Arkansas GIO GeoStor Platform Migration Study

Overall Assessment and Recommendations

Zekiah Technologies, Inc.

6/17/2011

The original document provided by Zekiah Technologies, Inc. was modified by AGIO staff on August 2, 2011. Modifications included the removal of individual industry provider's specific current and future capabilities. This was done to prevent confusion related to current or future capabilities as well as possible misunderstandings to questions that the industry providers graciously answered in the survey.

Contents

i Executive Summary

The Arkansas Geographic Information Office (AGIO) contracted with Zekiah Technologies to assess the feasibility of migrating its GeoStor platform to a cloud-computing environment. The purpose of this assessment activity was to provide AGIO with sufficient information, in terms of functional capability and potential costs, in order to support a decision regarding the potential migration of GeoStor.

The assessment activity consisted of the following activities:

1. A review of the current functional capabilities of GeoStor
2. Preparation of a report documenting the current state of GeoStor, including its users, the services it provides, and its annual costs.
3. Development and administration of an industry survey to allow vendors and service providers to self-assess the capabilities of their platforms against the functional requirements of GeoStor.
4. A physical review of the in-place assets that comprise the GeoStor infrastructure.

The industry survey was designed to characterize existing and emerging platforms across six broad areas:

1. Data Storage
2. Data Management
3. Data Distribution
4. Web Mapping
5. Application Support
6. General Platform Information

Questions on the survey were derived from the current functional capabilities of GeoStor and, to the greatest extent possible, required respondents to characterize their platform's ability to address these capabilities using the following descriptions: Current capability, Capability planned within 6 months, Capability planned within 1 year, Capability not planned. This information is designed to provide AGIO with not only the current state of each platform but also an idea of how each will evolve as AGIO moves through its decision-making process.

A total of eight responses were received from industry. The picture painted by these surveys shows mature capabilities in terms of data storage and management and less mature, but rapidly evolving, capabilities in the other areas. As a result, one of the key technical recommendations is that AGIO wait one year before pursuing any migration. During that time, most of the surveyed platforms will mature greatly and provide AGIO with greater choice. It is also recommended that AGIO continue hosting its raster data holdings in-house until storage costs of hosted platforms drop in the future.

Budget analysis, comparing current annual expenditures with voluntary cost/price data submitted by three respondents indicate significant potential cost savings can be achieved by migrating GeoStor to a hosted environment. The potential savings of between $110,934.00 and $153,934.40 per year lead to the recommendation that AGIO strongly consider moving GeoStor to a hosted environment within the guidelines of the technical recommendations detailed in this assessment.

1.0 Methodology

AGIO retained Zekiah Technologies to assess the feasibility of migrating the GeoStor platform to a cloud-based infrastructure. The assessment activity consisted of the following activities:

1. A review of the current functional capabilities of the GeoStor platform. GeoStor was examined from an end-user perspective to identify the various types of users and the services provided to each type. Supporting documentation from AGIO was reviewed to identify internal functions that may not be accessed by the typical end user.
2. Preparation of a report documenting the current state of GeoStor. This document described the user communities served by GeoStor, the various services provided by GeoStor, the annual expenditures related to GeoStor, and a review of the data and applications managed by GeoStor. This report, titled "Arkansas GIO GeoStor Platform: Current State Overview" was delivered as a separate document and will be referenced throughout the remainder of this assessment.
3. Development of an industry survey. The data collected in the first activity was used to develop a survey intended to collect information about the current state of cloud-based offerings from industry. The survey was constructed to assess the ability of each vendor/provider to meet the functional needs of GeoStor, without focusing on specifics of technology. The survey was conducted online after a general announcement inviting industry to participate in the survey.
4. Physical review of in-place assets of the GeoStor platform. This review consisted of a walk-through of the hosting site in Little Rock, Arkansas and a review of all vector and raster data holdings as well as the development of a high-level topology of the GeoStor platform.

2.0 Assessment

This assessment will be based upon the alignment of services defined in the "Arkansas GIO GeoStor Platform" report. Information from the eight vendor responses will be analyzed against the following service areas:

1. Data Hosting and Infrastructure
2. Web Mapping and Application Hosting
3. Geospatial Data Distribution

2.1 Overview of Survey Structure

The AGIO Platform Vendor Survey consisted of 53 questions divided into seven sections. The survey questions were based upon the current functional capabilities of the GeoStor platform. The survey was intended to allow interested vendors to self-assess their platforms/offerings against these functional capabilities. The majority of the questions asked vendors to assess their capabilities in the following terms:

1. Current capability
2. Capability planned within 6 months
3. Capability planned within 1 year
4. Capability not planned

Certain questions deviated from this format where responses were expected to be open-ended (such as providing a working URL) or multiple-response (such as indicating file formats supported for downloads).

