eTaxNG – Next Generation e-Services for Tax

October 2012
Ulaanbaatar, Mongolia

This publication was produced for review by the United States Agency for International Development. The views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

Project:	Mongolia Business Plus Initiative Project (BPI)
Report Title:	***eTaxNG – Next Generation e-Services for Tax***
Main Author:	Jerry P. Henzel
Contract No.	438-C-11-00001
Submitted by:	BPI Project/Chemonics International Inc., Express Tower, 12th Floor, Chingeltei District, Ulaanbaatar, Mongolia
Telephone and fax:	(976-11) 32 13 75 Fax: (976-11) 32 78 25
Contact:	Efrain Laureano, Chief of Party
E-mail address:	elaureano@bpi-chemonics.biz

ABBREVIATIONS AND ACRONYMS

BCP	Business Continuity Plan
BPI	Business Plus Initiative
COTS	Commercial of the shelf software
GDT	Government of Mongolia General Department of Taxation
GoM	Government of Mongolia
LUN	Logical Unit Number
MOSA	Mongolian Software Association
OSB	Oracle Service Bus
RAC	Oracle Real Application Cluster
SAN	Storage area network
SBR	Standardized business reporting
XBRL	Extensible Business Reporting Language
XML	Extensible Markup Language
XSD	XML Schema
USAID	United States Agency for International Development

TABLE OF CONTENT

ABBREVIATIONS AND ACRONYMS i
TABLE OF CONTENT i
EXECUTIVE SUMMARY i
OVERVIEW 1
SECTION I: E-GOVERNMENT & E-FILING 3
SECTION II: CURRENT STATE OF GDT ICT & DATA EXCHANGE 5
SECTION III: SERVICE BUS: A STRATEGIC ASSET 7
SECTION IV: DATA SEMANTICS 9
SECTION V: DATA INTERACTION PATTERNS 15
SECTION VI: WEB SERVICES CANDIDATES 21
SECTION VII: TECHNICAL ARCHITECTURE 25
SECTION VIII: SECURITY 33
SECTION IX: MOSA INVOLVEMENT 37
SECTION X: PROJECT MANAGEMENT STRATEGY 39
SECTION XI: PROJECT RISKS 43
SECTION XII: PROJECTED PROJECT TIMELINES 45
SECTION XIII: FURTHER STUDY 47
SECTION XIV: MISCELLANEOUS 49
CONCLUSION 51
ANNEX A: PROJECT MANAGEMENT REQUIREMENTS 55
ANNEX B: WEB SERVICES PROGRAMMER REQUIREMENTS 59
ANNEX C: QUALITY ASSURANCE CONSULTANT REQUIREMENTS 63
ANNEX D: TECHNICAL TEAM LEAD CONSULTANT REQUIREMENTS 67
ANNEX E: GDT WEB SERVICE CANDIDATE NOTES 71

EXECUTIVE SUMMARY

An assessment was performed in order to determine how to implement the next generation of e-services in Mongolia; specifically to enable the submission of tax returns via electronic means. This is in contrast with simply submitting tax returns via the current GDT website.

The current architecture and data flows at the GDT were examined and members of the Mongolian Software Association were also engaged in discussions. Based on those discussions, a mechanism for describing the semantics of the data was developed. Then an examination of the patterns of data interaction was undertaken and strategies developed to deal with the variety of interactions that needed to take place. Finally a logical and physical architecture was proposed based on the findings.

In combination, a project was developed that would permit the development of a system to perform electronic filing of tax returns to the GDT by 1 March of 2013. This strategy was also placed in a larger context so that any investment in time and infrastructure would be reusable in this broader context. In effect, it would allow the GDT to commit to a multi-year project and incrementally build a scalable, secure and performant system that could handle businesses and individual taxpayers in the future. Infrastructure and human resources were outlined, as were projected project timelines.

The project itself focuses on working with a small group of MOSA software members and GDT to implement the infrastructure needed for GDT to accept electronic returns. Both the GDT and MOSA would have to implement significant additional functionality. The timelines are very aggressive and put the project in a very high-risk category. Other risk factors also contribute to this risk categorization.

OVERVIEW

The General Department of Taxation (GDT) for the Government of Mongolia (GoM) has requested assistance via the USAID Business Plus (BPI) Initiative program to implement the next generation of e-services in Mongolia; specifically to enable the submission of tax returns via electronic means. This is in contrast to being able to enter and submit tax return data via a website, i.e. online. The former is a step toward full e-Service that will reduce the amount of paperwork and help increase data quality and ultimately boost Mongolia's ranking on the World Bank Doing Business Index.

In order to achieve these goals, a number of different issues need to be resolved.

First, there has to be an understanding of the data semantics; a common understanding of what the data looks like and means as it traverses from party to party. Second, there needs to be common understanding of the types and patterns of interactions as the data moves between parties. These patterns and interactions are largely defined and driven by functionality needed at each end of the communication tunnel. Third, the technical underpinnings that enable the data to move need to be understood and in place. This can be thought of as the technologies that make data orchestration possible.

The ICT situation at the GDT was examined closely with these issues in mind in order to determine how data exchange could be facilitated between MOSA accounting software and the GDT to enable full and complete e-Filing. Then the semantics of the data were defined, patterns of interaction identified and the technological underpinning specified. Finally, an action plan was put together showing how this type of project may be completed.

SECTION I: E-GOVERNMENT & E-FILING

Initial discussions with relevant parties revealed a wide variety of definitions, disagreement and misconceptions on the concept of e-Filing. Compounding the problem was a previous report by the Business Plus Initiative that used the term 'true e-Filing' which was interpreted negatively and gave vagueness of the term e-Filing itself. The GDT currently provides the ability to file a tax return via a website, i.e. online, and some view this as e-Filing.

It is easy to see how this confusion can arise given what is in the literature on e-Services and eGovernment. What GDT is currently doing falls into the realm of e-Services by many definitions present in the literature. The literature treats this concept broadly. J. Rowley, writing in An Analysis of the E-Service Literature: Towards a Research Agenda (2006) notes,

> "E-service is deeds, efforts or performances whose delivery is mediated by information technology (including the Web, information kiosks and mobile devices). Such e-service includes the service element of e-tailing, customer support and service, and service delivery"

Other authors extend this definition to mean an online service available via the Internet where a valid transaction is possible, as opposed to an online presence where a transaction is not possible. Gustav Aagesen of the Norwegian University of Science and Technology extends the definition to include an aspect of self-service. It is easy to see why the confusion arose. Those who say the GDT is doing e-Filing are not entirely incorrect.

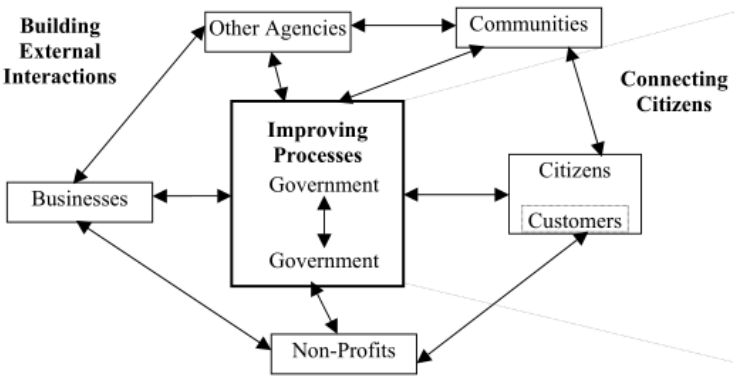

Figure 1: Domains of eGovernment, S. Nitro, 2000 eGovernment in Eastern Africa, KPMG

The Internal Revenue Service in the United States allows its citizens to file tax returns via online fillable forms or via brand-name software. It uses the term e-Filing liberally. An important nuance we learn from the IRS example is the ability to submit tax returns via brand-name software. This was a result of the Office of Management and Budget's 2001 Quicksilver Task Force that instructed the IRS to work and find a solution to improve government to business electronic capabilities. One of the practical outputs of this was the Free File Alliance, a consortium of 15 commercial tax preparation providers who offer brand-name tax software for free to enable electronic filing for lower income Americans. The significance of something like FreeFile is that the boundary of provision of service goes beyond the web browser. It is another and additional mechanism for interfacing with government services. This seemingly small distinction has great significance. It opens more than one type of interface to government services. While FreeFile is targeted at the consumer, the premise behind it is applicable to all types of taxpayers and classes of tax software. It means that the government provides multiple interfaces to its services; a business with IT capacity may be able to generate and submit a return without have to re-type that information. This is something that has been missing in Mongolia to date. It is something the current version of e-Filing does not do.

Perhaps a better way to define and understand what is to be accomplished here is to put it into the broader rubric of eGovernment. Nitro's report for KPMG (Figure 1) lays out the basic domains of eGovernment. These are: 1) improving government processes, 2) connecting citizens and 3) building external interactions. Improving processes includes cutting process costs, managing process performance, making strategic connections in government and creating empowerment. Some refer to this as e-Administration. Connecting citizens involves talking to citizens, listening to citizens and improving public services. Building external interactions or e-Society means working better with business, developing communities to build social and economic capacities, and building partnerships to achieve social and economic goals. What the GDT currently does is help improve public services by allowing the public and business to enter tax online vs. hardcopy. It can be considered to be "doing eGovernment".

However, the GDT has the possibility to more fully embrace the concepts of eGovernment and improve citizen and business services. It could significantly improve the interaction between government and business. It could take its online services and further improve them to make submitting taxes more convenient and less costly, i.e. by not re-keying data. It could make major strides in making strategic connections across government agencies and data stores that could lead to even better citizen and business services. This we could term as the next generation of the current eTax service the GDT offers. It is a quantum leap from a simple online electronic presence that more fully embraces all aspects of eGovernment. Hence, internally, this next step is already being called eTaxNG or eTax Next Generation as it represents going beyond a simple online presence and fully embracing eGovernment.

SECTION II: CURRENT STATE OF GDT ICT & DATA EXCHANGE

In order to fully understand the data and technology landscape, an assessment of the current structure, data flows and current systems was undertaken at the GDT. Figure 3 illustrates the current situation. Even in the limited time available for an assessment, a large number of projects can be noted. These include an Oracle Service Bus project that takes in data/sends out data from various other sources. A business intelligence project that takes data from an intermediary database server that pushes data to a data warehouse that then is used to provide the mechanism to generate reports. The Tax Administration Information System (TAIS) for actual processing of tax return data. A project called TRIPS funded by the World Bank that is to replace TAIS eventually. However, that project has been in existence since 2006 and still is not operational. Questions exist as to its long-term viability at the time of this writing. Further, there are at least two identifiable virtualization efforts: an older virtualization

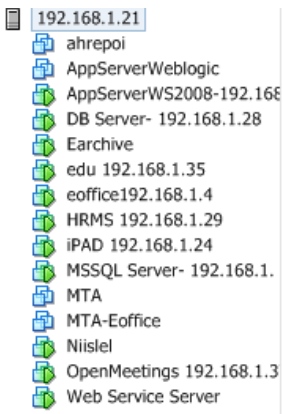

Figure 2: Virtualization Guests

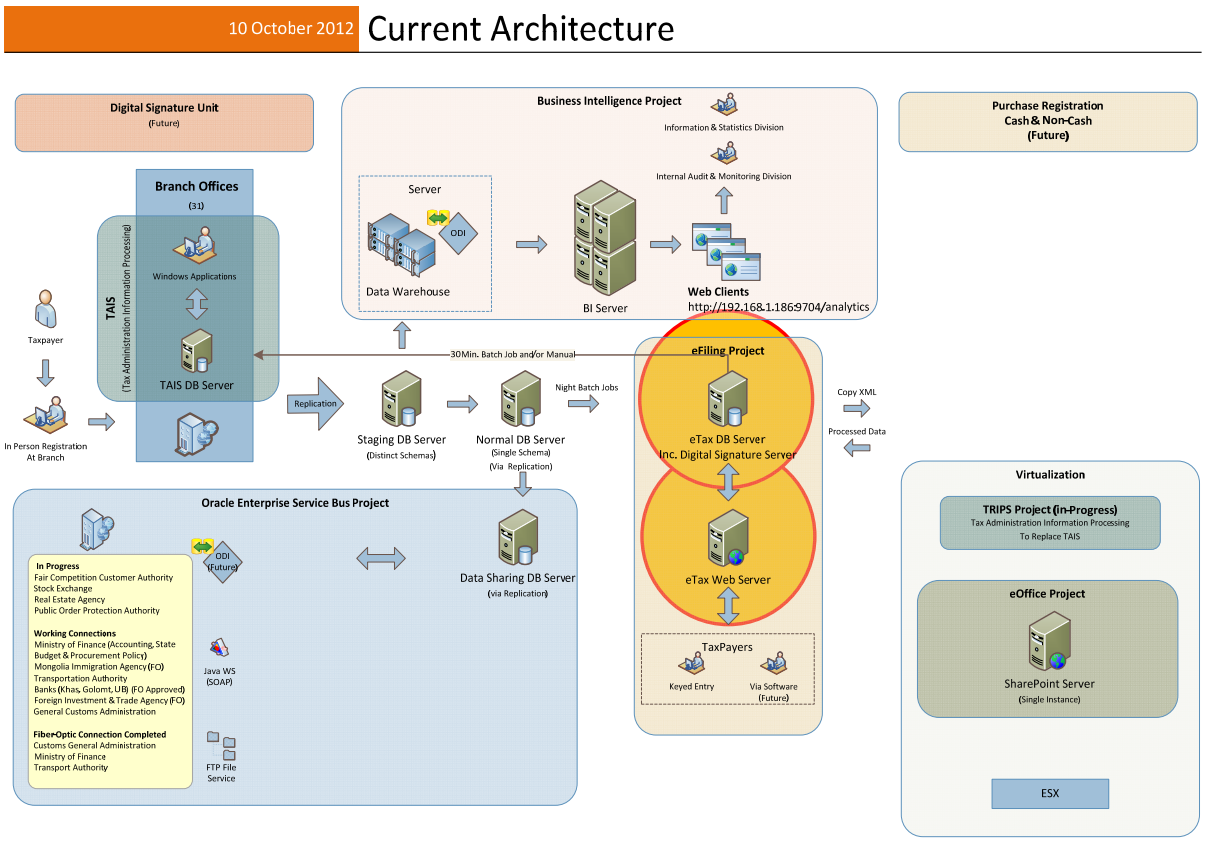

Figure 3: Current Architecture & Data Flows

environment and a newer one. Figure 1 shows the contents of the older initial virtualization project. This virtualization effort uses storage via another server using the OpenFiler Linux distribution. A second virtualization effort has been started and is larger and more sophisticated, including a storage area network or SAN. The latter is running

on 2 physical hosts both having 32 GB of RAM. This system was delivered in August of 2012 and is currently under test. The first host currently has guests that include, Microsoft Exchange Server, Active Directory, Lynx Communication server and several others. The other host is currently unallocated.

Within the current IT structures at GDT, how data moves varies widely. In some cases such as tax information there is tight database-to-database integration. While tightly coupled integration is often undesirable because of its brittleness, in this case it is an acceptable and reasonable solution. For example, data moves from tax branches and districts to a staging server, is then coalesced on a second server before being pushed onto the eTax database server. In other instances, data may be moved through the use of Oracle Database Integrator, web services, files, etc.

What is clear from the assessment done to date is that there are a wide variety of technologies and systems deployed at the GDT. Many of these are single machine deployments. Many do not have any redundancy or high availability. There is not much in the way of economy of scale in the GDT IT infrastructure. Whether this has been a conscious strategic decision by the GDT or has been exacerbated by various donor projects pushing their own agendas is beyond the scope of this work. The net effect is the creation of a variety of 'data islands' or silos that make IT management more challenging and costly.

Also of note is that there is need to do a detailed business analysis of the various data flows and potentially streamline the various data flows. While this is outside the scope of this particular study, it would be a very worthwhile exercise and have many potential benefits for GDT.

SECTION III: SERVICE BUS: A STRATEGIC ASSET

Within the current GDT infrastructure, there is one item that has architectural and strategic importance. That is the Oracle Service Bus (OSB) that has been implemented as part of the current data sharing initiative.

A service bus, architecturally speaking, is an abstraction that is designed for interaction and communication between distinct parties. It is particularly useful in situations where systems should be or are loosely coupled, where asynchronous communication is important, where guaranteed message delivery is vital and many systems need to communicate and share data. It is loosely analogous to an electrical bus that is common in residential electrical wiring setups. Power comes in from one source (the power company) and all the different circuits in a house tap off the bus to draw power. There is not a one-to-one connection between the power company's source and each individual circuit. The service bus, like its electrical counterpart, is a way to avoid one-to-one connections, manage, control, secure and monitor systems and traffic as they share data. One-to-one connections may not seem problematic when the number of connections is few but management quickly escalates out of control exponentially when the number of systems communicating rises.

Figure 3 shows a list of data connections that GDT currently has or is in progress of establishing. A survey of MOSA members indicates there are up to 2164 clients that will potentially communicate with the GDT.

eTaxNG will bring a significant load to the GDT infrastructure that will need to be enabled, monitored and managed by virtue of incorporating MOSA software. Adding additionally to the load are large taxpayers, who have their own proprietary systems, legacy systems, commercial off the shelf software (COTS) like Microsoft Dynamics, etc. This class of taxpayer is resistant to change or unable to change easily and the types of interfaces they may demand may be extremely varied. There needs to be some type of boundary or edge layer in the architecture that deals with this complexity.

