Improving Satellite Data Archiving Facility for Environmental R&D Purposes Based on Architecture of Information Approach

Romualdo Alves Pereira, Jr.[1] and Marcos Aurelio Ferreira dos Santos[2]
National Institute for Space Research – CRN/INPE, Natal, RN, Brazil

Mamede Lima-Marques[3]
University of Brasilia - UnB, Brasília, DF, Brazil

and

Fatima Mattiello-Francisco[4]
National Institute for Space Research - INPE, Sao José dos Campos, SP, 1222410, Brazil

In the twenty years of operation, the mission center software designed to process, archive and disseminate environmental data collected by satellites, developed by the Brazilian Institute for Space Research (INPE), have evolved under the influence of three generations of information architectures. From the standalone applications within individual organization, the mission center software evolved to integrated sets of client and server database components. With the migration of the Brazilian Data Collection System to INPE Northeast Regional Center (INPE-CRN), in 2008, and the growing interdependence among the organizations, that are the owners of the data collection platforms, the modernization of the environmental data archiving, processing and dissemination facilities were required by the Brazilian Environmental Data System (SINDA). Focusing on the view of information as corporate resources supported by IT tools and techniques, the third generation of information architecture actually guides the new version of SINDA, under the process of development supported by Brazilian Space Agency (AEB). Architecture of Information is a useful approach to separate technology and information architectures. The approach addresses knowledge management, the intelligent systems and a more holistic view of information, driven by the development of multi-dimensional architecture and open information patterns. This paper presents the architecture of information framework for SINDA, focusing on the construction of dimensional models, business process modeling (BPM), and customized IT tools/techniques, such as business intelligence, Rich Internet Applications (RIA), Semantic Web, mobile applications and Grid Computing and Constraints Programming. As a result, the new SINDA will improve the system governance based on the engine rules certified by end-users. Also the system scalability is improved by means of multi-telemetry media, space temporal system database, and environmental data integration supported by open patterns.

I. Introduction

SCD-1 and SCD-2 are the first Brazilian satellites designed, built and operated at the National Institute for Space Research (INPE) for environmental data collection mission. Launched in February 1993 and October 1998, respectively, both equatorial orbit satellites are still operational, and have had satisfactory performance even after 15 and 10 years' time in orbit. They were originally designed to last two years.

[1] Analyst in Science and Technology Msc, INPE' Northeast Regional Center (CRN), romualdo.alves@crn2.inpe.br
[2] Electronic Engineer Msc, INPE' Northeast Regional Center (CRN), aurelio@crn.inpe.br
[3] Professor Phd, Center for Research in Architecture of Information (CPAI), mamede@unb.br
[4] Computer Engineer Phd, INPE Regional Centers Coordination (CCR), fatima@iss.inpe.br

The SCDs satellite services were added by a data collection transponder (DCS), operating on polar orbit on board of the Sino-Brazilian satellite (CBERS-2B) from 2007 to 2010.

Presently, the Brazilian System for Environmental Data Collection retransmits the data collected by a network of approximately 750 Data Collection Platforms (DCPs), that are distributed throughout the Brazilian territory. When the satellite goes over the regions of Cuiabá (MT) and Alcântara (MA) visibilities where the tracking stations antennas are located, for contact with the satellite, the DCPs signals that are visible to the satellites are captured and retransmitted to the stations. The received data is recorded and transmitted to the Data Collection Mission Center after the passage of the satellite, for processing, archiving and distribution to users.

These environmental data are used in various applications such as weather forecasting, studies of ocean currents, tides, atmospheric chemistry, agricultural planning, among others. An application of great importance is monitoring of the watershed, and river and rain gauge data. The users are government entities and private companies. Nowadays, there is an average of fifty entities that are the owners of these 750 installed DCPs.

Along the twenty years of operation, the Brazilian Environmental Data Collection System went through three stages of information architecture that encompasses satellite infrastructure, information technology and business processes issues.

The diagram in Figure 1 presents the SINDA's three generations of Information Architecture.

