Paper PO-26

Relevant Data Intelligence and Enterprise Analytics: Grading Decision Support Based Upon Enterprise Data Maturity

Alan Mann, Albert Briggs, Emile Barnes
Deloitte Consulting, Arlington, VA

ABSTRACT

According to Gartner, "Analytics leverage data in a particular functional process (or application) to enable context-specific insight that is actionable." [1] The thesis of this poster seeks to define a maturity model that can be expressed as an enterprise's Relevant Data Intelligence, or RDI. Where does an analytic and data transformation product such as SAS® contributes to the overall build of an intelligent enterprise and lift the degree of RDI?

Can a functional driver of more accurate and reliable analytics be expressed as RDI within the enterprise, and continue to indicate more comprehensive, valuable analytics? Could the enterprise use RDI as a metric to gauge data maturity?

This poster seeks to chart an indicator of RDI over the maturity of data in an enterprise regarding sustainable insight through analytics based upon data governance, practical architecture, and sound integration. Inclusive in the treatment is a data maturity model essential to the basis of the poster's proposition of RDI. Research into data governance could provide an inferential measure of data maturity and analytic ability within enterprise data establishments.

.

DATA ANARCHY

At the very beginning of the Relevant Data Intelligence continuum organization's data resides in many disparate and unstructured sources, such as Sneakernets (i.e. O Drives, desktops, CD's) email's, spreadsheets, etc. In such an environment, the organization lacks real institutional control or knowledge of what data is being produced or accessed. We will refer to this stage of Relevant Data Intelligence (RDI) development as "Data Anarchy", which can be thought of as a "Wild Wild West" where no laws or structures exist to handle and manage data, which consists of unorganized facts and materials. In Data Anarchy organizations issue business questions, as they present themselves, to the select few who are able to answer them. This process or "fire drill" is reactive and task driven. Typically an issue or question will arise and a report or briefing must be put together in order to provide the information to answer it. The request for said report or presentation then falls on someone's desk and that individual is left to their own devices to respond.

Within the Data Anarchy model, data must be manually organized and structured in order to produce information or meaning. The transformation of data to actionable information is controlled by those who have the ability to find and access data (regardless of the source), and package its meaning to leadership. In this sense a person's value to the organization is their ability to provide information to the most esoteric of business questions. As a result, data becomes currency which should be hoarded and not shared. Organization then begin to breed segmented subject matter experts, who align themselves with their individual analytical tools of choice (i.e. SAS, SPSS, R, Excel) in the same manner a gunslinger forms a bond with his trusty pistol. Because data is shared on a need to know basis, the organization as a whole lacks a clear understanding of how information is gleaned from its data and must rely on the output from the handful of "gunslingers".

AWARENESS OF DATA GOVERNANCE

As organizations grow, the Awareness of Data Governance stage of the RDI Continuum emerges. At this level, organizations realize that their data is a key mechanism to proactively improve business processes. As John Lucker states: *"Companies have an opportunity to close the gap on inefficiencies and suboptimal processes. They're able to substantially boost their return on technology and business investments."*[2] No longer can the operation of gleaning valuable insights from data be designated to an esoteric few. If an organization wants to gain true value from its data,

[1] Jeremy Kirk, "Analytics' Buzzword Needs Careful Definition." Infoworld, February 7, 2006. Retrieved from http://www.infoworld.com/t/data-management/analytics-buzzword-needs-careful-definition-567.

[2] http://www.bestanalytics.net/2011/07/business-intelligence-and-analytics-grow-up-business-intelligence-news-reviews/

it must include the entire enterprise into the analytical process. However, such a mandate becomes very daunting. With the increasing collection of data from unstructured sources such as email, Blogs, Wiki's, Twitter, and traditional sources such as corporate transactional and operational systems, the scope of available data to the enterprise can be overwhelming. According to April Adams of Gartner *"data capacity on average in enterprise is growing at 40% to 60% year over year"*[3]. Lucker best summarizes the plight of organizations when he writes *"Data is available in ways that were unimaginable ……...The challenge is putting it to use."*[4]

As a result, this paper theorizes that the important concept of Awareness of Data Governance is not only the realization that its data is valuable, but also a plan must be put into place in order to best manage and execute against that data.

