Data, Visualization and Analytics

A Strategy for Business Intelligence

Executive Summary

The intent of this document is to provide a clear strategy and framework for various ideas and concepts under the wide umbrella of "Business Intelligence" (BI). This document describes how the three BI components of ETL (data extraction, transformation and loading), visualization, and analytics need to fit together and work for the City of Albuquerque over the next three to five years.

In order to arrive at a realistic strategy, we need to keep three underlying questions in mind:

1. What is the vision of a BI strategy in relation to the City of Albuquerque?
2. What is the optimal toolset required?
3. How should these tools be deployed to support an overall BI vision?

BI in the context of communications may take on some or all of 4 distinct stereotypes:

- Government as platform.
- Government as communicator.
- Government as sensemaker.
- Government as actionmaker.

Instead of attempting to find one single product to match these constraints, our proposed strategy is to match the 3 main functional areas of ETL, Visualization, and Analytics. We noted that there are sound business reasons for adopting this approach. Although it is likely that these reasons more than outweigh the savings that can be made from adopting a strategy centered around a single product, we must be careful to ensure that a justifiable total cost of ownership can be ascertained before proceeding.

Based on our experience with the FileNet replacement project, IPRA, Volunteerism, and Web Governance, we suggest that, as an initial step, a working group representing the BI community of practice within the City of Albuquerque be formed to take ownership of this effort.
Once formed, this leadership group would then:

- Ensure the validity of the overall strategy, making any necessary amendments.
- Split into working groups based around ETL/API, Visualization, and Analytics to further refine

requirements and establish possible solutions. In particular, these groups would decide on the level of expectations of internal and external users and, as a result, the extent that the two group overlap from an architectural standpoint.

- Identify funding opportunities.
- Recommend a sequence of implementation.

We should note that, once the overall strategy has been validated by the group, the working groups and subsequent requirement analysis can take place in parallel (subject to resourcing), before being resequenced again for implementation.

Data, Visualization and Analytics

A Strategy for Business Intelligence

Introduction

The intent of this document is to provide a clear strategy and framework for various ideas and concepts under the wide umbrella of "Business Intelligence." We can define Business Intelligence (referred to as "BI") as "the set of techniques and tools for the transformation of raw data into meaningful and useful information for business analysis purposes."[1] To reduce the scope even further, this discussion will focus on three areas of BI:

- Extraction, Transformation and Loading of data (commonly referred to as ETL). To this, we will also add the functionality of allowing end users to access raw datasets in machine readable format.
- Visualization of data (e.g. through graphs, charts and other visualization techniques).
- Analytics tools.

This document describes how these three BI components need to fit together and work for the City of Albuquerque over the next three to five years. In order to arrive at a realistic strategy, we need to keep three underlying questions in mind:

1. What is the vision of a BI strategy in relation to the City of Albuquerque?
2. What is the optimal toolset required?
3. How should these tools be deployed to support an overall BI vision?

Contextual Background

Data Sources within the City

As a municipal government with a history spanning over 300 years and serving a population estimated in 2014 at 557,169[2] (metro area of 907,679[3]), the City of Albuquerque has many places where data is stored (repositories), many uses for the data it produces, and many stakeholders who rely on City data. Although a full catalog of data and repositories is beyond the scope of this discussion, we can create an

1http://en.wikipedia.org/wiki/Business_intelligence
2http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF
3http://www.abq.org/Demographics.aspx

abstract map of significant data repositories and show data flows that will be important later.

