Computers & Geosciences 60 (2013) 23–33

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

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

Using GIS servers and interactive maps in spectral data sharing and administration: Case study of Ahvaz Spectral Geodatabase Platform (ASGP)

Mojtaba Karami, Kazem Rangzan*, Azim Saberi

Department of Remote Sensing and GIS, Faculty of Earth Sciences, Shahid Chamran University, Golestan Boulevard, Ahvaz 6135743135, Iran

ARTICLE INFO

Article history:
Received 30 April 2012
Received in revised form
7 May 2013
Accepted 10 June 2013
Available online 28 June 2013

Keywords:
Remote sensing
Spectral database
Metadata
Web-GIS

ABSTRACT

With emergence of air-borne and space-borne hyperspectral sensors, spectroscopic measurements are gaining more importance in remote sensing. Therefore, the number of available spectral reference data is constantly increasing. This rapid increase often exhibits a poor data management, which leads to ultimate isolation of data on disk storages. Spectral data without precise description of the target, methods, environment, and sampling geometry cannot be used by other researchers. Moreover, existing spectral data (in case it accompanied with good documentation) become virtually invisible or unreachable for researchers. Providing documentation and a data-sharing framework for spectral data, in which researchers are able to search for or share spectral data and documentation, would definitely improve the data lifetime. Relational Database Management Systems (RDBMS) are main candidates for spectral data management and their efficiency is proven by many studies and applications to date. In this study, a new approach to spectral data administration is presented based on spatial identity of spectral samples. This method benefits from scalability and performance of RDBMS for storage of spectral data, but uses GIS servers to provide users with interactive maps as an interface to the system. The spectral files, photographs and descriptive data are considered as belongings of a geospatial object. A spectral processing unit is responsible for evaluation of metadata quality and performing routine spectral processing tasks for newly-added data. As a result, by using internet browser software the users would be able to visually examine availability of data and/or search for data based on descriptive attributes associated to it. The proposed system is scalable and besides giving the users good sense of what data are available in the database, it facilitates participation of spectral reference data in producing geoinformation.

1. Introduction

Since NASA's launch of Earth Resources Technology Satellite (ERTS) in 1972, remote sensing of earth surface has gradually established its firm ground in different disciplines, such as agriculture, geology, environmental monitoring, cartography, forestry, land use planning, and oceanography. With development of technology and methods, spectral reference data are now essential to validate models and to maintain sensor calibration post-launch (Milton et al., 2009). Field spectroscopy measurements are vital for vicarious calibration of remotely sensed data (Thome et al., 2008) and may be used for atmospheric corrections (Smith and Milton, 1999). Remote sensing now benefits from field spectroscopy in various fields such as: geology (Kruse, 2012), soil sciences (Croft et al., 2012), vegetation analysis (Jago et al., 1999), hydrography (Legleiter et al., 2009) and oceanography (Ahn et al., 2008). By providing endmember spectra, reference spectral data can be used for spectral unmixing (Asner and Lobell, 2000), or to improve image classification (Roberts et al., 1998). Therefore, the need for high quality and well-documented spectral collections has been noticed by many researchers (Ben-dor et al., 1999; Clark, 1999; Clark et al., 1993). One of the first experiences in providing spectral collections was made by U.S. Army Engineer Topographic Laboratories which collected visible-near infrared, thermal and fluorescence signatures (Satterwhite and Henley, 1990). Currently, publicly available USGS spectral library (Clark et al., 2007) and ASTER spectral library (Baldridge et al., 2009) contain thousands of spectral signatures from different objects and continue to release newer versions to date. Most of the existing spectral collections are based on hierarchical storage of physical files and have serious drawbacks such as limited scalability, low performance in data

---

* Corresponding author. Tel.: +98 611 3339338; fax: +98 611 3339338.
E-mail addresses: m-karami@mscstu.scu.ac.ir (M. Karami), kazemrangzan@scu.ac.ir (K. Rangzan).

0098-3004/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
10.1016/j.cageo.2013.06.007

query, low flexibility of descriptive data (metadata) structure, and dependence on file format (Bojinski et al., 2003). Bojinski et al. (2002) and Ruby and Fischer (2002) used relational database models for storage and management of spectral data to curb these deficiencies.

However, no special attention has been given to taking advantage of spatial characteristics of ground-based radiometric measurements in spectral data management systems to date, despite the benefits that spatial solutions may provide for storage, visualization, and transfer of the data resulting from ground-based observations.

This paper investigates the possibility of incorporating GIS servers in spectral data management systems, and is intended to draw the limitations and possibilities regarding the use of GIS technology in spectral data sharing and administration. We present Ahvaz Spectral Geodatabase Platform (ASGP) as a new prototype system for the purpose of spectral data management. In the proposed system relational databases, GIS servers and spectral processing units collaborate in a distributed framework, with the aim of storage, processing, and sharing of spectral data and associated metadata. By using system's web-mapping interface, users are able to perform simple or complex queries on metadata, in order to gain access to desired portions of spectral data, for example: “All igneous rock samples”, or “All ASD fieldspec 3 soil samples”.

2. Concepts

2.1. Long term usability of spectral data

Although recent improvements in portability of spectroscopy instruments have increased the popularity of field spectroscopy, field spectroscopy campaigns and spectroscopy experiments in laboratory are still time consuming and relatively expensive. Enhancing the reusability of spectral data would prevent repeated measurements, thus reduces these costs. Enhancement of the spectral data lifetime consists of two main steps: (1) to make data usable for others through a clear definition of targets and environmental conditions under which the experiments have been carried out and (2) to improve data access for end users. The first goal can be achieved by providing good documentation for data, which necessitates the design of a metadata system to structure the descriptive data. Associating supplementary photographs of the target, environment and sampling process is also a common practice by spectroscopy researchers to improve the interpretations. The improvement in data access can be achieved by utilizing a data sharing environment. Such sharing environment should be equipped with effective search tools to maintain ease of access, and should also guarantee integrity and security of the data.

2.2. Metadata

The quality of spectral measurements can vary depending on the operator and the calibration of the instrument and standard panel, and the localized environmental conditions. Hence, extreme caution should be placed on using reference spectra without metadata, and common practice should be to collect and document metadata associated with the spectral response (Pfitzner et al., 2006). However, identifying what knowledge about spectroscopic measurements is critical to record is often controversial, due to absence of widely accepted metadata standards and highly individual preferences among researchers from different backgrounds (Rasaiah et al., 2011).

Another challenge regarding the metadata schema design is the level of details that metadata should bear. Duval et al. (2002) stated that although detailed metadata descriptions may benefit other users by improving searching precision and quality assessment, it requires higher investment in metadata creation. In contrast, using simple descriptions may result in more false search results, or more efforts on the part of searchers to identify most relevant results. In this context, Ruby and Fischer (2002) used a detailed hierarchical attribute system for metadata with sub-attributes arranged under more general attributes, thus provided a logical organization for their spectral database and made it easy for users to perform both narrow and broad searches.

2.3. Data sharing

Spectral data might be a subject of interest for various research groups with different backgrounds. These data are often stored on DVDs and hard disks. Therefore, remarkable geographical separation between researchers or lack of information on availability of data would seriously limit the access of researchers to available data. A simple and easy to understand sharing environment in which the users be able to search for, download, or upload data can improve the accessibility of data for research groups. In addition to having a good conceptual design, data sharing systems face some challenges in order to gain popularity including: good and handy Graphical User Interface (GUI), simplicity of the process, security and privacy of data (Curry et al., 2010). In this manner, interactive user interfaces are mainly advantageous over “menu-selections” or “form fill-in” methods due to faster learning, easier error recovery, immediate feedback, and less anxieties for inexperienced users. In case design limitations exist, such interfaces can easily be augmented with texts, menus, and form fill-ins (Galitz, 2007).