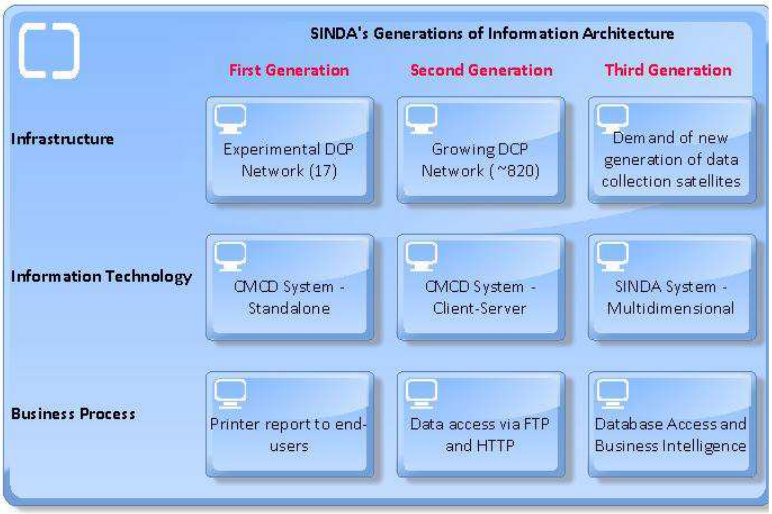

Figure 1. SINDA's Generations of Information Architecture

This paper is organized in five sections. This introduction concerns the first one. In the following three sections, we present the aspects of the three different generations of information architectures under the mission center perspective. The last section concludes the paper.

II. SINDA's First Generation: Standalone Architecture

Starting from launching of SCD1 until 1997, the system counted with an experimental DCP network installed by INPE users. They were composed by 17 DCPs for Amazon Program and Tide-gauge. The first aimed at studying the Ozone Layer, greenhouse effect, tropical forest regeneration, and biomass burning. The last enabled, for the first time, a systematic data collection of South Atlantic (coast and oceanic area), acquiring sea level in an hourly basis through the submarine pressure and water temperature. The DCPs rate of repetition varied from 180 to 215 seconds.

The DCP sensors data were stored and processed to engineering units at Data Collection Mission Center (CMCD) located in Cachoeira Paulista (SP). The data sensors were kept available for remote end-users, through the public Packet Switch Network (RENPAC), by means of communication line or a printer report to be mailed to the user [1].

The CMCD software counted with the existing national computational infrastructure (the Digirede DGR 8000) that employed a Motorola 68010 micro-processor, a UNIX-like multi-user operating system and a relational database package with C language interface. To perform all the tasks of data acquisition and treatment, the CMCD software was implemented as a set of functions interacting with external entities and operating over the data base. The main functions were: Acquire payload data; Check Acquired Data; Process Sensor data; Send User Requested Data; Update Directories; Generate Reports. Standalone applications were responsible to process, store and retrieve the processed data from the database, sending the obtained information via the communication line to a printer report to be mailed to the end-user.

A. DCPs Network Scenario

Until 1997, 260 DCPs were added by two governmental organizations to the experimental 17 active DCP network that was installed by INPE in the first year of SCD1 operation. The launching of the SCD2 guaranteed the continuity of the Data collection System service offered by INPE to society, increasing DCP coverage from 8 SCD1 passages per day to 16 satellite passages per day. Thus the gap of approximately 10 hours without satellite passes had diminished.

III. SINDA's Second Generation: Client-Server Architecture

The second stage started with the improvements on data dissemination by using RNP (National Research Network), Bitnet and other forms of access provided in the Data Collection Mission Center.

In order to support the growing DCP network (more than seven times in ten years), improvements on the original CMCD software were required to provide users with user-friendly interfaces and an efficient way to access environmental data collected by SCDs.

The development of the second generation of the mission center software, named SCMCD, was carried out based on the user centered principles. Three categories of users with different goals and different skills were identified: (a) Mission Center Operator; (b) Satellite Managers and (c) DCPs data end-users.

Regarding the architecture of information, the approach client-server had been used with the aim of separating the dialogue from both the data processing and the data storage components [2]. The SCMCD was developed in C language, using Linux Fedora Operating System and PostgreSQL Database Manager.

A new version of CMCD software oriented to fulfill the needs of the Mission Center operators and the users of the environmental data was deployed at the later part of the 90s.

B. DCPs Network Scenario at Cachoeira Paulista (SP)

Until 2008, 820 DCPs were installed in the Brazilian territory. Many of them were under responsibility of state governments. Among them, 10% were non-operational and the oldest required maintenance. With Internet advent, the access and the interest of individuals to the environmental data increased. Thus, the computational infrastructure updating was required in order to improve system availability during the seasons that natural disasters are more frequent such as raining periods.