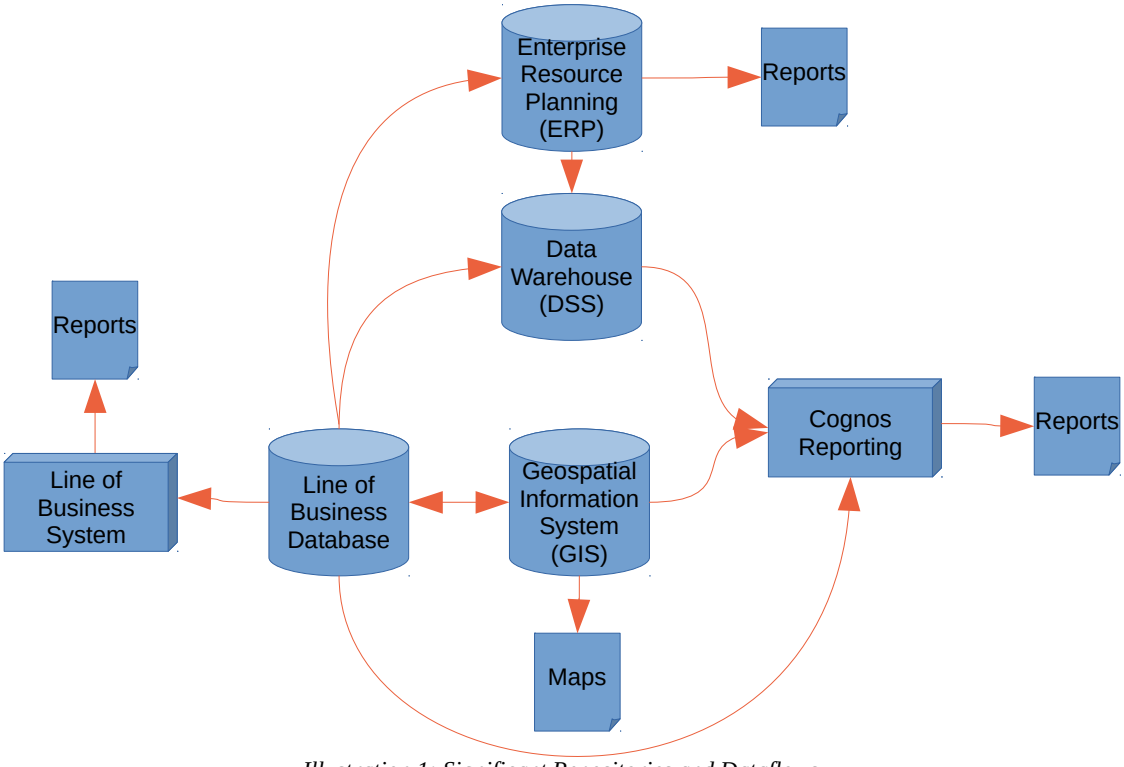

Illustration 1: Significant Repositories and Dataflows

We should note at this point that the Data Warehouse (known as within the City as "DSS") functions as an enterprise-wide data repository and a means of aggregating data for further analysis and reporting. It is important because it is often considered as the "go to" repository for the majority of internal financial reporting.

Emerging Roles of Government

To truly understand the context in which any BI strategy will need to exist and be successful within the City of Albuquerque, we need to consider four emerging roles of government:

- Government as platform.
- Government as communicator.
- Government as sensemaker.
- Government as actionmaker.

Each of these roles has a different level and kind of interpretation associated with it. Government might take on these roles at different times – and for different reasons. In turn, this requires different

kinds of BI tools – sometimes working on their own and sometimes working in conjunction with other tools and communication forms.

Government as Platform

Since its publication in 2010, Tim O'Reilly's articulate vision of "government as platform"[4] has become a compelling and popular framework that defines how government can remain relevant and re-engage with citizens. Part of the power in this vision is that it allows government to focus on one of the services that it does best: producing data for use by others. In the government as platform role, there is typically minimal interpretation or presentation – the focus is, instead, on the raw data itself. It is almost as if government is a sensor generating a stream of data for use and interpretation by others.

Government as Communicator

Of course, we know that government isn't simply about producing data. It also has an important role as a communicator. But "communication" in this context is not simply about broadcasting news. Instead it ensures that the communication can be provided in a format that can be easily consumed by the recipient. Providing a raw JSON data-stream, for example, might not be a suitable channel for communicating when advertising senior center events to a demographic aged over 55. However, making the identical events available as an attractive web page for the same audience would be considered as good communication.

Government as Sensemaker

Alongside the role of communicator is a frequently hidden (but no less vital) role of "sensemaker" for the community. The process of sensemaking can be defined as "the process by which people give meaning to experience"[5] and as the "collaborative process of creating shared awareness and understanding out of different individuals' perspectives and varied interests."[5] The sensemaker, therefore, helps stakeholders understand and make sense of the environment and issues around them. Whereas government as platform requires raw data with no interpretation, government as sensemaker is the reverse, with minimal raw data and the emphasis on interpretation.

Although community issues might appear on the surface to be obvious and might connect on an emotional level with the audience, the actual hard, objective data supporting the arguments involved can quickly become highly abstract and technical in nature and only truly understood by a minority. As a result, the challenge is to present and express issues in ways that inspire and lead instead of inflame and divide.