It is entirely possible to write web services to perform the role of the service bus. However, that will take time and effort. A service bus offers much functionality 'out of the box' and favors configuration over code writing. This make for faster and easier adaptation to changes and enables business agility. It is also meant to work in a distributed architecture and designed for horizontal scalability. Further, by acting as an intermediary layer it gives some protection to internal systems and vice-versa. For example, external changes in data format can be handled at the service bus layer and reduce the need for making changes to internal systems. The service bus with configuration changes can handle changes in transport mechanism and coded business logic does not have to change and be re-tested. Given the GDT's existing use of this system, it is advisable to leverage this skill and architectural pattern for eTaxNG. While initially eTaxNG may have its own service bus, strategically, the GDT should merge the existing project and eTaxNG into one enterprise wide service bus over time.

SECTION IV: DATA SEMANTICS

What is clear is that even in the initial phases there will be many different classes and types of clients potentially submitting data to the GDT electronically. This underlines the need for some type of standardized data format for the sharing of data. The common way to do this is to define a taxonomy. Ernst & Young (http://www.ey.com/GL/en/Services/Assurance/-Accounting-and-Financial-Reporting/Assurance-XBRL-What-are-Taxonomies) define taxonomies as "'vocabularies' or 'dictionaries' created by a group in order to exchange information". Currently, there is no taxonomy for data sharing at the GDT.

There are potentially many different ways data can be packaged and shared between systems. It can be in the form of ASCII text, CSV, Excel spreadsheets, Adobe Acrobat PDF files, etc. Another common format is XML or extensible markup language. XML is a text-based format that is human-readable, platform independent, based on international standards, supports UNICODE and allows for information to be structured and self-describing. The latter is perhaps one of its greatest strengths. It is also verbose, which can be a disadvantage for transmission over a network, but the verbosity permits parsers to more easily catch errors, e.g. each data element must have a start and end tag.

A variety of formats were examined to determine the optimal for the GDT tax taxonomy. Many of these data formats were brittle and small changes would cause parsers to break easily. This is common in CSV or ASCII submitted data for example. Others were proprietary or difficult to extract data from, e.g. PDF. Within the international community there has been increasing momentum toward using an XML-based format known as XBRL or Extensible Business Reporting Language. The XBRL.US (http://xbrl.us/Learn/Documents/FactSheet.pdf) factsheet notes that XBRL is

> "A technology standard, based on Extensible Markup Language (XML), that puts metadata around business information to give both creators and users of that data more context about the data, e.g., labels, definitions, references. XBRL makes information computer-readable which makes it portable, and easier to extract and analyze. In doing so, the use of XBRL enables greater accuracy, reliability and timeliness for users including regulators, investors, and analysts."

The April 2010 XBRL Adoption Survey (http://www.xblr.org) notes that over 30 countries are using or have adopted XBRL as a standard. The Securities and Exchange Commission and the Federal Deposit Insurance Corporation in the US mandate XBRL usage, in the UK, HM Revenue and Customs mandates XBRL usage from April 2011, the Ministry of Corporate Affairs is driving XBRL usage in India, etc. The International Accounting Standards Committee (IASC) Foundation through its standard-setting body, the International Accounting Standards Board (IASB) has developed an International Financial Reporting System (IFRS) XBRL-based

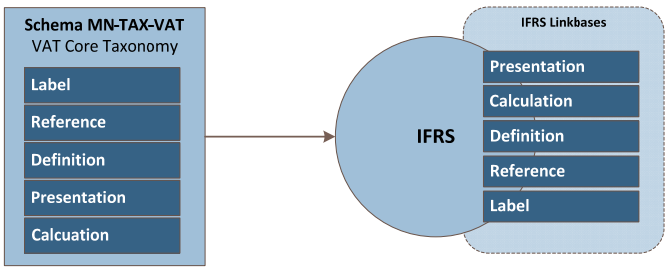

Figure 4: XBRL VAT Schema - IFRS Derivation

taxonomy. IFRS is used in over 120 countries worldwide for standards-based financial reporting. Mongolian accounting regulations mandate its (IFRS) use for financial reporting.

An assessment was made whether it was feasible to use XBRL in Mongolia for e-Filing. The VAT, corporate and personal withholding taxes were examined. Initially a sample XBRL schema (.xsd) was made in XBRL format derived from the IFRS taxonomy (Figure 4). The idea was to check the feasibility of doing this and to see how many elements could be carried over from the IFRS taxonomy to a GDT extended taxonomy. The results showed that for VAT not many elements were carried over as the GDT has many custom items in their VAT forms. However, the link between IFRS and the GDT taxonomy should be maintained for the time being. If it remains the same after completing modeling, the link could be dropped with minimal changes. Sample data was also generated (Figure 5) along with the schema as a test case. Additionally, some basic reports were made to test out where an XBRL-based taxonomy could be used in an XBRL-enabled reporting tool. The results were positive and, for example, languages (EN-MN & MN-EN) could be switched in the reporting tool by toggling an option in the reporting tool. The resulting files are included with this report as attachments.

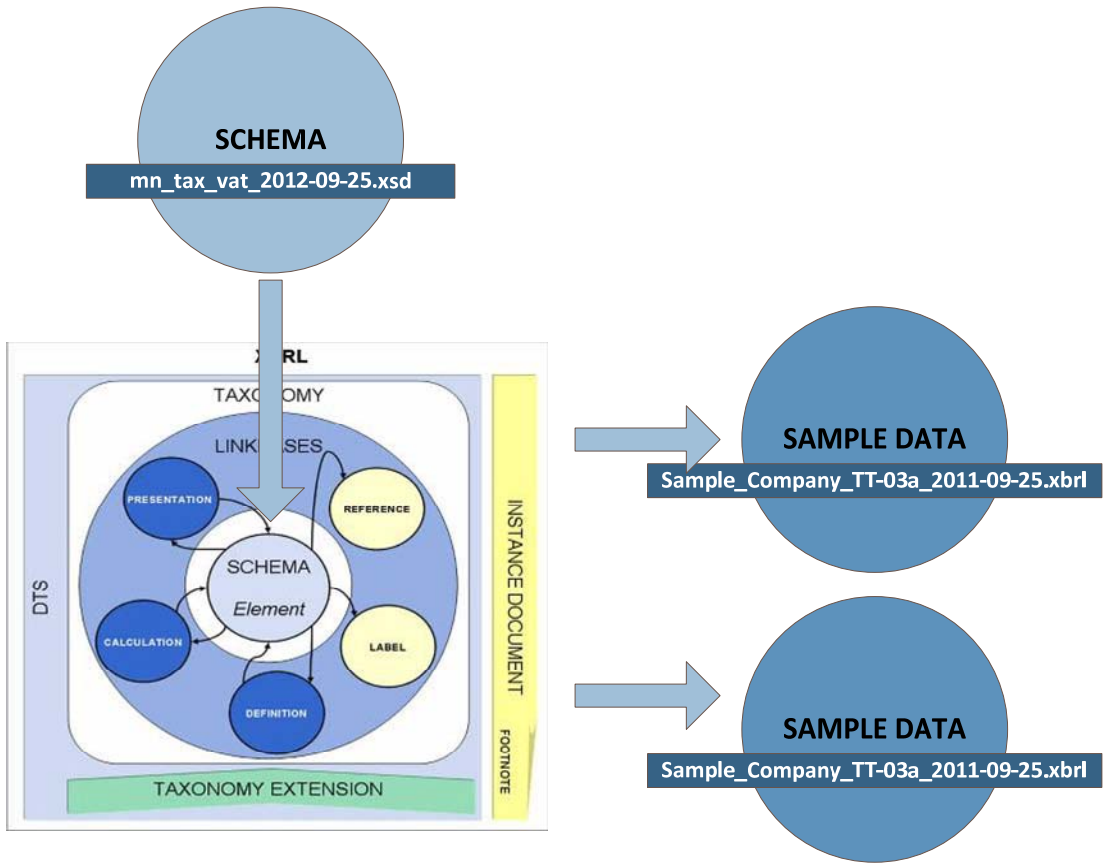

Figure 5: XBRL Schema and Instances

A potential issue with XBRL is the size of the specification. XBRL is made up of a number of constituent parts (Figure 6): the schema, an instance document and various linkbases. The schema is the core definition of the data. This can be thought of as a data dictionary for the data elements that need to be defined and present in the data transmitted. The instance

document is XBRL-speak for a separate file that contains the actual data. A third-party exports out their data into an instance document and then validates that data against the data dictionary or schema. If that validation process passes, the exported data meets basic requirements and can be accepted for further processing. The linkbases are additional files defined in the specification for a variety of purposes. There are 5 separate linkbases: label, presentation, reference, calculation and definition. The label linkbase allows for various labels

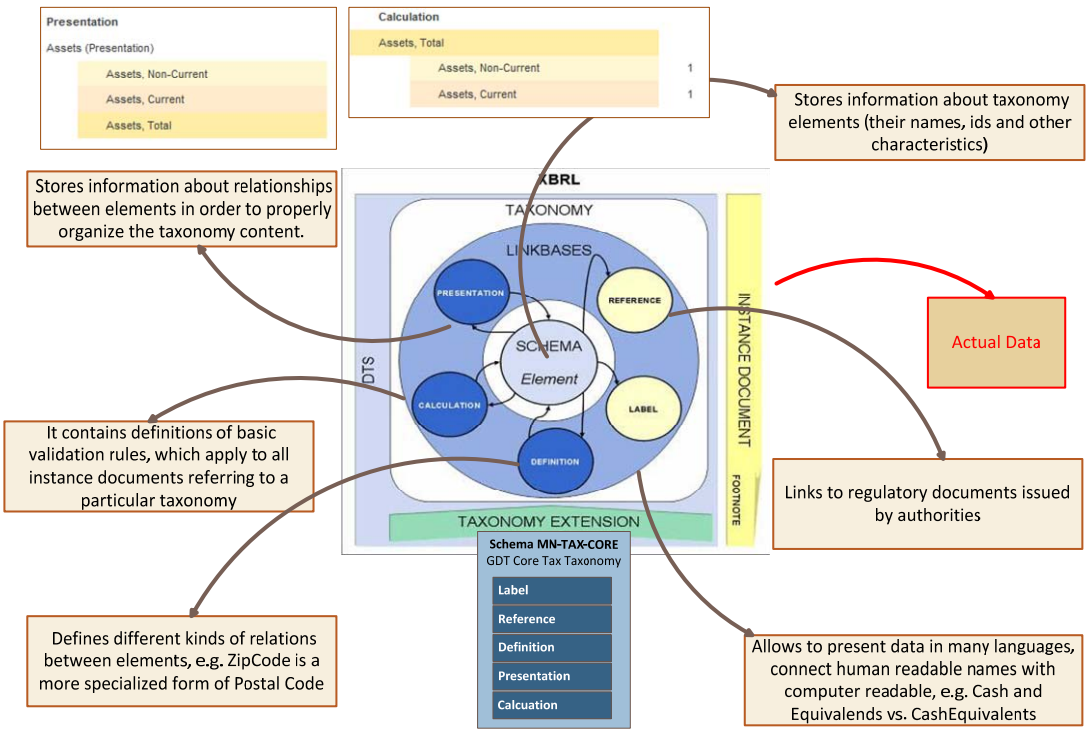

Figure 6: XBRL Specification

to be attached to a base data element. This could be a label in another language, in the language of accounting (vs. a programmer-defined term), etc. This is what permits XBRL-enabled reporting tools to switch languages by simple toggling a setting for a report. The presentation linkbase stores information about the relationships between elements in order to properly organize the information on a report, the calculations linkbase contains definitions of basic validation rules and the definitions linkbase defines different kinds of relationships between elements.

What is apparent very quickly is that the XBRL specification has a lot of power and flexibility. In this initial phase, enabling all this power and flexibility would complicate the project and potentially extend timelines. Most of this additional flexibility will not be used and the GDT does not currently use any XBRL-enabled tools internally. Therefore, it is recommended that the GDT focus solely on the schema and defining the elements that constitute the schema. This is the core minimum requirement to have a functional system. Later, if the GDT decides to move forward, their schemas could be enhanced to include these additional capabilities.

The above discussion raises an important issue: versioning of schemas. Without doubt the GDT will have to change schemas in the future for a potentially wide variety of reasons and be able to distinguish between versions. One of the best and simplest examples of how this is

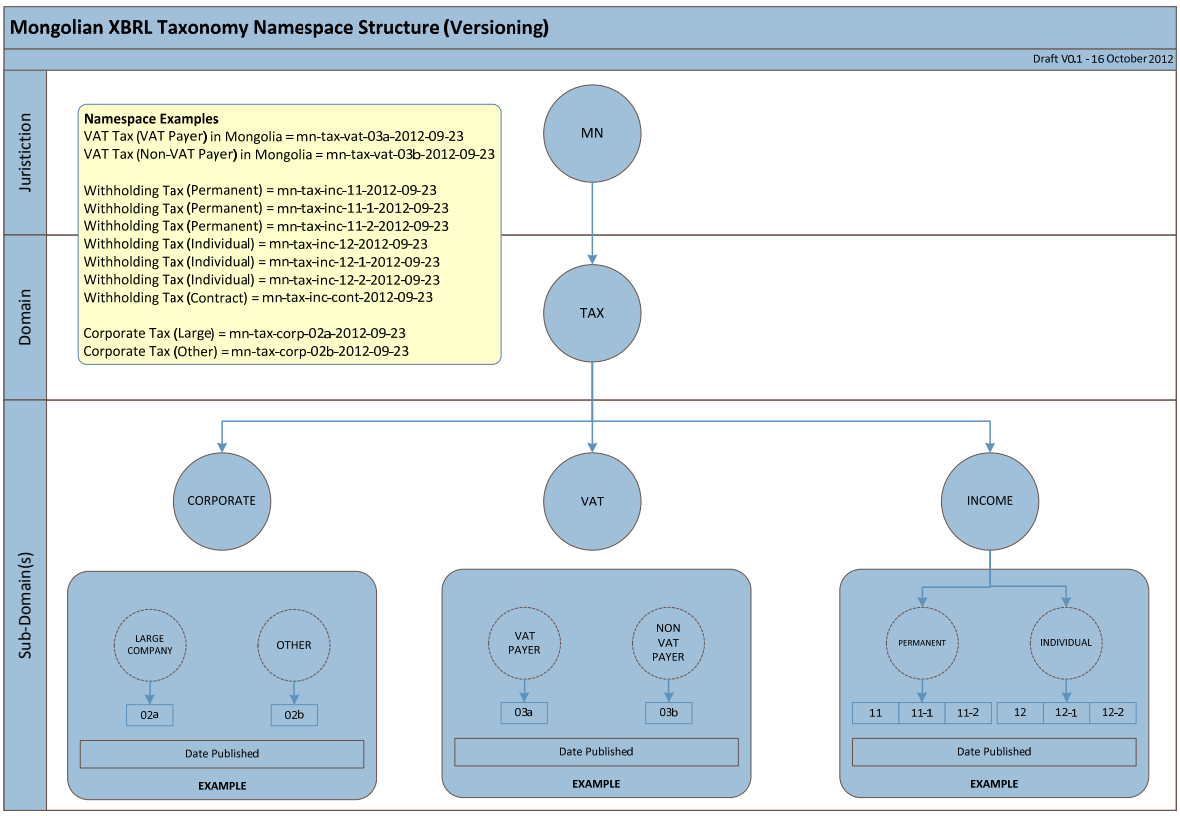

Figure 7: Namespacing

accomplished is in the XBRL adoption by Belgian tax authorities. They laid out a simple namespacing specification that prevents collisions between different versions of their schema across time (http://minfin.fgov.be/portail2/downloads/e-services/biztax/2012-04-30-be-tax-architectureguide.pdf). A sample of how this might be done at the GDT was also prepared (Figure 7). Ideally, this falls within the domain of a broader eGovernment initiative, which in Mongolia falls to the eGovernment Reform Team.

The sample that was prepared for this work is very basic and should only be considered as Version 0.1. It should not be viewed as a final definitive work by any means. Additional modeling needs to take place for corporate and personal (withholding) taxes, and for the application of full and proper technical modeling techniques. Figure 8 shows how a final schema might look. Normally, appropriate modeling techniques would involve looking at the entire tax domain, identifying common elements and breaking them out into a separate schema, then isolating elements common to VAT, corporate tax, etc. to their own schemas. Further, common lookup values, e.g. country codes, ISO codes, etc. could be isolated into their own schemas. All these schemas would be linked together but the modularity would permit greater flexibility and allow changes to be made to smaller portions so as not to potentially disrupt other parts of the taxonomy.

