A Hybrid Architecture for Mobile Geographical Data Acquisition and Validation Systems

Claudio Henrique Bogossian[1], Karine Reis Ferreira [1], Antônio Miguel Vieira Monteiro[1], Lúbia Vinhas[1]

[1]DPI – Instituto Nacional de Pesquisas Espaciais (INPE)

{bogo, karine, miguel, lubia}@dpi.inpe.br

Abstract. *Mobile devices, such as smartphones and tablets, are useful tools for in situ gathering information about spatial locations. Specialists need mobile geographical data acquisition and validation systems to be used in fieldwork and in places where there is limited or any network connectivity available. This paper presents an ongoing work on designing and implementing a hybrid architecture for this kind of systems able to work online as well as offline. The offline module of the proposed architecture is based on the OGC Geopackage standard. As part of this work, we tested and evaluated Geopackage documents as interoperable files between spatial data infrastructures and mobile geographical data acquisition and validation systems.*

1. Introduction

The recent advancements of GPS, wireless communication network and portable technologies have motivated the use of mobile devices for *in situ* gathering information about spatial locations and validating geographical data. Tsou (2004) defines the term *mobile GIS* to refer to an integrated technological framework for accessing geospatial data and location-based services through mobile devices, such as smartphones and tablets. He argues that there are two major application areas of mobile GIS, *field-based GIS* and *location-based services*. This work focuses on mobile field-based GIS, that is, mobile systems for geographical data collection and validation in the field.

Two examples of projects that need mobile field-based GIS are PRODES (Monitoring of Brazilian Amazon Rainforest) and DETER (Real Time Deforestation Detection System), developed by INPE [INPE 2014]. PRODES has been yearly monitoring deforestation since 1988 whereas DETER has been producing near real-time deforestation and forest degradation alerts for more than 5 million Km^2 in the Brazilian Legal Amazon. Specialists of these two projects require mobile systems to collect extra information about deforested regions (e.g. photos) and validate them in the field, including places where there is limited or any network connectivity available. Therefore, an essential feature of geographical data collection and validation mobile systems is the capability of working offline.

To meet this demand, this paper presents an ongoing work on designing and implementing a hybrid architecture for this kind of systems able to work *online* as well as *offline*. The offline module of the proposed architecture is based on the Open Geospatial Consortium (OGC) Geopackage standard [OGC 2014]. This work presents an evaluation of Geopackage documents as interoperable files between Spatial Data Infrastructures (SDI) and mobile geographical data acquisition and validation systems.

SDI is a sharing platform that facilitates the access and integration of multi-source spatial data in a holistic framework with a number of technological components including policies and standards [Rajabifard et al 2002] [Mohammadi 2008].

1.1. Related Work

Nowadays, mobile GISs have been widely used in different application areas, including location-based systems [Raper et al 2008], volunteered GIS applications (VGI) [Davis et al 2013] and field-based geographical data acquisition and validation [Tsou 2004] [Poorazizi et al 2008]. This work focuses on this last application area.

Tsou (2004) proposes a generic architecture for mobile GIS where there is a module, called "Geodata cache", responsible for storing geospatial data in a cache located in the mobile storage space or a flash memory card. The idea is to download customized datasets and synchronize them from GIS content servers. Poorazizi et al (2008) proposes two mobile GIS architectures for field geospatial data acquisition: one to work offline (Stand-Alone Client Architecture) and another online (Distributed Architecture).

Differently, we propose a hybrid architecture with two modules for geospatial data access, *online* and *offline*, based on OGC standards (Web Services and Geopackage). The compliance with OGC specifications assures spatial data interoperability between existing SDIs and the mobile systems. Nowadays, many data providers throughout the world have created their own SDIs, organizing and disseminating their geospatial data sets and metadata on the Internet via OGC web services. Accessing spatial data sets from distinct SDIs can improve the geographical data collection and validation task. For example, specialists from INPE's PRODES and DETER projects can use the protected areas disseminated by IBAMA via web services to help with deforested areas validation.