4http://chimera.labs.oreilly.com/books/1234000000774/ch02.html
5http://en.wikipedia.org/wiki/Sensemaking

Government as Actionmaker

The final role to discuss is that of actionmaker. Whereas the role of sensemaker is to present issues for others to make decisions, the role of the actionmaker uses analytics to implement and monitor outcomes of these decisions. Thus, typically, the actionmaker role involves monitoring and reporting metrics as they relate to a particular process.

If the time to issue a building permit, for instance, was considered to be a community metric, then an actionmaker role would be to:

- Define the steps required to issue a building permit.
- Provide baseline metrics.
- Provide metrics on a continual basis that show change in relation to the baseline.

Any measurable deviance, therefore, would prompt further action (e.g. remedial analysis) to either bring the process back into compliance or gain further understanding as to how the process can be improved.

Technical Trends

There are a number of technical trends of which we will need to be aware. This section provides a brief discussion of each trend at a very high level. Links are provided for further reading.

From Data to Services

The first trend to consider is that "open data" is undergoing a transformation from providing open data to also providing open services. Retrieving open data can be considered as a simple request for data, followed by a response from the server (usually retrieved data, but also potentially an error of some kind). Open services, however, imply some kind of transactional other than retrieving data: I log a call, I request and pay for a permit, and so on. Often, these transactions are multi-step and cross different services and repositories.

Making services open and freely accessible by any application has the advantage of:

- Lowering the cost of service delivery for the City.
- Enabling innovation.
- Allowing greater participation.

The move from data to services is an important consideration for BI. While we do not expect BI to become the major hub for the implementation of these services, it nevertheless has an important role in

either supporting (as part of additional functionality), or measuring them as part of reporting or analytics requirements.

Application Programming Interface

Application Programming Interfaces[6] (APIs) will continue to replace static files as the main way in which data is delivered from a repository. Here's why:

- **More granular security** – data can be filtered based on the content of that data instead of simply whether a file is accessible or not.
- **Smaller download** – being able to query and customize the data ensures that only the minimal data is returned instead of simply all rows regardless of whether they are needed.
- **More customizable** – many APIs allow developers to customize their queries so that they only get the data that they need.

The implication for a BI strategy is that it will no longer be sufficient (or even desirable) to depend solely on static file systems for a content delivery mechanism. The City has already begun to move in this direction through its ESRI's GeoServer[7], its implementation of SeeClickFix (which required the PeopleSoft CRM repository to publish an API that was compliant with the Open311 specification[8]) and, to a very limited extent, its Cognos[9] infrastructure. We should note at this point that although we anticipate that the City's investment in Geographical Information Systems (GIS) infrastructure will still continue to grow and show a positive ROI, it is possible that Cognos may no longer be the product of choice for the City.

Linked Data/Semantic Web

The increase in the number and kinds of data being published on the web means that we somehow need to ensure that we have consistent meanings across different datasets. Salary, for example, should have a consistent definition (whether hourly rate, hourly rate before deductions, hourly rate plus overtime or some other definition) across different reports within an organization (internal consistency) and when compared to other entities (external consistency).

Key technologies to enforce this kind of consistency fall under the collective heading of “semantic

6http://en.wikipedia.org/wiki/Application_programming_interface
7http://docs.geoserver.org/stable/en/user/data/database/arcsde.html
8http://www.open311.org
9http://www-01.ibm.com/software/analytics/cognos/

web”[10] or “linked data”[11]. They currently include RDF[12], OWL[13] and SPARQL[14]. Although many of these technologies may have been around for a while and received considerable market hype on introduction, they are now gaining more acceptance as good use cases and supporting tools emerge.

From Data to Knowledge

When deciding what datasets to release, one natural temptation is to simply release all data (also known as “pump and dump”). In one sense this is admirable as it is ultimate implementation of government of platform. However, it is also becoming clear that truly successful open data initiatives require an understanding and expert interpretation of the kinds of data that are being requested. More often than not, one of the key questions that developers and researchers now often try to answer is “how can I make things better for my community?” These questions are more abstract and therefore usually require a higher degree of analytic capability than can be provided by basic ETL functionality.

The traditional view of data is one in which data is passive and, largely, fixed and set in stone once delivered. The amount of pollen on a certain day, for example, might have been considered as unchangeable fact[15]. We are now seeing this view challenged as data is increasingly perceived as malleable and subject to change. The pollen values depends on the different kinds of pollen – and the ratio of species will change according to season and climatic conditions. So I might now see a high reading for pollen and ask “compared to when?” This then might lead me to a historical analysis of the data instead of an unquestioning assumption about the data which blocks my curiosity.

