AN INTEROPERABLE CLOUD-BASED GEOPORTAL FOR DISCOVERY AND MANAGEMENT OF EARTH OBSERVATION PRODUCTS

Dante D. Sánchez-Gallegos[1], J.L. Gonzalez-Compean[1], Victor J. Sosa-Sosa[1], Heidy M. Marin-Castro[2] and José Tuxpan-Vargas[3]

[1]Cinvestav Tamaulipas, Cd. Victoria, Tamaulipas, México
[2]Cátedras CONACYT -UAT, Cd. Victoria, Tamaulipas, México
[3]IPICyT, San Luis Potosí, SLP, México

ABSTRACT

This paper presents the design and development of an interoperable geoportal service for discovery and management of earth observation products (EOPs). In this service, the geoportal components are encapsulated into virtual containers that are launched into the cloud by using a microservice scheme to solve issues such as interoperability (with other systems) and implementation (over different platforms). A search microservice that analyses the preferences of end-users (settings of spatiotemporal and polygon shapes) and builds clusters of users sharing preferences was included into the geoportal for recommending/delivering, in advance, products matching with end-user preferences. The geoportal service also enables end-users to organize EOPs on-the-fly by using spatiotemporal parameters. A prototype of this service was implemented in a private cloud and connected to a satellite imagery repository of an antenna (ERIS) managed by Mexican Space Agency in a proof of concept. Learned lessons and performance assessments are described through an experimental evaluation with real users' participation.

KEYWORDS

Spatial data management, Satellite imagery repositories, User Preferences, Data Analysis & Geoportals

1. INTRODUCTION

Geoportals are key tools for space agencies to show and delivery earth observation products (EOPs) to end-users. Catalogs of satellite images, derivative products (corrected and thematic products) and maps [1], as well as maps created on-the-fly by end-users [2] are examples of EOPs. These products represent assets for agencies, scientific community and government instances to conduct missions, research and programs about earth respectively. EOPs are used in studies about climatic and disaster event management, territory management strategies as well as environment studies.

Dhinaharan Nagamalai et al. (Eds) : ITCSE, WiMo, ICAIT, ICDIPV, CRYPIS - 2018
pp. 01–13, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.80701

However, the development and deployment of these tools on the cloud/web when managing large satellite imagery repositories is not trivial because of technical and management issues.

The technical challenges are related to the interoperability of the geoportals with other systems. A geoportal is the last stage in a life cycle of satellite imaginary where additional software is also used depending on the stages of that cycle managed by agencies. In these stages EOPs are transformed into new versions (derivative products, maps, etc) or new products are created by grouping EOPs using spatial and temporal parameters. For instance, in the acquisition stage, services are required for parsing metadata files created by antenna operators, mapping with EOPs and make data and metadata indexes available for other services/applications and geoportals. In the preservation [3] stage indexed EOPs are ensured/stored in conservation infrastructure (I.e. Cloud/cluster storage services) for other systems to acquire these products. In manufacturing [4] stage the preserved products are used to create new EOPs that are indexed by geoportals and made available to other services.

These services, systems and applications are commonly created as either a monolithic application or independent solutions using different methods of access and data exchange. In the first case, the building of solutions passing through different life cycle stages is not trivial, whereas in the second case, designers must adapt their geoportal to each system with which it exchanges data/metadata. Moreover, these systems are commonly developed as web services deployed mainly on the cloud side by using virtual machines that, depending on the size of the geoportal, could become unmanageable [5].

In turn, the virtual container represents an alternative to virtual machines for solving issues of deployment on different types of infrastructures in an efficient manner [5]. When these containers are added to microservices including standardized input/output data exchange, interoperability issues are solved as well as different solutions including different stages of life cycle can be built. However, to the best of our knowledge, architectures of geoportals based on microservices and containers have not been enough explored.