2. The Hybrid Architecture

Figure 1 presents the proposed architecture showing the two modules for accessing geographical data, "Online Data Access" and "Offline Data Access". The "Online Data Access" module accesses geographical data from SDIs through two kinds of well-known OGC web services, Web Map Server (WMS) and Web Feature Server (WFS) [OGC 2006] [OGC 2010]. This module only works online and will be used when there is network connectivity available in the field.

WMS standard provides a simple HTTP interface for requesting geo-registered map images from one or more distributed geospatial databases. The response to the request is one or more geo-registered map images (returned as JPEG, PNG, and others) that can be displayed in a browser application. WFS document specifies the behavior of a service that provides transactions on and access to geographic features in a manner independent of the underlying data store. It specifies discovery operations, query operations, locking operations, transaction operations and operations to manage stored parameterized query expressions.

The "Offline Data Access" module works offline and is responsible for accessing geographical data in the mobile storage memory. We propose to store them in OGC Geopackage files [OGC 2014]. The Geopackage specification defines a SQL database schema designed for the SQLite software library. This schema contains a set of pre-defined tables with integrity assertions, format limitations and content constraints to store spatial data sets and their metadata. GeoPackage files are platform-independent SQLite database files that contain vector and tiled raster data sets as well as their metadata. They are interoperable across different platforms, including personal computing environments and mobile devices.

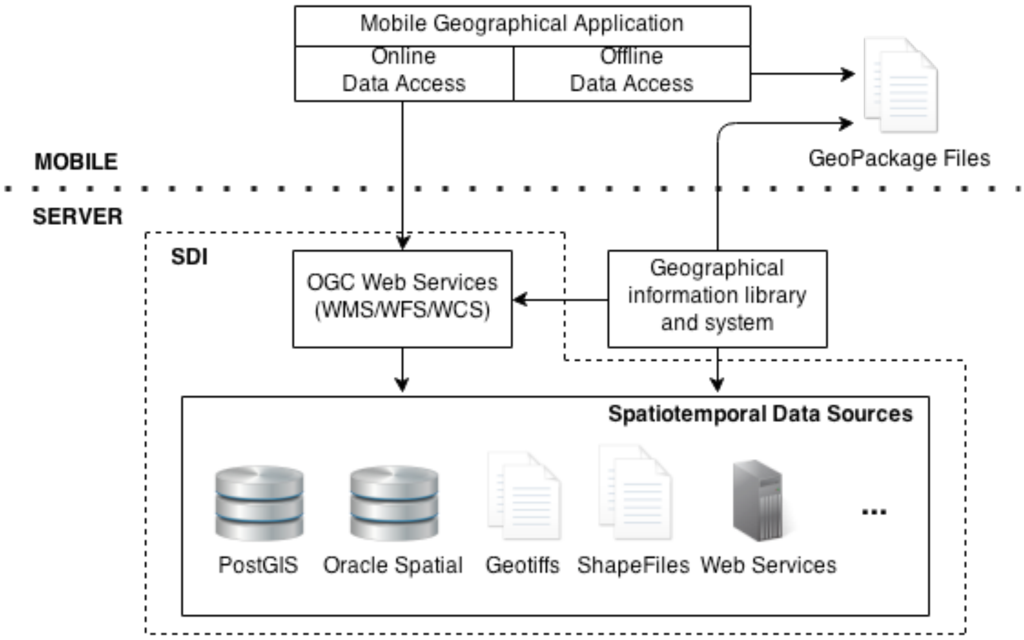

Figure 1. A Hybrid Architecture for Mobile Geographical Data Collection and Validation Systems

On the server side we need geographical information libraries and systems to decode and encode Geopackage files. These systems must be able to create Geopackage files from different kinds of spatiotemporal data sources, including files, spatial databases and web services. Users will use these systems to access different kinds of data sources, select the necessary spatiotemporal data sets from these sources to be used in the field, and export these data sets to Geopackage files. Users will execute this task of creating the necessary Geopackages files during the fieldwork planning phase.