The implication for any BI strategy, therefore, is that it must be able to:

- Answer the next question – if we present a total amount, for example, we must also be prepared for when the user asks for a breakdown of that total.
- Help and encourage the user to set up and test self-guided hypotheses. In our pollen example, a time-based approach would certainly be an early next step as an obvious hypothesis could be along the lines of “the pollen ordinances made a big change to the amount of pollen over time”.

Bifurcation of Datasets

As the City makes more and more open data available, two interesting problems have emerged:

10http://www.w3.org/standards/semanticweb/
11http://linkeddata.org
12http://www.w3.org/standards/techs/rdf#w3c_all
13http://www.w3.org/standards/techs/owl#w3c_all
14http://www.w3.org/standards/techs/sparql#w3c_all
15For a good discussion on the background epistemology behind this concept, see ARBESMAN S., 2013, “The Half-Life of Facts: Why Everything We Know Has an Expiration Date”, 2nd Edition, Penguin Books

1. The datasets are growing larger over time (pollen data, for example, now goes back to 2007). This has given rise to situations where download performance to retrieve all rows for a dataset has grown to unacceptable levels.
2. Users are requesting variations on a given dataset to match needs. It is important to stress that this is usually not because of "laziness" on the part of developers. The crime dataset, for example, is currently more than 330,000 rows for a 180 day period. A developer writing an app for a resource-constrained device such as a smartphone would be justified in requesting data that imposes less of an overhead in terms of bandwidth consumed, memory required and time taken to process.

On the one hand some users are requesting more data while, on the other, we see other users requesting less. Who is right? The answer, of course, is that they both are. Our BI strategy, therefore, needs to ensure that:

- End users can filter raw data where it makes sense.
- Data content repositories can handle large amounts of data (whether stored internally or accessed by reference to another repository).
- Different access strategies will be needed to ensure that data is delivered to users as quickly as possible. One example strategy could involve archiving to a different dataset or repository by date. This further implies that a more robust life cycle assessment will have to be considered for each dataset, and it is likely that the specific details (e.g. monthly archiving or yearly archiving) will vary based on the dataset under consideration and the expectations of the user community.

Improved Standardization of Capabilities

A final trend that we should anticipate is that, as services begin to gain in popularity, municipalities will begin to replicate and offer these services themselves. GTFS[16], LIVES[17] and OpenTrails[18] are all examples where this has happened.

The implications for a City strategy are twofold:

1. The City will, in some situations, act as a leader in defining and implementing standards for new services.
2. The City will, in other situations, act as a follower in implementing services that were

16https://developers.google.com/transit/gtfs/
17http://www.yelp.com/healthscores
18http://www.codeforamerica.org/specifications/trails/

developed and specified by others.

City BI tools, therefore, will need to not only be flexible enough to enable new specifications to be developed, but also be able to quickly implement services specified by others.

Use Cases

Reviewing the different stakeholders who will use any BI system is critical to the success of any strategy. Before we do so, we need to make one assumption quite clear: the success of a public service implemented by the City depends on making sure that it is designed from the perspective of the citizen. This means that we must not:

- Skip service and process design all together.
- Design from the perspective of what is easier for the City.
- Expect the citizen to have a detailed knowledge of City organization and culture.

The same thinking is true when we consider internal users: the success of an internal service implemented by the City will depend on making sure that it is designed from the perspective of the end user. This means that we must not:

- Skip service and process design all together.
- Design from the perspective of what is easier from a technology to accomplish (unless deploying quickly is one of the chief considerations).
- Expect the user to have a detailed knowledge of technology.

We will now list each stakeholder and examine their capabilities.

Stakeholders

Time constraints have prevented a full, and empirical analysis. Instead, a list of stakeholders was drawn up from discussion and experience and classified into different user groups. An estimate was made for each user group as to how much use was made of BI services ("Estimated Percentage Consumption"). Each user group was then given a score out of 100 based on whether experience indicated that they were very engaged and passionate (100) or disengaged (0).

The same process was repeated to estimate the technical ability of a group. The results are shown in Table 1, and in Illustration 2, below. We see that that regular users are estimated to be the biggest consumer of BI services, followed by internal operations (a finding backed anecdotally in Code for America discussions). This analysis also highlights the fact that although the media is an important and often vocal group, it is also only the third-largest group and is a smaller group than the citizens that it

represents.