The intent of the survey was two-fold. First, it provided a consistent structure to information from vendors, placed within a context relevant to GeoStor. Second, it will provide a roadmap with which AGIO can work if they choose to proceed with a potentially lengthy procurement process.

The survey sections were:

1. General Platform Information – This section collected company name, platform name, point of contact information and characterization of the vendor's offering against the NIST Definition of Cloud Computing.
2. Data Storage – Basic information about each platform's data storage capability.
3. Data Management – Information related to each platform's ability to support creation, editing and deletion of data sets and well as support for access control. Metadata support was also addressed in this section.
4. Data Distribution – Information about each platform's ability to support direct and indirect data downloads.
5. Web Mapping – Information about each platform's ability to support the hosting of interactive, web-based maps using data stored in the platform.
6. Application Support – Information regarding each platform's ability to support customization, both by AGIO and by GeoStor users.
7. Conclusion – Final information from each vendor.

2.2 Summary of Survey Responses

A vendor survey was produced and released online through SurveyMonkey.com. The survey was available from 13 May 2011 to 19 May 2011 and was open to vendors/providers who had expressed interest in participating. A total of eight (8) responses were received from the following participants:

1. Arc2Earth, LLC (www.arc2earth.com)
2. Environmental Systems Research Institute, Inc. (www.esri.com)
3. GeoIQ, Inc. (www.geoiq.com)
4. Google, Inc. (www.google.com)
5. Latitude Geographics Group, Ltd. (www.latitudegeo.com)
6. OpenGeo (www.opengeo.org)
7. Skygone, Inc. (www.skygoneinc.com)
8. WeoGeo, Inc. (www.weogeo.com)

The detailed survey responses were delivered to AGIO separately and are not included here. This document will use information from those surveys as the basis of this assessment.

2.2.1 Overview of Survey Responses by Section

The following sections provide a more detailed summary of the survey responses within each section.

2.2.1.1 General Platform Information

The primary information of interest in this section was the characterization of each platform/offering in terms of service and deployment models as defined by the NIST Definition of Cloud Computing available at http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf. AGIO had no specific requirement regarding service model but was specifically only interested in the "Public Cloud" deployment model.

All vendors, with the exception of one, indicated that their platforms/offerings supported the public cloud deployment model.

Conclusion: At the time of this writing, sufficient options for public cloud deployments are currently available to AGIO.

2.2.1.2 Data Storage

Section 2 collected information about the ability of each platform to store raster and vector data as well as limits on storage capacity and upload size. All vendors indicated that their platforms stored both raster and vector data. Two vendorsindicated limits on the size of data uploads into the platform. This can be mitigated by sending data to the vendor for loading. One vendor indicated a size limit for raster (5 gigabytes for an individual data set) that may be a limiting factor for AGIO if the decision is made to store raster data.

Conclusion: At the time of this writing, sufficient options for geospatial data storage in the cloud are currently available to AGIO.

2.2.1.3 Data Management

Section 3 collected information regarding the ability of each platform to support data management functions. These functions included the ability to upload, replace, edit or delete raster and vector data sets inside the platform. In the case of editing, information was collected related to the ability to edit geometries and attributes for individual features in a data set. Information was also collected regarding the ability to perform access control by user and group/role as well as the ability to create and edit metadata for each data set stored in the platform.

Two vendors indicated all of the functions in Section 3 as current capabilities of their offerings. Further analysis into the specifics of each vendor's implementation will be required if AGIO decides to move forward.

Two vendors indicated that individual capabilities in Section 3 were not planned. One responded that the ability to manage individual features within a data set was not planned. The other responded that the ability to delete individual raster data sets was not planned. If AGIO decides to move forward with a migration of GeoStor to the cloud, the importance of these capabilities will need to be assessed.

The aggregate responses from all other vendors indicate that each platform will support all of the Section 3 capabilities within one year, except as noted above.

