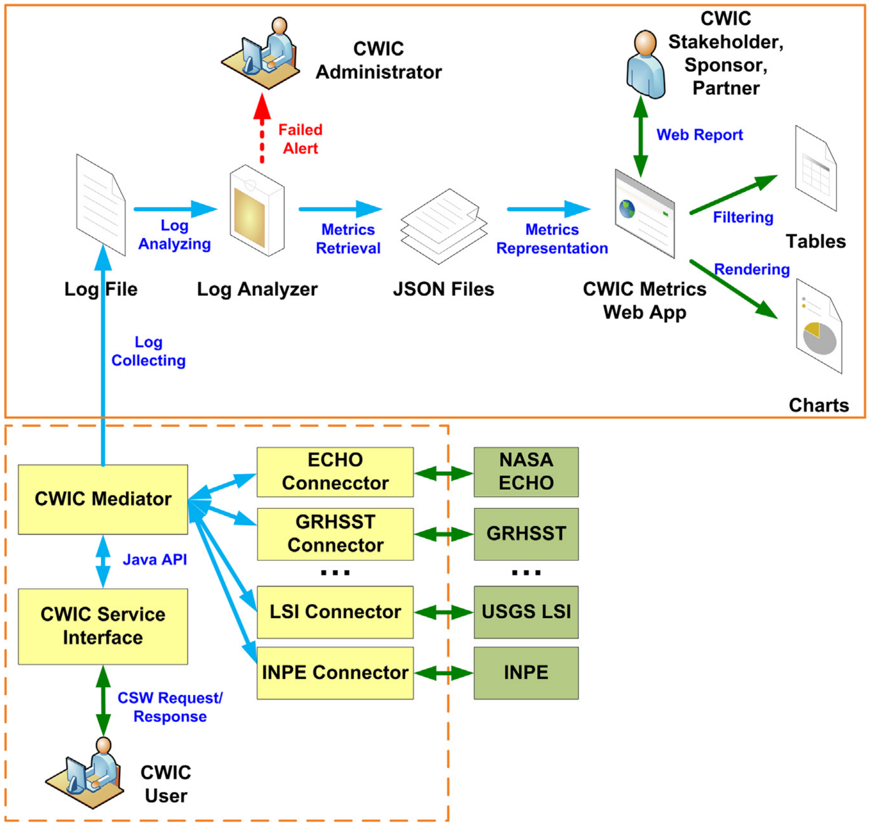


Fig. 1. Processing flow of CWIC metrics application.

3.2. *Metrics information collection*

Visit information of CWIC service are collected and stored in a separate log file by *CWIC Mediator*. *CWIC Mediator* is deployed within Web application server (e.g. Apache Tomcat). It is not practical to extract requesting information from the large log file of application server because the file includes more non-related entries. *CWIC Mediator* parses standard CSW request (via *HTTP GET* or *POST*) from client, then retrieves and transfers required information to the corresponding connector. The highly optimized open source logging Java library (i.e. *Apache Log4j*) is leveraged to track and collect processing information, it doesnot impose performance cost on *CWIC Mediator* at runtime.

When processing each CWIC request, the visiting information is organized as one complete record in the format (as seen in Fig. 2) to derives the metrics defined in Section 3.1. *CWIC_STARTDATETIME* and *CWIC_FINIHSDATETIME* are recorded as the time of beginning and ending of request processing. The information on operation, dataset identifier, output element set type name, output schema type, and catalog can be obtained from *CWIC_OPERATIONS*, *CWIC_DATASETID*, *CWIC_ELEMENTSET*, *CWIC_TYPENAME*, and *CWIC_CATALOG* respectively for aggregation by these categories. The Internet Protocol (IP) address from *CWIC_IP* is utilized to find user's approximate geographic location that is sufficient for the needs. *CWIC _STATUS* indicates whether the request is processed successfully or not.