Besides the interoperability and flexibility issues, the information management remains as a challenge for space agencies. The number and type of EOPs available in geoportals depend on the transformation depend on the transformations applied to EOPs, required by an agency, during the stages of the life cycle of the satellite imagery. In this context, product discovery schemes could enable geoportals to exhibit useful products and to improve the service experience of the end-users. Traditionally, geoportals include information searchers based mainly on a spatial criterion defined either by polygon shapes and/or parameters (e.g. Path and row, decimals or coordinates) [6]. This type of search management is mainly suitable for researchers and academics. Although recently geoportals are also incorporating searchers using natural language [6], there still is an opportunity area to improve the experience service of end-users by analysing preferences of end-users and their relationships with the preferences of other researchers/academics.

This paper presents the design and development of an interoperable cloud-based geoportal for the discovery and management of earth observation products. A microservice scheme was developed to encapsulate the applications used by geoportals in the life cycle of EOPs into virtual containers, which are launched into the cloud to create a single unified service of chained microservices.

This scheme enables agencies to create processing workflows as well as to solve interoperability and deployment issues. A search engine was created as a microservice connected to the geoportal to analyse the preferences of end-users (settings of spatiotemporal and polygon shapes) and to build clusters of users sharing preferences. This microservice was included into a geoportal for recommending/delivering, in advance, products matching with end-user preferences. The geoportal also enables end-users to create, on-the fly, products by using spatiotemporal parameters, which are also indexed by the geoportal and the search engine.

A prototype of the geoportal service was implemented in a private cloud and connected to a satellite imagery repository service of an antenna (ERIS) managed by Mexican Space Agency by using a federated service [3, 7]. A proof of concept was performed with this prototype and an experimental evaluation was conducted with the participation of real end-users. The evaluation revealed the feasibility of developing interoperable cloud-based geoportals by using microservices and containers. Learned lessons and performance assessments about management and discovery of EOPs are also described.

This paper is organized as follows: Section 2 introduces the design principles, including development details using microservices and provides an explanation of the method for detecting user preference; Section 3 and 4 present experimental evaluation and results respectively, and finally Section 5 provides some conclusions and future work.

2. AN INTEROPERABLE CLOUD-BASED GEOPORTAL: DESIGN PRINCIPLES

In this section, we present the general design of an interoperable geoportal service designed for the Mexican Space Agency (AEM), to exhibit EOPs through the cloud.

Figure 1 shows a conceptual representation of the containers of a geoportal implemented by using a microservice scheme. As it can be seen, the microservice structure includes three virtual containers (front-end, service and back-end) chained in a pipe and filter pattern. The front-end is in charge of receiving data or metadata from applications/services or other microservice and makes indirection of this information to service containers in a load balancing manner. The front-end orchestrates the launching and coordination of service and back-end containers in a workflow. The services container performs the choreography sent by front-end, which enables administrators to create software patterns. The service container of the geoportal includes three micro-services (Access, management and storage) that are chained building a stack structure defined by the orchestration configuration in Front-end container. Other patterns such as pipe and filter or master and slave can be deployed in the service container.

The containers in the service stack processes the requests sent through front-end and delivers the results through back-end container, which was designed for receiving and/or delivering data and/or metadata to/from either a client application or other microservice. In the case shown in Figure 1, geoportal service is connected to the service for the preservation of a satellite imagery, which is used to preserve the EOPs captured by the ERIS antenna.

In this scheme, each container also includes its dependencies and even a reduced operating system to ensure a regular operation over different types of platforms; as a result, the geoportal can be replicated to build clusters to improve the experience of end-users [8]. Moreover, the service container can be reused to change the geoportal by another application (e.g. preservation of images or manufacturing of derivative products) and could enable agencies to create workflows through the life cycle of the satellite imagery [3].