2.4. Spectral databases

Although file-based data management systems may perform very well in sharing static contents, serious drawbacks can be expected when relying on such methods for manipulating data or searching through constantly-changing datasets. Data redundancies, inconsistencies, low scalability, and lack of efficient methods for handling concurrent and/or failed editing sessions are among the issues that limit the efficiency of file-based methods for scenarios that encompass considerable traffics or large datasets (more information regarding the inefficiencies associated with the use of file-based methods can be found in Sears et al., 2006). Moreover, users of file-based spectral libraries may easily lose sight of the entirety of information available to them (Bojinski et al., 2000).

On the other hand, relational database models and Relational Database Management Systems (RDBMS) are traditionally used to handle large datasets and to maintain logical consistency and integrity of data. In addition to non-redundant storage of data, RDBMSs take advantage of mechanisms like caching in order to minimize physical disk reads and comprise sufficient algorithms to maintain data integrity in distributed nearly-scalable architectures (Hodak and Kumar, 2007).

Given the fact that in a sharing environment spectral data should be associated with descriptive data (Hueni et al., 2009) and the system might have a non-static nature that involves continuous updates, concurrent use, and large traffics, relational databases would be ideal candidates for spectral data management. Besides the fact that databases are known to provide unique capabilities for query-handling and maintaining concurrency of edits/updates, Sears et al. (2006) have shown that relational database's performance in transmitting files smaller than 256 KB can, interestingly, be comparable to that of a file-based system. That is to say, a relational database can perform gracefully as a storage for large number of small files (e.g., spectral signatures), as well as the metadata associated with them.

2.5. Spatial data

In-situ spectral measurements are spatial data in nature. In many cases, there are spatial variations among spectral samples. For example, spatial variations in soil conditions may influence spectral responses of a plant species (Mistele and Schmidhalter, 2008). Particularly for atmospheric corrections, the field spectra should definitely be associated with pixels on satellite images (Clark et al., 2002; Karpouzli and Malthus, 2003). A handheld GPS or an automatically operated GPS attached to the spectroradiometer instrument should be used in order to capture geographical coordinates of the samples. ASGP considers spectral data as geospatial data and mainly uses interactive web-maps in order to visualize spectral datasets. Appropriate point symbols are used for representation of sampling locations on a map, giving the users the ability to discriminate between samples related to different features and materials.

3. Implementation

3.1. Architecture

The ASGP is a distributed network of collaborating units working in a thin-client architecture. The system comprises database server, GIS server, web/application server, and spectral processing unit as illustrated in Fig. 1. Implementation of thin-client architecture exempts the users of ASGP from any form of local data storage and processing (Alesheikh et al., 2002). Instead, the spectral data processing algorithms are included in system's processing node and the data are stored in database with the aid of a spatial middle-ware. The users of ASGP mainly establish connections to web/application server in order to reach the web-mapping interface of the system. In this model, the GIS server bears the responsibility of providing spatial data services for users.

3.2. Metadata structure

A spectral database to be used in a multidisciplinary and multiuser environment should be able to adapt itself to data with different levels of details from different sources and in different formats. This feature of the system can be translated into independency of file format and flexibility to attribute changes (Bojinski et al., 2002).

Although standardization of metadata schema for in-situ spectral data is beyond the scope of this research, we used a "merged superset" of metadata elements from various spectral databases (including Baldridge et al., 2009; Clark et al., 1993; Hueni et al., 2009; Pfitzner et al., 2006) in order to form a robust metadata schema for ASGP. The selected metadata elements are structured in a two-level hierarchical metadata system, providing means of describing a target from low to high level of details. Robustness of the ASGP metadata system is maintained using (1) a two-level hierarchical metadata system to define samples, and (2) a non-constrained metadata policy.

3.2.1. Hierarchical metadata system

The proposed schema describes samples in two levels: (1) Generic descriptions about campaign, personnel, sampling geometry and sampling conditions in addition to feature-type, and (2) detailed feature-specific descriptions. Six main feature types, i.e., Vegetation, Rocks, Soil, Water, Snow/Ice, and Man-made (Impervious) form the coarse level of target description, wherein "the basic knowledge of investigator" is sufficient to describe the target. This initial basic classification also allows for implementation of detailed feature-specific schemas for each feature type separately. For example, species affect vegetation spectra and the texture affects soil spectra. It would be useless and confusing to assign texture values to vegetation or species to soil. Thus, the schemas of each of

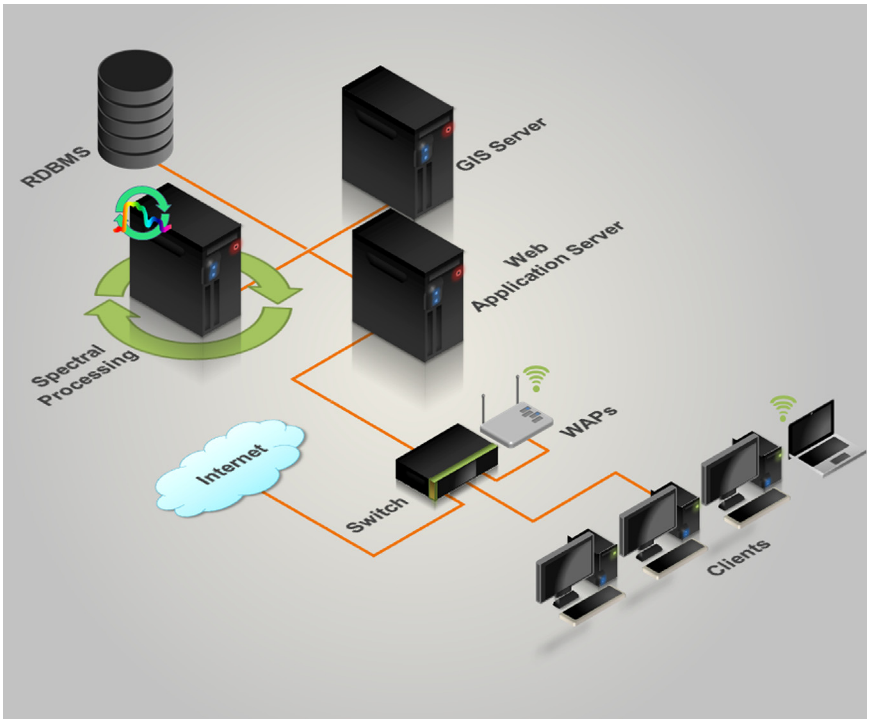


Fig. 1. ASGP system architecture.

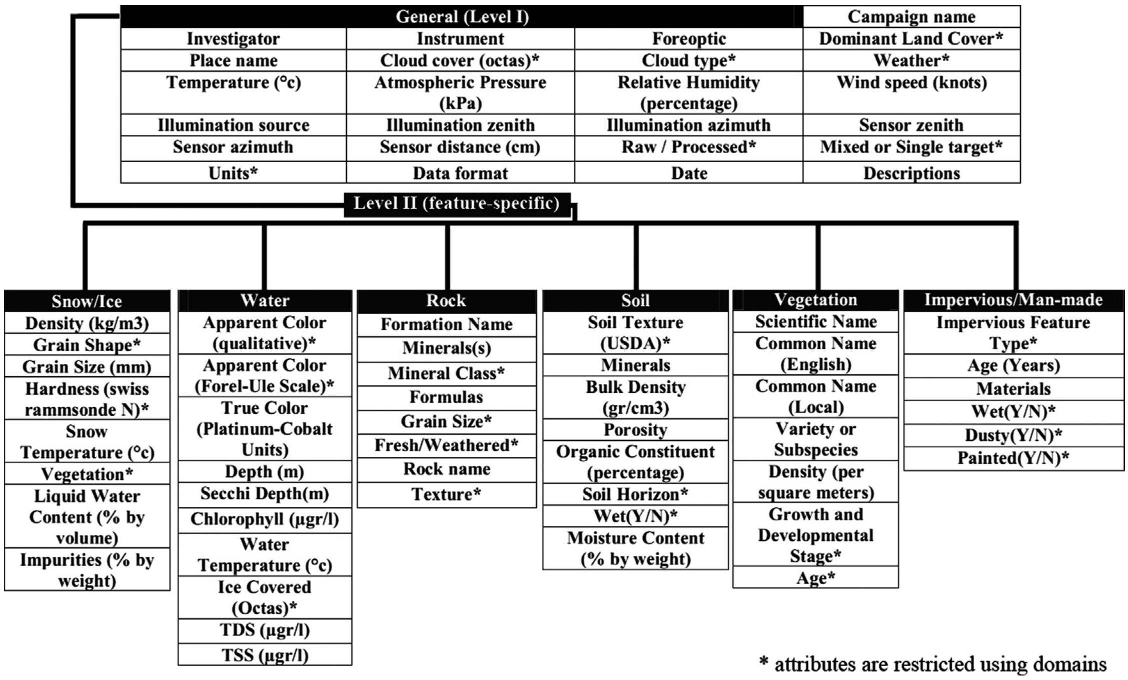


Fig. 2. ASGP metadata schema.