3.3. *Metrics information retrieval*

Analysis must be performed over time to determine trends and patterns of service requesting. A log file analyzer named *CWICLogParser* is developed to retrieve, store, and update metrics information from the log file. To avoid overloading memory and repetitive analysis, the *RandomAccessFile* class in the Java Input and

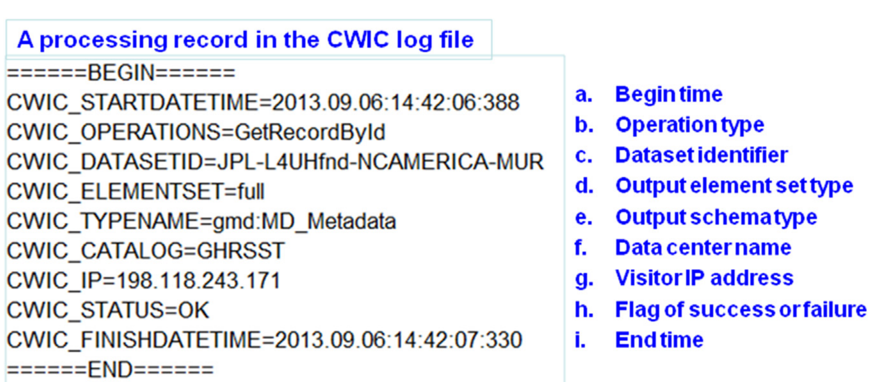


Fig. 2. Log file format.

Output (I/O) API is utilized to set the start position to the end one of last analysis and parse the logging record from this position.

In *CWICLogParser*, each log record representing a request is processed to retrieve the related metrics from the record elements shown in Fig. 2; the processing time can be calculated from the time of beginning and ending easily; *GeoIP Java* API (https://github.com/maxmind/geoip-api-java) is utilized to identify location information of CWIC client from its IP address. Several Java classes are defined to hold the requesting information for catalog (*NASA ECHO*, *USGS LSI*, *GHRSST*, *INPE*, and *CCMEO*), operation (*GetCapabilities*, *GetRecords*, and *GetRecordById*), output (output schema: *ISO 19115* and *ebRIM CSW*, output element set name: *brief*, *summary*, and *full*), country, basic dataset, and detailed dataset respectively.

JSON (JavaScript Object Notation) is a lightweight data-interchange language that can be easily parsed and created by different programming languages. JSON supports two kinds of structure, object (an unordered set of name/value pairs) and array (an ordered collection of values) (Crockford, 2011). In *CWICLogParser*, *JSON.simple* package is used to encode and decode metrics information in JSON format. Taking catalog access as an example, each data point is of the format like the following:


```
{
    "date": "2013.12.11",
    "USGSLSI": {
        "failure": 4,
        "success": 15,
        "visits": 19
    },
    "GHRSST": {
        "failure": 0,
        "success": 21,
        "visits": 21
    },
    "NASAECHO": {
        "failure": 0,
        "success": 68,
        "visits": 68
    }
}
```


Here, *date* indicates the day to which the record applies, *failure* means the numbered times of failed processing of that day, *success* represents the ones of successful processing of that day, and *visits* is the total times of accesses to the specified catalog of that day.

The utility of *CWICLogParser* is highly flexible and extensible. When new catalog from project partners is added to *CWIC Mediator* or new dataset is found from the log file, *CWICLogParser* will retrieve metrics information on this catalog or dataset and add them into the corresponding JSON files. This utility is executed automatically at 2:00 am EST every day on CWIC server to obtain metrics information during the past day, or is triggered manually to run at any time to get real-time updates. The latest metrics information will be merged with the existing JSON files. The JSON files will be loaded and visualized in a web graphic user interface (GUI) for review.

3.4. Metrics information representation

A web dashboard-like portal can provide useful insights to trends of accesses to a website or web application. CWIC metrics portal (http://cwic.wgiss.ceos.org/cwicmetrics/) is deployed to display metrics of CWIC service running on the same server. The interactive charts and tables in the portal help understand response time and other performance information correlated with user activities. Ext JS (http://www.sencha.com/products/extjs/), which is adopted in the implementation of CWIC metrics portal, provides a set of components, utilities, and APIs that help create web sites and applications more easily and flexibly. For example, the *GridPanel* objects show data stores in a tabular layout and sort the lists according to users' needs. The *Line Chart*, *Bar Chart*, and *Pie Chart* give quantitative information calculated from data stores in various types of graphs. In addition, Google *Geochart* is utilized to display geographic distribution of CWIC users in a world map. These components will be presented in the demonstrations of next section.

4. CWIC metrics demonstration

Metrics provides a set of quantitative indicators of how well a geospatial Web service serves current and potential applications. These quantity measures can be collected and calculated from each service request by resource monitoring and evaluation on the server side. CWIC metrics web portal shows metrics information during the past 30 days by default when loading it in the browser. Taking the period (2/1/2014–10/31/2014) as an example, the following sections demonstrate system capabilities using metrics information of this instance.