One of the sub-goals of this technical assistance was to identify a number of projects that could be done quickly in order to operationalize e-Filing. This is one of the areas that would

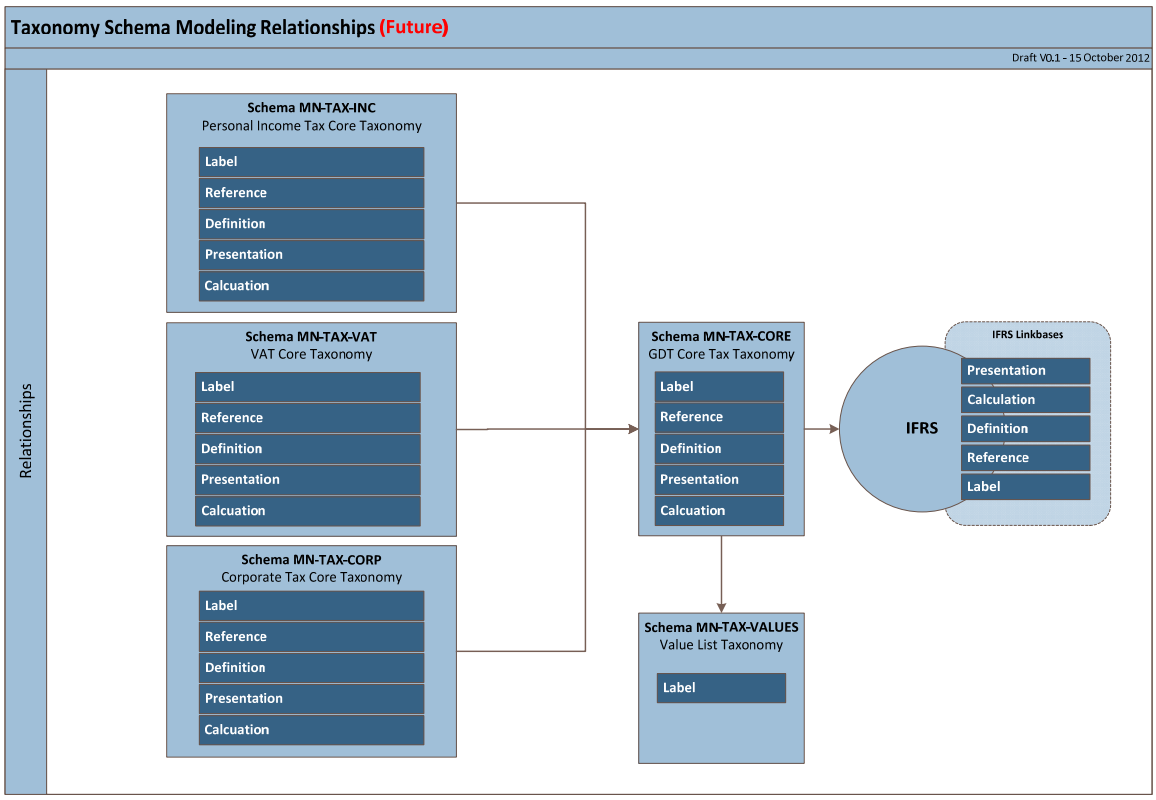

Figure 8: Additional Modeling

contribute substantially to getting e-Filing up and running. There remains a significant amount of work that needs to be completed in the area of data modeling as noted above. Figure 9 notes that in addition to working on the individual tax areas, there will be a need for

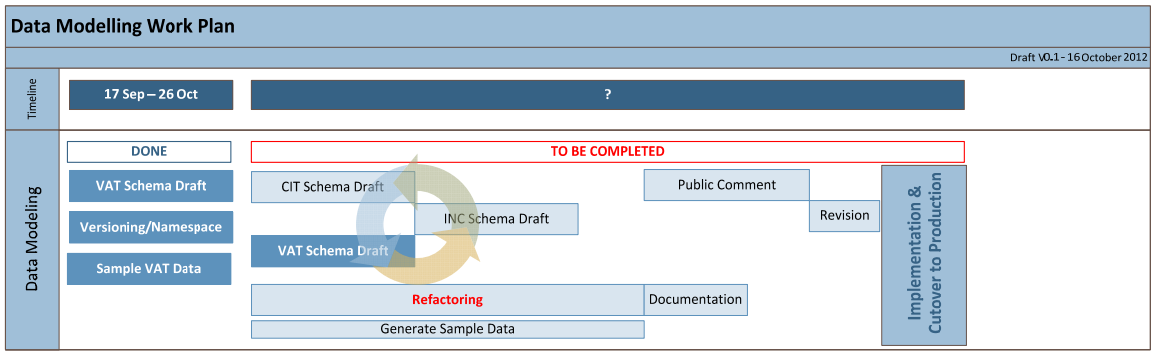

Figure 9: Data Modeling Work To Be Completed

refactoring (modifications) in each of the areas as changes occur and new elements are added in order to fit into data modeling best practices as noted above. Further, there should be a

period of public comment allowed and a further period of revisions based on those comments before the taxonomy is finalized. The timeline is aggressively tight. Deloitte, when they worked with the development of the Belgian XBRL implementation took 4 months to develop the taxonomy (http://www.deloitte.com/assets/Dcom-Ireland/Local%20Assets/Documents/-Consulting/ie_Con_XBRL_quals_Apr10.pdf, Page 2), not including any public comment period. A timeframe of 4 months is beyond the desired timeline here so all timeframes need to be compressed significantly, including the one for public comment. Obviously this elevates project risk. A more detailed discussion of timelines will be held later on.

SECTION V: DATA INTERACTION PATTERNS

Once the semantics of the data are known, attention must be given to how the data moves between parties. At this level, we are still not concerned with the technology that moves the data; that will come later. We are concerned with the different types and patterns of

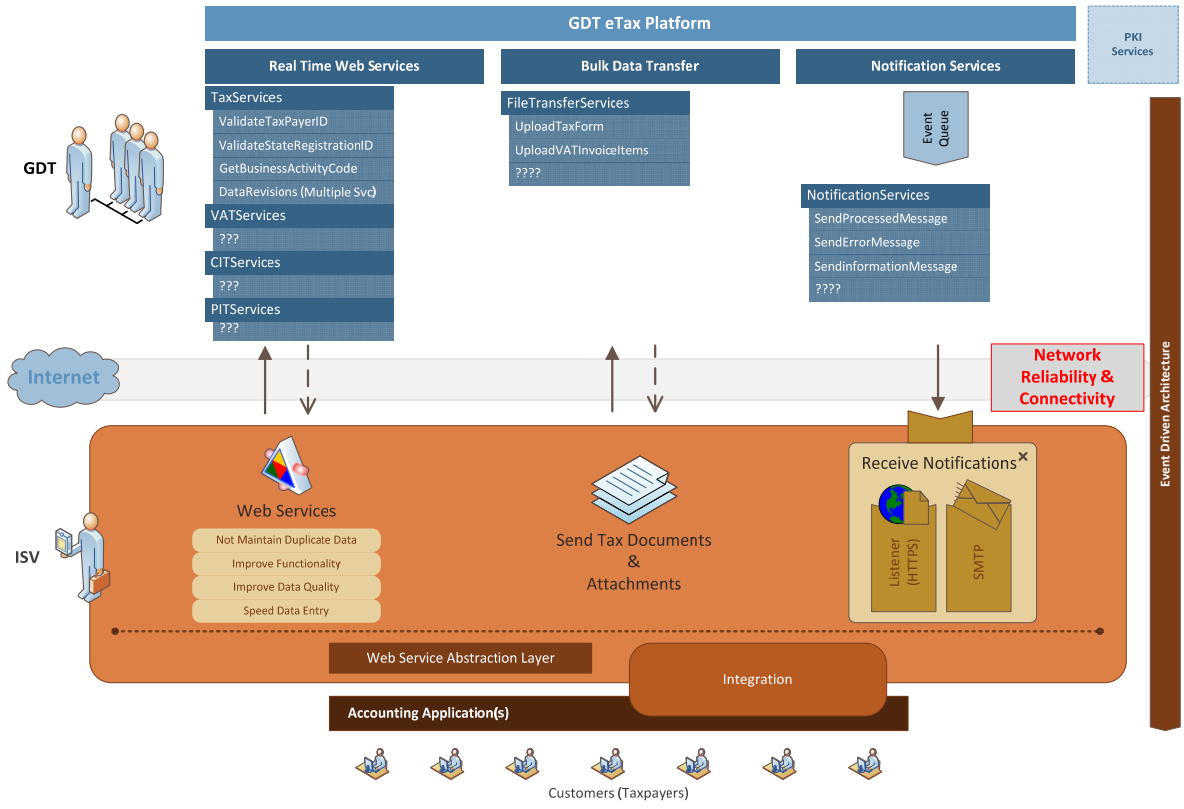

Figure 9: Data Interaction Patterns

interactions in order to move data in a manner that is consistent with what users expect.

Before a detailed discussion, it is essential to reiterate a basic core principle. The first is that of a "single source of truth" or the idea that for any data source there is one authoritative owner. By subscribing to this principle, governments greatly reduce data duplication and potentially increase data quality. This is a fundamental core of a solid eGovernment implementation. This was one of the first cores of eGovernment in Georgia. The Civil Registry Agency, the agency responsible for passport, birth, death & marriage data, began publishing web services (to authorized parties) that provided the ability to look up citizen information in approximately 2008. When Armenia implemented their business registry last year, the first capacity they built in was real-time data feeds from other agencies so as not to store duplicate data. In Mongolia, the GDT is one of a number of authoritative sources of data. There is no reason that vendors of accounting software should have to store duplicate data or not be able to look up data from the GDT in real-time. This would potentially have very positive benefits to e-Filers. They could present their business registration code and have their officially listed name and address sent back and populated in their interfaces, for

example. It would save time, improve data quality and make the general user experience much richer.

As organizations begin their experimentation with moving data from system to system they take a variety of approaches. Many of these tend to be tightly coupled, very technology dependent and often brittle, i.e. not very resistant to change. This is characteristically referred to as tight coupling. Writing in Enterprise Application Integration, Gregor Hohpe & Bernie Woolf state that a solution to tight coupling is loose-coupling. They write:

> "Mechanisms such as a common data format, queuing channels, and transformers help turn a tightly coupled solution into a loosely coupled solution. The sender no longer has to depend on the receiver's internal data format not its location. It does not even have to pay attention to whether the other computer is ready to accept requests or not. Removing these dependencies between the systems makes the overall solution more tolerant to change".

Ideally, systems need to be de-coupled. The architectural implication is that the systems put in place to make e-Filing happen (that will make up the eTaxNG system) need to be loosely coupled so that they minimize the amount of intimate knowledge each has about the other.

If one examines Hohpe's and Woolf's work on enterprise integration, they list common method of data integration:

- File transfer
- Database sharing
- Remote procedure invocation
- Messaging

File transfer is just what it sounds like, transferring files from system to system. Database sharing is the sharing of a common database across applications. With remote procedure invocation, "one application exposes some of its functionality so that it can be accessed remotely by other applications as a remote procedure. The communication occurs real-time and synchronously" (Gregor & Woolf). With messaging, Gregor and Woolf describe it as "one application publishes a message to a common message channel. Other applications can read the message from the channel at a later time. The applications must agree on a channel as well as the format of the message. The communication is asynchronous." Further, there is a variant that the GDT uses directly as previously noted, database replication. Depending on how replication is done, it could be in real-time or closely approximate file transfer. All of these methods are perfectly valid and have their uses. File sharing is simple but is not immediate and depends on precise formats but tends to be reliable and asynchronous. Database sharing reduces duplicate data but may not work well in highly distributed situations. Remote procedure invocation permits access to an application's functionality but again can be very complex to implement across boundaries. Messaging is more immediate, better encapsulated than database sharing, more reliable than remote procedure invocation, is asynchronous, is platform and language independent, etc. However, it can be complex to implement, have performance challenges and sometimes may lock you into a vendor-specific solution.

One more recent attempt to solve the challenges of messaging is the use of web services. The term itself is often misused or abused. Sometimes it is used analogously to represent a business process. Sometimes it is viewed as a technology itself. Web services, as defined by the World Wide Web Consortium (W3C) are "a software system designed to support interoperable machine-to-machine interaction over a network". They really are a set of

technologies and standards working together to implement data movement. They also attempt to be technology and platform independent. An informal survey of MOSA members indicates that they work with a wide variety of programing languages (C#, Delphi, PHP), platforms (Windows, Web) and technologies (desktop, cloud) in general. The GDT is a very heavy Oracle-based organization using the Java programming language. Abstracting these differences will be very important.

One of the 'lessons learned' from the Georgia experience has been that there needs to be capability for different types of data interaction. The general experience seems to be to start with real-time web services. Real-time web services are data feeds that can be called in real-time across a network and return back some type of information immediately. These are very useful in many situations. However, there often needs to be a movement of data that is not real-time. Further, there needs to be movement of data that involves transfer of a large amount of data. It might be possible to get this same type of data using a real-time web service but it tends to be slow and inefficient. Polling a web service with 5000 calls to pull back records and build a recordset is not a feasible technical strategy. In Georgia, for example, the Ministry of Health needed a list of all deceased persons from the Civil Registry every month in order to purge those persons from the benefits roll. There needed to be a way to make one call to the Civil Registry and return back this data in bulk. Commercially, eBay faces this same type of issue with large merchants. A typical use case would be a merchant that has separate catalogs for clothing, hardware, housewares, and electronics with 150 separate types of items in the clothing catalog, such as shirts, dresses, and hats. Most of those items have 24 SKUs each: gold, blue, brown, and green in x-small, small, medium, large, 1X, and 2X. (Each color/size/type combination will have a different SKU number.) Therefore, the merchant will add approximately 3,600 listings, for just the clothing catalog. Each listing will have multiple items of the same size, color, and type; for instance, 10 medium-size, gold shirts. Expanded out, a merchant may have between 14,000 and 15,000 listings to add to the eBay site, and each of these listings is a multi-quantity listing. Adding this individually to eBay would be complex and inefficient. The solution to this is a class of data interaction termed bulk-data services; Web services that enable the movements of large amounts of data at once with one call.

Finally, a key concern noted by MOSA members and already evident at the GDT are performance bottlenecks on the current eTax system. MOSA has a legitimate concern about scalability. Fundamentally, systems can scale in a vertical or horizontal fashion. Vertical scaling is simply adding more resources to a machine, e.g. CPU, memory, etc. Horizontal scaling involves partitioning the workload across multiple machines. Properly done, horizontal scaling can handle very large workloads. Horizontal scaling is one of the keys behind open-source systems like Hadoop that is meant to be used to data mine petabytes worth of information. Oracle real application database clusters (RAC) use the same principle to permit users to scale databases to high user loads. How does this relate to data patterns and data movement? As data moves in and out of GDT from taxpayer systems, the GDT needs to be able to have the capacity to scale horizontally at any point. This needs to happen for real-time data services. This needs to happen for bulk-data services. This is relatively simple to achieve architecturally. But to complete the picture, end users submitting data need to be notified of receipt of their tax returns, of processing, processing status, etc. This is why there needs to be another type of helper web service that works along side real-time and bulk-data services. These are notification services. As the GDT ingests data, bulk-data for example, and processes it across a number of different machines, a separate cluster of machines can push out notifications to those users. They can immediately inform the end user of receipt. Once the processing is done they can inform the user of status. These services also enable another

important architectural principle: that of event-driven architecture. Event-driven architecture relies on systems pushing out data and messages rather than having others ask for data. This is particularly helpful where systems must notify a large number of constituents. Using a publish-subscribe model is much more efficient than having all those systems continually poll for the change. A set of notification services would also allow the GDT to be proactive and push out messages to taxpayers, i.e. notifications of pending deadlines, changes, etc. Overall this has the effect of de-coupling parts of the system, letting them operate in an asynchronous fashion, letting them scale as needed and laying the groundwork for a full event-driven architecture.

Figure 9 lays out in a high level fashion what this type of system may look like for the GDT.

There are 3 different classes of web services: real-time, bulk and notification. Real-time services are used for immediate interaction, much like you would currently see on the web

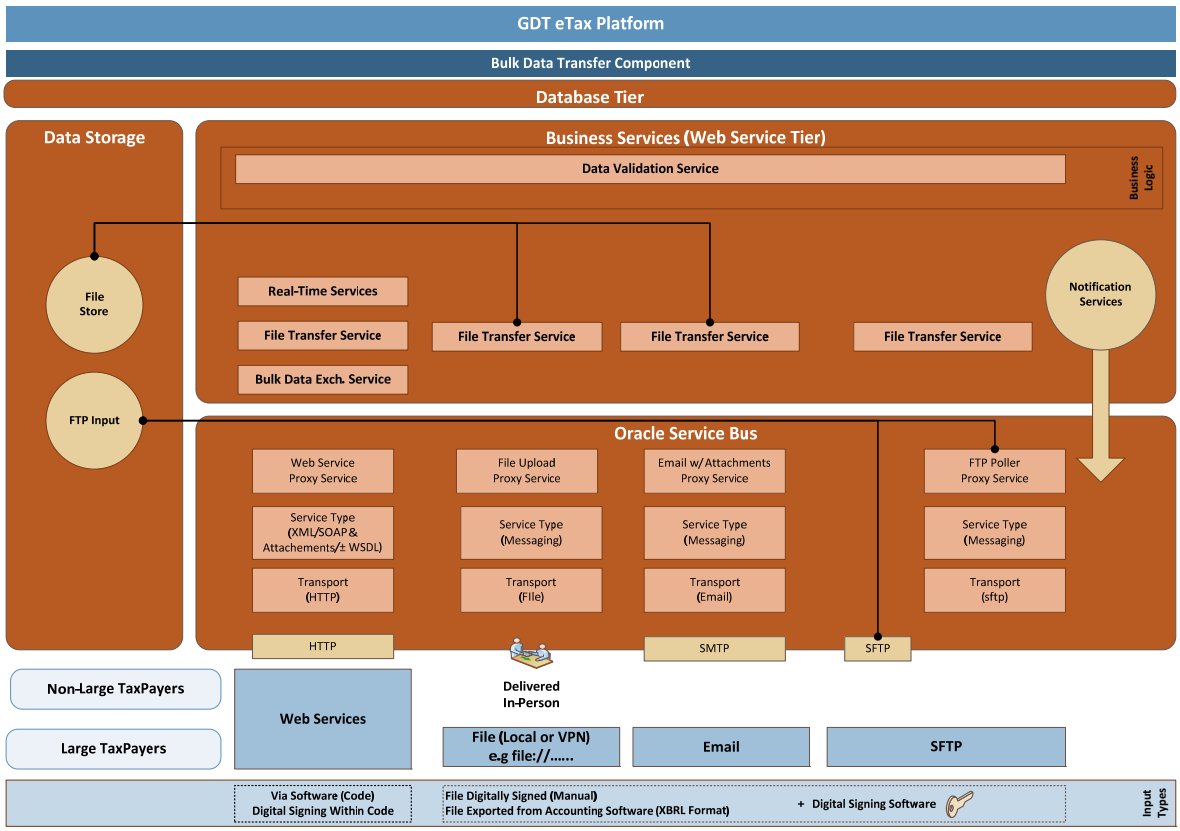

Figure 10: Bulk Data Services

right now. These could be incorporated into existing systems to enhance usability, data quality and accuracy and speed data entry. Bulk data services would permit the movement of large amounts of data to the GDT. For example, a VAT submission and all the accompanying invoices. Anecdotal evidence notes that some large taxpayers would have to submit 4000 – 5000 invoices per return. Finally, notification services would be used to keep users informed of successful receipt of submission, processing status, completion, etc.