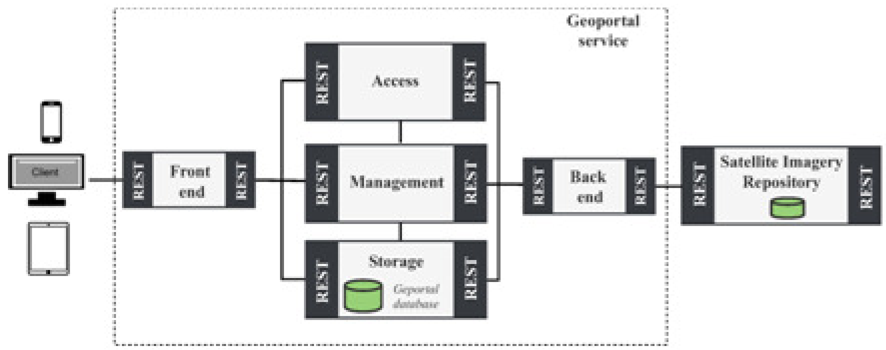


Figure 1: Conceptual representation of the containers of a geoportal implemented by using a microservice scheme

2.1. Geoportal microservices: development details

The development details of access, management, and storage microservices designed for Geoportal service are described in this section.

The access microservice allows geoportal to manage requests sent by either end-users or apps to the geoportal service through a web browser or invoking them by a REST API. An access control module provides end-users/applications with valid encrypted credentials and tokens to get access to the management microservice.

The management microservice includes a catalog management based on Pub/Sub system, a search engine, and on-the-fly building of EOP system. The catalog management system enables owners of EOPs to perform operations such as create, update, delete, modify and list catalogs. This system also enables users/APPs to share EOPs, found by the geoportal, with other active users. The search engine and the on-the-fly EOPs builder are described in next sections.

The storage microservice includes a retrieval system that enables to exchange data with repositories and cloud storage services.

2.2. Discovering EOPs by Analysing Activity and Preferences of End-Users

The search engine created for end-users to discover EOPs in the geoportal service includes modules such as recommendation and log activity.

The log activity module registers the actions performed by the end-users. It considers each request sent to the management microservice by applications as well as each click done by the end-users on the items of the geoportal web page.

The recommendation module creates suggestion lists of different versions of EOPs indexed by the geoportal for each user based on the user’s activity, which reveals preferences (a type of EOPs, catalogs, land cover, timelines, etc.).

A collaborative filtering technique following the memory-based approach was applied to the recommendation module and it is depicted in Figure 2. A recommendation is produced through phases such as logging historical activity of end-users, inference about the zone of interest, classification of active users, and delivering collaborative recommendations to the end-user through the exhibition web service of the geoportal. A profile for each end-user is created in this procedure, which includes the actions performed by the geoportal users, such as the number of views, clicking and downloading of each EOP, the configurations of each query that the user has done when performing searches as well as the user location. The first time that an end-user access to the geoportal, this automatically starts the construction of an activity profile, which is updated each time that user accesses to the geoportal service.

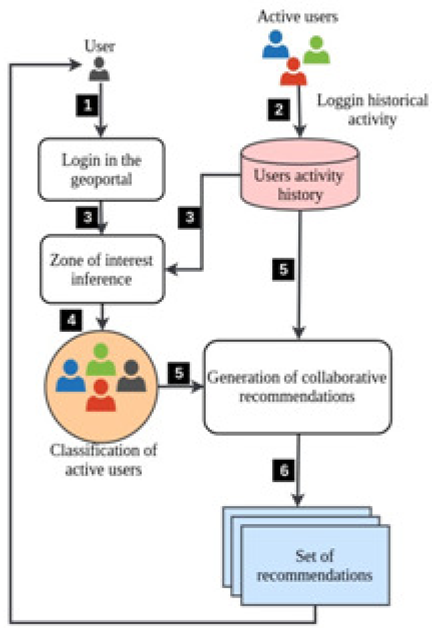


Figure 2: Workflow to create collaborative recommendations

In searches based on polygons, a set of geographic coordinates delimiting a land cover of interest is stored. In the case of using a circle to select an area, the radio and the coordinates defining the center of that area are stored. For points of reference, the latitude and longitude delimiting that area are considered and stored. An inference of the coverage land areas in which the users are interested in is performed as the end-users could be interested in specific coverage areas for long periods of time. In this context, the centroids of the areas where end-users perform queries for the geoportal represent a highly possible zone of interest for the end-user.

