

# THE RISE AND RISE OF THE DATA EMPIRE: HOW TO PREVENT THE FALL!

Data 'happens'. It used to be confined within our firewalls and systems but now it is 'happening' everywhere. Data is ubiquitous, but not glorious. It is a small word with big ramifications; the visible tip of a sprawling, submerged iceberg. It is an area which has been complicated in the past by too many words and unhelpful definitions. Data management has always been problematical, but if it was difficult before, it is perceived as nigh on impossible now. Most organisations are in a mess with data and no amount of bottom-up initiatives in data governance, data management, regulatory compliance or information exploitation are going to fix this. Yet, at the end of the day, data is a corporate asset. Why not adopt a simpler approach and treat it like any other corporate asset?

## THE ISSUES

The 'processing' of raw data to provide factual information, which could be used as a basis for reasoning and decision making, was taking place long before the first computer appeared on the scene. In times gone by it was the scientists and engineers who were most closely involved in the collection of data: experiments were performed (actions), data (results) were collected and these results were analysed. In the wider world, observations are made, the results are captured, patterns are identified, conclusions drawn and, from this, decisions can be made.

The problem for modern business is that the volume of data is growing at an exponential rate. Worse still, it is not just being generated by the 'actions' of the organisation, but is being fuelled by such phenomena as employee and customer 'self-service', social media and automated trading.

For an image of how 'helpful' data can become 'out of control', picture the Sorcerer's Apprentice sequence with the brooms in Walt Disney's Fantasia! Just like the chopped up pieces of Mickey's broom, initiatives driving this data spring up more often and in more and more places. The result is that the data is collected many times, for different purposes: trend analysis, regulatory compliance, performance and many others. For each, the raw data must be mapped, merged or manipulated to meet the needs of the stakeholders, internal and/or external. Once this is done, unfortunately, the lineage and provenance of the data is lost. So, when further questions are asked, the

additional data required to support the answers can often be difficult or downright impossible to trace.

## DATA MANAGEMENT TO THE RESCUE?

The 'How to Measure' data quality scale is well known in data circles and is founded on a number of fundamental measures:

- Accuracy: 'Correctness' of the data
- Consistency: Semantic standards applied
- Completeness: Gaps within a record
- Entirety: The quantity of entities or events captured versus those universally available
- Breadth: The amount of data captured about an entity or event
- Depth: The extent of entity or event history/versioning
- Precision: How exact the data is
- Latency: How current the data is
- Scarcity: How rare an item of data is

In its 'raw' form every datum scores 100% in each of the above. Once moved, changed, aggregated or manhandled in some other way, the scores begin to fall and in simple terms, a sliding scale effect is seen: the more handling points involved, the lower the score. Essentially we have data entropy.

There is a further complication for data managers, in view of the push to split data into two broad categories: structured and unstructured. This is allegedly designed to help us to manage data better! The premise is that data residing in each of these categories has different characteristics and so needs to be managed differently.

To be contentious, all data has structure - see any of the definitions! In this context, however, the meaning has been twisted so that any data not residing in a structured system with the explicit metadata trimmings has been designated 'unstructured'.

My own view is that it makes more sense to think of 'structured data' as a physical thing, place or person - a product, for example; and 'unstructured data' as a perspective. This perspective is valid regardless of the medium. So, irrespective of whether I use email, Twitter, Facebook, blog or database entry to rate my last holiday, the consumer of the data can understand my perspective and rationalise the information.

Regardless of how it is classified, elements of 'unstructured' data may still be required in addition to its 'structured' counterpart for trend analysis, compliance, performance statistics and more.

## DATA AS A CORPORATE ASSET

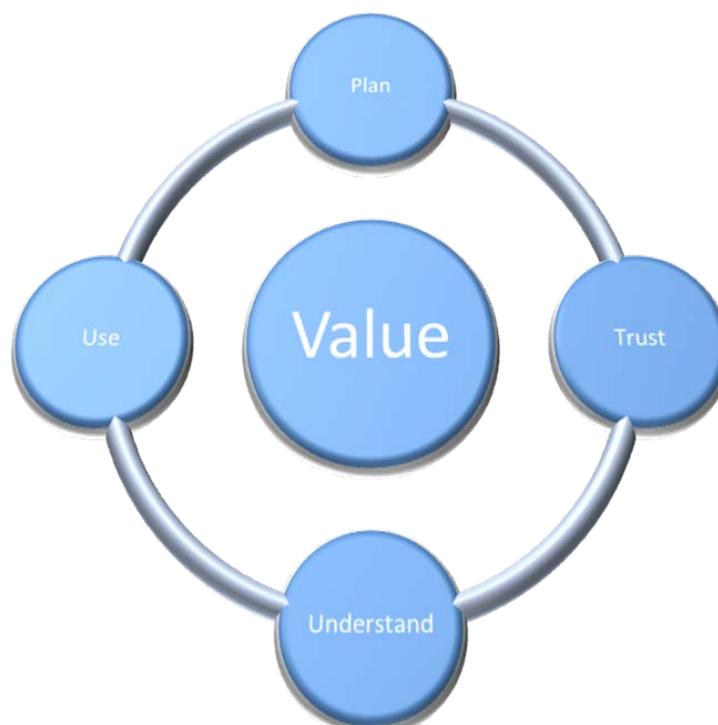
The whole point of data is to form the basis for decision making. In business, decisions are instrumental in achieving the business goals. With this in mind data needs to be:

Auditable: Accountable: Actionable

Data has a life-span, not a life-cycle. Unlike fine wines, the longer you keep it, the less use it is. Unfortunately, many organisations hold on to their data forever; what they should be doing is identifying the data sources, then holding the volume of data for no longer than required to inform decisions.

Organisations are struggling with data management, not because it is an inherently complex area, but rather because the data agenda is being driven 'bottom up' and the weight and volume of detail is crippling. External influences, particularly in the areas of compliance seem only to compound the data problem: in the effort to become compliant, new systems and processes are implemented and new data silos are created.

To prevent the data empire from collapsing under its weight, the data agenda must be driven from the top, with data managed and employed in a similar way to other corporate assets. Think of data simply. Manage it as you would manage other corporate assets: Plan; trust; understand; use.



**PLAN.** Whether we are talking about a new production plant, printer, or vehicle, physical assets are 'planned' in capability terms. In other words, what do we need the ability to do, how and for what purpose?

Planning for data should be no different: Why do we need it? What is the perspective of this need? What domains are involved? Who are the consumers?

There are a finite number of data sources in the organisation. Understanding which sources are both relevant and useful should not be overly onerous.

**TRUST.** Trust incorporates a number of concepts: governance, stewardship, quality assurance - these all play a part in building trust in the data. Consistency and purity are key. Think of data quality as a river: it starts pure and clean in the mountains but as it is moved and used (or even abused!) it becomes muddy. Typically, most Enterprise Data Warehouses are built on the banks of the estuary. Along the way, how many entry and exit points there are for the data? Do these have checks and balances? If data is stored or moved, you can check it!

**UNDERSTAND.** Analytics, business intelligence, semantic modelling, information aggregation...each of these has a role to play in understanding data and information.

It is important to understand what really make a difference to the business. What are the business drivers and goals? Which decisions drive value? Many reporting systems are geared up to present data in a predefined form. Over time, the repetitive nature of presentation leads to the meaning and value being diminished. Understanding should be a dynamic and frequent activity and not one which simply relies on a monthly reporting pack.

**USE.** Information exploitation, real-time decision making, compliance... all need data. But beware: reporting and compliance does not mean using! Most compliance takes a side swipe at data and drives reporting initiatives that aren't used for business decisions but serve only to complicate an already complex environment.

Proper use of data and information requires a feedback loop to the data plan; modelling and questioning are fundamental to the effective use of data.

In using data as a basis for business decisions, the dimensions (time, geography, speed, etc.) should be used in equal measure:

- History : What has happened?
- Future : What could happen?
- Present : What is happening?

Arguably the 'Present' poses the greatest business/technical challenges and risks: having the confidence to intervene in near real-time requires a detailed understanding of the business taken from a position of trust.

## CONCLUSION

The data empire continues to grow. Plans to manage and exploit it must be driven from the top if the ensuing 'fall' is to be avoided.

Treat data like any other corporate asset: plan for it; ensure you can trust it; understand how it can help the business; and really use it in decision-making.

Keep it simple: you can start by asking the following key questions:

- What data and perspectives really make a difference to the business?
- Can I be sure this data is correct and valid?
- Do I know where the data has come from, what it means and what has to be done with it?
- Does this data help me improve the value of my business?

And don't forget that even good data still relies on competent people for its interpretation. It has been estimated that most good managers make decisions based on about 25% of the available data; unfortunately, even with 100% data availability, poor managers would not be prevented from making the wrong decision.

More does not always compensate for less!



Glue Reply is UK's leading consulting services organisation focused exclusively on optimising IT/Business alignment and minimising the cost of business and IT technology change. Our core proposition is to help organisations maximise the value from their change and technology investments by helping them define, design, implement and resource best practice:

- Enterprise architecture and business/technology change management processes, roadmaps and competencies;
- Business design and process management initiatives;
- SOA, integration and data management platforms.