C. DCPs Network Scenario at Natal (RN)

The computational infrastructure had been improved with the transfer of the Brazilian Environmental Data Collection System to INPE/Northeast, counting with two (2) redundant servers configured as: two (2) processors Intel Xeon E5506 Quad-Core de 2.13 GHz (technology EM64T); eight (8) GB memory RAM DDR3 with 800 MHZ; four (4) Hard Disk 250GB each with Serial Ata2, 07.200 rpm; Integrated array control SAS 3Gb/s, with 256 MB cache memory ECC with battery (PERC6/i); four (4) network interfaces 10/100/1000 UTP Onboard; Redundant fans and hot plug.

The SINDA system became more stable during high demand periods; even though, the same data processing software components and database have been kept. In addition, it is important to notice that at the moment, the satellites SCD1 and SCD2 still operate on degraded mode for power sake. With 16 passages in total over Cuiabá (MT) Station, the DCPs coverage of received messages per day was 408 by SCD1 and 415 by SCD2 on April, 19th, 2012. A total of 8926 processed messages were stored in the database.

IV. SINDA's Third Generation: Multidimensional Architecture

The third generation is related to the transfer of the Brazilian Environmental Data Collection System to Natal (RN), the INPE Regional Center at Northeast of Brazil, in 2008. The increased number of users and the demand for investments on the DCP maintenance led to the need for new generation of data collection satellites.

The growing demand of interested users in the environmental data for R&D purposes motivated the conception and development of the new version of the mission center system.

SINDA Project, started in 2010, with financial support of the Brazilian Space Agency (AEB), follows the World Meteorology Organization – WMO guidelines that intend to provide Meteorological and Hydrological Centers with information on best practice of climate data management. The WMO guidelines many countries to make the transition from older databases to the kind of environmental system that needs to provide much greater utility, security and robustness [3]. According to these guidelines, desirable properties recommended for Climate Data Management Systems are considered in SINDA. They are organized in seven abstraction layers, as follows:

1. *Data Source* - key entry capabilities, input options;
2. *Data Structure* - data extraction, data manipulation and quick looks;

3. *Data Engineering* - ability to support WMO Core Metadata Standard [4]. WMO Core Metadata Profile for SINDA Project is structured in the following main parts: Content Description (what the data describes), Geographic Extent (where the data is relevant), and Temporal Extent (when the data is relevant). Particularly in the SINDA Project, the use of metadata is a way to guarantee scalability without reprogramming the system when new data collecting devices or platforms are included in the system. To do this, we also have to consider the Platform Description (what are the characteristics of the data collecting platforms, devices, related sensors and its valid ranges and related constraints);
4. *Data Quality* (what is the data accuracy), Responsible Party (who created the data), and Access Constraint (when the data was created, when it will be updated and what are the limitations on the use of the data).
5. *Data Validation* - scope of quality checks on observation values;
6. *Data Distribution* - security issues, user management and distribution policy. Data can be delivered by means of FTP/HTTP services or Web Services in any of the following formats: xls, csv or OPeNDAP;
7. *Data Management* - database management and monitoring, data multidimensionality and database queries.

Architecture of Information is a useful approach that addresses the knowledge management, the intelligent systems and a more holistic view of information, driven by the development of multidimensional architecture and open information patterns. The new version of SINDA system for the Brazilian Environmental Data Collection System operation, at INPE/Northeast, is based on the architecture of information framework. It focuses on the construction of dimensional models and business process modeling (BPM), separating technology from information architectures. The diagram in Figure 2, page 5, presents the business process model related to the seven SINDA layers which are driven by events from one specific layer to the next one.

The seventh layer, Data Management, is characterized by the customization of IT tools/techniques, such as Business Intelligence (BI) with Rich Internet Applications (RIA) Technology and versions for mobile devices; Grid Computing; Temporal Databases and a Help Desk system.

A. SINDA-BI

This component was developed with a combination of two Open Source technologies: Pentaho Community Edition for the system core and Rich Internet Applications - RIA (HTML5 and/or Adobe Flex) for user interface, making the solution at the same time very effective and user-friendly. Some modules of BI are available for mobile devices.