Figure 9 shows bulk-data services as a ubiquitous category. In reality, the situation is more complex and there are a variety of ways bulk-data services need to work (Figure 10). This is exacerbated by the large-taxpayers that the GDT must service. These tend to have large

commercial software applications that are customized to their needs or legacy systems or custom in-house systems. Some of their systems may not be able to take full advantage of the web service interfaces under consideration. There needs to be some alternatives that still align with the general architectural philosophy. As Figure 10 shows, any taxpayer software can integrate with the bulk-data web services. These would necessitate integration at the code level of that software that may not always be feasible for some. Alternatively, submissions could be made securely via email or via secure FTP; or even dropped in at GDT in person. These alternative methods would provide e-Filers with a variety of methods to submit tax returns and still would permit the GDT to use a unified architecture at deeper levels.

With these alternative methods of submission, the issue of digital signatures would need to be addressed. Where web services are integrated into accounting software, the digital signing process would also be integrated (more later in the security discussion). However, with the other bulk-data submission methods, a tax return would still need to be digitally signed so the taxpayer could not repudiate it. In Armenia, the private body EKENG (http://www.ekeng.am) is responsible for implement digital signatures for the Government. It is the only company authorized to issue digital signatures in Armenia. They have implemented a commercial system from ARX (http://www.arx.com). The solution from ARX includes the capacity to sign a variety of different document types (Microsoft Word, Adobe Acrobat, etc.) using downloadable client software that is integrated with the broader solution. As can be seen from the following manual (http://www.ekeng.am/wp-content/uploads/2010/05/doc_eSign_-usrmanual_v3.1.pdf), client software can be downloaded and used to sign documents. Something very similar would need to be built and distributed to e-Filers opting for either email or SFTP type submissions. Returns could then be signed and submitted via non-web service methods.

There is one further benefit for GDT of enabling real-time web services. The functionality that is programmed by into those services means that those services can be used to enhance the existing GDT online website. For example, one real-time web service might be to return back a company name and address given a taxpayer ID number. Once developed, it would be relatively simple to add code to the existing eTax website and enrich its functionality.

SECTION VI: WEB SERVICES CANDIDATES

With the assistance of 2 national consultants, the GDT has put together a list of web service candidates. These are potential web services that might be implemented. Original notes can be found in Annex E: GDT Web Service Candidates. These have been refactored from a technical perspective to make them more 'programmer friendly'. The original candidates have been generalized more to make the various web services more flexible and robust. For example, instead of bulk data services having to deal with the details of various types of data that is being transferred, they have been generalized to act as a transport mechanism regardless of the data type that is being moved. The definition of the various data types that are being transmitted should be defined in the data model assembled for the project. This maintains a strong separation of concerns (modularity) and flexibility. Figures 11-13 outline the various web service candidates based on type. There is significant refactoring that still needs to happen before these become final. These candidates should **not** be viewed as final but as a starting point. The development team technical lead will need to revise these before beginning development. The data modeler should also consider these in his/her work.

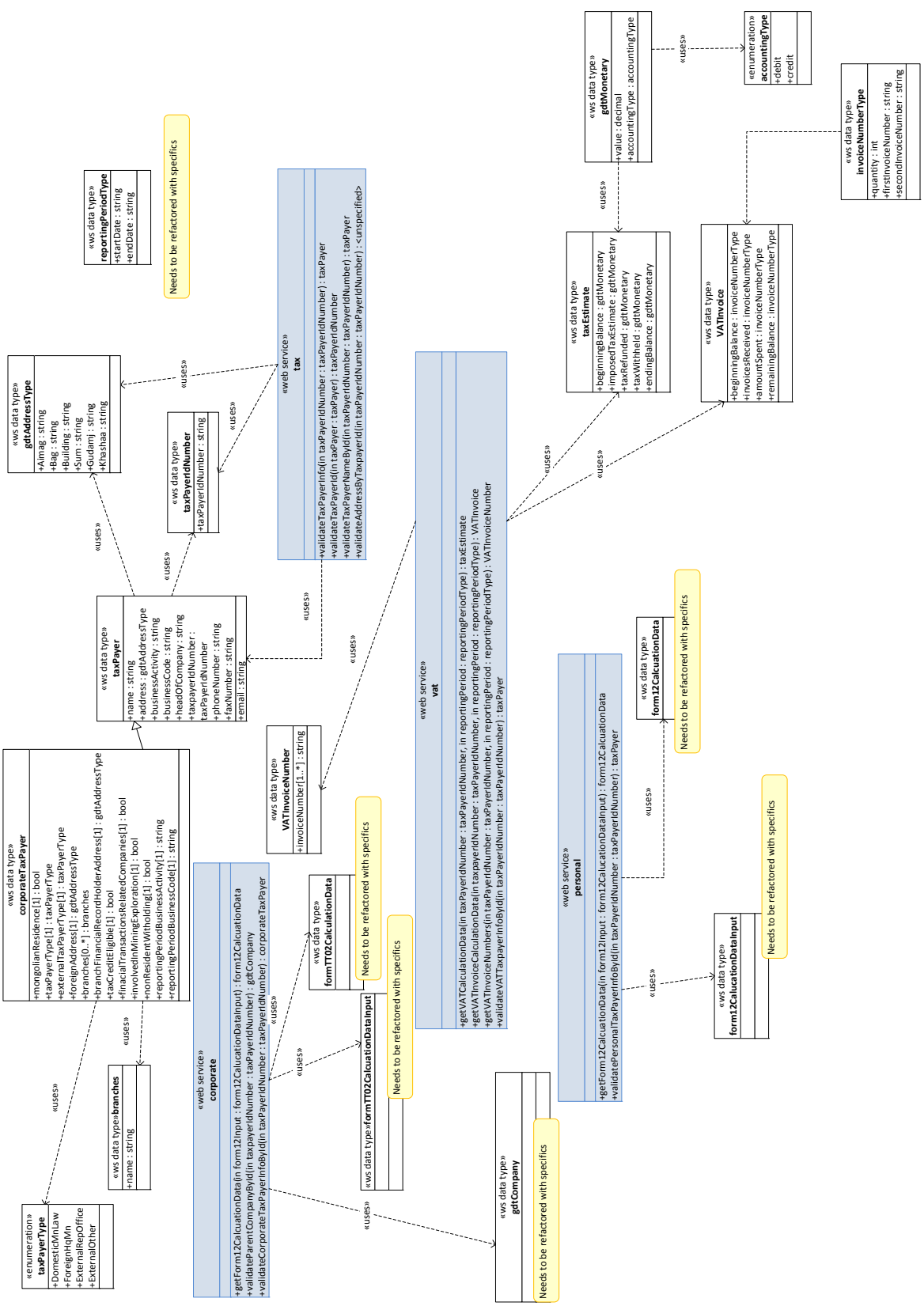

Figure 11: Real-time Web Service Candidates

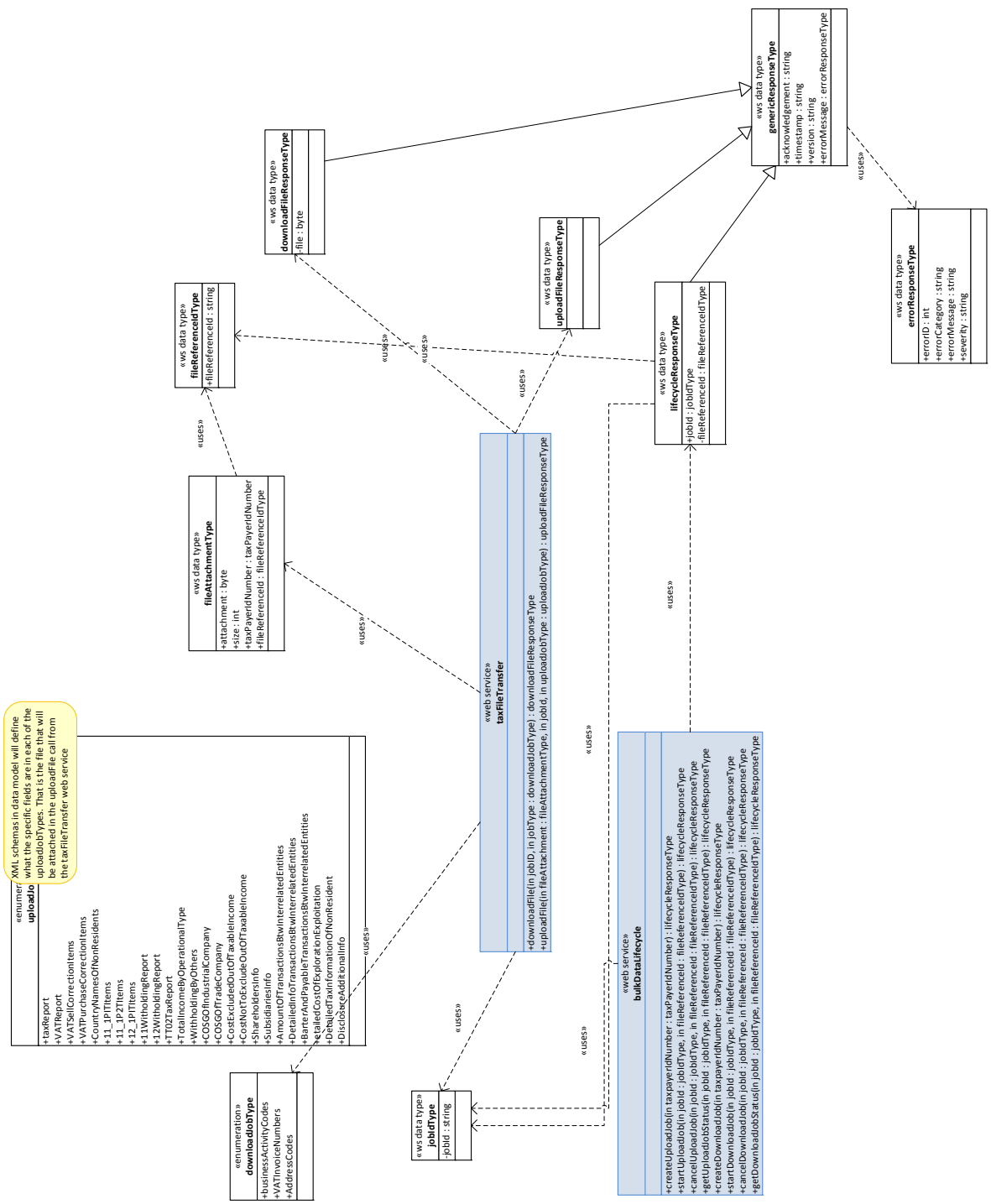

Figure 12: Bulk Data Web Service Candidates

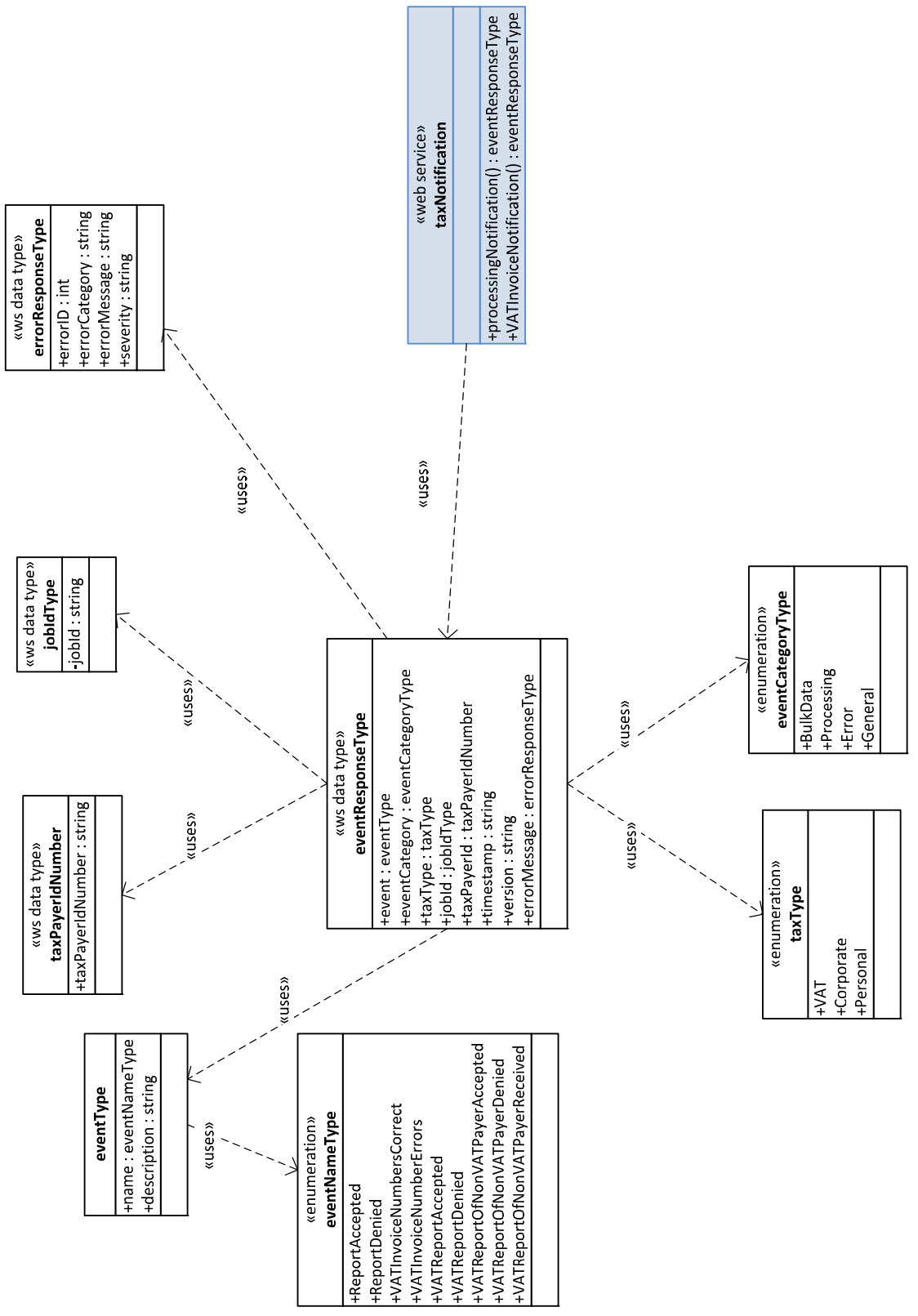

Figure 13: Notification Web Service Candidates

SECTION VII: TECHNICAL ARCHITECTURE

To this point we have outlined what the data will look like as it traverses from party to party

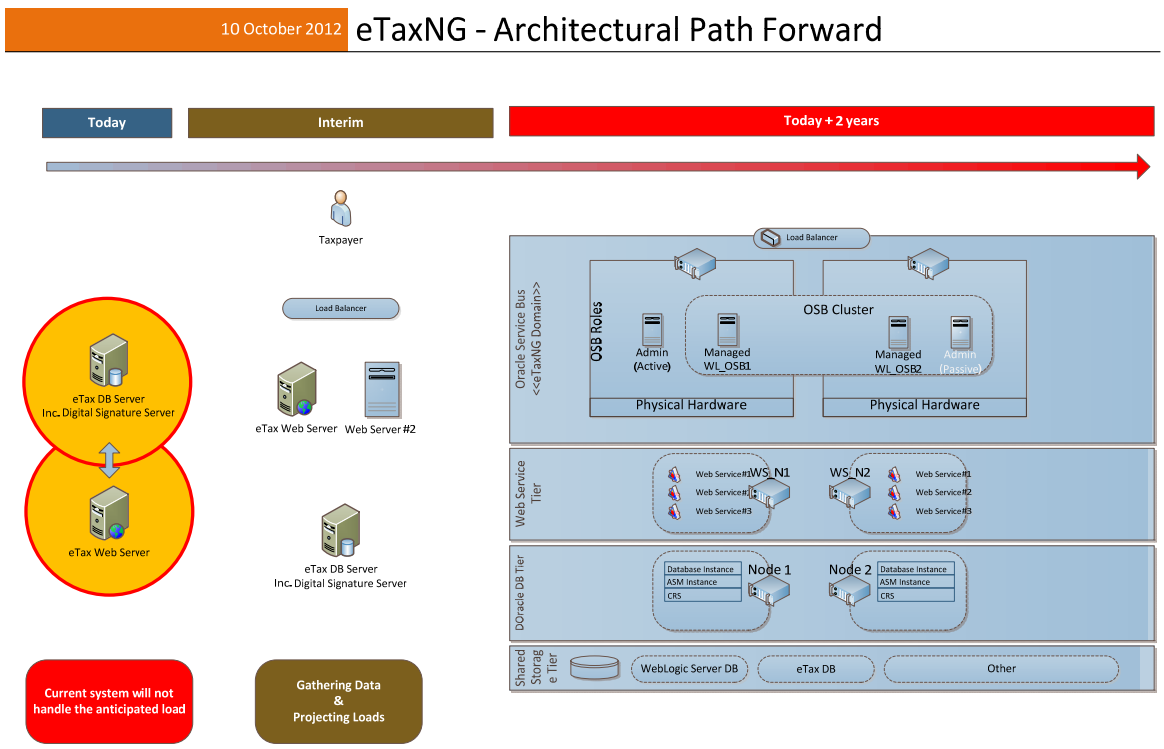

Figure 14: Architectural Path Forward

and the patterns in which it will move. What remains to be defined is the technology that will enable these patterns to function in a reliable and scalable fashion.