Regarding supported formats for data uploads, there was a wide range of support but Shapefile and KML were almost universally supported for vectors, with the exception of one vendor, which supports vector uploads through its API directly from ArcMap using any format that is supported by ArcGIS. Raster support was more varied but GeoTIFF and JPEG2000 were commonly supported. The majority of the data currently managed by GeoStor in not stored in these formats so data conversion may be a significant consideration in a cloud migration.

Conclusion: At the time of this writing, sufficient options for geospatial data management in the cloud are currently available to AGIO but available options will expand within the next year. By that time, multiple platforms should support all of the capabilities listed in Section 3.

2.2.1.4 Data Distribution

Section 4 collected information regarding the ability of each platform to support the functions that comprise the current indirect download and direct download capabilities of GeoStor.

Within this section, one vendor indicated the strongest support for the capabilities related to indirect downloads. This can be attributed to the fact that they use the same technology, Safe Software FME, as is currently used by GeoStor to perform this function. As a result, they support all of the output formats, both raster and vector, currently supported by GeoStor.

One vendor listed all of the Section 4 functions as current capabilities. Their response indicated that they may leverage one or more other supported platforms to provide full capability. If AGIO proceeds with migrating GeoStor to the cloud, this would need to be investigated more fully.

The primary function related to direct downloads, the ability to download a full data set via hyperlink without authentication, is currently supported by four vendors. The download formats supported by these vendors are not as comprehensive as those currently supported by GeoStor for indirect downloads but are comparable to those currently supported for direct downloads. One vendor indicated comprehensive data format support with the ArcGIS Data Interoperability Extension, which is also built on Safe Software technology. One platform focuses on strong OGC support augmented by the ability to download data and maps in many formats.

Two vendors indicated that they do not plan to support the ability to choose the spatial reference of downloaded data sets. It should be noted that one platform was still under development at the time of

the survey and the vendor's response listed none of the Section 4 functions as current capabilities. Their response indicates most of the Section 4 capabilities will be available within one year.

Conclusion: At the time of this writing, data distribution options, as compared to the current capabilities of GeoStor, are somewhat limited. Options should be greatly enhanced within one year.

2.2.1.5 Web Mapping

Section 5 collected information regarding the ability of each platform to support hosting of interactive web mapping, similar to the current web mapping and application hosting capabilities of GeoStor. This is a key service that GeoStor currently provides for other State agencies and county/municipal governments.

Seven vendors indicated the current ability to host interactive web mapping applications. One indicated this capability within six months. With regard to support for client-side technologies, all vendors, with the exception of one, indicated support for JavaScript. One indicated only support for Java applets. There was varying support for other technologies such as Flash, Silverlight and HTML5.

All vendors indicated the ability to define interactive maps using data hosted within the platform as either a current capability or available within the next six months to one year. Two vendors indicated no plans to support defining interactive maps from ArcMap documents. This is a key workflow factor for data and maps received by AGIO from counties and other State agencies which must be considered if AGIO moves forward. All other vendors indicated this as a current capability or available within six months. The survey sought information about the following mapping capabilities:

- Pan, Zoom, Identify
- Basemap switching
- Toggling layer visibility
- Displaying map legend
- Address lookup/geocoding
- Spatial bookmarks
- Overview map
- Map markup
- Map printing

Many vendors indicated support for all of these functions. The rest offered varying degrees of support. In any case, this support should be interpreted as "out of the box." It should be possible to extend the interfaces to support all capabilities via application development.

Support for the OGC "WxS" family of specifications varied greatly. In general, WMS is currently the most widely supported. Current support for WFS and WCS was indicated only by one vendor. Seven vendors indicate plans to support each of these specifications within the coming year. One has no plans to support WCS. Much broader current support was indicated for KML, with only two vendors not indicating current support. They indicated support within six months to one year.

Not surprisingly, there was limited current support for the GeoServices REST Specification with only three vendors indicating current support, but all other vendors indicated pending support within six months to one year.

The most strongly supported commercial mapping platforms were Google Maps, Bing Maps and ArcGIS.com. Since the survey was issued, Yahoo! has announced that its product will be discontinued.

Conclusion: At the time of this writing, support for web mapping is strong across all of the vendors. Currently, this seems to take the form of native mapping tools and APIs but interoperability options should increase over the next six months to one year. If AGIO decides to move forward with migrating GeoStor to the cloud, it should find sufficient support for its mapping requirements. During that time, AGIO should the evolution of support for interfacing with ArcMap.