4.1. Service request trends

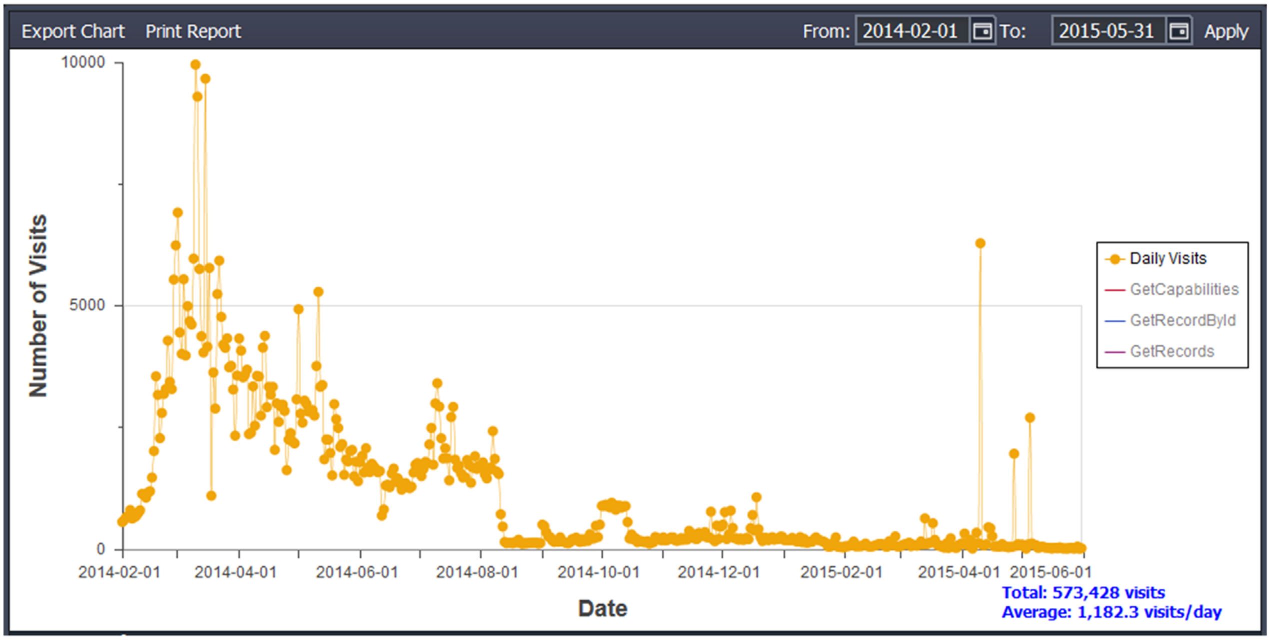
GWS metrics portal provides a more complete picture of how many average requests the service handles every day, where users are from and how many times they request each service operation.

Overview graph provides an overall view of request peaks and valleys of geospatial Web service during the specified time period. Fig. 3(a) reveals CWIC request trend from February 1, 2014 to May 31, 2015: CWIC service processed more than 570,000 CSW requests in these sixteen months and about 1180 requests per day. The request trend line for each operation can be displayed or hidden in the chart. The unexpected spike on early March, 2014 is caused by concurrent testing from CWIC client applications. An obvious decrease from middle August, 2014 should be investigated further for possible reasons.