First and foremost it is essential to note that the process towards moving forward with eTaxNG is not a one-off short-term project. It should be viewed as a continuum with a target date of at least 2 years from present. There is a significant amount of infrastructure and base technologies that have to be implemented and coded. A fully scalable system that follows proper reference architecture is also costly. The immediate goal is to make an investment in the core technology and infrastructure and then build on them over time to finally arrive at the goal. There are also a wide variety of taxpayer types – from individual to large-scale corporate. The latter have very special requirements and integrating them may take significant time and effort. Some of them may have to be integrated on a case-by-case basis. It is recommended that initially only MOSA software be integrated with the system. It may even be necessary to implement only selected MOSA members due to the very aggressive timelines demanded.

It is also important to be strategic. In other words, it would be less than ideal to put together a short-term solution to meet immediate needs without consideration to a longer-term strategic vision. Following, this paper will first lay out the architecture for a long-term strategic vision and then separately identify what the initial steps would be to make e-Filing happen in the immediate future with the idea that any investments in hardware, software and technologies

would be the initial building blocks of the longer-term vision. This also serves to identify the second immediate project that BPI could undertake to make e-Filing happen.

The eventual goal of a technical architecture would be one that is horizontally scalable and highly available. Oracle publishes a reference architecture that is designed to meet these particular goals (http://www.oracle.com/technetwork/middleware/service-bus/overview/osb-

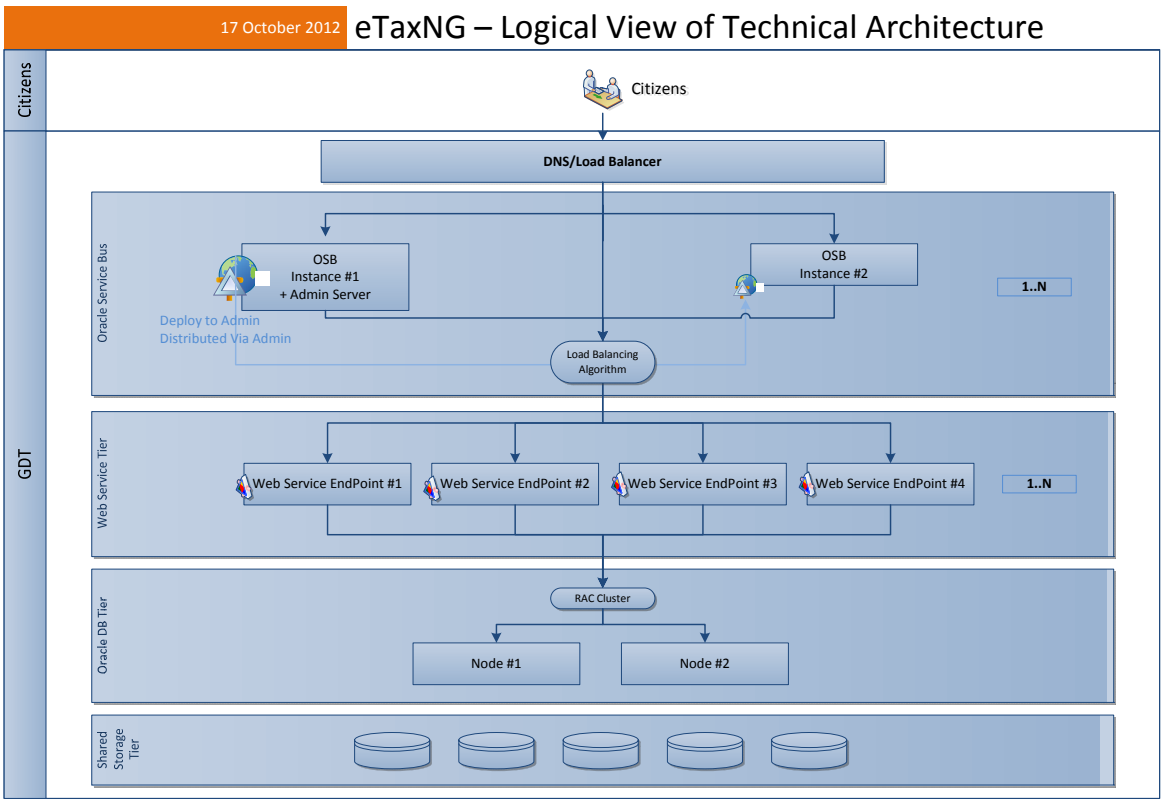

Figure 15: Logical View of the Technical Architecture

wp-ha-final-draft-134330.pdf).

Based on that reference architecture (see Page 4 in that document) a logical view of the architecture was developed (Figure 15). At the bottom of the stack would be enterprise wide storage. This would be a storage area network (SAN) with specific logical unit numbers (LUNS) specifically designed and allocated to this application. A LUN can be thought of as a "virtual hard drive". Other LUNS could be provisioned for other needs. The GDT has currently taken possession of a SAN that is dedicated to the use of a virtual environment previously mentioned. That SAN appears to be allocated as per the information from the GDT. Ideally, this SAN would be purchased with enough storage capacity to offload direct attached storage (local hard drives) that is used in many servers currently at the GDT. Figure 3 shows many of the current systems at the GDT. Many of these systems use direct attached storage, i.e. hard drives. According to the GDT, these hard drives are single SATA drives and often not configured in a RAID setting. RAID is a hardware configuration that provides some degree of redundancy. This future SAN would provide highly available and redundant storage for these systems. It would of course necessitate a long-term progressive migration plan. That is something that would need further study.

The next tier would host database systems. The most scalable solution based on Oracle technology would be to implement an Oracle Real Application Cluster or RAC. A RAC is a

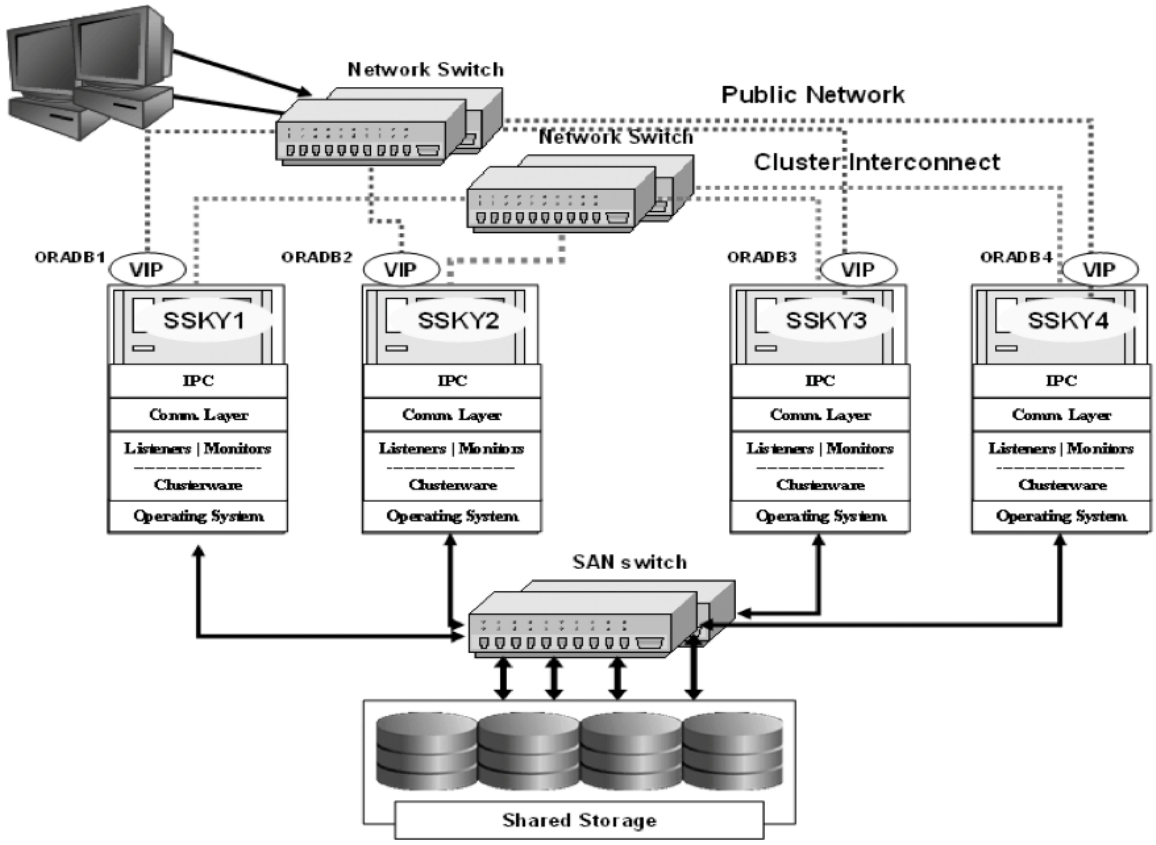

Figure 16: Oracle Cluster Components

Source: Oracle® 10g RAC Grid, Services & Clustering, Murali Vallath, Elsevier, 2006

very scalable nothing-shared solution from Oracle that permits growth over time. Figure 16 shows the basic components of such an architecture. It is also possible to extend the architecture across distance and have another cluster that is mirrored for disaster recovery purposes. This is an area at GDT that needs further study.

Of note, upon seeing an architecture as presented in Figure 16, clients often ask about specific hardware and storage specifications and setup, e.g. what type of RAID setup do I use for my SAN. Specifics for this type of setup are difficult to specify at this point in time. Commonly, different storage solutions and different servers tend to have recommended best practices on how to specify and setup the hardware for their specific equipment in such a setting. For example, HP publishes its reference architecture for setting up a RAC on HP storage solutions using HP servers (http://h71019.www7.hp.com/ActiveAnswers/downloads/4AA1-0086ENW.pdf). The GDT will need to go further along this path before more specific answers can be provided and then will need to work with experts and Oracle vendors to find the right setup, e.g. disk layouts, RAID configurations, etc.

The next layer in the stack would be an application server layer that runs the various web services needed to operationalize e-Filing. By having a horizontal tier of servers running web

services, the GDT would eliminate any potential performance bottlenecks. How the services would be allocated is also flexible depending on the type of service, i.e. how memory and processor intensive that service is and what other services are on that server. In one scenario all servers could be running the same set of services with traffic being load balanced from the tier immediately above. In another scenario, services could be dedicated to a specific set of machine(s).

At the top of the logical architecture would be the Oracle Service Bus (OSB). OSB is a mediation layer based on the enterprise service bus (ESB) design pattern. Kerrie Holley and Dr. Ali Arsanjani, writing in 100 SOA Questions Asked and Answered describe an ESB as follows:

> "The ESB mediates characteristics of interactions between service requestors and service providers. The ESB enables the substitution of service providers or implementations transparent to service requesters. The ESB supports a variety of means to attach requesters and providers, and it allows intermediary services to be sequenced between requesters and providers. The ESB may provide a broad set of capabilities dependent on business needs and implementation in several areas, including network communications, integration, security, message processing, quality of service, and service management."

The OSB is a product that permits the above. At its core, the OSB offers loose coupling, location transparency, mediation, schema transformation, service aggregation, load balancing, monitoring, and configuration over coding. Given the wide variety of clients e-Filing may have to support and the wide variety of potential methods of transport, there must be a layer between the outside world and the inner GDT world that mediates and protects internal systems from excessive modification. Figure 10 on Bulk Data Services illustrates this point precisely.

Of particular interest is also OSB capacity for configuration over coding. With the high degree of functionality built into OSB, there is very little coding that needs to be done and most of the work is configuration based. This will help GDT be more responsive to change and reduce their time to market. Fundamentally, there is nothing in an ESB that could not be custom coded. However, that code would need to be maintained and extended with each new use case. OSB can also be scaled horizontally by adding new nodes or machines that would permit it to handle a very large number of concurrent clients. It is a solid, proven, future-proof technology and the GDT is already using it in another project. Strategically, over time, these projects should unite.

At the very top of the stack would be a hardware load balancer. This is designed to distribute traffic across the ESB seamlessly. It would complete the scalability solution.

Further work was done to show how this type of logical architecture could be placed on physical hardware. Figure 17 shows how physical servers map to the logical layers.

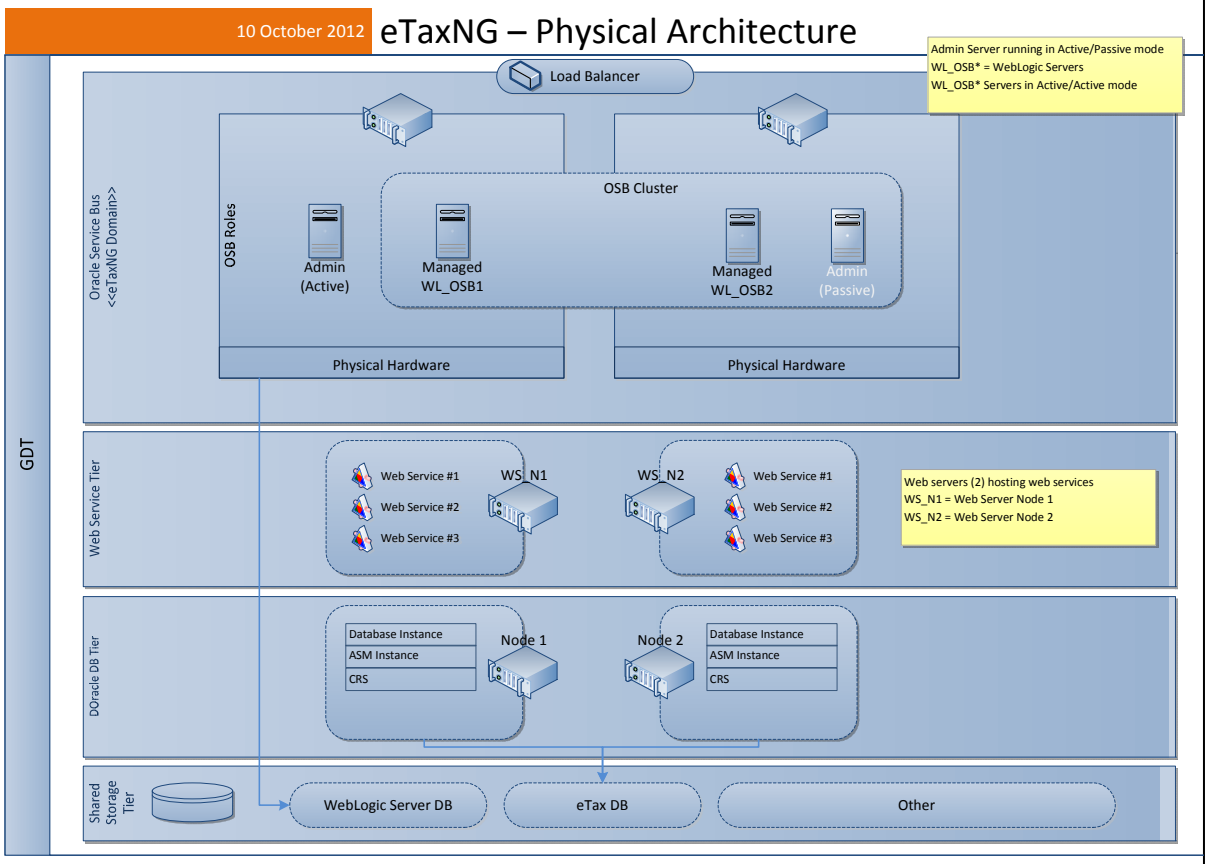

Figure 17: Physical View of Technical Architecture

The most pressing question is what needs to be implemented in the short-term to begin the progression on the continuum. In other words, how does e-Filing get started?