2.2.1.6 Application Support

Section 6 collected information regarding APIs and other methods of customization supported by the platforms. Cloud platforms must be viewed in the same way other "out-of-the-box" solutions are viewed. Since it is highly likely that not all of GeoStor's requirements will be met at the outset, the ability to customize and extend a platform is a key consideration.

With the exception of one, all vendors indicated the current availability of open (supported, documented and free-of-charge) APIs for their platforms. That vendor indicated that their platform would have open APIs available within six months. There was almost universal support for RESTful APIs on the "server" and JavaScript on the client. Support for specific languages such as .Net, Ruby, or Python was more varied but the widespread support of REST mitigates this greatly. All of these languages can be used to leverage RESTful APIs so specific language support is less of an issue.

Support for various open transport protocols such as Atom, Atompub, OpenSearch, GeoRSS, GeoJSON and OGC CWS varied widely. These protocols are not currently part of GeoStor so this is less of a consideration.

Of greater importance was the ability to support 1) geocoding and 2) the use of a custom geocoder. This is a critical function within the current GeoStor platform. With the exception of two, all vendors indicated current or pending support for geocoding. Those two vendors indicated no plans to support custom geocoders, whereas all other vendors indicated current or pending support.

Conclusion: At the time of this writing, sufficient options for customization of cloud platforms are currently available to AGIO, although support for custom geocoders will bear further analysis moving forward.

2.3 Technical Conclusions and Recommendations

Based on the observations detailed above, it can be concluded that it is currently possible for AGIO to migrate GeoStor to a cloud environment. If AGIO were consider undertaking such a migration at present, a multi-vendor/platform approach could address all current GeoStor capabilities, emulating the current GeoStor separation of capabilities as described in Arkansas GIO GeoStor Platform Current State

Overview. This approach, however, may limit efficiencies to be gained by the move. For example, such an approach may require rewriting the current GeoStor search interface in order to provide a seamless user experience. In a multi-platform approach, the application would need to be written to interface with multiple APIs. In addition, data loading and management, and especially access control, would be cumbersome and potentially redundant.

Analysis of the survey results makes clear that all of the responding vendors are rapidly expanding and enhancing their platforms. The responses indicate that a majority of GeoStor's current capabilities will be available on most platforms within one year. Development schedules can slip and features can be reprioritized so the actual state of each platform must be monitored, but it is clear that much additional capability will be available to AGIO by the first quarter of 2012.