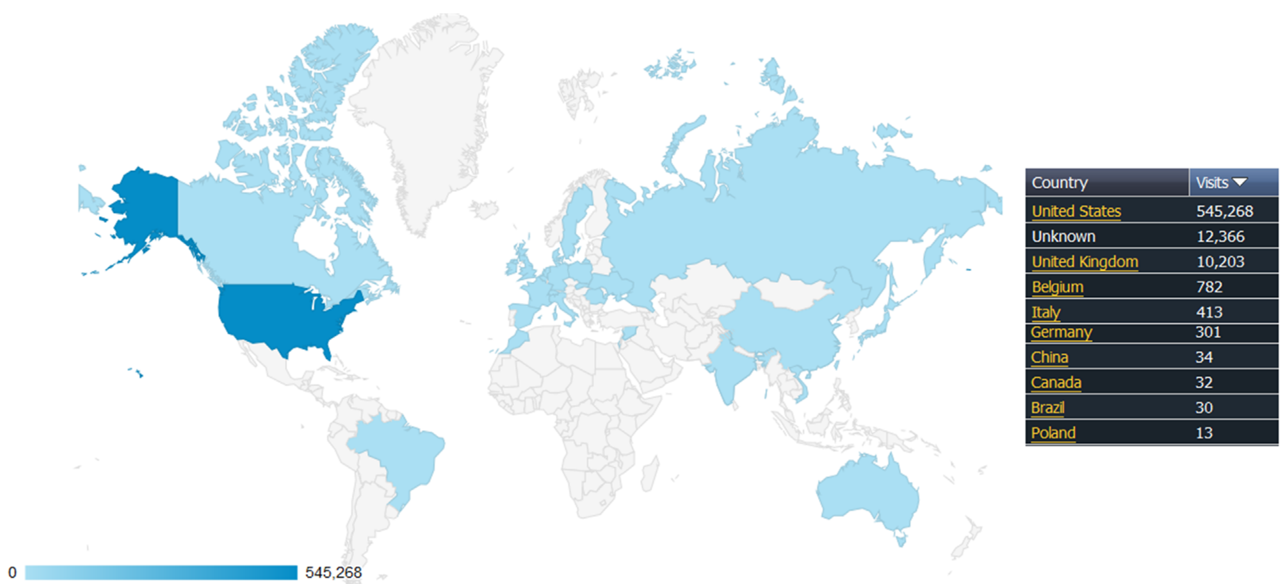
User distribution map and source country list in Fig. 3(b) illustrate that CWIC users are mainly from CEOS member countries, including United States, United Kingdom, Belgium, Italy, Germany, China, Canada, and Brazil (Unknown represents those countries that cannot be identified from IP addresses)., CWIC service should be promoted more by team members within and beyond the CEOS community in the future. In addition, user distribution can be displayed in pie chart and bar chart. Users can drill down into each catalog or dataset to view detailed information on its visits through the publicly accessible CWIC Metrics portal.

4.2. Operation performances

CWIC Metrics portal also delivers indicators of service performance, like success and failure count, average, minimum, and maximum processing times, which measure the effectiveness of each service operation, and offers a trend chart by day of week to visualize on what days of the week CSW requests for each operation are submitted and how much average processing times CWIC service process these requests on a weekly basis to detect CWIC users' behavior patterns.



(a).CWIC Request Trend



(b). CWIC user distribution

Fig. 3. CWIC Request Overview in 2013. (a) CWIC Request Trend (b) CWIC user distribution.

Table 1
CWIC operations performance.

Catalog	Datasets*	Granules*	Average processing times			Total requests	Successful ratio (%)
			GetRecords	*GetRecordById*	*Overall*		
NASA ECHO	1781	64,258,572	7.82s	5.68s	6.53s	284,726	99.2
USGS LSI	64	6,916,116	2.81s	1.33s	2.01s	93,186	82.19
GHRSST	65	1,047,094	3.37s	1.68s	2.39s	77,799	87.88
INPE	15	831,006	1.55s	0.73s	1.10s	37,263	97.17
CCMEO	2	1,419,047	12.57s	N/A	12.57s	60	35.0

* *Note*: All numbers as of 10/31/2014.

AS shown in Table 1, the columns of datasets and granules provide the numbers of datasets and granules which catalog providers register in Global Change Master Directory (GCMD) with the tag of CWIC. The metadata of each dataset can be fetched from *GetCapabilities* request like http://cwic.wgiss.ceos.org/cwicv1/discovery?service=CSW&request=GetCapabilities&version=2.0.2. Table 1 also gives average processing times of *GetRecords* and *GetRecordById*, total requests, and successful ratio during the defined period. It should be noted that the processing times mainly depend on the processing capacity of the accessed catalog servers referring to the CWIC federated catalog architecture as seen in the lower part of Fig. 1. These statistical results can be dynamically calculated and obtained from CWIC Metrics Web page for each individual catalog. Although average processing times of NASA ECHO requests are a little longer than ones of other catalogs, this catalog offers 1781 datasets and more than 64 million granules along with the highest successful ratio. The number of failed requests of USGS LSI and GHRSST are higher. CWIC development team and catalog providers should coordinate and check the issues out together. The INPE catalog delivers the fastest response time as well as a better successful ratio. The CWIC interface of CCMEO is still in trial phase, only a few requests were handled during this period. These results can be used to inform resource allocation decisions and monitor service improvements.