In order to better understand what the system requirements may be for e-Filing, the current patterns of traffic were examined at GDT for users submitting data online. These were hampered by lack of initial instrumentation on GDT servers. However, some data was obtained. The maximum load occurs on the 20th of each quarter with approximately 5000 different IP addresses hitting the GDT eTax site. Peak hours are between 1000 and 1600. Between 800 and 1000 users connect during the peak hours (5000 IPs / 6 hours). The distribution appears flat during that time which would average out at approximately 830 users during any hour. In theory at any given point during that hour the user load may range from 0 all the way to 830 concurrent. In reality, concurrent usage is somewhere between these extremes.

There are some smaller peaks during the 10th of the month but they do not approach the level of usage noted above. Load on the server at non-peak times is almost nothing.

The current web server with a PHP application has 4 GB of RAM and is suffering performance issues during this peak period. Adding additional RAM or adding another server and load balancing would help greatly.

What the current patterns show is that traffic to GDT for the purposes of filing tax reports is very 'bursty', i.e. comes in very high spikes for a very short period of time.

Conversations with MOSA members indicate that they will add approximately 2100 potential users. Large taxpayers may add another 300 – 400 users for a potential total of 2500 users. The GDT has future plans to add up to 1 million individual taxpayers. There is no reason to believe that usage patterns will be any different unless the Government makes some radical changes to deadlines or staggers deadlines. The current user load from the current eTax server may become an issue; the backend database system will be shared. Front-end traffic will remain on the current web servers. There will be some shift of load from the frontend web servers as some filers move to eTaxNG. What that shift will be is difficult to predict.

Given a possible of 2500 users and the same basic traffic patterns, that would project out to a possibility of approximately 0 to 400 users on at any one time for e-Filing. Note that instituting systematic monitoring and then performance tuning bottlenecks can also enhance performance. Stress tests are also recommended before the system goes to production.

Therefore, the basic requirements for the interim system are:

- High availability during peak tax times with failover
- 400 concurrent connections for e-Filing
- Database must handle 400 concurrent e-Filing users plus existing workload from eTax web server
- Generally be able to handle payloads of under 500k to 1M. Payloads should be compressed.

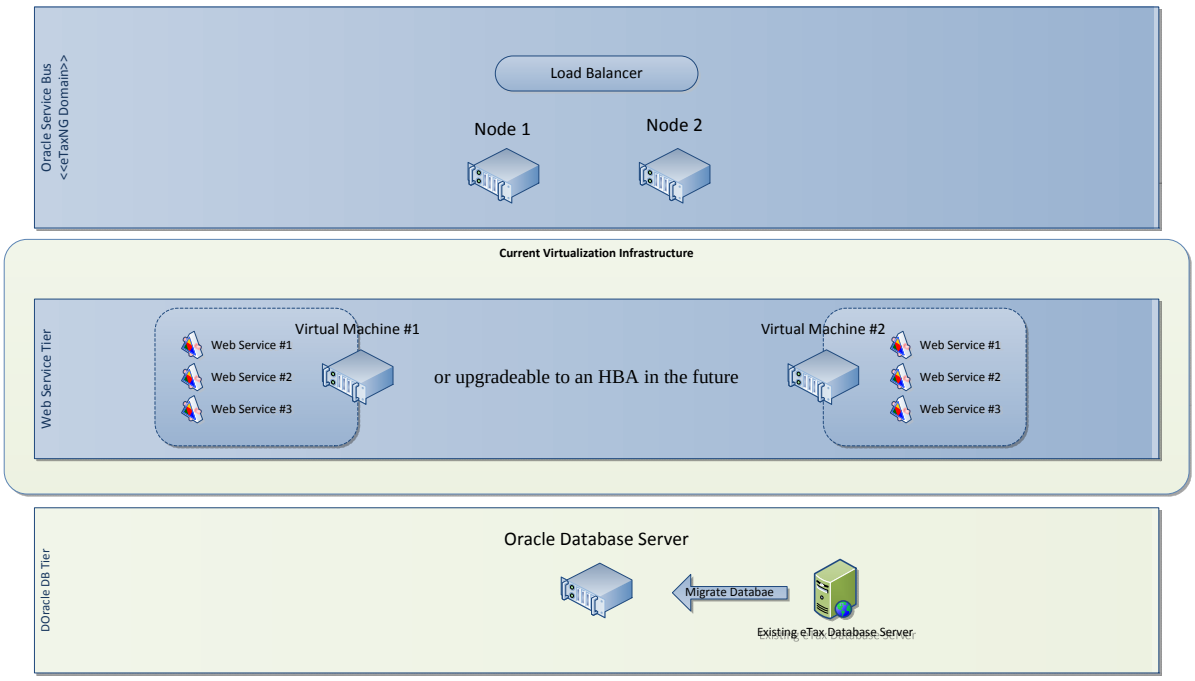

Figure 18: e-Filing Technical Architecture Physical View

Sizing hardware for Oracle systems is challenging simply because Oracle does not easily share sizing tools and prefers clients go through Partners and Vendors. However, as a benchmark, the sizing recommendations provided by HP for sizing Oracle middleware on their hardware

was used (http://hporacle.com/-/documents/view.asp?id=499).

Based on the above discussion and assumptions, Figure 18 outlines the basic physical hardware needed for the startup of e-Filing. The recommended path forward is to add an Oracle Service Bus cluster, virtualize the web services tier and replace the current database server with a larger and more robust machine. That existing database server would have its data migrated to the new database server and then be retired. The Oracle database license would be repurposed to the new machine (some additional licenses may need to be purchased depending on the agreement between Oracle and GDT for the current database server). The new database server would handle traffic for both the eTax website and for e-Filing. It would have a faster and more fault-tolerant disk drive subsystem and would permit GDT the opportunity to layout the disk subsystems as per Oracle recommended practices. This is slightly less than the ideal solution; implementing an Oracle Real Application Cluster for the database. However, implementing a cluster would add significant complexity and cost at this time. Not only would more servers and licenses be required, high-end networking equipment would also be required. The next logical step for GDT would be to add another machine and subsume the new database server into a cluster. Then, as need arose, additional nodes could be added to the cluster. This is an area that warrants further study. Base hardware and license specifications are noted below.

Item	**Qty**	**Specifications**
Load Balancer	1	Should support: • Basic link load balancing • Active-Active link load balancing • Active-Backup link load balancing • Linktypes: ADSL, Ethernet, Fibre, etc. • Switching between active and failed links, while guaranteeing 100% of active connections will remain active • Balancing modes: weighted, least traffic, least bytes, least users, response tie, hashing & round robin • Protocols: HTTP/HTTPS, SFTP/SSH, SMTP, JMS, DNS, FTP, IMAP, SMTP, TCP/IP, UDP/IP • Direct Routing LB Method (aka Direct Server Return or N-Path) • SSL Pass-through mode • Minimum 850Mbps throughput & upgradable Routing protocols: RIP, RIP V2: OSPF, etc. • Must be capable of load balancing Oracle Service Bus • Ports – 8 x 1GB minimum
OSB Tier Servers	2	Oracle Enterprise Linux – 64 Bit X86_64 Bit Architecture 2 x 6 Core Processor Xeon @ 2.6 GHz or greater 15K RPM SAS Drives in a RAID 5 Configuration. SSD drives preferred Redundant Power Supplies 32 GB RAM Minimum Dual Gigabit Ethernet Cards

		HBA for Shared Storage (if GDT has specified a SAN solution to connect to) or upgradeable to an HBA in the future
Database Tier Servers	1	Oracle Enterprise Linux – 64 Bit X86_64 Bit Architecture 2 x 6 Core Processor Xeon @ 2.6 GHz or greater 15K RPM SAS Drives in a RAID 10 Configuration. SSD drives preferred Redundant Power Supplies 48 GB RAM Minimum Dual Gigabit Ethernet Cards HBA for Shared Storage (if GDT has specified a SAN solution to connect to) or upgradeable to an HBA in the future Adapter upgradeable to integration with an Oracle RAC
Oracle Service Bus License	-	Per-Processor License & Support as per http://www.oracle.com/us/corporate/pricing/sig-070616.pdf . Must support the 2 OSB Tier servers specified.
Oracle 11g Database License	-	Repurpose existing eTax database license to new database server. Additional licenses may need to be purchases depending on the type of license GDT has for the existing database system. Must support the Database Tier server specified.
Oracle Enterprise Linux	2	Support and service patches

Please note that servers on the Web Services Tier will be virtualized on the current VMware VSphere virtualization environment. At the time of this writing there was full availability on one of the hosts in the virtualization cluster and this issue has been discussed with the Head of IT at the GDT. If that changes, then either additional servers need to be purchased or the virtualization cluster expanded. The latter is preferable if physical space remains in the data center.

Ideally, there would be 2 load balancers to provide full high availability. For this interim step one will suffice to reduce costs.

It is also anticipated that web services will be deployed to either Apache Tomact or Oracle Glassfish servers. Tomcat is an open-source project and Glassfish is an open source Java EE 5 application server. GlassFish is a robust, commercial, production quality, compatible application server that is free for development, deployment, and redistribution. No additional licensing costs are anticipated.

SECTION VIII: SECURITY

Security for the e-Filing system is a vital and critical component. Implementing security correctly can be very challenging. Security is also a moving target as threats are constantly evolving and new ones being created. An assessment of how the current online eTax system works in terms of security was done. It was observed that no transport layer security was implemented. Logons are done without any encryption. All data is currently transmitted without any encryption. Upon submitting and digitally signing tax returns, the tax filer was asked to submit their private key. That was also done without encryption.

The US National Institute of Standards and Technology or NIST has some very specific recommendations regarding this practice. NIST is the US National Institute of Standards and Technology. The E-Government Act in the US [Public Law 107-347] passed by the 107th Congress and signed into law by the President in December 2002 recognized the importance of information security to the economic and national security interests of the United States. Title III of the E-Government Act, entitled the Federal Information Security Management Act of 2002 (FISMA), included duties and responsibilities for the Computer Security Division in Section 303 "National Institute of Standards and Technology." Their duties include providing assistance to comply with FISMA.

NIST Special Publication 800-57, Recommendations for Key Management (http://csrc.nist.gov/publications/nistpubs/800-57/sp800-57_part1_rev3_general.pdf) states:

> "Information protected by cryptographic mechanisms is secure only if the algorithms remain strong, and the keys have not been compromised. Key compromise occurs when the protective mechanisms for the key fail (P 59)"

Compromise of a key has the following implications according to NIST SP 800-57:

> "In the case of the unauthorized disclosure of a key used to provide integrity protection (e.g., via digital signatures), the integrity protection on the data may be lost. For example, if a private signature key is compromised, the unauthorized entity might sign messages as if they were originated by the key's real owner (either new messages or messages that are altered from their original contents), i.e., non-repudiation and the authenticity of the information are in question.
>
> The unauthorized disclosure of a private signature key means that the integrity and non- repudiation qualities of all data signed by that key are suspect. An unauthorized party in possession of the private key could sign false information and make it appear to be valid. In cases where it can be shown that the signed data was protected by other mechanisms (e.g., physical security) from a time before the compromise, the signature may still have some value. For example, if a signed message was received on day 1, and it was later determined that the private signing key was compromised on day 15, the receiver may still have confidence that the message is valid because it was maintained in the receiver's possession. Note that cryptographic timestamping may also provide protection for messages signed before the private signature key was compromised. However, the security provided by these other mechanisms is now critical to the security of the signature. In addition, the non-repudiation of the signed message may be questioned, since the private

> signature key may have been disclosed to the message receiver, who then altered the message in some way.
>
> The disclosure of a CA's private signature key means that an adversary can create fraudulent certificates and Certificate Revocation Lists (CRLs). (Page 60)"

What is clear is that security could be more effectively implemented and must be more stringent for e-Filing. Compromise of a digital signature destroys trust in the process and institutions.

In addition to the above issue, electronically filed tax returns need to be digitally signed. This is common practice. However, the process of digitally signing a tax return is complex and needs to be made simple and transparent to the end user. Failure to do so will likely result in poor adoption rates due to excessive complexity.

In the context of the proposed architecture, the various patterns of interaction were analyzed and implications for information security noted. These are presented below in Figure 19.

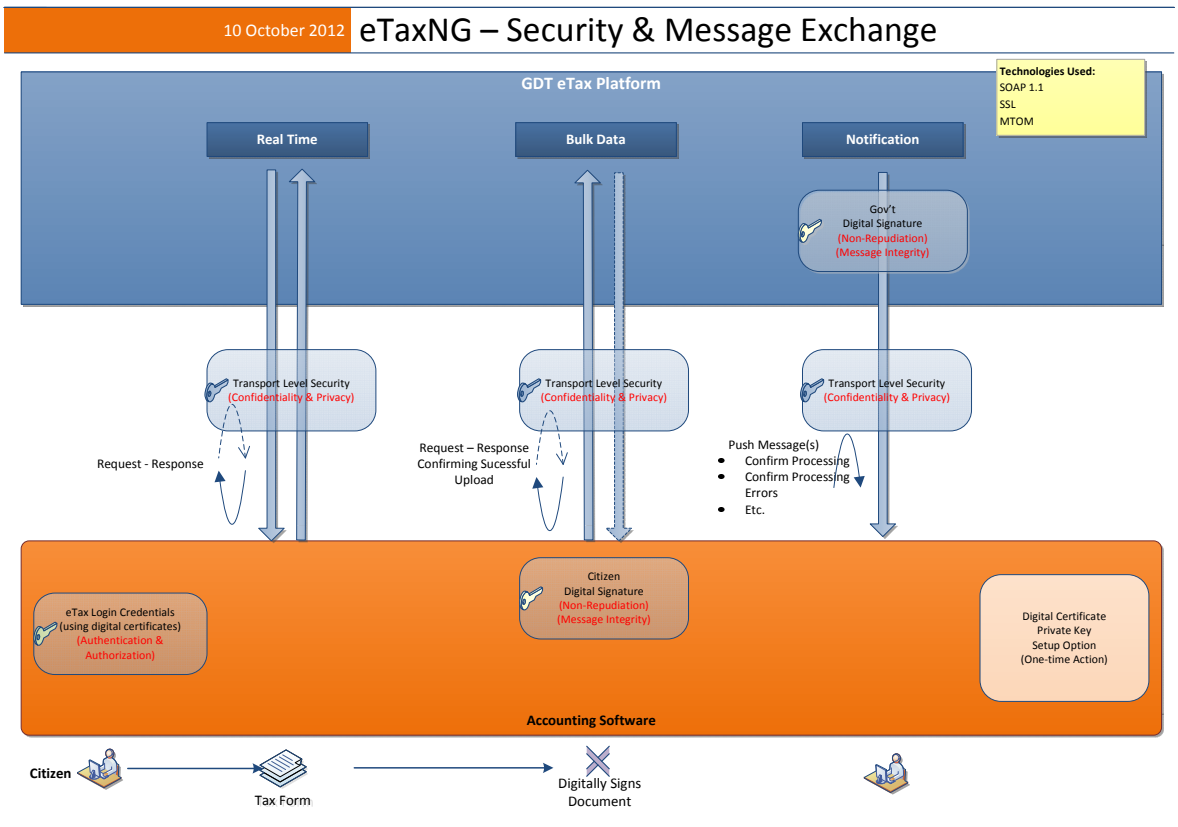

Figure 19: Information Security and Data Interaction Patterns

In order to ensure security the following must be present:

- Authentication
- Authorization
- Confidentiality
- Message Integrity

- Non-repudiation

Authentication is used to identify the origin and identity of the information. If verifies the identity of the user or the system. Authorization is concerned with providing permission to perform a security function or activity. This is granted after authentication. Non-repudiation provides assurance of the integrity and origin of the data. This must be verifiable by the receiving party. Message integrity shows that the data has not been altered in any way since it has been created, stored or transmitted. Confidentially is the assurance that the data has not

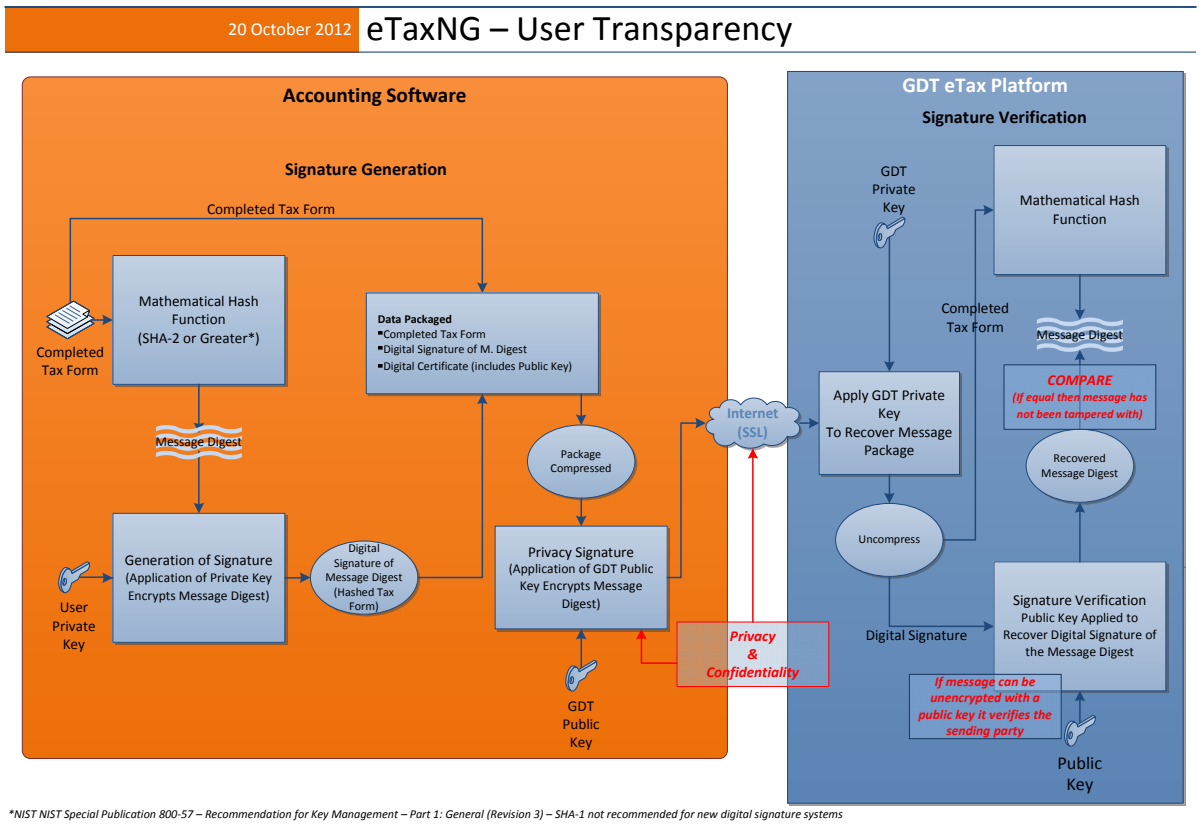

Figure 20: Security Implementation

been disclosed to any third party.