2.2.1. Classification of Active Users

A classification of active users is performed as the recommendations for a given user are created by using the information of other end-users exhibiting similar preferences to that users. The centroids of the zones of interest previously calculated are passed to a clustering algorithm called DBSCAN that was included in the recommendation module to create groups of users that share preferences by using the Haversine distance.

When the clusters of users are created, the recommendations for a specific user are built and sent to the geoportal. This task is performed by the Algorithm presented in Figure 3, which receives as input the cluster in which the user was grouped, and the set of EOPs with which the user has made contact (clicked, viewed, downloaded).

Algorithm 1 Collaborative filtering recommender

Require: $N = \{n_1, n_2, n_3, ..., n_n\}$: group (cluster) of neighbor users of user u;
$I = \{i_1, i_2, i_3, ..., i_m\}$: set of current EOP queried by u.
Ensure: $R = \{r_1, r_2, r_3, ..., r_k\}$: set of recommended items (EOP) for u.

$items = \{\}$
for all $n \in N$ **do**
  $clicked = n.getClickedEOPs()$
  $downloaded = n.getDownloadedEOPs()$
  for all $EOP \in clicked$ **or** $EOP \in downloaded$ **do**
    if $EOP \notin I$ **then**
      $incrementCounter(EOP.id)$
      $items = items \bigcup EOP$
    end if
  end for
end for
sort($items$)
$R = selectTop(items)$
return R

Figure 3: Collaborative filtering recommender

The algorithm selects the most popular images previously queried by using the historical activity of the users in the cluster. A table of items is constructed with the EOP identifier (a counter indicating the times that the EOP appears in the group) and the EOP that was clicked or downloaded. Finally, the table is sorted by the number of occurrences of a given EOP, and a configurable number of top records are returned as recommended items.

2.3. Organizing EOPs on-the-fly

The users can organize EOPs on-the-fly, through the geoportal, in the form of mosaics (spatial settings), overlaps (temporal settings), and mosaics overlapped (spatiotemporal settings).

Mosaics are EOPs created for a given land cover selected by an end-user. The images that are within the area, are ordered sequentially according to their coordinates (see Figure 4). The overlaps of images are built with a group of images with the same Path/Row value and sorted by their acquisition date and a timeline for this coverage area is created for end-users to observe changes in coverage land over the time. Mosaics overlapped are created for end-users to observe images when using spatiotemporal parameters.

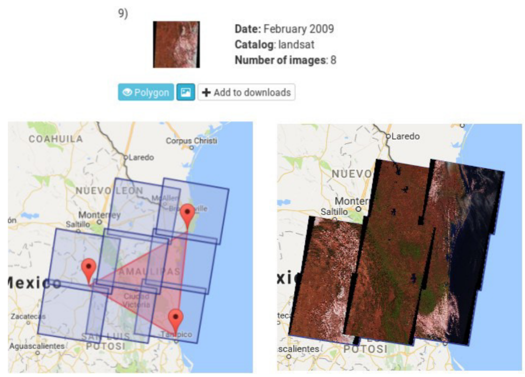


Figure 4: Organizing EOPs in the form of mosaics for a search delimited by an end-user

3. EXPERIMENTAL EVALUATION METHODOLOGY

In this section, we describe the methodology used to conduct a proof of concept and an experimental evaluation based on a satellite imagery of an antenna.

3.1. The Geoportal deployment on a private cloud

The Geoportal service was deployed on a private cloud built by using OpenStack Mitaka. This service was deployed by using a master/slave pattern (See Figure 5) where the microservices of the geoportal were cloned five times and launched into the cloud to build a cluster of five Geoportal slave services and one dispatcher intermediary microservice representing the master of the cluster. The master was in charge of distributing the requests/users to the cluster of slaves in a load balancing manner. The dispatcher also included a centralized database that is accessed by each geoportal microservice, which also includes a front-end container to receive data from the dispatcher and a back-end container to get access to the repository of EOPs and the metadata files as well as to deliver this information to the end-users through the front-end container.