the 6 feature-types are limited to a pre-defined sub-schema, only containing metadata elements related to that particular feature-type. ASGP has 6 different feature-specific tables each containing items that could affect spectral response of that particular feature type (Fig. 2). It is also preferable that metadata schemas match available classification or identification schemas in related fields of research (Lanz et al., 2007; Satterwhite and Henley, 1990).

In this regard, some of the existing spectral databases use standard land-cover classification schemes for target description (e.g., Hueni et al., 2009). This approach is useful since it facilitates the adoption of spectral datasets in automatic or semi-automatic classification and analysis of satellite and airborne hyperspectral products. However, in some cases land cover classification schemes do not encompass the spectral characteristics of the targets and tend toward the description of land use. Moreover, there is a concern over the sufficiency of the details for description of the targets when solely relying on a land cover scheme. Thus, ASGP separately captures land cover properties as well as target descriptions. The feature-specific items are carefully selected based on presence of at least one case in available literature at the time of study that confirms such effect.

3.2.2. Non-constrained metadata policy

Although ASGP uses a very detailed and feature-specific descriptive system, users are allowed to fill-in any metadata element available to them. This non-constrained strategy is intended to facilitate absorption of spectral data with low-quality and incomplete metadata or in situations where metadata schema of the data only partially matches the database schema. However, despite the increased chance of data absorption, this strategy is prone to be a source of data inconsistency. Diversity of vocabularies used to fill-in metadata elements and the overall decrease of metadata quality in the database due to incomplete records would negatively affect users' searching capabilities and usability of spectral datasets.

In order to minimize the negative effects of typographical errors and diversity of vocabularies on search process, the allowable metadata attributes for some elements are restricted, forcing users to select from a predefined list when adding data. However, these restrictions are imposed with caution and only on metadata fields where the number of imaginable attributes is finite (e.g., Dominant Land Cover or Soil Texture).

Furthermore, a quantitative estimation of metadata quality is useful for filtration of low-quality or unwanted data by both administrators and end-users. In ASGP, the “Processing node” is responsible for automated calculation of metadata quality (completeness) indexes during metadata updates (see Section 3.6.1). This quantitative estimation can be used by administrators or automated algorithms to periodically move or remove the low-quality data, or simply be used by users to limit search results to a desired level of metadata quality.

The participating users of ASGP are also provided with pre-defined field sheets, in order to help them organize the acquisition and documentation of proper metadata in-situ.

3.3. Data model

As one needs to keep record of spectral samples, every spectral sample consists of some elements: sampling location, description of target and environment, the spectral file, and associated media. With respect to these elements and given the fact that the design of the system is based on spatial identity of the samples, GIS friendly formats for example the widely used ESRI shapefile would be a natural choice. However, poor performance of shapefiles for large amounts of data is a limiting factor. Moreover, the system should be editable and viewable by many users concurrently. Thus, ESRI Geodatabase data model (Zeiler, 2010) is used for storage of spectral data and associated metadata. In the proposed spectral geodatabase, 6 feature classes represent 6 feature types, with each point feature representing a spectral sample obtained from field campaigns. This way, the captured metadata could be imported into designated attribute fields, while the associated photos and spectral files are attached to the feature. Furthermore, the attribute domain concept is implemented by employing the built-in functionality of the Geodatabase model to define domains and subtypes.

3.4. Storage

ArcSDE technology (ESRI, 2004a) serves as a connector between GIS software and RDBMS software, and manages storage and access of a geodatabase within a RDBMS. Hence, the system benefits from scalability, security and performance of RDBMS software in storage, query and access. ASGP uses ArcSDE for Oracle 11g (ESRI, 2010) and Oracle Database 11g R2 (Oracle, 2011) for the storage tier.

3.5. GIS server

ArcGIS server (ESRI, 2010) is used as GIS server software for ASGP in order to support a web mapping application. ArcGIS for server supports several Open Geospatial Consortium specifications in data publishing (e.g., WMS, WFS, KML) and provides a built-in tile cache engine.

3.6. Spectral processing node

A stand-alone application consisting of data parsers and spectral data workflows form the ASGP's spectral processing node, bearing the responsibility of estimating metadata completeness, preparing spectral plots, and performing instrument-specific corrections. The spectral processing algorithm, which is running periodically according to a pre-set schedule, is the only component of the system that is aware of spectral data contents. The flowchart of the ASGP spectral processing algorithm is shown in Fig. 3.

By scanning the most recent GIS server logs every 5 min, the program makes separate lists of new metadata updates as well as file attachment tasks. Further, the algorithm downloads the required data and/or metadata directly from SDE database tables using an Open Database Connectivity (ODBC). The results of metadata quality calculations are inserted back into the database tables by establishing an ODBC connection. However, ArcGIS Server's REST interface is being used in order to upload compressed ZIP files generated by the spectral data processing workflows. Briefly, the ArcGIS server's REST interface provides means for consumption of GIS services through logically organized URLs. That is, all resources exposed by ArcGIS server in form of REST-ful web services are accessible through a hierarchy of endpoints or URLs where clients can retrieve or manipulate them using GET and POST methods. For example, attributes and coordinates of the feature number 1045 from Rocks class can be retrieved in JSON format by sending a GET request to:

`http://<server name>/ArcGIS/rest/services/<name of the service>/FeatureServer/1/1045?f=json&pretty=true`

This logical structure enables the ASGP algorithms to dynamically build URLs, send GET requests to the ArcGIS server, fetch JSON data, or upload processed files via POST method.

In order to visualize spectral data for users, the generated spectral plots are uploaded to Web/Application Server's shared directory, while the database tables are being updated with corresponding URLs. Further, the web-mapping application visualizes spectral data plots for users in form of pictorial data by using URL attribute of each feature.

3.6.1. Metadata quality flags

The quality of metadata records in public databases is subject to serious concern, since the lack of supervision on data acquisition in addition to lack of widely accepted metadata schemas increase the heterogeneity of data. It becomes impractical for users and administrators of large databases to individually assess quality of