4.3. Dataset popularity

Metrics also answer the question of how popular integrated catalogs and datasets are. These information are very useful and helpful for deployment choice of CWIC service and future enhancement of the participant catalog services. More than 85% datasets requested by CWIC service are from NASA ECHO, which ranks No. 1 among CWIC data providers during the specified period, followed by USGS LSI and GHRSST, as seen the column of Total Requests in Table 1.

The top 10 datasets requested by CWIC users are listed in Table 2. MODIS satellite data products are most popular among CWIC users. The MODIS/Terra Snow Cover Daily L3 Global 0.05° CMG (MOD10C1) data set, which provides a global snow cover at a 0.05° resolution in Hierarchical Data Format-Earth Observing System (HDF-EOS) format, is the most requested one from NASA ECHO during the months of February to October, 2014. GHRSST's sea surface temperature data products from several organizations are the second most popular ones after MODIS datasets. The imagery products covering South America Region from INPE are the third most requested ones in the same period.

5. Discussion

Standardization and federation become highly important in the age of Information (Al-Hazmi et al., 2012). CWIC offers an interoperable catalog federation of EO data within CEOS community. Metrics monitoring is one of the fundamental parts of such a federated system. Measurements and information on capacity, performance, and availability of CWIC and its integrated catalog services are important to both stakeholders and partners.

CWIC partners have their own solution to catalog service and monitoring, and some of catalog services (for example, NASA ECHO) support multiple projects. It is not practical for them to build a system to monitor catalog facility for CWIC, or extract access information of CWIC requests. A comprehensive CWIC metrics application offers catalog providers an efficient solution regarding their service utilization in CWIC. It helps them manage, track, and analyze running catalog services requested by CWIC users. The web-based graphic user interface that displays metrics information at the catalog and dataset levels provides useful insights to service utilization through Web portal.

As a web analysis system of the federated system, it must allow easy plug-in of new systems and components. In our implementation, the defined metrics for newly added catalog or dataset can be easily analyzed and presented when its options are specified in the configuration file. Metrics information for the federated system should be aggregated by different categories. The aggregation and presentation of metrics at the catalog level give CWIC partners a holistic view on their service usage during the specified period. Some aggregators are implemented through components of Ext JS on the browser client side.

A federated system should ensure health, availability, interoperability of the whole system and the integrated systems. In CWIC metrics, the fault message is gathered and sent to CWIC administrator to determine the root cause, like the client request is bad, the requested catalog is offline, the requested dataset is removed, or CWIC is overloaded.

As described in Section 2.1, fewer operational monitoring systems are available for geospatial catalog service. The introduced approach in this paper can be reused or borrowed by other CSW service or similar federated system to evaluate its quality. In contrast to those GWS monitoring systems mentioned in Section 2.2, CWIC metrics provides service quality information from the perspective of a federated service provider. It should be stated that current CWIC metrics focus on monitoring availability, processing capability, and transfer reliability of CWIC service and its integrated catalog services, the domain-specific information, like bounding box and temporal range of the requested dataset from CSW request, are not tracked and analyzed. This functionality will be implemented in the future release. In addition, source code improvement and configuration optimization should be done to make the proposed framework more general and reusable in other standard-compliant federated catalog services.

Table 2
Dataset ranking.

Dataset	Description	Catalog	Requests
MOD10C1V5	MODIS snow cover daily L3 global 0.05° climate modeling grid product	NASA ECHO	8847
UKMO-L4HRfnd-GLOB-OSTIA	Sea surface temperature daily L4 global 0.054° grid product	GHRSST	6060
GES_DISC_ML1RADD_V003	EOS Aura Microwave Limb Sounder daily L1 global radiances product	NASA ECHO	5758
EUR-L2P-NAR18_SST	EUMETSAT sea surface temperature L2p North Atlantic regional 2 km grid product	GHRSST	5529
INPE_CBERS2_CCD	INPE high resolution CCD South America regional camera imagery product	INPE	5529
MYD10A2V5	MODIS snow cover 8-day L3 global 500 m grid product	NASA ECHO	5422
MCD43A35	MODIS directional hemispherical reflectance and bi-hemispherical reflectance 16-day L3 500 m grid product	NASA ECHO	4504
JPL_OUROCEAN-L4UHfnd-GLOB-G1SST	Sea surface temperature daily L4 global 0.009° grid product	GHRSST	4122
INPE_LANDSAT5_TM	INPE Landsat-5 TM South America regional 30 m imagery product	INPE	3851
MCD43A45	MODIS NADIR BRDF-adjusted reflectance 16-day L3 global 500 m grid product	NASA ECHO	3802