DATA MANAGEMENT

This Data Management stage of the RDI is when an organization develops a plan and strategy to formally organize its data. However, this process can be filled with many questions. What is the most efficient way to collect and organize data? How does an organization account for all its data particularly from unconventional unstructured sources?. *Griffin defines four steps to effective data management: migration, maintenance, quality control and governance*.[5] **Data migration** is the transferring of data from its native legacy sources (structured and unstructured) into a new platform. During the transfer process data is cleansed of redundancies and errors. SAS' own David Barkaway, believes the following;

> "the data migration process should be robust and resilient, meaning it should manage all aspects of the migration and be able to cope with high data volumes; execute efficiently by leveraging existing facilities; provide progress updates during the migration process…."[6].

Once data is transferred to new platform, **Data Maintenance** is updating the master data or important reference information regarding the business operations of the organization (i.e. customer data, office locations, vendors, etc) with changes that have occurred such as the addition of new customers or addition of more offices. **Data Quality Control** is the continuous process of ensuring the organization data is a correct, and consistent with no gaps. Quality Control determines the credibility of the data which will be used to provide analytical insights to the enterprise. Finally **Data Governance** represents the point in which an organization institutes formal policies and procedures to further ensure data security and quality. The Data Governance step of Data Management signifies the beginning of an organizations maturity toward the final stage of RDI, Formal Data Governance.

FORMAL DATA GOVERNANCE

The Formal Data Governance signifies the creation of the truly data centric organization. Griffin explains the importance of data governance:

> Without data, there is no governance, but effective governance goes way beyond bits and bytes. It calls for no less than a cultural shift away from thinking about data as a commodity toward thinking of data as one of the company's most valuable assets, and creating an organizational mindset of accountability. [7]

Griffin's definition elevates Data Governance beyond the administrative procedures of implementing.
Data management policies become part of the organization's culture, in which everyone is held accountable for establishing and upholding the integrity and security. No longer is information finding pushed the niche "gunslingers" with their analytical weapons of choice. In the Data Governance structure, all have a vested interest to ensure data is aligned to provide information which directly addresses with the business rules and needs of the organization. At the

[3] http://www.computerworld.com/s/article/9194283/Data_growth_remains_IT_s_biggest_challenge_Gartner_says
[4] http://www.bestanalytics.net/2011/07/business-intelligence-and-analytics-grow-up-business-intelligence-news-reviews/
[5] http://www.ehow.com/info_7936962_steps-data-management-process.html
[6] http://www.ehow.com/info_7936962_steps-data-management-process.html
[7] http://www.information-management.com/issues/14_6/data_governance_people_management_BI-10020270-1.html

ideal state of RDI, analytical tools and analysis are not the drivers of sound decision making. Instead they serve as the by- product of an organization's Data Governance structure which institutionalizes the accountability of data stewardship by everyone within the enterprise.

SCORING THE RDI

Central to good data governance is the need for a non-verbal, numeric representation of the degree of data governance expressed in the propensity of data to be adequate for certain levels of data analytics and administration. Master Data Management offers a maturity model, but is by necessity too general in its presentation of the capabilities of the organization to leverage existing data in providing measurements of data intelligence manifested in more accurate, directed analytics.

As previously stated, a 4 stage model to illustrate the 4 stages of data evolution reflective in the degree of relevant data intelligence (RDI) in an enterprise are:

1. **Data Anarchy**
2. **Emerging Data Awareness**
3. **Data Management**
4. **Formal Data Governance**

ASSUMPTIONS

There are 2 models of data architecture:

- Departmental, where data organization is different and cut off from other similar or dissimilar information management structures
- Enterprise, where data organization is cross-departmental and shared-purpose driven, expressed in the four architectural dimensions:
  - Data Integration
  - Analytics
  - Intelligent Data Services, such as the SAS SPd Server
  - Business Intelligence through interactive dashboards and reporting

The Enterprise model is assumed from Stage 2 forward, as silos fall.

