

'BIG DATA' – IT'S JUST DATA, ISN'T IT?

INTRODUCTION

Google 'big data' and you get over 700 million references: there are dedicated conferences; multiple reports and studies; it even has its own proposed ecosystem. All of the technology 'big hitters' seem to have a view on it - most of them designed to panic the CIO by highlighting the insurmountable, escalating, sheer overwhelming scale of the 'big data' phenomenon.

But fundamentally, we're just talking about data aren't we? Bigger volumes – definitely; a greater variety – certainly; more complexity – undoubtedly; and coming at us with ever-increasing speed – yes indeed. Still, basically, it's data, with all of the management issues that data has always had.

The fact is: 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. 'Big data' is just an evolution of this trend.

Data management has always been problematical, but if it was difficult before, it is perceived as nigh on impossible now. Assigning a new terminology and 'bigging up' the consequences of a failure to deal with data is only helpful in so far as it has put the whole issue back at the top of the CIO's agenda. Data, 'big' or otherwise, needs to be addressed at the highest level with a 'top-down' approach.

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 manage it as you would any other corporate asset?

THE ISSUES

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 also being fuelled by such phenomena as employee and customer 'self-service', social media and automated trading.

For an image of how 'helpful' data can quickly 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 and for different purposes: trend analysis, regulatory compliance, performance, marketing 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 provenance of the data is lost. So, when further questions are asked, the additional data required to answer these can be difficult or downright impossible to trace.

'BIG DATA' – IT'S JUST DATA, ISN'T IT?

DATA MANAGEMENT TO THE RESCUE?

Data management is not just a question of handling large volumes.

The 'How to Measure' data quality scale is well known 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.

DATA AS A CORPORATE ASSET: THE FOUR-STEP MANAGEMENT PLAN TO DATA VALUE

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 for far too long; what they should be doing is identifying the data sources, then holding the volume of data for no longer than required to enable them to make informed 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*

'BIG DATA' – IT'S JUST DATA, ISN'T IT?

PLAN

Whether we are talking about a new production plant, a printer, or a vehicle, physical assets are 'planned' in capability terms. In other words, what do we need to be able to do, how will we do it; and to what purpose?

Planning for data should be no different: Why do we need it? What perspective are we coming from? 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 abused, even!) 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 are there for the data? Do these have checks and balances? If data is stored or moved, you can check it!

UNDERSTAND

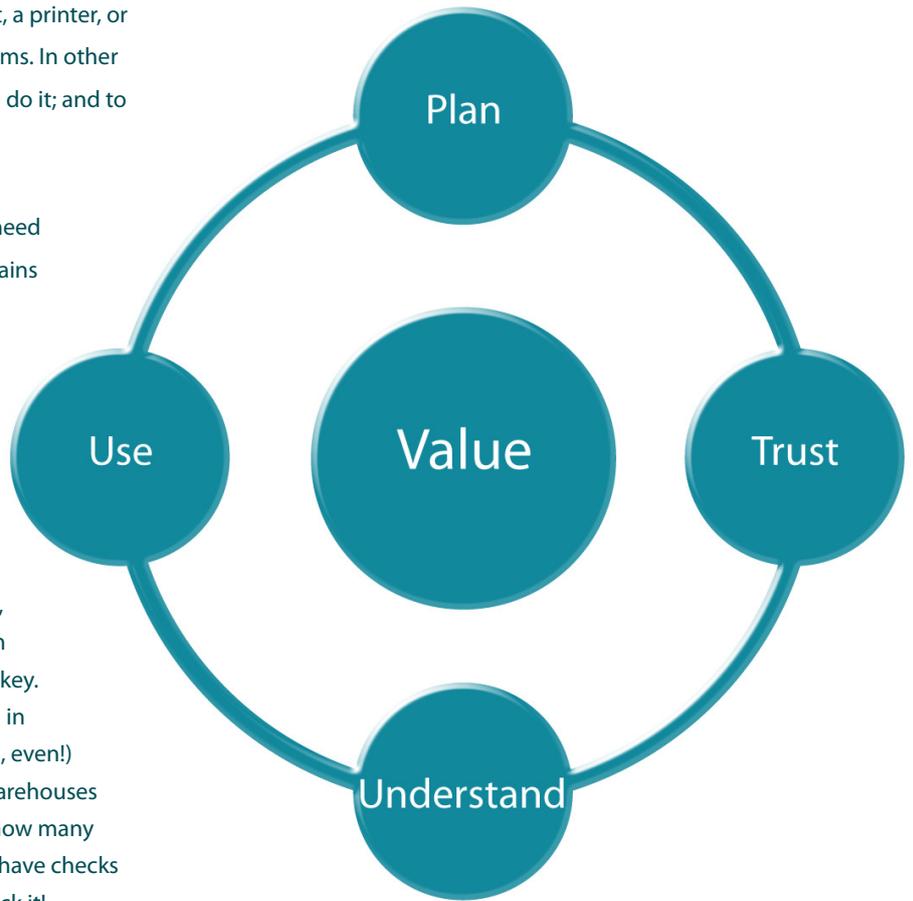
Analytics, business intelligence, semantic modelling, information aggregation...each of these has a role to play in understanding data and information.

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.

'BIG DATA' – IT'S JUST DATA, ISN'T IT?

CONCLUSION

Managing data is a 'big' problem; but does giving it the new 'big data' name help today's CIO to understand and deal with it? Or does it obfuscate an already complex area?

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?

According to Gartner, "Big Data" is only the beginning of extreme information management". I would contend that data is data. It's a continuing battle; and we have to deal with it.

And don't forget that even the best-managed data cannot provide the whole answer. Good data still relies on competent people to interpret it. 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.



Glue Reply UK is a cutting-edge, independent IT consultancy and part of Reply Limited, a Reply S.p.A company. Not bound to any technology provider in the market and not afraid to challenge its clients, Glue Reply provides high level, independent advice on the technology solutions that bring clients' business objectives to life.

Glue Reply works with blue chip and smaller companies as a trusted advisor as well as being known for getting stuck into the nuts and bolts of an IT problem / project to ensure the desired outcome.

Its consultants work with heads of architecture, programme security, CIOs and CTOs across many sectors from retail, telecoms, financial services, utilities and defence to media and entertainment. Set up in 1999, its clients include: Marks & Spencer, AXA UK, Boots, Cable & Wireless, Sainsbury's, Ministry of Defence and National Grid.

Contact: **Glue Reply UK**

Telephone: **01628 481553**

Email: **info@replyltd.co.uk**

www.replyltd.co.uk