4. Geopackage: Tests and Evaluation

As part of this work, we tested and evaluated the suitability of Geopackage files to support the interoperability between SDIs and the offline module of mobile systems for geographical data collection and validation. Mobile devices have limited storage capacity and processing power. Thus, we evaluated Geopackage documents by comparing their sizes and query processing times with other data sources.

We tested Geopackage files with *vector data* and *pyramid raster tiles* [OGC 2014]. Vector data is necessary for applications that need to handle feature geometries and attributes, for example, applications that execute spatial and attribute queries or edit the feature geometries. Pyramid raster tiles are useful for applications that display geographical information only as background layers. A pyramid structure organizes the raster tiles in a series of reduced/increased resolutions and is mainly used to improve the display performance.

4.1. Geographical Data Sets

For testing vector data, we created two data sets using the deforested regions detected by the projects DETER and PRODES, as shown in Figure 2. The first data set from DETER contains 439,596 regions or polygons detected from 2004 and 2012. The second one from PRODES contains 1,350,652 polygons detected from 2001 to 2012. From these two data sets, we generated three data sources; Geopackage vector files (GPKG), shapefiles (SHP) and PostGIS database (PG); and compared their sizes and query processing times. All these data sources were created with spatial indexes.

Since mobile devices have limited storage capacity, the files used in the "Offline Data Access" module must be as small as possible. Two possible options for this module are shapefile and Geopackage files; but we included PostGIS database running on local machine to illustrate our comparison. Figure 3 presents the sizes in megabytes (MB) of the three created data sources. The sizes of Geopackage files are 326 MB (DETER) and 1321 MB (PRODES), while shapefiles are 850 MB (DETER) and 2600 MB (PRODES). Geopackage vector files are around 50% smaller than shapefiles.

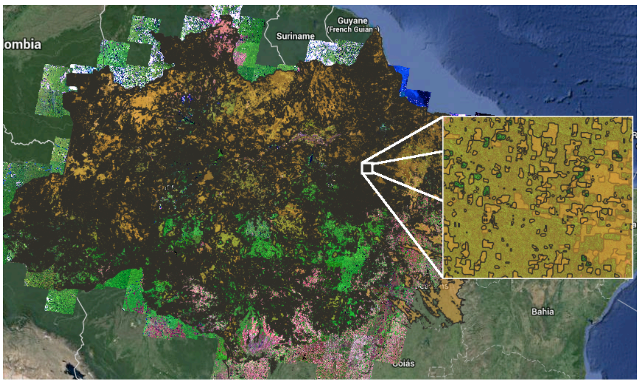

Figure 2. Data sets.

Figure 3. Data set sizes.

4.2. Attribute and Spatial Filter

Geographical data collection and validation applications can require operations that involve attribute and spatial filters. Examples of these operations are "*given a spatial location, return the attributes to be edited of the region that contains this location*" or "*return the regions whose areas are smaller than x*". Thus, we executed these two filters in the three created data sources and compared their runtimes. Runtime is the period for executing the query and fetching all returned records.

For attribute filter, we used the queries "*select all deforested regions detected by DETER project in 2003-01-01*" and "*select all deforested regions detected by PRODES project in 2005*". The first query returned 138,955 polygons and the second one, 241,439. The runtimes (in milliseconds) of these queries in the three data sources are presented in Table 1 and Figure 4.

For spatial filter, we used the bounding box of the Legal Amazon and selected all deforested regions inside this box. This query returned 430,044 polygons in the DETER data set and 1,300,552 polygons in the PRODES data set. The runtimes (in milliseconds) of these queries are presented in Table 2 and Figure 5.