Figure 20 describes the best practice application of security to eTaxNG. The goal is to make the security process transparent to the end-user of the accounting software. It must be! The tax form that is to be digitally signed must first have a mathematical hash function applied to it using SHA-2 or greater. Weaker keys leave the system more vulnerable to incursion.

The details of implementation are spelled out clearly in Figure 20. Salient points are that the outgoing tax submission must be signed and encrypted by both parties by the time it leaves the tax filers computer system and enters the network. Private keys should also never be transmitted over the general Internet. If the above is followed the GDT will be able to assure tax filers that privacy and confidentiality has been maintained, message integrity is intact and non-repudiation is guaranteed.

In order to establish authorization and authentication the GDT will need to issue digital certificates based on PKI (Figure 19). These will be used to establish that a particular user is

authorized to connect to GDT web services. As per best practice these certificates should not be the same as the ones used for digital signature. While technically using the same keys is possible, it couples authentication with non-repudiation, message integrity, etc. If a key is compromised then everything is compromised. If separate keys are used then the damage can be contained somewhat and may not spread as easily. A compromised authorization key will likely not compromise a digital signature.

The GDT will also digitally sign outgoing messages from the notification services to authenticate those messages.

Figure 19 illustrates the division of labor between GDT and the accounting software. This is one area that adds project risk in that it adds burden to MOSA software vendors to implement this functionality into their software. They will also likely need to add additional functionality to setup the various digital certificates and keys.

Large taxpayers also potentially present a security challenge. If they are not going to avail themselves of the security services via web service calls, they will need a standalone application that can digitally sign documents before they transmit them via other means. This will have to be custom developed. This is a project risk as it impacts timelines.

One of the MOSA members also noted that some accounting software runs in a cloud environment. That is, users enter accounting information via a web browser and the data is stored in a cloud-based data center. There are obvious security implications for the transmission of this data to the cloud data center and storage of this data at the cloud data center. A more basic question is how to implement the functionality of digital signing from within a browser environment since a browser is sandboxed from the base operating system, i.e. prevented from doing certain functions. Several options have been investigated. Some seem promising but all require further investigation. In brief, the following options are presented for further study:

- Cyrpto.js library from Sencha (http://www.sencha.com/forum/showthread.php?80415)
- Stanford JavaScript Crypto Library (http://crypto.stanford.edu/sjcl/)
- A method used by the coSign Cloud service from ARX (http://www.arx.com/files/-DOCUMENTS/CS-Cloud-fact-sheet.pdf). They use their signing software on the users system to create a document hash and send it to their security appliance via an encrypted link. Their security appliance digitally signs the documents and sends it back to the user. The user then proceeds with their normal business. It is feasible for a cloud provider to adopt this type of approach. Buying and implementing the ARX solution may also be an option.
- ActionScript 3 Cryptography library (http://code.google.com/p/as3crypto)
- ActionScript 2 Cryptography (http://www.yuniti.com/As2RSA)
- Custom ActiveX control. Microsoft formerly provided the CAPICOM library but that has been deprecated. To make the alternatives work (http://msdn.microsoft.com/en-us/library/windows/desktop/cc778518(v=vs.85).aspx) with scripts Microsoft recommends writing a custom ActiveX control. Deeper investigation is warranted if an ActiveX plugin is viewed as an option.
- Store private keys securely on a secure appliance at the cloud data center
-

SECTION IX: MOSA INVOLVEMENT

During the Mission, MOSA members have been contacted and invited to 3 separate discussions. The first was a general introduction and broad overview. The second was a discussion of data semantics and XBRL format. The third was a discussion of the technical architecture. In large part these meetings serve to attempt to involve MOSA members in the process and give them advance notice of the directions that the project will be taking so that they know what to expect, when and how to accomplish their tasks. MOSA members were very receptive and provided valuable feedback. MOSA involvement is a project risk as noted below so their involvement is critical throughout the process.

SECTION X: PROJECT MANAGEMENT STRATEGY

Key to success of any IT project is project management. If this project is to succeed then it must follow proper project management techniques. Risks are very high in this project.

A study by the Standish Group in the early 2000's classified 66% of IT projects as being unsuccessful. Failure was defined as not being delivered on time, exceeding budget, delivering fewer features than planned, fail completely, etc. A Gartner research study (Figure 21), looking at the top reasons for project failure found that undisciplined project management practices are a major contributor to failure. Together with inexperience, these account for 35% of project failure.

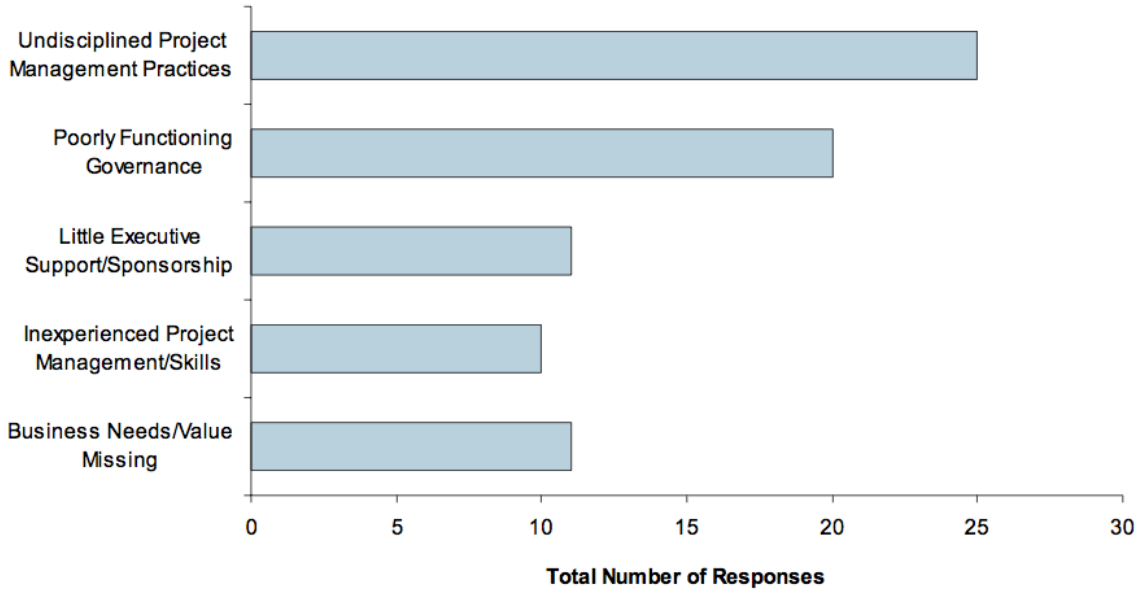

Figure 21: Reasons for Project Failure

Source: http://condor.depaul.edu/dmumaugh/readings/handouts/SE477/Gartner%20Reports/from_the_cio_trenches_why_so_151721.pdf

Discussions with GDT reveal that GDT does not follow a recognized project management practice like the ones from the Project Management Institute or PRINCE from the UK. There is awareness of these methodologies but they are not used. Anecdotal evidence reveals that approximately 30% of projects are delivered on time at the GDT. Exacerbating the situation is the fact that this project has very tight timelines. This along with the industry's poor general performance record indicates that this project is potentially very high risk and project management will be a key factor in its success or failure.

The GDT currently is accustomed to working in a waterfall style of project methodology. Think of this as a traditional approach with a heavy requirements gathering phase followed by development and finally deployment.

An Oracle 2011 White Paper (http://www.oracle.com/us/solutions/018860.pdf) identifies strategies to avoid common pitfalls. These are:

- Constituent alignment (engagement & commitment of stakeholders)
- Proactive risk management

- Performance Measurement
- Project Scope Definition and Management
- Critical Project Communication
- Methodology Usage.

A key observation in the above work is that it is essential to select a methodology and enforce its usage. This leads to the question of how the GDT should approach this project from a management perspective? Does it keep its current approach? Does it try something new? A major driving factor in this project is a very aggressive time deadline. There is great pressure from many quarters to deliver very fast – unrealistically fast. The topic at hand is complex, there is little documentation of business processes, many new challenges may be discovered as the project moves forward and there are many stakeholders, some of which are outside the scope of GDT. MOSA is a key stakeholder and they need to start work in a lag-parallel fashion with GDT if they are to deliver to their customers. If MOSA does not deliver then the project as a whole does not deliver. The methodology chosen has to be very flexible and resilient to change. Characteristic of a big design up front project is sufficient lead-time to do design. That lead-time is practically non-existent in this situation.

The recommended approach that GDT should consider is to use an agile-style project approach to delivering eTaxNG. The main principles (http://www.agilemanifesto.org/principles.html) behind an agile approach are:

- Customer satisfaction by rapid delivery of useful software
- Welcome changing requirements, even late in development
- Working software is delivered frequently (weeks rather than months)
- Working software is the principal measure of progress
- Sustainable development, able to maintain a constant pace
- Close, daily co-operation between business people and developers
- Face-to-face conversation is the best form of communication (co-location)
- Projects are built around motivated individuals, who should be trusted
- Continuous attention to technical excellence and good design
- Simplicity- The art of maximizing the amount of work not done - is essential
- Self-organizing teams
- Regular adaptation to changing circumstances

It is very clear that the project's most essential goal is to deliver a working product in a very tight amount of time. It is also clear from the agile principles that most if not all of them will need to be adhered to for the project to be successful. It is also clear that following these principles will also help the GDT deal with the project management risk avoidance strategies noted by Oracle above.

A risk of using this methodology is that GDT is not familiar with it. However, GDT does not follow formal methodologies and adopting agile is no more risky than adopting another more traditional approach. Maintaining the current GDT approach is also not an option – it does not have a track record in delivering software on time.

It is anticipated that the team developing eTaxNG will be comprised of:

- Project Manager
- Technical Lead / Programmer
- 4 x Programmers
- Quality Assurance Expert

Individual statements of work are noted in the Annex.

SECTION XI: PROJECT RISKS

The above discussion has already noted that this is a very high-risk project. There are a number of very specific risks that can be identified currently. These are:

- The timelines are extremely aggressive given the complexity. A possible mitigation of this risk to keep scope control extremely tight. This will greatly depend on excellent project management being exercised. Further, it is a strong argument for limiting this phase to a small group and treating this as a pilot phase. This might mean limiting the initial rollout to only MOSA members and then tackling large taxpayers later.
- As already noted several times, very strong project management will be essential. Failure is this area will likely result in overall failure. The only mitigation is to enlist the services of an experienced project manager who can work in an agile-style. Acceptance of this risk is not a viable strategy.
- The implementation of security is a risk in that if not done properly will result in potential loss of reputation to the Government. This risk can be mitigated by close monitoring of how security is being implemented during the project implementation, by closely following best practices of groups such as NIST, by performing security testing before deployment and by including security as part of the architecture and not as an 'add-on' feature that is bolted onto the system post production.
- Performance and scalability are a risk to the project. A completed system that does not perform will result in reputation loss for the Government. A non-scalable solution will result in much additional cost and complexity in the future. The mitigation for this is for GDT to commit to the long-term vision noted in this document and work incrementally towards it. The architecture specified has great capacity to grow if followed.
- A serious risk to this project is that of skill-gaps. These can occur in many different areas with many different parties. To mitigate this risk technologies have been chosen that the GDT is very familiar with and has worked with in the past. There may be some additional learning curve in the area of security. MOSA members may also have an additional learning curve here. Support can be provided to MOSA in the form of code samples, extensive documentation, whitepapers, etc. to assist with this learning curve. This is typical of companies like Microsoft who provide extensive developer support.
- Due to the aggressive timelines there is no additional time for any 'learning curves'. That is to say that the team chosen to implement this project must know the technology to be implemented. There is no other mitigation.
- Even with the scaled back infrastructure proposed, funding will still need to be arranged. The current systems at the GDT will not scale to handle electronic filing. They are already experiencing performance problems. The fully scalable and future-proof system proposed will require much additional funding. For example, moving to an Oracle RAC database will require expensive additional networking plus additional servers. The GDT has committed long-term to use of Oracle technologies and this risk must be accepted.
- The cost of Oracle licenses is a risk to the project. Even in the interim phase there will have to be a significant outlay for licenses that will dwarf the cost of hardware. As noted above, this risk must be accepted. The other solution would be to move to less-costly or open-source technologies in the long-term. While upfront costs in the case of open-source

technologies is reduced or eliminated, other costs must be factored in such as expertise, cost of maintenance, etc.

- MOSA members have a major impact on this project and hence are part of the risk pool. They not only have to implement the changes into their software but then have to roll it out to their customers. In part this can be mitigated by having MOSA members work in parallel, with a lag behind GDT. In part providing additional support as noted above can mitigate this.
- Tax reform legislation is also a risk to the project. It is a moving target and significant changes in legislation may result in significant changes to tax forms and hence systems. Depending on the type of change, it could potentially domino throughout the entire system. This is a risk that must be accepted.
- Large taxpayers are a risk to the project. While they constitute a large percentage of the tax paid, they tend to have large and complex systems that are not amenable to change easily. Some will never be able to integrate functionality into their software. Some have legacy systems that will not change in the foreseeable future or ever. The business processes that GDT has to use with these taxpayers are also the most complex. Exploratory work also finds that the additional workflows will need to be built in the system to handle these types of taxpayers. Further, integration of many of these systems may almost have to be done on a one by one basis. The risk to the project is very great from them. They greatly endanger the timeline. Further, architecturally speaking, with all the edge use cases they present, it may overwhelm the team developing the system in the short time frame they have to work. The core foundation may end up being sacrificed for them. The most logical mitigation strategy is to move large taxpayers into separate second phase.
- A possible risk to the project is re-engineering backend GDT business processes. As envisioned, this phase of the project attempts to put in the middleware to enable e-Filing. Some backend engineering is foreseen. However, a full re-write of backend processes will not be possible. In fact, it would seriously impact the timeline. Where possible, the existing systems must be re-used. On GDT's part there must be commitment to a long-term strategy where backend processes are first optimized and then re-engineered. The best way to mitigate this is through strong project management and scope control.
- Currently the GDT issues its own certificates based on public key infrastructure. There is a competing government initiative that is meant to be a government-wide digital signature initiative. There may be some potential effects when this wider initiative comes online. This bears close watching for potential effects.

SECTION XII: PROJECTED PROJECT TIMELINES

A master project plan has been developed outlining the major tasks and resources needed to complete this project. It is presented as a separate Microsoft Project 2010 file and an Adobe Acrobat file attached to this document.

SECTION XIII: FURTHER STUDY

During the course of this study a number of additional areas were noted where further study might be of benefit to GDT. These were not done as part of this study simply because they were out of scope. Some were tangential to the implementation of e-Filing. Some have partial impact on e-Filing but can remain in their current state until subsequent phases. These are outlined below – the list is not exhaustive.

- An element that needs further discussion and elaboration at GDT is that of disaster recovery. Currently backups are done as expected on machine-by-machine basis. This provides some safety but not high availability and certainly not a formal, structured approach to business continuity in case of disaster. Currently there is no secondary or backup data center either. Ideally, the GDT, on a strategic level, would go beyond just backing up data and frame the issue within the broader context of business continuity planning (BCP). A BCP takes a broader view than just disaster recovery. It involves getting critical systems online in another environment, getting the right people in place, being able to perform core business in an emergency mode while still dealing with citizens & partners, etc. The National Institute of Standards and Technology (NIST) in the US, for example lists the following broad categories of BCP (Special Publication 800-34):

    1. Develop the continuity planning policy statement. Write a policy that provides the guidance necessary to develop a BCP and that assigns authority to the necessary roles to carry out these tasks.
    2. Conduct the business impact analysis (BIA). Identify critical functions and systems and allow the organization to prioritize them based on necessity. Identify vulnerabilities, threats, and calculate risks.
    3. Identify preventive controls. Once threats are recognized, identify and implement controls and countermeasures to reduce the organization's risk level in an economical manner.
    4. Develop recovery strategies. Formulate methods to ensure systems and critical functions can be brought online quickly.
    5. Develop the contingency plan. Write procedures and guidelines for how the organization can still stay functional in a crippled state.
    6. Test the plan and conduct training and exercises. Test the plan to identify deficiencies in the BCP, and conduct training to properly prepare individuals on their expected tasks.
    7. Maintain the plan. Put in place steps to ensure the BCP is a living document that is updated regularly.