User Group	Estimated Percentage Consumption	Estimated Engagement Score (0 = none, 100 = max)	Estimated Technical Ability Score (0 = none, 100 = max)
Citizen Users	28	70	40
Internal Operations	19	80	80
Media	15	80	60
Commercial	15	75	60
Community Advocates	11	85	50
Elected Officials	5	60	30
Educational	5	60	70
Other Cities	2	40	50
Total	100		

Table 1: BI Stakeholder Summary

Although this is a non-scientific analysis, some of the findings have also been echoed by others – but there is clearly more work that needs to be done. In the meantime, Illustration 3 shows the estimated ability and estimated engagement for each user group. We can see that, overall, the estimated engagement is usually greater than the estimated ability. In other words, although people might be engaged and passionate about data, they do not usually have the ability to use complex tools to manipulate the data to meet their needs.

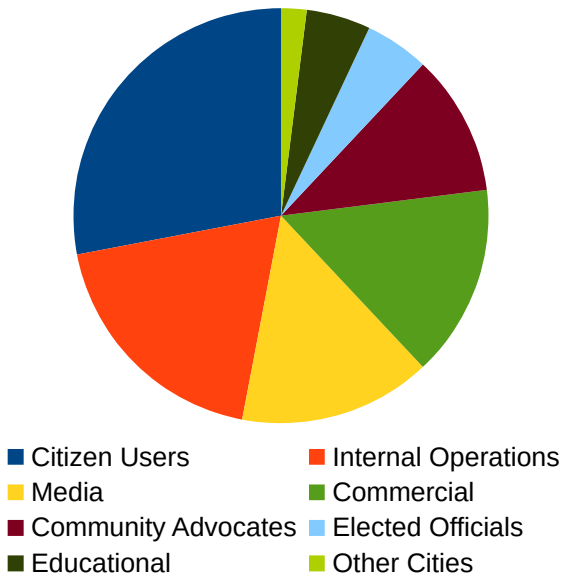

Illustration 2: Estimated Percentage Consumption by Different BI User Groups

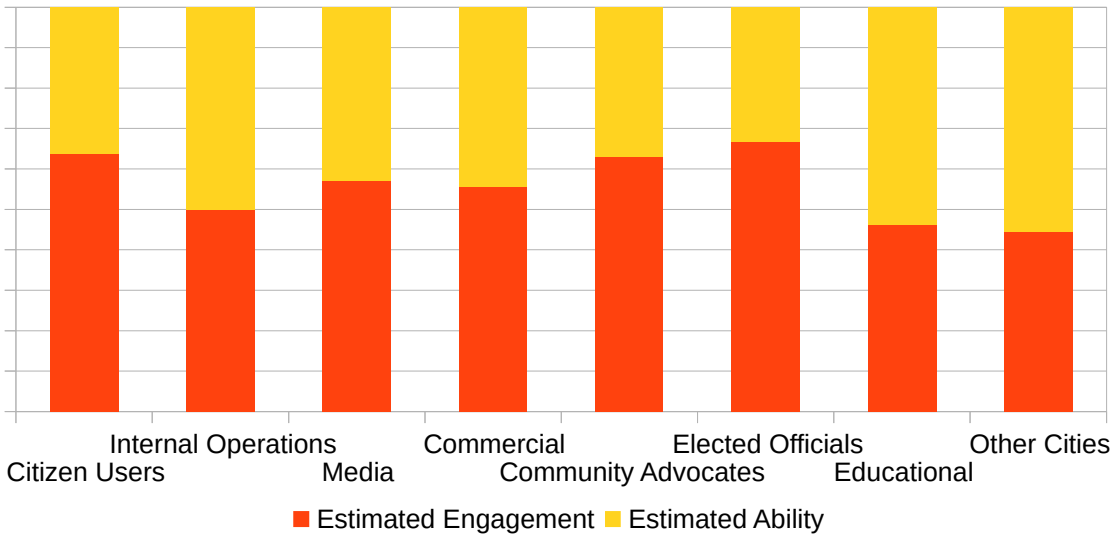

Illustration 3: Estimated Engagement and Estimated Ability

Even though this is an unscientific study, our BI strategy needs to note that:

- Any tools intended for use by end users will need to be very intuitive as most of the users that we encounter are outside the City organization. This means that, at a minimum, training

delivery will be complex.