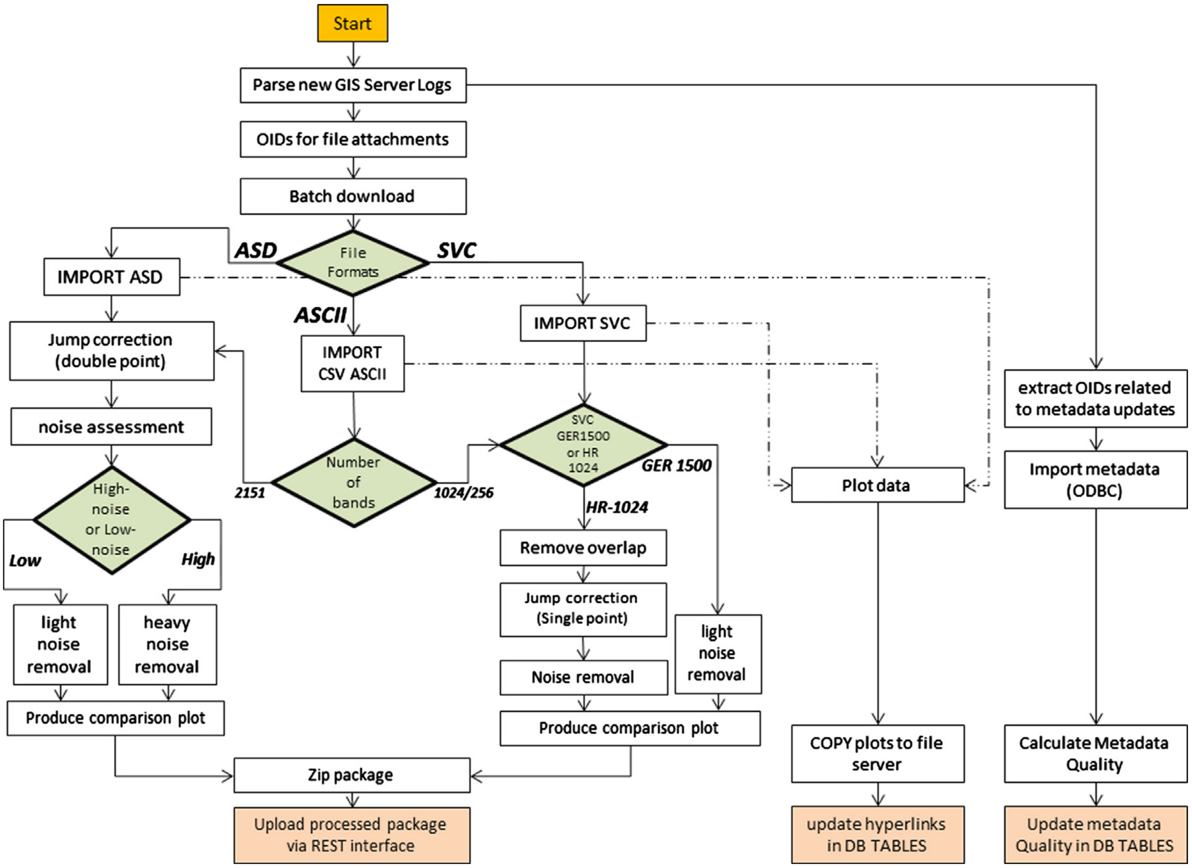


Fig. 3. Flowchart of ASGP spectral processing algorithm. The algorithm performs three main tasks: calculation of MCI, plot generation, and data processing.

metadata. In this vein, implementation of quantifying methods for assessment of metadata quality allows for more automation in quality maintenance, and makes it easier for users to identify high quality data. Metadata Coverage Index (MCI) is a straightforward metric for evaluation of metadata richness. The MCI is simply defined as the number of metadata fields in a record for which the information is provided, expressed as a percentage of the total fields available (Liolios et al., 2012). ASGP processing algorithm calculates MCI for metadata update tasks, and inserts the results back into database tables. Users are able use this index for refinement of search results. MCI could also be used by ASGP administrators in automated or manual filtration of data, based on a given quality threshold.

3.6.2. *Spectral data processing*

Data resulting from field spectroscopy are often noisy data. Small amounts of noise on the data can be removed by the use of smoothing methods. However, certain regions of the spectra known as atmospheric water absorption bands are dominated by noise, and in many applications, these bands should be entirely removed from the spectral data. Field spectroradiometers are also known to produce systematic errors, which should be corrected prior to data use. Therefore, there is a high potential for increasing the automation in routine spectral data processing tasks, e.g., jump correction for ASD and smoothing for noisy data. Several spectral data workflows are included in ASGP data processing algorithms in order to perform routine corrections on spectral data upon arrival (see Fig. 3).

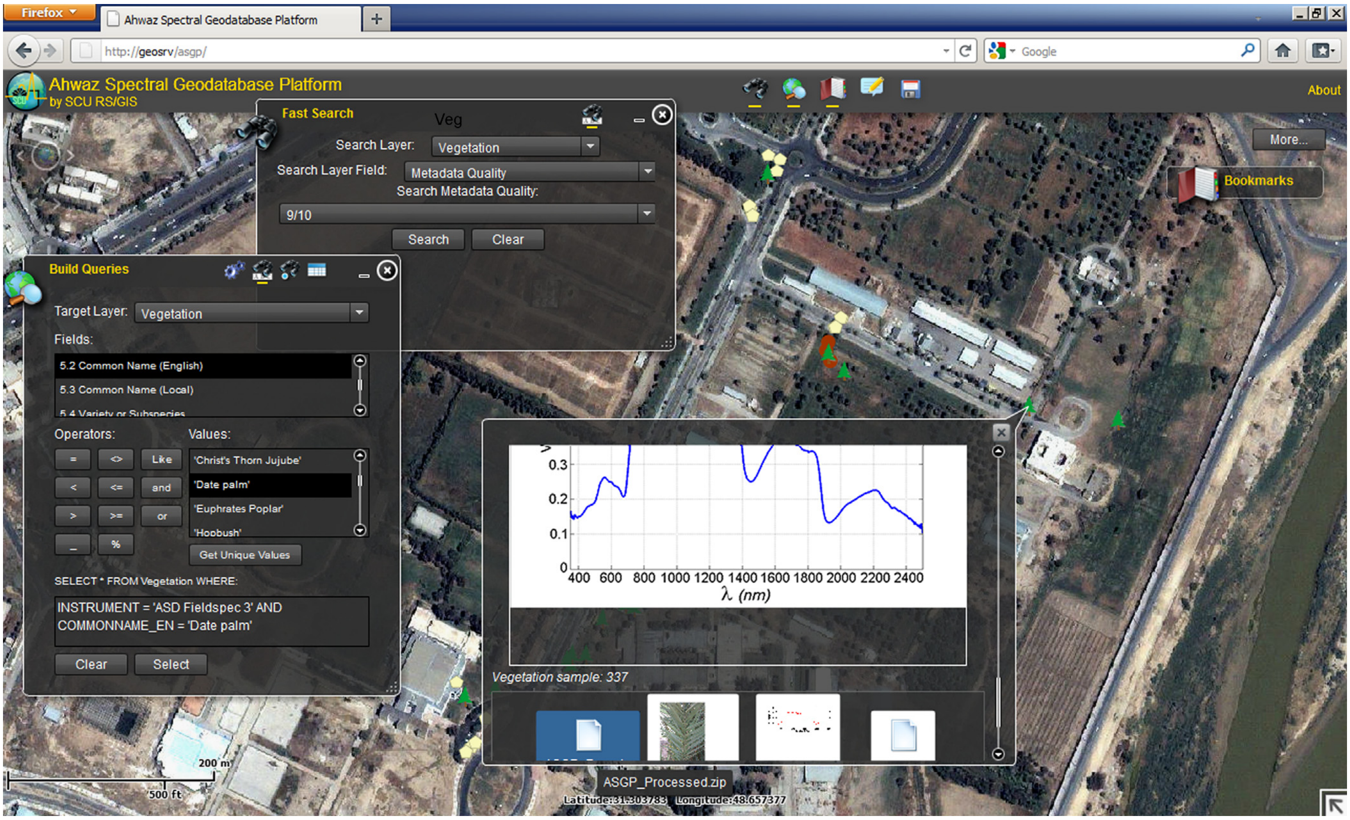
After performing additive and multiplicative jump-corrections (Rueda and Wrona, 2003), a modified version of Savitzky–Golay smoothing filter applies to the data. For ASD Fieldspec 3 data, mean Root-Mean-Square Error (RMSE) of the data over the best third-order polynomial fit is being used in order to assess the amount of noise on the data. Further, high-noise data receive higher degrees of smoothing in noise-dominated water absorption bands.

The final product of ASGP processing task is a compressed zip file, which is automatically attached to the feature using ArcGIS server's REST interface, and is stored in the database tables as BLOB. These packages usually contain the results of additive and multiplicative jump corrections in addition to the results of noise removals for each one, all in the form of comma delimited text files.