- The point of reference is from the Analytics dimension.
- To achieve the best grade of Relevant Data Intelligence, an enterprise needs to score a 4; to achieve the minimal, an enterprise achieves a grade of 1. This differs from the 5 stage Master Data Management maturity model of Strategy, Organization, Process, Governance, and Technology in its application in the both the departmental model of data architecture, while Master data management recognizes its relevance in the enterprise data architecture.
- RDI seeks to exist in both the Departmental and Enterprise models, but higher grading would reflect enterprise-oriented information management..
- All enterprises start at 1, and if not dysfunctional, work toward 4.
- Data always exist in enterprises.
- Enterprises consist of functional groups.
  - Data anarchy may be unique to one or more functional groups
- Logical data models begin from Stage 2 forward.

DATA MANAGEMENT EVOLUTION SCORING, OR THE RELEVANT DATA INTELLIGENCE (RDI) SCORE PROPOSAL

SCORING ANARCHY

Data does not always begin chaotically. It may be organized in logical, subject-matter relevant classes, but if it falls short of organized coordination within the functional group, it is subject to chaos, which could be sudden loss, change due to unforeseen additions of data attributes by colleagues, and spontaneous ad-hoc changes. The medium of exchange may be over email and legitimate access to databases, but at the analyst's entry point, if most data gets distributed by thumb drives, the network is effectively foot-based, or, 'sneakernet.' Analytics are totally ad-hoc.

Other manifestations of anarchy have been previously listed, but in general, unevenly distributed tools and data, organizational coordination of data resources, and absence of consistent data sharing would be scored at an RDI of 1 on a 4 point scale.

SCORING NECESSITY, OR DATA AWARENESS

Necessity and pragmatic direction lead to organized management. However, while this stage is Emerging Data Awareness , data is still in silos, organized in repositories rather than relational, modeled databases, and leadership moves from the few to those in touch with the many and the business. Data becomes more global, but there is still a risk of unknown upgrades of data, followed by frequent crashes of the group's automated jobs. However, as gloomy as this seems, the renaissance within the functional group has started, with a push to organization. This stage would be scored and RDI of 2 out of 4.

DATA MANAGEMENT FROM NECESSITY

Once the organization moves toward formalized management, and it can be a difficult road, resources are dedicated toward organizing data resources, identified by organized transactional and master data tables signify more structured systems such as hierarchic and relational databases, with normalization progressing under the Boyce-Codd (BCNF) normal form model:

- When a relation has more than one candidate key, anomalies may result even though the relation is in 3NF.
- 3NF does not deal satisfactorily with the case of a relation with overlapping candidate keys i.e. composite candidate keys with at least one attribute in common.
- BCNF is based on the concept of a *determinant*.
- A determinant is any attribute (simple or composite) on which some other attribute is fully functionally dependent.
- A relation is in BCNF is, and only if, every determinant is a candidate key. [8]

Data awareness now becomes more apparent as data management. The data anarchists still groan, but find they can get more work done with a way to easily find data sources for research and analysis. More formalized organizational structures take hold, and while creating and posting tables becomes more constrained by rules, the audience has moved from divergent subject-matter experts to data management specialists. Discussions of formal data governance become part of the managerial and analytic conversation. Information infrastructure, logical and physical becomes very apparent. Data warehouses along the Kimball/Inmon [9]schools of design appear as demand for specialized, summary data increase across the enterprise, or within the workgroup. Review and auditing become cultural in good practice.

An RDI score of 3 out of 4 would be assigned to this stage.

MATURITY, OR FORMAL DATA GOVERNANCE

This may be impossible for many organizations within enterprises to achieve, due to relative autonomy of analytic projects, but in the interest of easier research and coordination of data, all forms of data, from transactional to historic/archival would be organized into relational and data warehouse/operational data store architectures. The major audience for data would be the entire enterprise, including the functional work group where the anarchy started. Master Data Management, comprised of master databases managing Customer Master, Item Master, and Account Master tables and databases coordinated across the organization would by necessity lower redundant records and guide data along more functional lines. While more rules on access and security are apparent, data is more easily retrieved and available for analytics, thus saving more time in research and development of analytic reports and visualizations. [10] More decision-making assets are directly tied to data rather than seat-of-the-pants analyses. Metadata management becomes more aligned with business rules in the functional groups comprising the enterprise.