- Estimated engagement outstrips ability in most categories. This means that, more often than not, tools will need to "save users from themselves". Although somewhat tongue-in-cheek, we must still be concerned about problems that can quickly arise from:
  - Misunderstanding tool capabilities and limitations.
  - Misunderstanding data (e.g. definitions, assumptions and limitations).
  - Overzealous application of decoration at the expense of the data (also referred to as "Chart Junk"[19]).
- We must support a wide range of users with differing abilities and levels of engagement. Most of them do not have time to figure out the complexity in obtaining or understanding data. We must respect their choice.
- Internal and external users are separate and may require different support from the same data repository. Three examples will provide clarity:
  1. The ABQ-View transparency portal and the ABQ-Data open data portal serve different external needs, but rely on much the same underlying data (e.g. salary information).
  2. The internal DSS data warehouse and the external ABQ-Data open data portal have similar technical requirements, but differing expectations on data privacy and sensitivity.
  3. The GIS system extracts crime data from an internal architecture and publishes it as an external REST[20] endpoint.

19http://en.wikipedia.org/wiki/Chartjunk
20http://en.wikipedia.org/wiki/Representational_state_transfer

A Blended BI Strategy

Having examined the context for our BI strategy, it is now time to discuss the strategy itself. For purposes of clarity, we will firstly discuss an overall strategy. We will then discuss any specific strategies for data Extraction, Transformation and Loading (ETL), Visualization and Analytics separately.

Overall Considerations

It is clear from our discussion so far that our needs for BI tools are complex and cover a great range of requirements – from simple report creation to complex analytics and powerful ETL and API requirements. With the additional requirements of rapid response to business needs, we can see also that meeting these needs in a single product is going to require some trade off and compromise because it is doubtful that a single product will meet all of these needs perfectly.

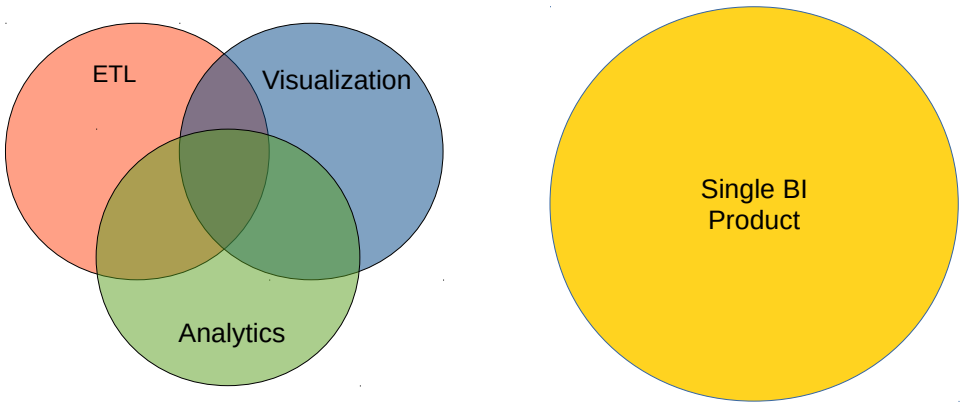

Illustration 4: Comparison of Portfolio-Based vs Single Product Architectures for BI

Instead of a single monolithic BI system, our strategy is built up into a "portfolio" of different systems for each of the groups within BI: ETL (including API access and repository), Visualization, and Analytics. It is important to note that although we are not excluding the selection of a single solution for the whole BI strategy, we anticipate that a "component-based" approach has the following benefits:

- Best of breed choices can be made for each component.
- Better phasing for project management.
- Flexibility in phasing may translate to more funding options.
- User training can be more modular and constructed to match need more effectively – a citizen, for example, should not be made to go through training on analytics and ETL if they only need

to download a report.

- Diversification of risk.

Validation of this approach can be seen in the way that the Pentaho[21] BI suite have been similarly split into separate product lines under the main product banner.

It could be argued that there are substantial savings of scale that could be made if a single product was selected. While this is certainly a risk to some degree, the exact amount of savings is unclear. In particular, we need to ensure that the total cost of ownership (TCO) for the product as a whole (including licensing, maintenance, training, hardware, IT resourcing, business unit resourcing etc) is lower for a single product than for a portfolio-based strategy.

More importantly, a strategy built around a single product is, necessarily, a strategy of compromise. A portfolio-based approach, by contrast, allows us to select the best tool for the need. Although there will always be some degree of overlap (see Illustration 4, above), it is far from clear as to whether the costs incurred by this redundancy exceed the savings made by a single product strategy.