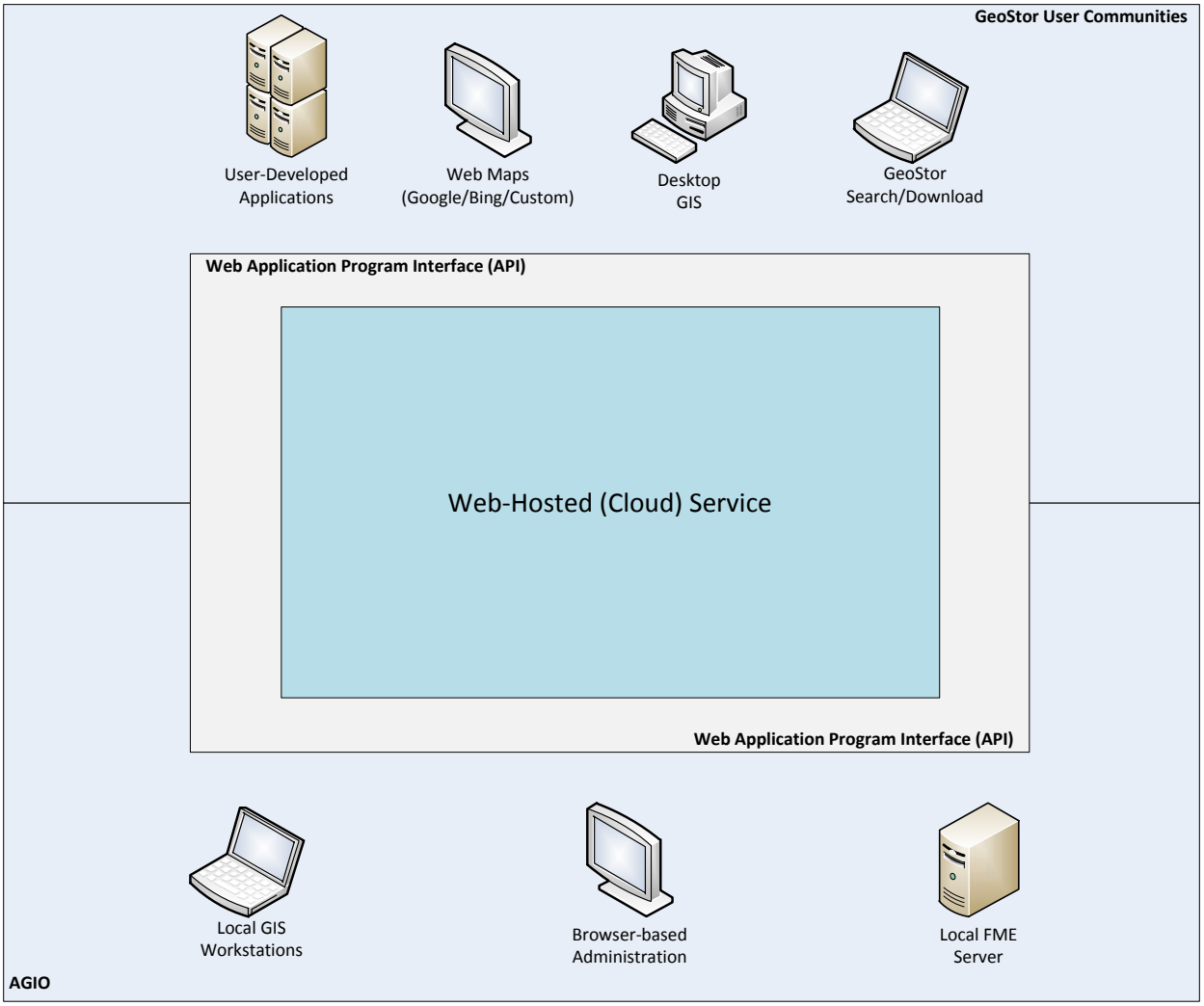

Figure 1: Notional Post-Migration View of GeoStor

Based upon these observations, the following recommendations are made:

1. It is recommended that AGIO wait one year before undertaking a migration of GeoStor to the cloud. This recommendation is based primarily on the current state of data distribution capabilities. Data distribution is a key service of GeoStor and support for the capabilities currently possessed by GeoStor varied widely. In general, all of the required capabilities should be available within the next six months to one year. Given the critical nature of data distribution to GeoStor, it would be prudent for AGIO to allow the market to mature in this area. Additionally, only four of the eight survey respondents indicated support for custom geocoders, a key requirement of AGIO. Within one year, six of the platforms will support this requirement.
2. It is recommended that AGIO review the data distribution formats currently supported by GeoStor. Common vector formats such as shapefile, KML are widely supported but most others were less so. Raster format support was particularly poor, except in terms of "web formats" such as GIF, JPEG and PNG. If AGIO drops support for less-requested formats, a transition a cloud platform will be smoother.
3. AGIO should initially only host vector data in the cloud. This recommendation is based upon two key factors:
    a. Data storage costs – budgeting considerations are discussed later in this document but data storage is one of the key price drivers. AGIO currently stores approximately 7 terabytes of raster data. This amount of storage will greatly inflate the cost of a cloud solution.
    b. Limited raster support – As mentioned in item 2, raster support, especially for data distribution is limited. This may change as platforms evolve but this factor, combined with data storage costs, suggests continuing with local storage and FTP.
4. AGIO should maintain at least one FME Server locally. As cited in the "Current State" document, GeoStor currently supports some data processing activities for other State agencies. An example of this is the processing of Streamline sales tax information. Support for this type of activity is very limited across the platforms surveyed. Those platforms that do support such custom functions generally do so at an additional fee. Additionally some of these processing activities are dependent upon non-AGIO data that would not reside in the hosted environment. It is recommended that AGIO update its scripts to pull data from the hosted environment as needed and perform such functions locally.

Figure 1 above depicts a notional view of GeoStor based upon these recommendations.

2.4 Budget Analysis

This effort was not an official request for proposal so submission of pricing/cost information was strictly voluntary. The estimates in this section are based upon an analysis of all information received and are intended to support future budget planning.

The approach to pricing cloud architectures can vary greatly across vendors. In general, however, data storage and throughput are common factors in the pricing models received as part of the survey.