---

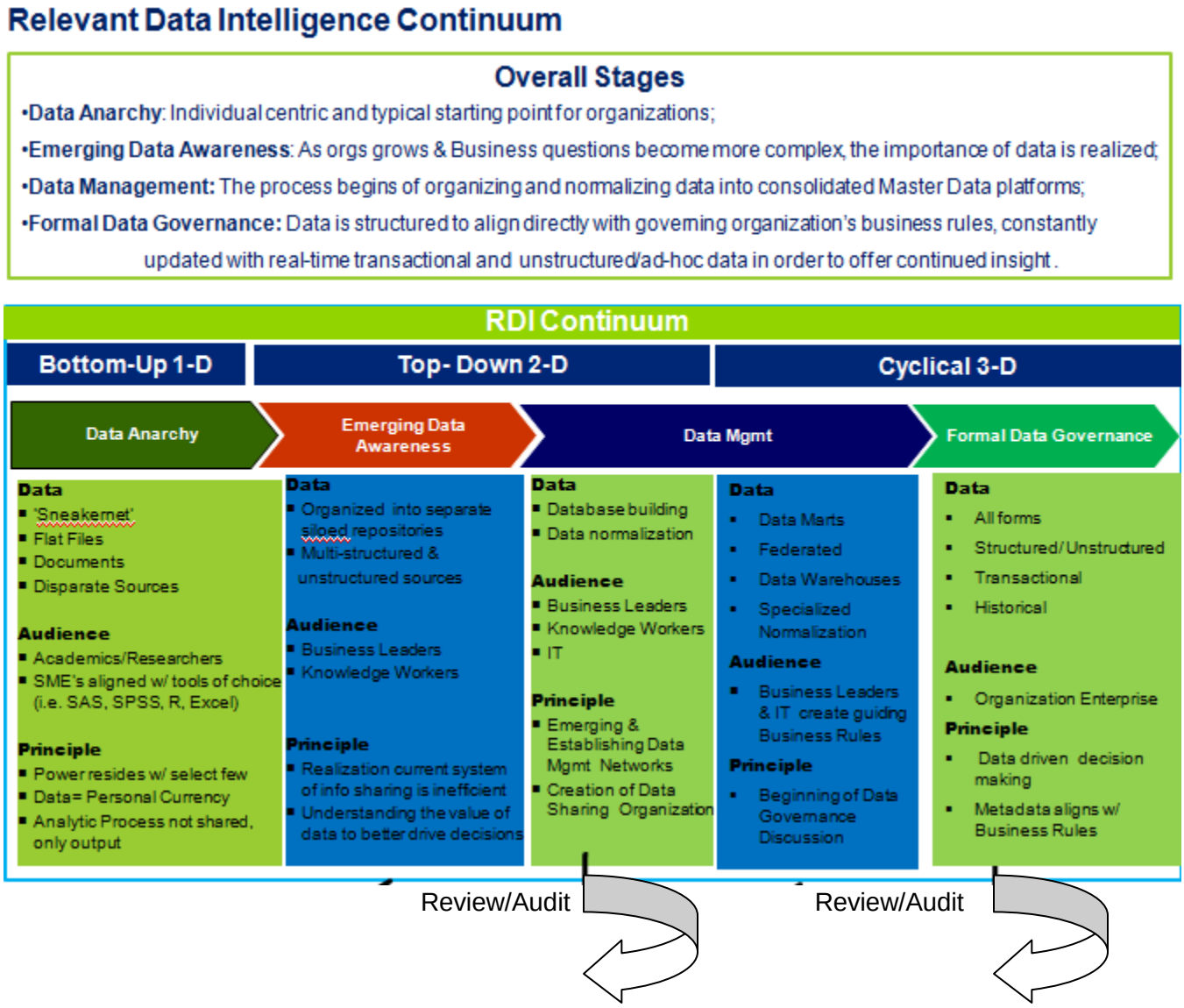
[8] Retrieved 31 August 2011 from: http://db.grussell.org/section009.html#_Toc67114457

[9] Retrieved 11 September 2011 http://www.intranetjournal.com/features/datawarehousing.html

[10] Retrieved 1 September 2011 from: http://msdn.microsoft.com/en-us/library/bb190163.aspx

A score of 4 out of 4 would be ideal, but realistically, a 3.5 would be excellent. Perfection should be an ideal, not a stopping point. 4/4 is therefore theoretically plausible but given that a factor of uncertainty and limited resources mark reality, this score would be unlikely. It would be something to work toward.