3.7. *User interface*

ArcGIS viewer for Flex (ESRI, 2011) is an open-source interactive web-mapping application, primarily based on ArcGIS API for Flex. The interactive web-mapping solution provides source-codes for developers as well as compiled out-of-box packages. New functionalities, e.g., map navigation and searching, can be added to the viewer in form of widgets.

A web-mapping application based on ArcGIS viewer for Flex 2.5 is used in ASGP application-tier, providing regular functionalities and ease of access for end-users (Fig. 4). Three separate versions of the application are prepared including: Offline mapping, Online mapping, and Online data-editing. The mapping services aim at providing means of search and data export, while the editing service provides means for adding new data and can be protected by password. Mosaic of High resolution World View-1 panchromatic satellite images of entire Ahvaz city along with high resolution Quickbird satellite images of Shahid Chamran University of Ahvaz are used as basemap layers for the offline mapping application. However, the online mapping application takes advantage of internet for its underlying basemap service.



Fast Search

Search Layer: Vegetation

Search Layer Field: Metadata Quality

Search Metadata Quality: 9/10

Build Queries

Target Layer: Vegetation

SELECT * FROM Vegetation WHERE:

INSTRUMENT = 'ASD Fieldspec 3' AND COMMONNAME_EN = 'Date palm'

Vegetation sample: 337

Fig. 4. Screenshot from ASGP interface. Users can search through database using either FastSearch or QuaryBuilder tools. Metadata attributes, associated files, and plot of the spectra are accessible from pop-ups.

Users may click on sampling points in order to access a pop-up containing pictorial and textual descriptive data as well as spectral data files associated with the samples. In addition, two separate search tools are available in order to help users to perform simple or complex queries on the available data (Fig. 4).

When using the editing application, after placing sampling points on the map, ASGP users complete pop-up forms in order to enter metadata for the new records. Pictorial and spectral data files could be uploaded as attachments to the feature. The processing algorithms, being located on spectral processing server, automatically process new records and upload the results to the database. Hence, the processed spectral data become accessible as new file-attachments for the sampling point.

3.8. *Performance tests*

In order to test the performance of the system in data transfer, synthesized random datasets are generated, and the speeds of reading the records from different storage options are measured. The synthesized random test set consisted of total number of 38,400 records with rich metadata (MCI=1). The schema of the test set were the exact similar of the ASGP schema, and its records were equally distributed among 6 feature-types of ASGP with each feature-type having a unique table. Moreover, the synthesized data were randomly distributed around the globe, in order to minimize the possible confounding effect of spatial distribution of samples on the performance of ArcSDE. The tests were carried out on a windows server machine equipped with a 2.8 GHz Intel Pentium G6950 and 8 GB of RAM at a clock speed of 1333 MHz.

The file-based and the database methods used for storage of the synthetic test sets include: ESRI shapefile (.shp, .shx, and .dbf) with total size of 58.1 Mbytes, Microsoft Excel spreadsheet (.xls) with the size of 16.3 Mbytes, Matrix data format (.mat) stored as structured arrays with the size of 1.493 Mbytes, Geodatabase feature-classes being accessed through ArcSDE middle-ware, and the database tables being accessed through ODBC using the following SQL queries:

```
SELECT * FROM <SDE.FeatureclassName>
```

Another test was carried out, however, for comparing the performance of different methods of storage for spectral files, regardless of the metadata storage method being used. That is, in order to evaluate the performance of file storage strategies, a total number of 10,000 spectral files—which will be referred to herein as attachments—were stored in either a file server or a database as BLOBs, while the metadata were stored in a database for both of the test cases. In the mentioned file-based attachment storage, the client's pathway to access the attachments consists of fetching the attachment URL from the attribute fields through REST interface, and subsequently sending GET requests directly to the file server. The SDE-based pathway also consists of obtaining the attachment URL from REST. However, in this case the attachment URL is indeed an indirect call to ArcSDE for extraction of the attachment from database tables. Finally, in line with our objectives, the performances of the RDBMS and IIS in batch-exporting large number of spectral signatures without mediation of ArcGIS server are also measured.

4. Results and discussion

4.1. *Metadata schema*

Although some suggestions have been made regarding metadata schema of spectral data collections (e.g., Milton et al., 2009; Hueni et al., 2009; Pfitzner et al., 2006), determination of metadata requirements is still subject to ongoing research (Rasaiah et al., 2011).

Many researchers propose allowing flexibility into metadata schema of scientific data management systems, in situations where lack of standardization or presence of multiple metadata schemas increases the data heterogeneity (Curry et al., 2010). Such systems allow users to define the metadata schema of their own data collections, thus expedite participation of researchers in data-sharing. However, the resulting plethora of metadata elements would seriously harm data consistency, and makes it time-consuming—if not impossible—for users to find relevant data.

Another workaround for the de-facto problem of heterogeneous spectral data is merging current sets of schemas in a structured, logical and comprehensive manner with minimal replications, to form a robust metadata structure. The resulting similarity between schemas could minimize the amount of metadata loss during data exchanges.

The strategy of ASGP is to absorb data as much as possible regardless of quality, but to provide good search capabilities to minimize pain on behalf of users during search process. This strategy, if implemented correctly and accompanied with quality checks, would provide good opportunities for up-scaling of ASGP to the extent of a crowd-sourced spectral data management system. Good search capabilities are achieved via utilization of search tools (both single-attribute and complex multi-attribute) as well as putting restriction on allowable attribute values of each field. Such restrictions would help to maintain data integrity, therefore reduce user's effort during search process. The authors chose these strategies due to lack of a widely accepted standard method to define what descriptive data should be collected during spectroscopy measurements.

Additionally, by utilization of a MCI calculator in spectral processing unit, administrators and users may overcome the problem of low quality metadata in such a collaborative database. Nevertheless, it is important to note that metadata flags represent the quality of metadata contents and the quality of spectral data can only be partially assessed by metadata. In addition, calculation of MCI by giving an equal weight to various metadata elements should only take place in situations where there is no information available regarding the criticality of metadata elements. Besides the quality of metadata records, the MCI can also be evaluated for each metadata element through the entire dataset. This can particularly be useful in assessing the practicality of metadata schemas and metadata elements. Unfortunately, due to small number of datasets this concept is not implemented in analyzing ASGP schema yet.

4.2. *Architecture and software components*

A summary of the current functionalities of ASGP together with the capabilities that can be improved are shown in Table 1. It should be emphasized that this is an exploratory study with all ensuing limitations. Currently, the processing unit supports a limited number of prevalent spectral file formats and instruments. In addition to drawing spectral plots for visualization of spectra in the web-mapping application, the algorithm performs jump corrections with both multiplicative and additive approaches (see Fig. 3). A comparison plot which is automatically generated by spectral data processing algorithm for an extremely high-noise ASD Fieldspec 3 data is shown in Fig. 5.

Current GIS solutions take a step further than only providing a geometrical calculation engine and means of visualization for spatial data. GIS server solutions are now capable of efficiently storing data for multi-user editing and offer a variety of options to maintain data security (ESRI, 2005) and consistency (ESRI, 2004b; ESRI, 2007), thus provide a safe and efficient framework for

Table 1
Current functionalities of ASGP system and possible improvements.