Data Extraction Transfer and Loading (ETL) Operations

Basic ETL Tools

In order to deliver on concepts such as open data and open services, we still need to consider a good mechanism to ensure the basic, unattractive tasks of extracting, transforming and loading data. Basic ETL, in some respects, may be the easiest component to implement – albeit if we assume that, even here, we will not limit ourselves to a single solution. Although there is a wide range of ETL products out (such as Refine[22], Talend[23], KETL[24] and JasperSoft[25], Pentaho Data Integrator[26] and Scriptella[27]), it is also possible that sufficient functionality already exists in a combination of Python[28] and (for example) the Django framework[29] to meet the needs of the City in extracting and transforming data. No further costs (other than training) would be required.

21http://en.wikipedia.org/wiki/Pentaho
22http://openrefine.org
23http://www.talend.com/resource/free-etl.html
24http://www.ketl.org
25http://www.jaspersoft.com
26http://community.pentaho.com
27http://scriptella.javaforge.com
28https://www.python.org
29https://www.djangoproject.com

Data Repository and API Tools

However, these ETL tools are really “back office” tools and are not intended to act as a front end user interface or API access to a data repository. API's for general data repositories (excluding geospatial data) appear to be coalescing around two main API platforms: CKAN[30] (or its DKAN[31] derivative) and Socrata's SODA[32]. At this point, we remain neutral and agnostic as to which of these platforms should be used provided that:

- An ROI can be demonstrated (including total cost of ownership).
- Data can also be referenced by the repository (i.e. residing in another repository, but referenced as a view or a filter) as well as held within the repository.
- Datasets can be filtered by users in realtime by additional parameters (e.g. using the REST transaction model).
- Data updates can be scheduled automatically.
- DCAT[33]-compliant catalogs of the datasets themselves are generated on demand.
- Potentially separate repositories to store City data for internal uses (e.g. within the context of DSS) and external use (open data). These have separate needs, requirements and expectations. (see use case discussions, above).

There is a current trend for open data repositories to be built quickly using the github[34] version control system. However, we will not recommend this as a strategy for the City due to the current lack of API support.

We should also note at this point that the City's investment in GIS infrastructure will continue to play a critical part in this strategy going forward – both as a final repository for data and also as a source location for data. It is also important to realize that GIS services will play an increasingly important role in ETL operations (e.g. moving data from one repository to another).

Visualization

Based on our analysis of users and their expectations, we can again see that it is unlikely that a single solution exists that:

30http://ckan.org
31http://nucivic.com/dkan/
32http://dev.socrata.com/consumers/getting-started.html
33http://www.w3.org/TR/vocab-dcat/
34https://github.com

- Is easy to for non-technical end users to master.
- Allows inline embedding into regular web pages using the “Sparklines” concept described by Tufte[35].
- Allows advanced visualization[36] techniques such as word clouds, network graphs and heatmaps to be used in web pages, dashboards and presentation tools[37].
- Allows realtime updates as the underlying data changes.

Accordingly, to meet these needs, we will need the following classes of tools:

- **Web-centered charting library** such as Google Charts[38], Chart.js[39], D3.js[40] and Flot[41] that can be quickly embedded and can update in realtime. Data for these charts must be available via the open data repository (whether through API access or CSV static files). This allows other users to create their own visualizations from the same data for purposes of validation or further research.
- **Web-centered geospatial tools** for online maps. At this point, we assume that Google Maps[42] will continue to be used by the City for quick deployment, relatively simple maps, while the City's investment in ESRI and Latitude Geographics[43] will suffice for more complex projects.
- **Strong visualization tools** such as Tableau, Visual.ly[44], Fusion Charts[45], Dipity[46], ManyEyes[47] and Processing[48]. These tools are intended for more expert users with strong design skills. They may allow realtime updates, but are intended for larger, more complex projects requiring a greater degree of design input. We should also not discount the utility of standalone apps such

35http://en.wikipedia.org/wiki/Sparkline
36http://en.wikipedia.org/wiki/Data_visualization
37Powerful examples of this include Minard's map showing the progress of Napoleon's attempt to invade Russia (http://www.martingrandjean.ch/historical-data-visualization-minard-map/)
38https://developers.google.com/chart/
39http://www.chartjs.org
40http://d3js.org
41http://www.flotcharts.org
42https://maps.google.com
43http://www.latitudegeo.com
44http://create.visual.ly
45http://www.fusioncharts.com
46http://www.dipity.com
47http://www-969.ibm.com/software/analytics/manyeyes/
48https://www.processing.org

as Microsft Office[49] and LibreOffice[50] to also create these kinds of diagrams.
It also bears repeating that if the data is intended for public display, good data science practice would ensure that the data is also available for all to use.