Table 1. Attribute filter runtimes

Attribute Filter (Time ms)		
Data Source	DETER	PRODES
GPKG	1866	13700
PG	6245	12264
SHP	4163	29298

Table 2. Spatial filter runtimes

Spatial Filter (Time ms)		
Data Source	DETER	PRODES
GPKG	4197	23229
PG	11883	47902
SHP	3554	37423

Each runtime presented in Tables 1 and 2 is the average of ten iterations and all data sources were stored in the local machine. We can note that, in most cases, attribute and spatial filter runtimes in the Geopackage vector files are smaller than in the other data sources.

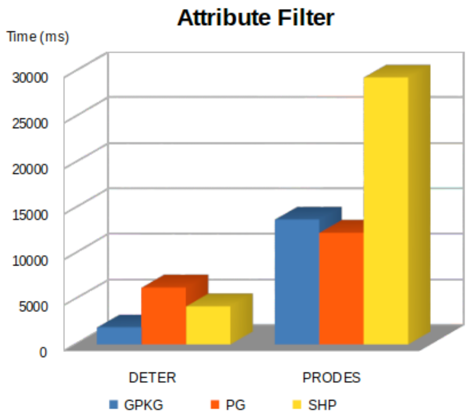

Figure 4. Attribute filter

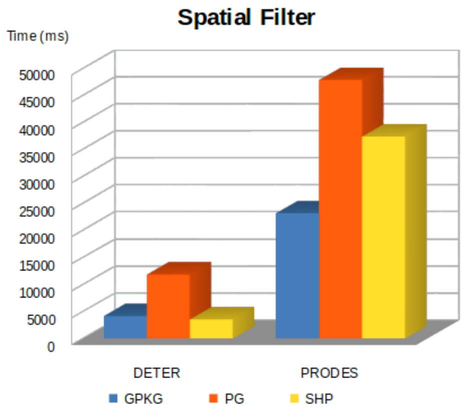

Figure 5. Spatial filter

4.3. Geopackage Tiles

For testing pyramid raster tiles, we created a Geopackage raster file from the deforested regions detected by PRODES project and Landsat 5 satellite images covering the Brazilian Legal Amazon. We created this file using tiles of 256 x 256 pixels and a pyramid with 10 levels of resolution. Figure 6 presents the Geopackage raster file, showing its level 7 in the big picture and its level 10 in the zoon area. Figure 7 shows the sizes in megabytes (MB) of the Geopackage vector and raster files. The vector file size is 1321 MB, while raster file is 552 MB.

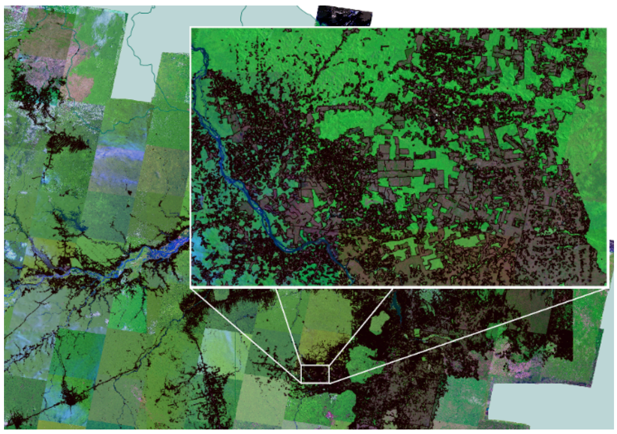

Figure 6. Geopackage tiles

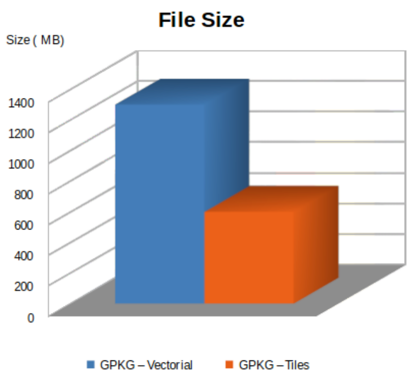

Figure 7. Geopackage file sizes

5. Final Remarks and Conclusions

This paper presents an ongoing work on designing and implementing a hybrid architecture for mobile geographical data acquisition and validation systems that can operate *online* as well as *offline*. We propose the use Geopackage documents as interoperable files between SDI and these mobile systems and show that this is viable through tests using Geopackage files with vector data and pyramid raster tiles.