Functionality	Current state	Possible improvements
User interface	– Web mapping interface with pan, zoom, go to, fast search, query builder – Desktop GIS software supporting ArcSDE	Adding the ability to import non-spatial spectral datasets
System's availability	Local (institutional)	Internet-based (crowd-sourced)
Quality checks	– MCI calculation for metadata updates	To improve MCI calculation to distinguish between critical and optional metadata elements
Data export	– Web-mapping application – REST interface – Geodatabase XML export (exports both data and schema using ESRI ArcCatalog) – GML (a form of XML dialect) – ArcSDE command line	Batch export functionality for Web-mapping application, e.g. XML export for selected features
Data import	– Web mapping interface – Desktop GIS software supporting ArcSDE – REST interface or ODBC connectivity to RDBMS (suitable for automated algorithms)	– Import or update metadata from the spectral file headers – Batch attribute update
File formats	All file formats are accepted, however only these file formats are readable for ASGP: – ASD – SIG – TXT and ASCII	ENVI spectral library
Instruments	All instruments are accepted and plots are generated for supported file-formats. However, data processing and corrections only apply to these instruments: – ASD Fieldspec 3 – SVC GER1500 – SVC HR-1024	Other SVC and ASD products
Available data	Total number of 182 spectra, mostly obtained from vegetation and rock samples	Collaboration with research groups should be improved

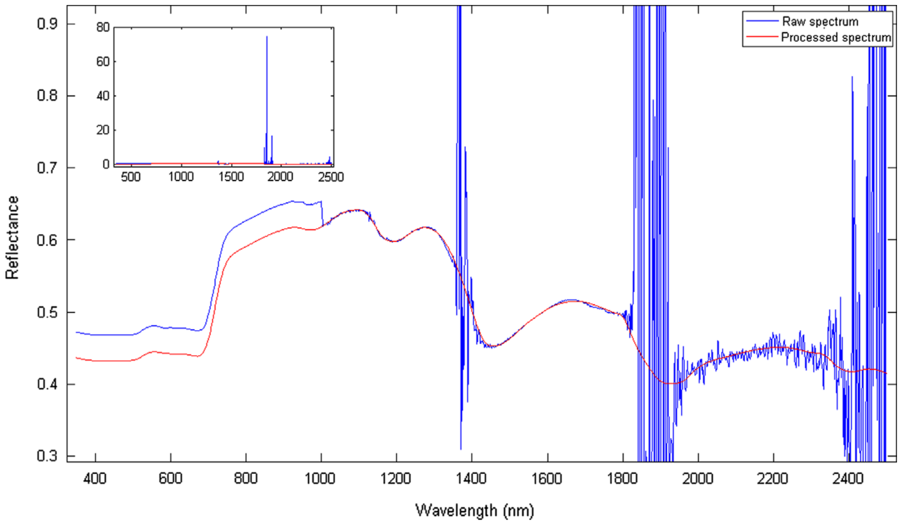


Fig. 5. Comparison plot for ASD Fieldspec 3 data obtained from a vegetation sample.

concurrent input and output flow of spatial data in a multi-user environment (Peters, 2009). The scale of such GIS services range from local networks to cloud-based GIS services (Muzafar et al., 2011).

The spatial information abstracted into map objects give a better view of the entirety of data, and aid users in identifying spatial patterns. Although spatial characteristics of spectral samples might sometimes be neglected, the data resulting from field spectroscopic measurements are geospatial in nature, and can be managed by geospatial information systems. In comparison to pure textual representation of data, users of geospatial information systems perceive different aspects of information in different forms and directly interact with these aspects in a manipulative manner. Remote sensing researchers often use field spectroradiometric data alongside other types of geodata for a variety of multidisciplinary analyses. The use of geospatial technology in storage and presentation—which is increasingly becoming reliant on standard specifications in order to facilitate interoperability—provides means of exchange and adaptation of spectral datasets in larger geospatial information systems.

Although the ASGP database is accessible via its web-mapping application, data can also be accessed using other software if needed. For example ArcGIS desktop (ESRI, 2010) or any web-mapping application which uses WMS and WFS specifications can establish connections to ASGP. Such compatibilities provide several benefits, such as directly incorporating spectral datasets alongside other geodata, or a possibility to develop special environments to make use of data in spectral data mining (Forestier et al., 2009; Shepherd and Walsh, 2002; Zi-li et al., 2011).

Regarding the choice of storage for such a GIS-based system, available strategies can be put into three main categories: file-based storages, central databases, and distributed databases. The database storage methods may also involve the use of spatial database extensions or spatial middle-wares (e.g., ArcSDE) in order to better cope with spatial characteristics of large datasets. With respect to the scale of the current study, a central database seems to be sufficient, since no performance issues observed in ASGP storage-tier during the tests.

With regards to the choice of the database, Oracle 11g is used as RDBMS in this research due to its scalability, multi-level security (Huey, 2011) and good performance (Chan and Ashdown, 2011). Oracle 11g has the record of fastest online transaction processing (TPC-C) at the time of study (TPC, 2012).

4.3. *Performance test results*

CPU usages were high for all 5 test cases, and no RAM saturation was observed during the tests. The results of metadata exports and the regression equations derived from each one are shown in Fig. 6. The results indicate that file-based storage methods have significantly lower performances in comparison with database storages. With respect to the observed near-linear increase of time-to-read for all five methods in this study (min $R^2 > 0.95$), it can be said that each record requires a minimum average of 0.0007 s to be read in case of using file-based storages. However, performing queries on database tables without the use of middle-wares may reduce this time to average of 0.0005 s for each record, providing 1.4 times faster data export on average. According to the data resulting from data export on this particular hardware setup, the use of ArcSDE could increases this speed to 7 times faster on average, in comparison with the fastest file-based storage in the test, i.e., ESRI shapefile.

Time-to-read values for three file-based methods also show a significant difference, which is likely due to difference in logical structure of the file formats. In case of XLS files, however, the relatively high value of offset (approximately 3.69 s) is presumably related to prerequisite processes associated with data parsing methods, hence does not reflect the innate performance of this file-format.

Interestingly, the results of ArcSDE exports also show a relatively high offset (approximately 3.717 s), rendering the database query method superior for reading data, wherein the dataset contains less than 12,300 records. However, insignificant amount of time required for accessing small SDE datasets by other client software (e.g., less than 0.1 s required for ArcGIS desktop) indicates that this delay time is not related to the increased communication path between client and data when bypassing a mandatory process, and is mostly caused by the test method itself. Overall, it can be concluded that in this particular setup, ArcSDE is approximately 5 times faster in data export in comparison to directly querying tables residing in the same RDBMS. This effect is presumably due to ArcSDE's utilization of Oracle Call Interface (OCI) and bind variables in contrast to using traditional ODBC.

Regarding the attachment export tests, the file server method could provide an average export speed of 0.839 s per attachment, which is slightly higher than the average of 0.814 s for SDE-based feature attachments. By removing the required time for acquiring the URLs, however, it revealed that IIS was able to export each file in 0.0198 s on average, while the export of SDE attachments in non-cached and cached states took 0.1074 and 0.0883 s per file, respectively.

Nevertheless, more benefits can be expected from database in terms of performance during batch export scenarios, e.g., during data migration or backups. The performance measurements for batch export of 10,000 files from file server, and 10,000 similar BLOBs through ODBC indicate that the results of the database (0.0027 and 0.0016 s per BLOB for non-cached and cached access, respectively) are quite comparable to that of the file server (0.0198 s per file). That is to say, the database was able to adequately reduce overheads per attachment in batch export scenarios. It should be emphasized here, that we used the term "batch export", which by that we mean the BLOB data are not turned into real files during the process. Conversion of BLOBs to files, which is necessary for the end-users, would definitely require client-side or server-side process and perhaps yield results similar to that of the attachment exports discussed earlier.