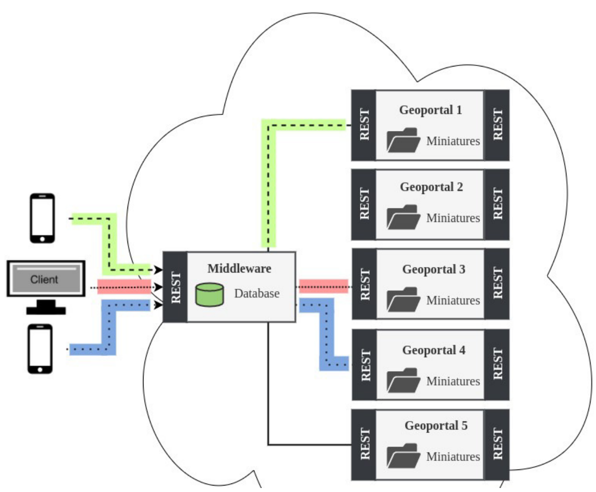


Figure 5: Geoportal distribution

In the storage microservice of the slaves, modules retrieve/delivers EOPs from catalogs preserved and transported by using a federated cloud storage service called FedIDS[3]. The catalogs of satellite images of three sensors, Terra, Aqua of the MODIS platform and LandSat5, were considered in this evaluation.

The containers were built by using Docker. A bot client, was deployed on a virtual machine in the cloud to create workload of end-users that the microservices of the geoportal assumed as a real workload.

3.2. Metrics

The metrics chosen to evaluate the performance of the prototype were:

- *Service time*: Represents the time spent in completing a task by a component of the geoportal, such as master, slave or microservice.

- *Response time*: When performing experiments with real users, this metric represents the time observed by end-users when sending requests to the geoportal. When the bot produces requests in automatic manner, this metric considers the time spent by the bot client to send requests to the dispatcher, the time required by the geoportal slave to serve a given request and the time spent by the dispatcher in spending the response to the bot.

4. EVALUATION RESULTS

In the evaluation performed with client bot, the tests were carried out to measure the performance of the geoportal service when the geoportal slaves produce recommendations for the end-users. The service spent 0.61 seconds to serve a simple query drawing a polygon over some areas of Mexico and 0.93 seconds required to generate the recommendations to one user. This last time is composed by the clustering time (0.68 seconds), creation of recommended mosaics (0.10 seconds), creation of recommended images (0.08 seconds) and creation of recommended overlaps (0.07 seconds). In order to evaluate concurrent queries, the client bot sent to the geoportal microservice 10 and 100 simple of concurrent queries/requests. In the first experiment was observed that, on average, each request was made in 0.75 seconds, whereas when the number of clients was 100, the service time average per query was 0.94 seconds.

When end-users create mosaics and overlapped EOPs on-the-fly, the response time is the metric to be observed because in this type of EOPs the geoportal returns several results to end-users. This time depends on the time spent by a geoportal slave to serve the queries sent by the end-users through their browsers, the volume of results found by the geoportals and the type of shape used to create EOPS based on spatial/temporal parameters. In an experiment, 24 real users sent different shapes to build mosaics and overlapped EOPs and both cases we measured the mean response time observed by these users for each experiment.

Table 1 shows the number of results obtained when the coverage radius of search is changed by the end-users for each catalog considered in this evaluation (Aqua, Terra and LandSat5). For each catalog, we asked end-used for changing the radius until all the raster (individual) images were obtained as a result.