    As can be seen from the above NIST plan of action, developing recovery strategies is just part of a broader set of issues that would need to be addressed in case of disaster. The GDT would benefit greatly from having these broader issues addressed and then being able to precisely formulate its disaster recovery strategy.

- Currently the GDT is acting as its own Certificate Authority (CA) and issuing digital signatures based on public key encryption. There is a competing initiative at the government level that was originally expected to take effect in January of 2013, although may be delayed. That is a project risk itself. However, currently the eTax Oracle-based database server also runs the PKI infrastructure used to issue digital certificates. GDT has

plans to expand the role of this function to involve, for example, issuing certificates for devices. This is a positive plan and could have many management benefits to GDT. However, there needs to be discussion and study of how to properly structure and implement this PKI infrastructure.

- As noted, there needs to be extensive planning for the next phase of eTaxNG and how GDT moves strategically along the continuum towards a fully scalable system. By this, we mean what are the specific next steps. Included in this is the GDT long-term vision of rolling out e-Filing to an additional 1 millions individual taxpayers. There needs to be study on the specifics of how and when this will be accomplished.
- As discussed in project risks, there are potential changes to tax legislation. This can impact business processes and flow inside the GDT. It might be helpful to carefully study the business processes in the GDT and strategically optimize them. From an ICT perspective, it makes sense because it is easier to build systems on optimized business processes than to build them on current processes, have those processes change and then incur system changes.
- Noted in the discussion will be the existence of 2 Oracle Service Bus projects. The current existing one and the one for eTaxNG. Also noted is that these 2 should be combined into an enterprise level OSB. There needs to be study on the specifics of how this will be accomplished.
- As noted, the implementation of security at GDT would benefit from a more systematic treatment and greater general awareness of information security issues. It is vital for a government agency like the GDT to have a proper and effective information risk management strategy and to implement information security very well. The GDT would benefit from even something as simple as a workshop where they would learn about development of an effective information security risk management program and obtain practical guidance necessary for assessing and mitigating risks. This would include:
    - Risk Management
    - Risk Assessment
        - System Characterization
        - Threat Identification
        - Vulnerability Identification
        - Control Analysis
        - Likelihood Determination
        - Impact Analysis
        - Risk Determination
        - Control Recommendations
        - Results Documentation
    - Risk Mitigation
    - Evaluation & Assessment

    On a deeper level, performing the above methodology of GDT systems would also benefit the GDT.

SECTION XIV: MISCELLANEOUS

Monitoring and operations management will important as the system is deployed and assumes full operational status. Without doubt, there will be additional work that needs to be done to improve performance of the system. Key to this is effective instrumentation of the systems running eTaxNG. Key also is that GDT is proactive and not solely reactive.

At the database level, it is recommended that monitoring be expanded using Oracle Enterprise Manager. That utility is already in place and in use at GDT. The main tool in Oracle Enterprise Manager that should be used is the Oracle Automated Workload Repository or AWR. As guidance, GDT should follow at minimum the Oracle tuning methodology noted at http://docs.oracle.com/cd/B28359_01/server.111/b28275/tdppt_method.htm#CIHIEGEC.

The Oracle Service Bus should be monitored using Oracle recommended methods presented at http://docs.oracle.com/cd/E21764_01/doc.1111/e15867/monitoring_ops.htm. There will need to be further study on what metrics the GDT wants to monitor and what type of service level agreements it desires.

All systems should be monitored on a systems level using standard Linux-based tools:

- SAR
- Uptime
- Iostat
- Vmstat
- Top or Htop
- PS.

It should also be noted that GDT has implemented a business intelligence solution based on Oracle Business Intelligence Enterprise Edition. This is a set of tools for making reports. These tools are available in the future for business intelligence report generation.

It is assumed that the GDT will continue using its current backup solutions for eTaxNG until they complete a business continuity plan as noted above.

CONCLUSION

In conclusion, 3 projects have been identified that would have immediate results in implementing e-Filing and helping Mongolia advance its World Bank Doing Business Index ranking. These are:

- Model the data semantics to describe the format of the data that will move between the GDT and taxpayers
- Develop a core set of web services that would enable data movement. These web services would offer real-time functionality to enrich user interfaces, bulk data ones to move large amounts of data around and a set of notification services that would enable an event-driven architecture
- Upgrade infrastructure to sufficient levels to be able to handle the anticipated load.

A project timeline complete with resource allocation has been drawn up that integrates the various stakeholders and put completion of the project at or near the 1st of March 2013. There would permit at least 2 VAT tax submissions and 1 Corporate/Personal submission before the World Bank Doing Business deadline in May 2013.

It should be noted that the project has been put into the context of a long-term strategic vision for the GDT that would result in a fully scalable and performant system approximately 2 years form present if followed.

To assure success infrastructure requirements have also been identified and presented. The investments in this infrastructure have purposely been put into the context of the long-term strategic vision. This ensures that any investments made now are fully part of the long-term plan, i.e. they are the initial building blocks.

Key to both the reputation of the Government and the GDT, and the safety of tax filers, a vision has also been presented for the best practice integration of security into the eTaxNG system. This vision follows recommended practices that will assure privacy and confidentiality, authorization and authentication, message integrity and non-repudiation.

Finally, given the very high-risk nature of this project, project risks have been identified and where appropriate, mitigation strategies suggested.

ANNEX A:

ANNEX A: PROJECT MANAGEMENT REQUIREMENTS

Scope of Work

The Business Plus Initiative (BPI) Project intends to hire an individual consultant to perform the duties delegated by BPI for the purposes managing the development of the eTaxNG system (System).

The consultant will carry out the following tasks:

- Familiarise themselves with all technical matters relating to the System
- Represent BPI on all day-to-day matters relating to the System or arising from the contract, including, but not limited to:
    - Overseeing all project management tasks related to the System
    - Leading in the establishment and facilitation of project management best practices
    - Insuring that project methodologies, project delivery processes and implementation management processes are followed, including:
        - Project Schedules
        - Risk and Issue Management
        - Scope & Change Management
        - Stakeholder Engagement & Communication
    - Giving/receiving notices on behalf of BPI;
    - Developing & maintaining working relationships with external parties (MOSA) upon which the project has dependencies

Deliverables

- Monthly reports covering:
    - The progress made by the development team with focus on achievements, challenges, quality in implementation of the System
    - Recommendations to BPI whenever needed
    - Follow up of recommendations provided earlier
- Final consultancy report summarizing all technical aspects mentioned above

Duration

Expected starting date of assignment is 1st of November 2011. Project is expected to last until 1 March 2013.

Required Qualifications

General Qualifications:

- Has at least 5 years experience working in information and communications technology including but not limited to quality assurance, data centers, web technologies and architecture, relational databases, LAN/WAN & network, server platforms, messaging, service oriented architectures

- Has at least 5 years demonstrable experience in a Project Management role including significant experience managing large projects
- Bachelor's degree in Computer Science, Management Information Systems, Business Administration or equivalent educational or professional experience and/or qualifications. An advanced degree is also preferred.
- Has documented experience with Agile-style approaches to software development
- Very strong written communication skills
- Very strong oral communication skills

ANNEX B:

ANNEX B: WEB SERVICES PROGRAMMER REQUIREMENTS

Scope of Work

The Business Plus Initiative (BPI) Project intends to hire an individual consultant to perform the duties delegated by BPI for the purposes developing programming code for the eTaxNG system (System).

The consultant will carry out the following tasks:

- Familiarise themselves with all technical matters relating to the System
- Develop code for the implementation of web services in the Java programming language

Deliverables

- Production-quality code for eTaxNG

Duration

Expected starting date of assignment is 1st of November 2011. Project is expected to last until 1 March 2013.

Required Qualifications

General Qualifications:

- Has at least 5 years experience working in information and communications technology including but not limited to development of web services in Java
- Strong, senior level Java coding skills
- Solid understanding of web protocols and distributed application architectures
- Knowledge of SOAP, MTOM, WSDL, etc.
- Experience with Oracle technologies, especially relational databases
- Must be able to code with loose specifications.
- Familiar with service-oriented architectures
- Has documented experience with Agile-style approaches to software development
- Very strong communication skills
- Bachelor's degree in Computer Science, Management Information Systems, Business Administration or equivalent educational or professional experience and/or qualifications.

ANNEX C:

ANNEX C: QUALITY ASSURANCE CONSULTANT REQUIREMENTS

Scope of Work

The Business Plus Initiative (BPI) Project intends to hire an individual consultant to perform the duties delegated by BPI for the purposes performing quality assurance for the eTaxNG system (System).

The consultant will carry out the following tasks:

- Familiarise themselves with all technical matters relating to the System
- Gathers testing requirements and designs testing plans for all parts of the eTaxNG
- Schedules and executes testing plans
- Performs customer-facing tests, integration, system, performance, load, stress and security tests
- Works closely with software development team and GDT to identify and correct deviations from the product software specifications
- Mentors GDT staff on quality assurance best practices

Deliverables

- A quality assurance plan detailing how QA will be performed on all aspects of the eTaxNG
- On an as-requested basis by GDT and/or BPI, the development team will provide access to a working defect tracking system and summary reports on issues, defects, resolution of same, etc.
- A final quality assurance report detailing compliance of the software delivered with specifications.

Duration

Expected starting date of assignment is 1st of November 2011. Project is expected to last until 1 March 2013.

Required Qualifications

General Qualifications:

- 3+ years experience in Quality Assurance at the enterprise level
- In depth knowledge of QA test methodologies and strategies
- Ability to demonstrate a systematic approach for identifying and rendering complex test cases
- Expert understanding of the SDLC
- Experience in rendering and reviewing metrics for assessing readiness of product release
- Performance monitoring and analysis
- Excellent analytical skills and attention to detail
- Experience with service oriented architectures and web services
- Experience with Java
- Strong communication skills and is able to communicate effectively with developers, project managers and stakeholders
- Experience working in an agile-style approach to software development
- Internet security experience a plus
- Bachelor's degree in Computer Science, Management Information Systems, Business Administration or equivalent educational or professional experience and/or qualifications.

ANNEX D:

ANNEX D: TECHNICAL TEAM LEAD CONSULTANT REQUIREMENTS

Scope of Work

The Business Plus Initiative (BPI) Project intends to hire an individual consultant to perform the duties delegated by BPI for the purposes of acting as technical team lead for the eTaxNG system (System).

The consultant will carry out the following tasks:

- Familiarise themselves with all technical matters relating to the System
- Provides leadership for other developers on the team
- Responsible for all design and implementation decisions during the eTaxNG construction phase
- Works "hands-on" with other members of the development team to build eTaxNG
- Follows the eTaxNG architecture designed by the GDT and implements it
- Acts as a key resource on the development of web services in an SOA environment
- Follows GDT development standards
- Follows industry best practices
- Mentors GDT staff on development activities

Deliverables

- Production-quality code for eTaxNG

Duration

Expected starting date of assignment is 1st of November 2011. Project is expected to last until 1 March 2013.

Required Qualifications

General Qualifications:

- 5+ years developing web services
- Experience developing XML interfaces, creating schemas, pipelines, orchestrations, doing message routing, integrating web services
- Experience using service-oriented architectures including web services, SOAP, WSDL, WS-I Basic Profile, etc.
- Strong understanding of Internet application security, optimization, best practices and Internet protocols
- Database experience using Oracle or Microsoft SQL Server 2000/2005/2008
- Experience with Oracle Service Bus a plus
- Experience in writing code in Java
- Ability to learn new technologies very quickly
- Strong analytical and organizational skill
- Very strong communication skills
- Capable of mentoring junior staff in development best practices
- Bachelor's degree in Computer Science, Management Information Systems, Business Administration or equivalent educational or professional experience and/or qualifications.

ANNEX E:

ANNEX E: GDT WEB SERVICE CANDIDATE NOTES

Notes

Reporting period will only be the period of calculation.

Any rules, which can be locally checked, should be validated locally, for example, sums out of VAT invoices classified by income type, calculation of different items in VAT report etc.

Logical validation should be also at the end of accounting software. Even then, GDT system will check validation of whole report.

We need also some web services to handle uploading of bulk data properly.

An entity can submit its tax reports many times in the submission period, but the last one is the one to be counted on.

General web services

ValidateTaxPayerInfo – Real time. Accountant may want to check ALL information.

ValidateTaxPayerIDByName – Real time. If ID fits to name.

ValidateTaxPayerNameByID – Real time. If ID fits to name.

UpdateBusinessActivityCodes – Bulk. Just to update an integrated list of business activity codes if there have been some changes from GDT coding. As of now, GDT has no direct channel to State Registration Agency, but some works on that have been done and it may go live within next month. Even there will be direct connection to SRA, GDT (Batbayar) prefers to have this web service implemented.

I may better rename it to GetBusinessActivityCodes -> accounting software initiates a call to update business activity codes if there are some changes made in GDT system.

UploadTaxReport – Real time

RestartUploadAtPercentage – Bulk

AbortUpload – Real time

SendUploadProgressMessage

SendReceiptMessage – tax report is well received at GDT.

SendAcceptanceMessage – tax report is checked and accepted at GDT.

SendUploadReportFailedMessage – uploading of tax report is failed.

SendUploadBulkDataFailedMessage – uploading of bulk data is failed.

VAT web services

There are over 20K entities, which have actually to submit their VAT reports, of which

16.5K entities actually submit their VAT reports

5K entities submit X-VAT reports (X means no VAT payment)

around 11K have actually to attach VAT invoice list, of which

12K attach VAT invoice list.

Report TT-03A

ValidateVATPayerInfoByID

UploadVATReport – Real time – it will upload a complete VAT table, excluding correction items and invoice items.

VAT rules will be validated at the end of accounting software. So there is no need of validating each and any rule of VAT calculation.

SendVATUploadProgressMessage – Real time.

GetVATCalculationData – Real time. Data in section D of VAT report. Almost in each

GetVATInvoiceCalculationData – Real time. Data in section YE (E) in VAT report

SendMessageReceiptOfVATReport

SendMessageAcceptedOrNotAcceptedVATReport

There will be some business process implemented in the back-end of GDT system about the receipt of VAT report saying all formal requirements are fulfilled and about the acceptance of VAT report after it has been checked and accepted by tax officer.

UploadVATSellCorrectionInvoiceItems – Bulk

UploadVATPurchaseCorrectionInvoiceItems – Bulk

UploadVATInvoiceItems – Bulk

GetVATInvoiceNumbers – Bulk.

CheckVATInvoiceNumbers –

SendBadVATNumberErrorMessage – to inform wrong numbers of VAT invoices

Because whole content of VAT report is based on VAT invoices, the list of VAT invoices should be at first checked on their correctness and then the calculation and submission can start.

Report TT-03B

ValidateVATPayerInfoByID

PushTotalPurchaseAmountFromNonResident – Real time

UploadCountryNamesofNonResident – Bulk

SendMessageReceiptOfVATReportOfNonVATPayer

SendMessageAcceptedOrNotAcceptedVATReportOfNonVATPayer

Personal Income Tax web services

Report TT-11(1)

Upload11_1PITItems

Report TT-11(2)

Upload11_2PITItems

Report TT-12(1)

Upload12_1PITItems

Report TT-12

Uploat12WithholdingReport

Get12CalculationData

Report TT-11

Uploat11WithholdingReport

Corporate Income Tax web services

Report TT-02

ValidateTaxPayerInfoByID – another option is to deliver a database of all related information of all tax payers with accounting software

Upload02TaxReport

Get02CalculationData

Report TT-02 (Customer of Department of State Budget Income, Monitoring)

ValidateAddressByTaxPayerID – For GDT it is important to receive right code/numbers for address parts. So GDT proposes to deliver a small database of coding data with accounting software to end customers. After some time, end customers may need to update any alteration in the coding of address parts. Other data of 1-17 need to be filled in by accountant.

UpdateAddressCoding – Bulk data. If we handle address coding otherwise, then this web service won't be needed.

ValidateParentCompanyByItsID – Real time. GDT has data on interrelation between legal entities.

Get02CalculationData

Report TT-02 Disclosure (Customer of Department of State Budget Income, Monitoring)

UploadTotalIncomeByOperationalType – Real time

UploadWithholdingByOthers – Bulk

UploadCOSGOfIndustrialCompany – Real time

UploadCOSGOfTradeCompany – Real time

UploadCostExcludedOutOfTaxableIncome – Real time

UploadCostNotToExcludeOutOfTaxableIncome – Real time

UploadShareholdersInfo – Bulk

UploadSubsidiariesInfo – Real time or Bulk

UploadAmountOfTransactionsBtwInterrelatedEntities – Real time

UploadDetailedInfoTransactionsBtwInterrelatedEntities – Bulk – data will be separated by each company in interrelation

UploadBarterAndPayableTransactionsBtwInterrelatedEntities – Real time

SubmitAdditionalInfo74

UploadDetailedCostOfExplorationExploitation – Bulk or Real time

UploadDetailedTaxInformationOfNonResident – Bulk

All real time web services can merge into one single real time web service called Upload02Report. This service will be called after all bulk data is been uploaded.