The vision of "*starting small, but thinking large and scaling fast*" stated by William McKnight [5], is followed by SINDA, taking into account the existing DCPs demand and the increasing environmental data to be collected in the near future. Usually, there are immediate needs to be met all time and a Business Intelligence System provides ways to deal with this kind of demand. To illustrate, there is the need for rapid development that can be built upon over time; availability of qualified data; good query performance resulting in increased interactive usage; ability to get to real-time and/or quick look of acquired data; a platform to support advanced workload management; a scalable path forward as devices (or platforms) for data collection, users, and application modules.

B. SINDA-Grid

Grid Computing applies the resources of many (possible different) computers in a network to a single problem at the same time, usually requiring a great number of computer processing cycles or access to large amounts of data. SINDA-Grid uses Hadoop implementation in order to guarantee a high level of performance in database queries, providing transparent access to computational and storage resources geographically distributed. Hadoop provides a reliable shared storage and analysis system. With this approach, questions that took too long to get answered before can now be answered quickly, which in turn leads to new questions and new insights [6], providing a natural evolution of the system during its lifetime.

C. T-SINDA

Temporal Databases usually include valid-time and transaction-time, forming a bitemporal data [7]. Valid time denotes the time period during which a fact is true with respect to the real world. Transaction time is the time period during which a fact is stored in the database. T-SINDA is a temporal database that allows the maintenance of historical DCPs configurations with no needs to change its Identification (id) as a primary key. Moreover, it turns available Temporal Queries with special clauses like: Before, After, During, Between and so on.