Finally, it is worth pointing out that the results of our performance tests do not necessarily represent the ultimate performance of the software we used. As one can infer even from

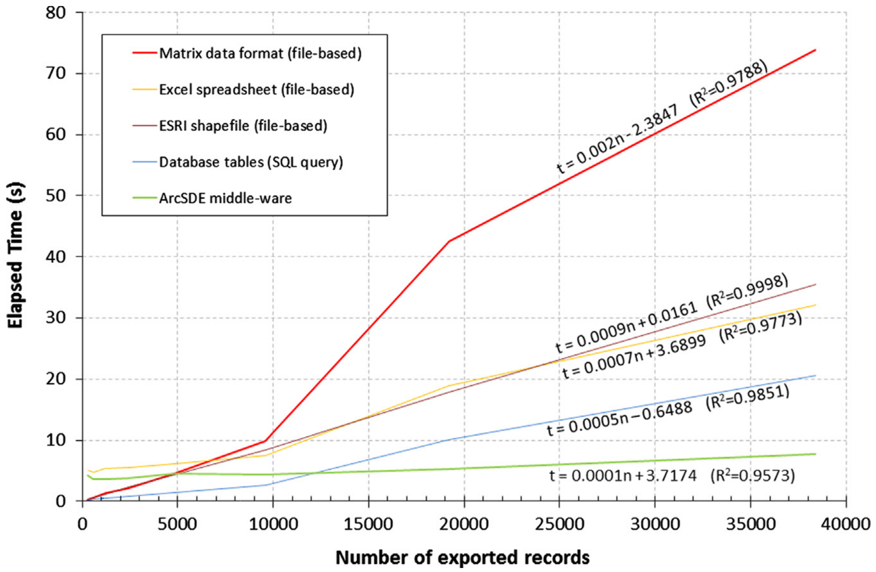


Fig. 6. Export rate for synthetic random data in different storage scenarios.

these results, the data access performance can be drastically affected by the way the data are structured and stored, the requests are generated, and the way the data is being accessed.

5. Conclusions

The purpose of the current study was to investigate the possibility of utilizing GIS technology in spectral data management systems. The proposed system consists of a central geodatabase stored within a relational database, a spectral processing unit, and an interactive web-mapping application powered by a GIS server. In this system, the robust design of the metadata system aims at storage of spectral data and metadata from variety of sources and with different levels of quality together in a central database. Implementation of RDBMS in the storage tier provides scalability, performance and security, while geographical visualization, use of search tools, metadata quality flags, and restricted vocabularies for attribute values aid users in using the database.

As mentioned earlier, despite its potential implications, it should be kept in mind that this is an exploratory study, which should form the basis of future research. Participation of spectral databases in spatial data infrastructures is definitively conditional upon specification of metadata standards, as well as methods for translation of spectral metadata between different schemas with minimal information loss.

Nevertheless, metadata is only one aspect of the creation and management of scientific data collections. The main conclusion to be drawn from this study is that GIS technology is able and mature enough to be incorporated in spectral data management systems. Our proposed system, i.e., ASPG, is capable of achieving the tasks of a spectral database, including efficient data storage and transfer, utilization of multiple sub-schemas, metadata quality evaluation, performing error corrections, and providing interfaces for searching through or adding new records. In addition, it stores and represents spectral measurements as geospatial objects, which makes it more convenient to access these data on web-maps or alongside other geodata.

It is worthily to note that many software components of ASGP are Commercial Off-The-Shelf (COTS) products. While allowing for easier setup and maintenance as well as achieving higher performances, the dependence on COTS and resulting cost implications could limit the feasibility of the exact software setup. It should be expressed that ASGP is a distributed system comprising multiple software components, which work separately and mostly communicate with standard-based solutions. This would allow for substitution of system components with open-source solutions, e.g., Geoserver as GIS server, PostgreSQL as DBMS, etc. However, it would typically require higher amounts of “glue” codes to integrate a fully open-source set of components.

Although the following functionality can be restricted for security reasons, the use of multi-tier architecture for ASGP allows for intentionally bypassing the system tiers. As a result, the data is simultaneously available to a variety of users and automated algorithms in form of different services, e.g., database records, ArcSDE service, REST interface, GML, and interactive web-mapping interfaces. The authors conclude that this feature would broaden the usability of the system.

Acknowledgments

The present work was sponsored by Grant No. 636410 dated 26-06-2012 from the Vice President for Research and Technology of Shahid Chamran University, for which the authors are grateful. Thanks are due to the anonymous reviewers, whose detailed and constructive comments throughout the review process led to vast improvements in this work.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.cageo.2013.06.007.

References

Ahn, Y.H., Shanmugam, P., Moon, J.E., Ryu, J.H., 2008. Satellite remote sensing of a low-salinity water plume in the East China Sea. Annales Geophysicae 26, 2019–2035.

Alesheikh, A.A., Helali, H., Behroz, H.A., 2002. Web GIS: technologies and its applications. In: Proceedings of the Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada, 9 pp.

ArcGIS Desktop, 2010. Version 10.0, ESRI, Redlands, CA, USA, ⟨http://resources.arcgis.com/content/arcgisdesktop/10.0/about⟩ (accessed 23.03.12).

ArcGIS Server, 2010. Version 10.0, ESRI, Redlands, CA, USA, ⟨http://resources.arcgis.com/content/arcgisserver/10.0/about⟩ (accessed 23.03.12).

ArcGIS Viewer for Flex, 2011. Version 2.5, ESRI, Redlands, CA, USA, ⟨http://help.arcgis.com/en/webapps/flexviewer/⟩ (accessed 23.03.12).

ArcSDE, 2010. Version 10.0, ESRI, Redlands, CA, USA, ⟨http://www.esri.com/software/arcgis/arcsde/index.html⟩ (accessed 17.04.12).

Asner, G.P., Lobell, D.B., 2000. A biogeophysical approach for automated SWIR unmixing of soils and vegetation. Remote Sensing of Environment 74, 99–112.

Baldridge, A.M., Hook, S.J., Grove, C.I., Rivera, G., 2009. The ASTER spectral library version 2.0. Remote Sensing of Environment 113, 711–715.

Ben-Dor, E., Irons, J.R., Epema, G.F., 1999. Soil reflectance. In: Rencz, A.N. (Ed.), Remote Sensing for the Earth Sciences. John Wiley and Sons, New York, USA, pp. 111–188.

Bojinski, S., Schaepman, M., Schläpfer, D., Itten, K., 2000. SPECCHIO: a web-accessible spectral database for the administration of heterogeneous campaign data. In: Proceedings of the 2nd EARSeL-Workshop on Imaging Spectroscopy, Enschede, 8 pp.

Bojinski, S., Schaepman, M., Schläpfer, D., Itten, K., 2002. SPECCHIO: a Web-accessible database for the administration and storage of heterogeneous spectral data. ISPRS Journal of Photogrammetry and Remote Sensing 57, 204–211.

Bojinski, S., Schaepman, M., Schläpfer, D., Itten, K., 2003. SPECCHIO: a spectrum database for remote sensing applications. Computers & Geosciences 29, 27–38.

Chan, I., Ashdown, L., 2011. Oracle Database Performance Tuning Guide, 11g Release 2 (11.2), Technical Note E16638-06, Oracle, Redwood City, CA, USA, 560 pp.

Clark, R.N., 1999. Chapter 1: Spectroscopy of rocks and minerals, and principles of spectroscopy. In: Rencz, A.N. (Ed.), Manual of Remote Sensing, vol. 3, Remote Sensing for the Earth SciencesJohn Wiley and Sons, New York, USA, pp. 3–58.

Clark, R.N., Swayze, G.A., Gallagher, A., King, T.V.V., Calvin, W.M., 1993. The U.S. Geological Survey, Digital Spectral Library: Version 1: 0.2 to 3.0 μm. U.S. Geological Survey, Open File Report 93-592, 1326 pp.

Clark, R.N., Swayze, G.A., Livo, K.E., Kokaly, R.F., King, J.B., Dalton, J.S., Rockwell, B.W., Hoefen, T., McDougal, R.R., 2002. Surface Reflectance Calibration of Terrestrial Imaging Spectroscopy Data: A Tutorial Using AVIRIS. ⟨http://speclab.cr.usgs.gov/PAPERS.calibration.tutorial⟩ (accessed 22.03.12).

Clark, R.N., Swayze, G.A., Wise, R., Livo, E., Hoefen, T., Kokaly, R., Sutley, S.J., 2007. USGS digital spectral library splib06a: U.S. Geological Survey, Digital Data Series 231, ⟨http://speclab.cr.usgs.gov/spectral.lib06⟩, (accessed 22.03.12).