An illustration of the relevant Data Intelligence continuum follows:



CHAOTIC ANALYTICS VS. SOUND ANALYTICS: CONCLUDING REMARKS

Functional groups within enterprises enter their monthly numbers in spreadsheets, PDF files, PowerPoint slides, and compare them to last month's numbers, calling these comparisons Analytics. Often this process is cultural. Without any formal business analysis, functional groupsrs will be forced to rely on gut instincts for nearly anything analytic.

To move toward a data management/governance, a strategy needs to be developed on how to produce meaningful analytics. A framework needs to be established and criteria set for what outcomes will add value to the analysis, tempered by a timeline for completion. Additionally, the group should be organized in a team structure. The overall strategy should be open to change as analytic needs change.

Along the lines of Business Process Management (BPM), change needs senior management approval and involvement, or, 'buy-in.' Analysis and development would be an iterative process.

In this writer's opinion, a rarely used element in analytics is ***requirements gathering***; determining the needs or conditions necessary to bring about the desired outcome or result. Studies have revealed that requirements gathering are often lacking when organizations perform analytical functions. One study, "The Role of the Business Analyst" from Web analysis firm Voke Inc., found that the average Fortune 500 analysis project costs $3.2 million and takes 1,290 months to complete. Yet only 37% of respondents answered "yes" when asked if their projects met users' needs.

There is a consensus among experienced analysts that more time for developing requirements is needed. Theresa Lanowitz of Voke argues**, "**Too much of the life cycle is spent in rework. The development and testing phases dominate the lifecycle in less mature organizations.' These phases are in inverse proportion to the requirements gathering and design phases, which make up larger chunks of the Software Development Life Cycle in optimized organizations. "When the design is laid out, there is so much more you can do in testing; you're so much more efficient in coding," said Lanowitz. "There's less room for error and interpretation." For business analysts who find themselves mired in rework, Lanowitz is confident that attitudes are changing in upper management and organizations are realizing how important requirements definition and business analysts are. Organizations "should treat their business analysts as paid consultants," she said. Until then, Lanowitz recommends BAs "survey their own organizations." They should "enlist support, lobby for enhanced communication, and advocate automation to eliminate waste," according to Lanowitz11[1].

Sound analysis and the analytic products the analysis produces often suffer from a lack of adequate tools to do the job. Many Business Analysts (BA), whose roles are central to sound analytics, are armed with traditional tools like Microsoft's Word and Excel. Under anarchy, they would attempt to glean useful meaning from rows and rows of numbers read from a typed or written notebook, spreadsheet or graph.

MOVING TO ORDER FROM CHAOS

Using tools such as SAS Business Intelligence suite based on data integration staged through products such as SAS DI Studio, and upon database architecture designed under business process management models such as Master Data Management, the BA can find facilities within the Microsoft and SAS product stacks to leverage nuggets of business intelligence. OLAP and formalized, organized data architecture means fewer development hours and a quicker delivery of data assets.

SUMMARY

Chaotic, reactive analytics is a product of inefficient business practices that can be repeated month after month without culture change and sustainable data architecture. This is largely due to a lack of a comprehensive strategy and the tools to carry out that strategy. Inadequate analysis would be the result when requirements gathering utilizing data assets in an enterprise rated with an RDI of greater than or equal to 2.5 are exploited. Sound analytics that deliver a reasonable representation of reality depend upon an RDI approaching 3 and hopefully, 4. RDI would help deliver the cultural shift from reactive inefficiencies to wiser analytics founded in solid requirements gathering.

Alan Mann can be contacted at: 540-336-7873.

11[1] Mullaney, Retrieved 19 August 2011 from:
http://searchsoftwarequality.techtarget.com/news/1326034/Requirements-gathering-resources-practices-lacking-at-Fortune-500-companies?mboxConv=searchCIO_RegActivate_Submit&