6. Conclusions and future work

GWS have been extensively utilized during the last decade to provide open and interoperable geospatial data and information to users around the world. The detailed metrics on availability, accessibility, performance, compliance, and reliability of these services are very vital to providers and consumers, but how to collect, analyze, structure, and deliver these metrics from the perspective of service providers have not been deeply explored. This paper gives details on the design and implementation of metrics information collection, analysis and representation of a federated catalog service that integrates heterogeneous and distributed systems from multiple data providers. Detailed CWIC service metrics can be filtered and calculated on-the-fly and displayed in the interactive tables and graphs. The periodical report can be generated within seconds to offer complete traffic analysis. These metrics present service providers with a range of detailed measures for both high-level and technical analysis of their operational services and assist the development team and the stakeholders in making technical and business decisions. The described measurement strategy and analytics framework offer GWS providers a guidance and reference to monitor and evaluate their service metrics in an easy and convenient way.

CWIC development team has implemented an OGC OpenSearch interface (http://cwic.wgiss.ceos.org/opensearch/datasets/osdd.xml?clientId=cwicClient) for searching EO metadata and data from integrated catalogs across CEOS agencies,, and also reused current CWIC Metrics for the CSW interface to develop the metrics application for the OpenSearch interface (http://cwic.wgiss.ceos.org/cwicosmetrics/). Additional features have been extended and implemented in the new metrics application, including tracking and displaying client information (e.g. Web browsers, CWIC clients, Java applications, etc.) and their activities, identifying unique visitors and their organization names from their IP addresses, supporting metrics information display by time of day, offering failure statistics by types (such as invalid request, wrong dataset identifier, overtime responding, and unavailable catalog).

In next coming years, CWIC development team will build an integrated CWIC Metrics application to track visiting information on both CSW and OpenSearch interface together. Future improvements include producing spatial and temporal extent ranking of client requests, sending detailed information exceptions and errors to developers and project partners in a timely manner, adopting lightweight standalone database such as SQLite as a replacement to the log file, and so on. In collaboration with project partners, the team will make the presented application or framework more applicable to other similar Web catalog services of CSW and OpenSearch like the Federated Earth Observation Missions (FedEO) of European Space Agency (ESA). Moreover, the team plans to make this framework as an open source package for evaluating CSW and OpenSearch services in the community of Earth science information.

Acknowledgments

The authors would like to thank other CWIC team members, particularly Yonsook Enloe, Martin Yapur, Archie Warnock, Calin Duma, and Yuanjie Li, for their valuable comments and suggestions on the CWIC Metrics application. The authors appreciate the constructive comments from two anonymous reviewers for improving the manuscript.