Figure 6 shows, in the vertical axis, the response time in seconds spent by the geoportal to show results of EOPs such as raster images, overlaps, and mosaics when end-users select one, two and three catalogs of satellites (horizontal axis) in the geoportal web page

Table 1 Number of results for different coverage radius of search

# Individual (raster)	# mosaics	# overlaps	Coverage radius (km)
LandSat			
111	29	5	100.421
324	33	13	329.953
1261	35	103	1000.595
1662	36	254	3987.391
Terra			
203	52	42	500.466
656	58	158	1003.251
1805	61	343	2857.952
Aqua			
41	17	22	500.749
406	31	77	1000.302
1033	32	181	2417.834

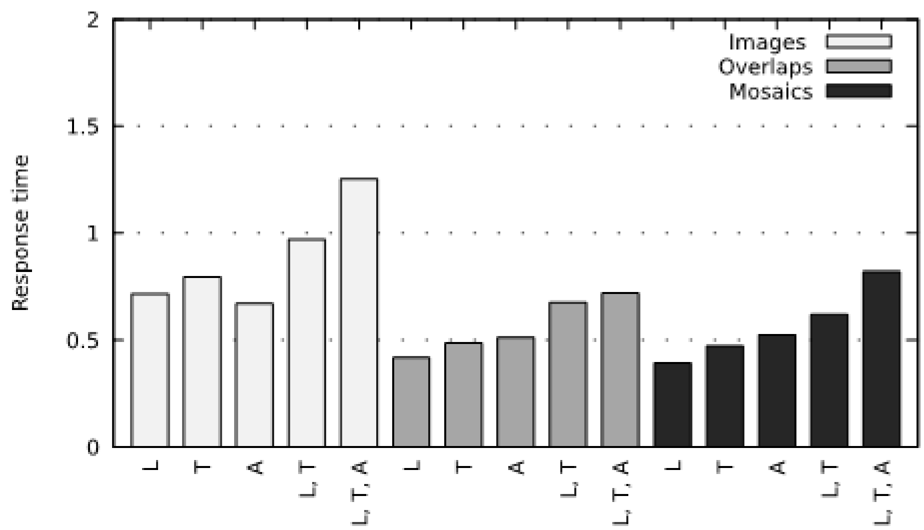


Figure 6: Response time of geoportal serving queries to different satellite sensors: LandSat5(L), Aqua(A) and Terra(T)

Basically, Figure 6 shows three effects affecting the service experience of end-users. The first one is the size of the data source used by the searcher of the geoportal. This means the number of results increases when increasing the number of EOPs in a given catalog, which is increased end-users choosing more sensors in each search. The second effect detected in these tests was that, as expected, the shape chosen by end-users affects the number of results offered by the geoportal to the end-users. For instance, it was observed shapes based on polygons produce increased response times in comparison with other shapes (i.e. circle and rectangle) as the geoportal produces more results of EOPs when these end-users select this type of shape.

Figure 7 shows the resultant classification of the 24 participating users in the previously described experiment, which were organized by the geoportal in clusters of preferences (see users with the same color). The circle shape for each user shows the EOPs of the regions recommended by the geoportal to the end-users. As it can be seen, agencies can identify the volume (in terms of EOPs per cluster) and density (the frequency of these EOPs are requested by end-users) of the consumption of the geoportal service end-users as well as their interests and preferences. This enables agencies to automatically create EOPs. It also enables agencies to either distribute EOPs among their users or build EOPs in advance and make them available in the geoportal depending on the end-user profile described by the clustering of the classification.

Figure 7: Classification of users in clusters of preferences

5. RELATED WORK

Geoportals has been deployed for end-users to search for EOPs with specific characteristics by using simple queries and parameters as the acquisition date, coverage coordinates, and other keywords included into metadata associated to EOPs. Different geoportals have been proposed to manage the information about disaster risks [9], as well as for monitoring rice fields by using the satellite imagery and its derivative products [10].

In order that the end-users can find and download EOPs, retrieval mechanisms as web crawlers were proposed to extract metadata from files found on the web [11]. Also, the information can be retrieved directly from geospatial databases, the metadata is parsed extracting the image information and stored in a database [12]. Geoportals also exhibit derivative products [4, 13] from satellite imagery, which add value to raster images and provide more information to the end-user [14].