Croft, H., Kuhn, N.J., Anderson, K., 2012. On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems. Catena 94, 64–74.

Curry, R., Kiddle, C., Simmonds, R., Pastorello, G.Z., 2010. An on-line collaborative data management system. In: Gateway Computing Environments Workshop (GCE), New Orleans, LA, pp. 1–10.

Duval, E., Hodgins, W., Sutton, S., Weibel, S.L., 2002. Metadata Principles and Practicalities. D-Lib Magazine8, 10.1045/april2002-weibel.

ESRI, 2004a. Understanding ArcSDE. ESRI, Redlands, CA, USA, 56 pp. ⟨http://downloads.esri.com/support/documentation/sde_/706Understanding_ArcSDE.pdf⟩, (accessed 22.03.12).

ESRI, 2004b. Versioning. An ESRI Technical Paper, Redlands, CA, USA, 64 pp. URL: ⟨http://downloads2.esri.com/support/whitepapers/ao_/Versioning_2.pdf⟩ (accessed 29.03.12).

ESRI, 2005. ArcGIS Enterprise Security: Delivering Secure Solutions. An ESRI White Paper, Redlands, CA, USA, 47 pp. URL: ⟨http://www.esri.com/library/whitepapers/pdfs/arcgis-security.pdf⟩ (accessed 29.03.12).

ESRI, 2007. An Overview of Distributing Data with Geodatabases. An ESRI Technical Paper, Redlands, CA, USA, 22 pp. URL: ⟨http://support.esri.com/en/knowledgebase/whitepapers/download/fileid/4438⟩ (accessed 29.03.12).

Forestier, G., Inglada, J., Wemmert, C., Gancarski, P., 2009. Mining spectral libraries to study sensors' discrimination ability. In: Michel, U., Civco, D.L. (Eds.). Remote Sensing for Environmental Monitoring, GIS Applications, and Geology IX, Proceedings of the SPIE, vol. 7478, pp. 74782O–74782O-9.

Galitz, W.O., 2007. The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques, 3rd edn John Wiley & Sons, Indianapolis, USA p. 857.

Hodak, W., Kumar, S., 2007. Oracle Database 11g Performance and Scalability, an Oracle White Paper, Redwood city, CA, USA, 16 pp. URL: 〈http://www.oracle.com/technetwork/database/focus-areas/performance/db-perfscale-11gr1-twp-129806.pdf,〉 (accessed 21.03.12).

Hueni, A., Nieke, J., Schopfer, J., Kneubühler, M., Itten, K., 2009. The spectral database SPECCHIO for improved long term usability and data sharing. Computers & Geosciences 35 (3), 557–565.

Huey, P., 2011. Oracle Database 2 day+Security Guide, 11g Release 2 (11.2), Technical Note E10575-07, Oracle, Redwood City, CA, USA, 126 pp.

Jago, R.A., Cutler, M.E.J., Curran, P.J., 1999. Estimating canopy chlorophyll concentration from field and airborne spectra. Remote Sensing of Environment 68, 217–224.

Karpouzli, E., Malthus, T., 2003. The empirical line method for the atmospheric correction of IKONOS imagery. International Journal of Remote Sensing 24, 1143–1150.

Kruse, F.A., 2012. Mapping surface mineralogy using imaging spectrometry. Geomorphology 137, 41–56.

Lanz, A., Brändli, M., Baltensweiler, A., 2007. A large-scale, long-term view on collecting and sharing landscape data. In: Kienast, F., Wildi, O., Gosh, S. (Eds.), A Changing World—Challenges for Landscape Research. Springer, Netherlands, pp. 93–111.

Legleiter, C.J., Roberts, D.A., Lawrence, R.L., 2009. Spectrally based remote sensing of river bathymetry. Earth Surface Processes and Landforms 34, 1039–1059.

Liolios, K., Schriml, L., Hirschman, L., Pagani, I., Nosrat, B., Sterk, P., Field, D., 2012. The Metadata Coverage Index (MCI): a standardized metric for quantifying database metadata richness. Standards in Genomic Sciences 6 (3) (accessed 18.11.12)〈http://standardsingenomics.org/index.php/sigen/article/view/sigs.2645953〉.

Milton, E.J., Schaepman, M.E., Anderson, K., Kneubühler, M., Fox, N., 2009. Progress in field spectroscopy. Remote Sensing of Environment 113 (Suppl. 1), S92–S109.

Mistele, B., Schmidhalter, U., 2008. Estimating the nitrogen nutrition index using spectral canopy reflectance measurements. European Journal of Agronomy 29, 184–190.

Muzafar, A.B., Razeef, M.S., Baader, M., 2011. Cloud computing: a solution to geographical information systems (GIS). International Journal on Computer Science and Engineering 3, 594–600.

Oracle Database, 2011. 11g Release 2, Oracle, Redwood City, CA, USA, 〈http://www.oracle.com/technetwork/database/enterprise-edition/overview/index.html〉 (accessed 04.04.12).

Peters, D., 2009. System Design Strategies, ESRI Technical Paper C11144-12, CA, USA, 252 pp.

Pfitzner, K., Bollhöfer, A., Carr, G., 2006. A standard design for collecting vegetation reference spectra: implementation and implications for data sharing. Spatial Science 52, 79–92.

Rasaiah, B., Jones, S., Bellman, C., Malthus, T., 2011. Building Better hyperspectral datasets: the fundamental role of metadata protocols in hyperspectral field campaigns. In: Proceedings of the Surveying and Spatial Sciences Conference. Wellington, New Zealand, pp. 35–48.

Roberts, D.A., Gardner, M., Church, R., Ustin, S., Scheer, G., Green, R.O., 1998. Mapping Chaparral in the Santa Monica Mountains using multiple end member spectral mixture models. Remote Sensing of Environment 65, 267–279.

Rueda, C.A., Wrona, A.F., 2003. SAMS User Manual Version 2.0. Center for Spatial Technologies and Remote Sensing, University of California, Davis, 40 pp.

Ruby, J.G., Fischer, R.L., 2002. Spectral signatures database for remote sensing applications. In: Proceedings of SPIE, vol. 4816, Seattle, WA, USA, pp. 156–163.

Satterwhite, M.B., Henley, J.P., 1990. Hyperspectral Signatures (400 to 2500 nm) of Vegetation, Minerals, Soils, Rocks, and Cultural Features: Laboratory and Field Measurements, Report ETL-0573, U.S. Army Engineer Topographic Laboratories, Fort Belviour, Virginia, USA.

Sears, R., Van Ingen, C., Gray, J., 2006. To BLOB or Not To BLOB: Large Object Storage in a Database or a Filesystem? Microsoft Research Technical Report MSR-TR-2006-45. Microsoft Corporation, WA, USA.

Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society of America Journal 66, 988–998.

Smith, G.M., Milton, E.J., 1999. The use of the empirical line method to calibrate remotely sensed data to reflectance. International Journal of Remote Sensing 20, 2653–2662.

Thome, K., Czapla-Myers, J., Kuester, M., Anderson, N., 2008. Accuracy assessment for the radiometric calibration of imaging sensors using preflight techniques relying on the sun as a source. In: Proceedings of SPIE, vol. 7081, San Diego, CA, USA, pp. 708118–708118-12.

TPC, 2012. TPC-C—Top Ten Performance Results: Version 5 Results. 〈http://tpc.org/tpcc/results/tpcc_perf_results.asp〉(accessed 23.03.12).

Zeiler, M., 2010. Modeling our World: The ESRI Guide to Geodatabase Concepts, 2nd edn. ESRI Press, Redlands, CA, USA p. 308.

Zi-li, Z., Chun-Feng, J., Di, W., Yong, H., Xiao-li, L., Yong-ni, S., 2011. Discrimination of varieties of rice using near infrared spectral by PCA and MDA model. In: Proceedings of the 6th International Conference on Computer Science & Education (ICCSE), Singapore, pp. 38–40.