In the tests, we used the GDAL/OGR library[1] to generate the Geopackage vector files and to access and execute the spatial and attribute filters in the three data sources. We also used the Mobile Atlas Creator[2] program to create the pyramid raster tiles from WMS server and the GeoTools library[3] to implement a JAVA code that stores these tiles as Geopackage raster files. The codes used in the tests are available in the GitHub: https://github.com/claudiobogossian/geopackage-test and https://github.com/claudiobogossian/GPKGTilesTest

As future work, we intend to develop a TerraView plugin that allows users to delimit an interesting area, access spatiotemporal data sets from different kinds of data sources and generate Geopackage files from these data sets. This plugin will be used in the server side and will play an important role in the proposed architecture, as described in Section 2. TerraView is a general-purpose GIS developed using the TerraLib GIS [Camara et al. 2008]. TerraView supports the development of plugin to enhance its functionalities.

References

Camara, G.; Vinhas, L.; Queiroz, G. R.; Ferreira, K. R.; Monteiro, A. M. V.; Carvalho, M. T. M.; Casanova, M. A. (2008) "TerraLib: An open-source GIS library for large-scale environmental and sócio-economic applications". *Open Source Approaches to Spatial Data Handling*. Berlin: Springer-Verlag.

Davis, C. A., Vellozo, H. S., Pinheiro, M. B. (2013) "Framework for Web and Mobile Volunteered Geographic Information Applications". *In: Proceedings of XIV Brazilian Symposium on Geoinformatics (GeoInfo 2013)*, November 24-27, 2013, Campos do Jordão, Brazil.

INPE (2014) Monitoramento da Floresta Amazônica Brasileira por Satélite (Monitoring the Brazilian Amazon Forest by Satellite). Available at www.obt.inpe.br/prodes.

Mohammadi, H. (2008) "The Integration of Multi-source Spatial Datasets in the Context of SDI Initiatives" *PhD thesis*, University of Melbourne. Available at: http://www.csdila.unimelb.edu.au/publication/theses/hossein-PhD.pdf (accessed in July 2014)

Open Geospatial Consortium – OGC. (2006) "OpenGIS Web Map Server Implementation Specification". Available at: http://www.opengeospatial.org/

Open Geospatial Consortium – OGC (2010) "OpenGIS Web Feature Service 2.0 Interface Standard". Available at: http://www.opengeospatial.org/

[1] Available at: http://www.gdal.org/

[2] Available at: http://mobac.sourceforge.net/

[3] Available at: http://www.geotools.org/

Open Geospatial Consortium – OGC (2014) "GeoPackage Encoding Standard". Available at: http://www.opengeospatial.org/

Poorazizi, E., Alesheikh, A. A., Behzadi, S. (2008) "Developing a Mobile GIS for Field Geospatial Data Acquisition". *Journal of Applied Sciences*, 8(18), 3279-3283.

Rajabifard, A., Feeny, M. E., Williamson, I. (2002). "Future Directions for SDI Development". *International Journal of Applied Earth Observation and Geoinformation* 4 (1), 11-22.

Raper, J., Gartner, G., Karimi, H., Rizos, C. (2008) "Applications of location–based services: a selected review". *Journal of Location Based Services*, 1(2), 89-111.

Tsou, M. H. (2004) "Integrated mobile GIS and wireless internet map servers for environmental monitoring and management", In: Special issue on Mobile Mapping and Geographic Information Systems, *Cartography and Geographic Information Science* 31 (3): 153–165.