References

Al-Hazmi, Y., Campowsky, K., Magedanz, T., 2012. A monitoring system for federated clouds. In: Cloud Networking (CLOUDNET), 2012 IEEE 1st International Conference on, IEEE. pp. 68–74.
Bai, Y., Di, L., Chen, A., Liu, Y., Wei, Y., 2007. Towards a geospatial catalogue federation service. Photogramm. Eng. Remote Sens. 73 (6), 699–708.
Bai, Y., Di, L., Nebert, D., Chen, A., Wei, Y., Cheng, X., Shao, Y., Shen, D., Shrestha, R., Wang, H., 2012. GEOSS component and service registry: design, implementation and lessons learned. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5 (6), 1678–1686.
Cepicky, J., Becchi, L., 2007. Geospatial processing via Internet on remote servers - PyWPS. OSGeo J. 1 (5).
Crockford, D., 2011. Introducing JSON. Available: ⟨http://www.json.org⟩.
Enloe, Y., Yapur, M., 2011. CEOS WGISS Integrated Catalog, a catalog for Earth Observation satellite data. In: AGU Fall Meeting Abstracts. Vol. 1, p. 1595.
Farley, J. D., Mitchell, A., Nakamura, E., Duma, C., Enloe, Y., 2011. ECHO's Reverb as a client to the CEOS CWIC Catalog Portal using OGC's CSW protocol. In: AGU Fall Meeting Abstracts. Vol. 1, p. 1433.
Giuliani, G., Dubois, A., Lacroix, P., 2013. Testing OGC web feature and coverage service performance: towards efficient delivery of geospatial data. J. Spat. Inf. Sci. 7, 1–23.
Gorgan, D., Bacu, V., Mihon, D., Stefanut, T., Rodila, D., Cau, P., Abbaspour, K., Giuliani, G., Ray, N., Lehmann, A., 2012. Software platform interoperability throughout enviroGRIDS portal. J. Sel. Top. Appl. Earth Obs. Remote Sens. 5 (6), 1617–1627.
Gui, Z., Yang, C., Xia, J., Liu, K., Xu, C., Li, J., Lostritto, P., 2013. A performance, semantic and service quality-enhanced distributed search engine for improving geospatial resource discovery. Int. J. Geogr. Inf. Sci. 27 (6), 1109–1132.
Han, W., Di, L., Zhao, P., Li, X., 2011. GeoBrain Online Analysis System: An SOA-Based Geospatial Web portal. In Geospatial Web Services: Advances in Information Interoperability. IGI Global, Hershey, PA, pp. 455–474.
Horák, J., Ardielli, J., Růžička, J., 2011. Performance Testing of Web Map Services. In New Challenges for Intelligent Information and Database Systems. Springer Berlin Heidelberg, pp. 257–266.
Khaled, R., Tayeb, L.M., Servigne, S., 2010. Geospatial web services semantic discovery approach using quality. J. Converg. Inf. Technol. 5 (2), 28–35.
Kiehle, C., Greve, K., Heier, C., 2007. Requirements for next generation spatial data infrastructures-standardized Web based geoprocessing and Web service orchestration. Trans. GIS 11 (6), 819–834.
de la Beaujardière, J., 2006. OGC Web Map Server Implementation Specification. Open Geospatial Consortium Inc, Wayland, MA.
Miller, J.A., Cardoso, J., Silver, G., 2002. Using simulation to facilitate effective workflow adaptation. In Simulation Symposium, 2002. Proceedings. 35th Annual, IEEE. pp. 177–181.
Nativi, S., Craglia, M., Pearlman, J., 2013. Earth science infrastructures interoperability: the brokering approach. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6 (3), 1118–1129.
Percivall, G.S., Alameh, N.S., Caumont, H., Moe, K.L., Evans, J.D., 2013. Improving disaster management using earth observations—GEOSS and CEOS activities. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6 (3), 1368–1375.
Scholten, M., Klamma, R., Kiehle, C., 2006. Evaluating performance in spatial data infrastructures for geoprocessing. Internet Comput. IEEE 10 (5), 34–41.
Schut, P., 2007. OpenGIS Web Processing Service. Open Geospatial Consortium Inc, Wayland, MA.
Shao, Y., Di, L., Bai, Y., Wang, H., Yang, C., 2013. Federated catalogue for discovering Earth Observation data. Photogramm.-Fernerkund.-Geoinf. 2013 (1), 43–52.
Sun, Z., Yue, P., Di, L., 2012. GeoPWTManager: a task-oriented web geoprocessing system. Comput. Geosci. 47, 34–45.
Voges, U., Senkler, K., 2005. OGC Catalogue Services Specification 2.0-ISO 19115/ISO 19119 Application Profile for CSW 2.0. Open Geospatial Consortium Inc, Wayland, MA.
Vretanos, P.A., 2002. OGC Web Feature Service implementation specification. Open Geospatial Consortium Inc, Wayland, MA.
Whiteside, A., Evans, J.D., 2008. OGC Web Coverage Service implementation standard. Open Geospatial Consortium Inc, Wayland, MA.
Wu, H., Li, Z., Zhang, H., Yang, C., Shen, S., 2011. Monitoring and evaluating the quality of web map service resources for optimizing map composition over the internet to support decision making. Comput. Geosci. 37 (4), 485–494.
Zhang, C., Zhao, T., Li, W., Osleeb, J.P., 2010. Towards logic-based geospatial feature discovery and integration using web feature service and geospatial semantic web. Int. J. Geogr. Inf. Sci. 24 (6), 903–923.