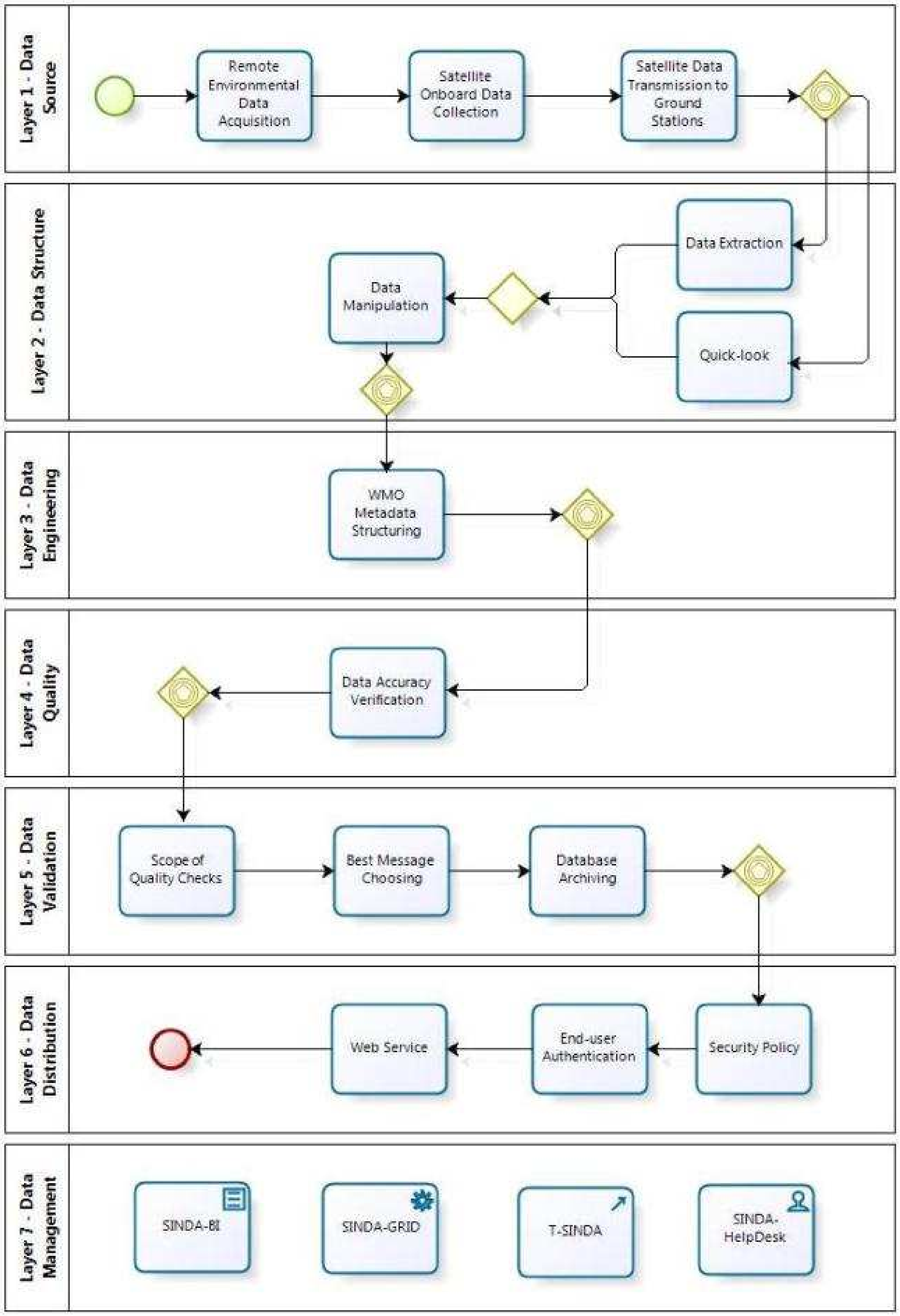

Figure 2. SINDA Business Process Model in Abstraction Layers Driven by Events

D. SINDA-HelpDesk

HelpDesk is a service that support customers in their technical problems and work as a Web Based Contact Center aiming to troubleshoot problems related to the use of the system and its potential. It is based on queues of service and Service Level Agreements (SLA). "The HelpDesk is often the first impression a prospective client will experience, and it is imperative all issues be resolved with experience and professionalism"[8].

SINDA-HelpDesk has a queue supervisor (a person or team responsible for managing the issues) who assign an issue to one of the specialized teams based on the type of issue (network, server, development, BI, database etc).

V. Conclusion

Considering the degraded status of the SCD1 and SCD2 and the new technologies available for environmental data collection, the modernization of the Brazilian Environmental Data Collection System has become mandatory. Wireless sensor network via Internet and cellular communication, which are complementary to data collection satellites, added to the INPE policy to allow users to have freely access to data, have contributed to increase the usage of environmental data for research and developmental purposes.

To deal with this demand, the environmental data center of INPE/Northeast in Natal (RN) became responsible for the data collection system modernization that includes DCP data processing, archiving and dissemination by SINDA.

Architecture of Information framework contributes to deal with the challenges of archiving and retrieving environmental data in a multidimensional database. In addition, it is suitable to address the different needs of the SINDA stakeholders: DCP end-users, managers of the Brazilian Data Collection System, SINDA operators and researchers. The framework increases both the system governance based on the engine rules certified by end-users and the system scalability by means of multi-telemetry media, space temporal system database, and environmental data integration supported by open patterns.

The modernization of SINDA will improve the existing satellite data archiving facility at INPE/Northeast for Environmental R&D Purposes. In the near future, the third generation of SINDA software will be available for the Brazilian Environmental Data Collection System.

.

Acknowledgments

The authors would like to thank the Brazilian Research and Development Council (CNPq) and the Brazilian Space Agency (AEB) for the financial support for SINDA Project.

References

Periodicals

[1]Yamaguti, W., Itami, S. and Becceneri, J.C., “Collection and treatment of the environmental data with the Brazilian satellite SCD1”,*RBCM Journal of the Brazilian Society of Mechanical Sciences*, XVI Special Issue, 1994.

Proceedings

[2]Mattiello-Francisco, M. F. and Gatto, R.C. “Environmental data system for Brazilian data collection satellite”, *Proceedings of the Annual European Symposium on Satellite Remote Sensing II*, edited by SPIE Vol. 2583, 25-28 September, Paris, France, 1995, pp. 515-525.

Technical Reports

[3]N. Plummer, Guidelines on climate data management, Tech. rep., World Meteorology Organization (October 2005).

[4] J. Tandy, WMO core metadata profile version 1.2: Guidelines on the use of metadata, Tech. rep., WMO (2010).

[5] W. McKnight, Starting small, but thinking large and scaling fast, Tech. rep., Teradata (2009).

Books

[6] T. White, Hadoop: The Definitive Guide, 1st Edition, O'Reilly, 2009.

[7] R. T. Snodgrass, Developing Time-Oriented Database Applications in SQL, Morgan Kaufmann Publishers, San Francisco, July, 1999.

[8] R. Stefany, In a Perfect World, A Survival Guide for the Help Desk Technician, Create Space, New York, 2011.