Solutions are focused on improving the quality of the results obtained by using semantic web tools such as RDFs, ontology, and SPARQL [14, 15]. Nevertheless, there is an opportunity area to improve the end-user experience by analysing the end-users' activity and grouping them based on their preferences [16]. Moreover, improving the deployment and interoperability of geoportals enable agencies to build clusters to improve the response and service times produced by geoportals when showing EOPs (derivate EOPs as well), which are aspects that were included in the geoportal proposed in this paper.

6. CONCLUSIONS

This paper presented the design and development of an interoperable geoportal service for discovery and management of earth observation products. In this service, the geoportal components are encapsulated into virtual containers launched into the cloud by using a microservice scheme to solve issues such as interoperability (with other systems) and implementation (over different platforms).

A search microservice that analyses the preferences of end-users (settings of spatiotemporal and polygon shapes) and builds clusters of users sharing preferences was included into the geoportal for recommending/delivering, in advance, products matching with end-user preferences. The geoportal service also enables end-users to create, on-the-fly by using spatiotemporal parameters.

A prototype of this service was implemented in a private cloud and connected to a satellite imagery repository of an antenna (ERIS) managed by Mexican Space Agency in a proof of concept. Learned lessons and performance assessments are described through an experimental evaluation with real users participation.

ACKNOWLEDGEMENTS

This work was partially supported by the sectoral fund of research, technological development and innovation in space activities of the Mexican National Council of Science and Technology (CONACYT) and the Mexican Space Agency (AEM), project No.262891.

REFERENCES

[1] Bernard, L., Kanellopoulos, I., Annoni, A. and Smits, P. (2005). The European geoportal—one step towards the establishment of a European Spatial Data Infrastructure. Computers, Environment and Urban Systems, 29(1), pp.15-31.

[2] Geoplatform.gov. (2018). Home. [online] Available at: https://www.geoplatform.gov/ [Accessed 21 Apr. 2018].

[3] Gonzalez-Compean, J., Sosa-Sosa, V., Diaz-Perez, A., Carretero, J. and Marcelin-Jimenez, R. (2017). FedIDS: a federated cloud storage architecture and satellite image delivery service for building dependable geospatial platforms. International Journal of Digital Earth, pp.1-22.

[4] Zhang, Y., Wang, B., Zhang, Z., Duan, Y., Zhang, Y., Sun, M. and Ji, S. (2014). Fully automatic generation of geoinformation products with chinese zy-3 satellite imagery. The Photogrammetric Record, 29(148), pp.383-401.

[5] Bernstein, D. (2014). Containers and Cloud: From LXC to Docker to Kubernetes. IEEE Cloud Computing, 1(3), pp.81-84.

[6] Hu, Y., Janowicz, K., Prasad, S., & Gao, S. (2015). Enabling semantic search and knowledge discovery for ArcGIS Online: A linked-data-driven approach. In AGILE 2015 (pp. 107-124). Springer, Cham.

[7] Vazquez-Martinez, G., Gonzalez-Compean, J., Sosa-Sosa, V., Morales-Sandoval, M. and Perez, J. (2018). CloudChain: A novel distribution model for digital products based on supply chain principles. International Journal of Information Management, 39, pp.90-103.

[8] Gonzalez-Compean, J., Sosa-Sosa, V., Diaz-Perez, A., Carretero, J. and Yanez-Sierra, J. (2018). Sacbe: A building block approach for constructing efficient and flexible end-to-end cloud storage. Journal of Systems and Software, 135, pp.143-156.

[9] Molina, M. and Bayarri, S. (2011). A multinational SDI-based system to facilitate disaster risk management in the Andean Community. Computers & Geosciences, 37(9), pp.1501-1510.

[10] Granell, C., Miralles, I., Rodríguez-Pupo, L., González-Pérez, A., Casteleyn, S., Busetto, L., Pepe, M., Boschetti, M. and Huerta, J. (2017). Conceptual Architecture and Service-Oriented Implementation of a Regional Geoportal for Rice Monitoring. ISPRS International Journal of Geo-Information, 6(7), p.191.