2.4.1 Analysis of Potential Impacts to Current Expenditures

The following table is taken from the document titled “Arkansas GIO GeoStor Platform: Current State Overview” and itemizes the current annual expenditures related to the GeoStor platform. An additional column has been added to show estimated impact to current costs as a result of a potential cloud migration. This table does not include potential annual costs of cloud hosting, which is discussed later in this document.

Item	Current	Estimated
Software	**$ 64,102.40**	**$ 5,168.00**
Hardware		
Servers/Storage (5 year spread)	$ 50,000.00	$ 4,000.00
Consumables/Replacements	$ 25,000.00	$ 2,000.00
Hardware Subtotal	**$ 75,000.00**	**$ 6,000.00**
Support	**$ 10,000.00**	**$ 0.00**
Hosting	**$ 36,000.00**	**$ 22,000.00**
Labor		
AGIO staff	$ 75,000.00	$ 75,000.00
Labor subtotal	**$ 75,000.00**	**$ 75,000.00**
GeoStor Total Annual Expenditures	***$ 260,102.40***	***$ 106,168.00***
Difference		***$ ($153,934.40)***

The estimated costs are based upon the following assumptions:

1. All AGIO staff currently in place will remain in place.
2. Hosting costs from Arkansas DOIT will be reduced by half due to smaller hardware footprint.
3. Retention of one server to support storage/FTP of imagery data. AGIO currently refreshes hardware approximately every five years. This estimate assumes a hardware refresh of the FTP server spread over a five-year period.
4. Migration of FTP server to a Linux operating system to eliminate current Microsoft license.
5. Retention of one server to support local instance of FME Server. AGIO currently refreshes hardware approximately every five years. This estimate assumes a hardware refresh of the FME server spread over a five-year period.
6. Migration of FME server to a Linux operating system to eliminate current Microsoft license.
7. Retention of current Symantec BackupExec License to continue backup of raster data and other retained systems.

Based upon the elimination of current GIS and database software and associated hardware and operating system licenses, the migration of GeoStor to a cloud-hosted geospatial platform will reduce current outlays by an estimated $153,934.40 per year. **This cost difference is the maximum amount that can be applied to cloud hosting fees in order to remain at current expenditure levels.**

2.4.2 Analysis of Voluntary Vendor-Provided Cost Information

Voluntary cost/pricing information was received from three vendors. Based upon the voluntary cost data provided by these vendors, current projections indicate that a migration of the GeoStor platform could potentially achieve annual cost savings of between $110,934.00 and $153,934.40. These projections are based upon current pricing data received from 3 participating vendors. As the various platforms surveyed mature over coming year, pricing may change for existing ones and be established for those that are still under development so continued monitoring is warranted. Based upon the findings using current data, it is recommended that AGIO strongly consider migrating GeoStor to a hosted cloud platform, within the guidelines of the technical recommendations above.

2.5 Key Considerations for Potential Migration

The cost analysis above considers only the end state of a migration of GeoStor to a cloud computing environment. AGIO will also need to consider costs and level-of-effort for performing the migration itself. The following would be key factors in the migration effort:

1. Data conversion – Many platforms have limited support for input formats, raising the potential need to convert data sets from the formats in which they currently reside in GeoStor to another format, such as the shapefile. If AGIO elects not to migrate raster data, then conversion requirements are greatly reduced. Vector data conversion may potentially introduce schema changes that would need to be assessed. With sufficient up-front planning, the impact of data conversion would be reduced. Much of the labor involved with data conversion could be mitigated by scripting, automated jobs run during off-hours and the use of existing FME software.
2. Data upload – Once data is in an acceptable format, data upload may be time-consuming. This could be easily mitigated by scripting and much of the upload could be performed during off-hours. With sufficient up-front planning, the impact of data upload would be reduced.
3. Application re-hosting – The GeoStor search interface, geocoding service and interactive mapping applications will need to be re-factored to work against the APIs of any selected platform. This aspect of migration will potentially be the most labor-intensive of all the migration activities.

Good project planning and execution will mitigate the impact of these considerations, especially the first two. With good use of automated tools, labor hours associated with conversion and upload will be minimized.

Application re-hosting should focus on technologies that are widely supported across the responding platforms. If AGIO follows the technical recommendations above, then staff members who support GeoStor will need to be proficient with JavaScript (for client-side customization and maintenance) and an appropriate server-side technology such as Python or PHP.