- **Dashboarding tools** such as D3.js[40], Crossfire[51] and Socrata[52], InformationBuilders[53] and iDashboards[54]. These tools may require some time and technical expertise to plan and integrate but, once in place, should be fairly low maintenance. This is because the kind of data will not change even when data values do. Some of these tools (e.g. D3) require a greater level of coding skills, whereas Socrata is aimed at business users wishing to implement "dashboard" style management metrics.

Analytics

Our final category under consideration is business analytics. We have already discussed how the role of business analytics within government is changing to ask deeper questions about process. As this transformation continues, BI will need to support these kinds of queries. The market space for business analytics tools is mature, but undergoing somewhat of an expansion in the government space as some of the concepts are re-invented and re-purposed with government in mind. Possible alternatives for tooling include LogiAnalytics[55], Cognos[59], InformationBuilders[53], Socrata[52], KNIME Analytics[56], Google Fusion Tables[57] and RapidMinder[58]. Again, we recognize and support the utility and convenience of "solvers" found most advanced spreadsheet packages.

Specific strategy requirements for analytics include:

- Support of ad-hoc queries and reporting. In the case of citizens, this will need to be web-based.
- Expert functionality such as scenario or goal solvers.
- Statistical capabilities.
- Support for data visualization – either directly through a range of graphs and charts, or indirectly by integrating with a 3rd party product.

49https://products.office.com/en-US/
50https://www.libreoffice.org
51http://square.github.io/crossfilter/
52http://www.socrata.com
53http://www.informationbuilders.com
54http://www.idashboards.com
55http://www.logianalytics.com
56http://www.knime.org/knime
57https://support.google.com/fusiontables/answer/2571232
58https://rapidminer.com

- The ability to link the workflow of a process with outcomes and measures.
- Separate repositories for internal and external use (see discussion on ETL API repositories and use case discussions, above).

We recognize that much of the functionality required of an analytics systems exists at the intersection of the Visualization and ETL operations (see Illustration 4, above). Of all the components described within this strategy, Analytics may be the one that can be easily merged within solutions for the other components. However, a prudent analysis realizes that growing communities both inside and outside the City who are interested and willing to experiment with data (e.g. setting up and testing hypotheses) will be to the greater benefit of the Albuquerque community at large.

Conclusions

This document has covered a huge amount of ground within a few short pages. We will begin to draw all the parts together by firstly summarizing some of the key points of our discussion so far:

- We defined BI as consisting of three main areas:
  - ETL (including open data functionality) – required to get the data into a repository in a convenient format.
  - Visualization – required to provide end user visualization.
  - Analytics – required to dig deeply into operational data.
- We need to support a wide range of users with a wide range of expectations, engagement and abilities. This means that, at a minimum, we need to be able to handle internal and external communities.
- BI in the context of communications may take on some or all of 4 distinct stereotypes:
  - Government as platform.
  - Government as communicator.
  - Government as sensemaker.
  - Government as actionmaker.

Instead of attempting to find one single product to match these constraints, our proposed strategy is to match the 3 main functional areas of ETL, Visualization, and Analytics. We noted that there are sound business reasons for adopting this approach. Although it is likely that these reasons more than outweigh the savings that can be made from adopting a strategy centered around a single product, we

must be careful to ensure that a justifiable TCO can be ascertained before proceeding.

We have also been extremely careful to avoid prescribing specific products at this stage. Instead, we have simply described the conditions that a product needs to fulfill. This not only allows our strategy to be more inclusive given the wide range of products (known and unknown) that may meet our needs but, more importantly, allows this strategy to remain relevant as products and business models change. We anticipate that multiple procurements will be made underneath the overall BI strategy.

Initial Thoughts Regarding Prioritization and Next Steps

Our final question is to consider how a portfolio strategy can be implemented. After all, a strategy is next to useless if there are constraints that prevent it from being used. Based on our experience with the FileNet replacement project, IPRA, Volunteerism, and Web Governance, we suggest that, as an initial step, a working group representing the BI community of practice within the City of Albuquerque be formed to take ownership of this effort.

Once formed, this leadership group would then:

- Ensure the validity of the overall strategy, making any necessary amendments.
- Split into working groups based around ETL/API, Visualization, and Analytics to further refine requirements and establish possible solutions. In particular, these groups would decide on the level of expectations of internal and external users and, as a result, the extent that the two group overlap from an architectural standpoint.
- Identify funding opportunities.
- Recommend a sequence of implementation.

We should note that, once the overall strategy has been validated by the group, the working groups and subsequent requirement analysis can take place in parallel (subject to resourcing), before being resequenced again for implementation.