[11] Skluzacek, T., Chard, K. and Foster, I. (2016). Klimatic: A Virtual Data Lake for Harvesting and Distribution of Geospatial Data. 2016 1st Joint International Workshop on Parallel Data Storage and data Intensive Scalable Computing Systems (PDSW-DISCS).

[12] Steiniger, S., de la Fuente, H., Fuentes, C., Barton, J. and Muñoz, J. (2017). BUILDING A GEOGRAPHIC DATA REPOSITORY FOR URBAN RESEARCH WITH FREE SOFTWARE – LEARNING FROM Observatorio.CEDEUS.cl. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W2, pp.147-153.

[13] Institute, W. (2018). Interactive Map | Global Forest Watch. [online] Globalforestwatch.org. Available at: https://www.globalforestwatch.org/map [Accessed 21 Apr. 2018].

[14] de Andrade, F., de Souza Baptista, C. and Davis, C. (2014). Improving geographic information retrieval in spatial data infrastructures. GeoInformatica, 18(4), pp.793-818.

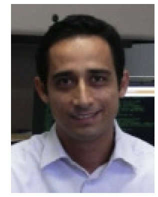
[15] Aditya, T. and Kraak, M. (2007). Aim4GDI: Facilitating the Synthesis of GDI Resources through Mapping and Superimpositions of Metadata Summaries. GeoInformatica, 11(4), pp.459-478.

[16] Dareshiri, S., Farnaghi, M. and Sahelgozin, M. (2017). A recommender geoportal for geospatial resource discovery and recommendation. Journal of Spatial Science, pp.1-23.

AUTHORS

Dante D. Sánchez-Gallegos is a master student at Information Technology Laboratory, Center of Research and Advanced Studies of the National Polytechnic Institute, (CINVESTAV), Ciudad Victoria, Mexico. His research areas of interest are data analysis and cloud computing.

J. L. Gonzalez-Compean received his Ph.D. in Computer architecture from UPC Universitat Politècnica de Catalunya, Barcelona (2009). He was a visiting professor at Universidad Carlos III de Madrid, Spain and researcher at Cinvestav, Tamaulipas, Mexico. His research lines are Cloud-Based Storage systems, linguistic archival systems, federated storage networks. His expertise areas are the design of fault-tolerant, adaptability and availability strategies as well as task scheduling and storage virtualization.



Victor J. Sosa-Sosa is a full-time research-professor at CINVESTAV-Tamaulipas - IPN. He has a Ph.D. in Computer Science from Technical University of Catalonia (UPC-Barcelona). He was professor and researcher in the Distributed Systems Group at CENIDET-Mexico until 2006. Participant and responsible of research and development projects funded by the National Council for Science and Technology of Mexico (CONACYT) and private companies. His research interest and specialization areas are related to Distributed Information Systems and Databases. Recently, his work is more focused on distributed information at Web scale, especially on topics related to efficient information search, storage and analysis, integration of Web query interfaces (DeepWeb) and knowledge harvesting.

José Tuxpan-Vargas was born in Puebla, Mexico, in 1982. He received the B.S. degrees in electronic engineering from the Technological Institute of Puebla, in 2005 and the M.S and Ph.D. degrees in telecommunications from Centre for Research and Advanced Studies of the IPN, Jalisco in 2010 and 2014 respectively. His interests include the design and implementation of new algorithms for imagery systems, adaptive signal /image processing, monitoring and interpretation of geophysical data using remote sensing techniques.

Heidy M. Marin-Castro received the Ph.D. degree in Computer Science from Information Technology Lab in Center for Research and Advanced Studies of the National Polytechnic Institute, Mexico (Cinvestav) in 2014, the M.Sc. degree in Computer Science from the National Institute for Astrophysics, Optics and Electronics (INAOE) in 2008 and the B.Sc. degree in Computer Science from the Autonomous University of Puebla. Currently she is a professor researcher at Engineering and Sciences Faculty in the Autonomous University of Tamaulipas, Mexico. Her research areas include Web Data Management, Databases, Data Mining and Information Retrieval.