

# The Big Data Revolution

## Big bang or loud noise?

By Tony Lock, October 2012

Originally published on



<http://www.theregister.co.uk/>

Anyone currently employed in any area of the IT business will be aware, however reluctantly, of the considerable amount of effort being put into marketing 'Big Data'. During August and September of 2012 Freeform Dynamics surveyed 502 readers of the Register (for sample breakdown see [here](#)) to gauge how far organisations have advanced in their adoption of Big Data solutions and the current state of play concerning business analytics. This article looks at where organisations report they currently hold business data, how well they exploit the value locked in their data stores and their thoughts on how things may develop in the near future.

Today the words 'Big Data' are currently used and abused in a myriad of ways. From these it is possible to distil the term to be shorthand to refer to a number of advanced data storage, access and analytics technologies aimed at handling high volume and/or fast moving data in a variety of scenarios. These typically involve low signal-to-noise ratios, including but not limited to brand monitoring, log file analysis, high volume transaction monitoring for fraud detection, to name just a few.

### Data volumes are growing

It is no news that the amount of data being generated by organisations continues to grow rapidly and this survey illustrates just how much of a challenge this is with over half of respondents reporting that they are experiencing high, or extremely high, rates of data growth and very few experiencing no growth at all (Figure 1).

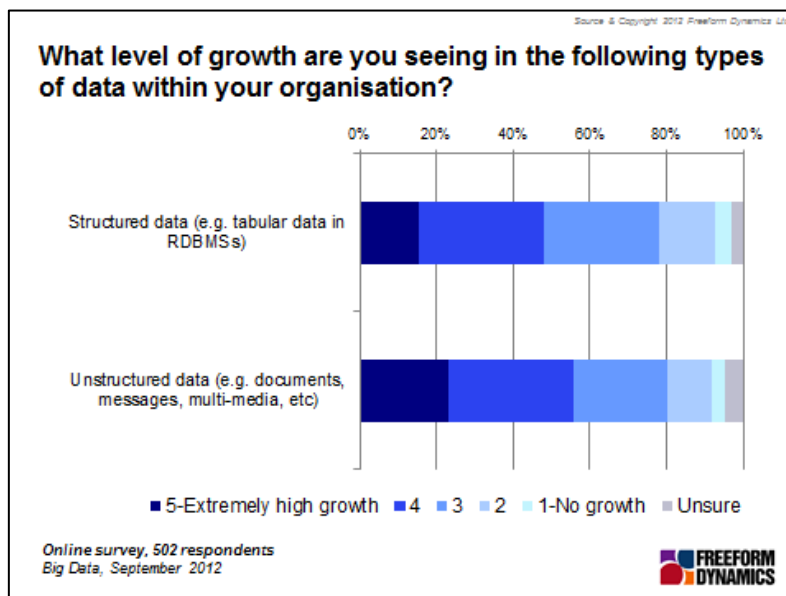


Figure 1

Of considerably more importance with the vendor focus firmly directed at “Big Data”, is the fact that whilst the expansion of volumes associated with unstructured data sources such as email, office productivity documents, video files etc. is very strong, the reports of growth in structured data sources, such as those associated with relational database systems, is not that much lower.

This begs the question, just how valuable or important are the types of information held in structured sources compared with unstructured repositories?

**The value of data**

It is evident that the majority of organisations today still keep most of their business critical information in structured repositories, typically relational database management systems (RDBMS). Far fewer hold such data in unstructured sources (Figure 2).

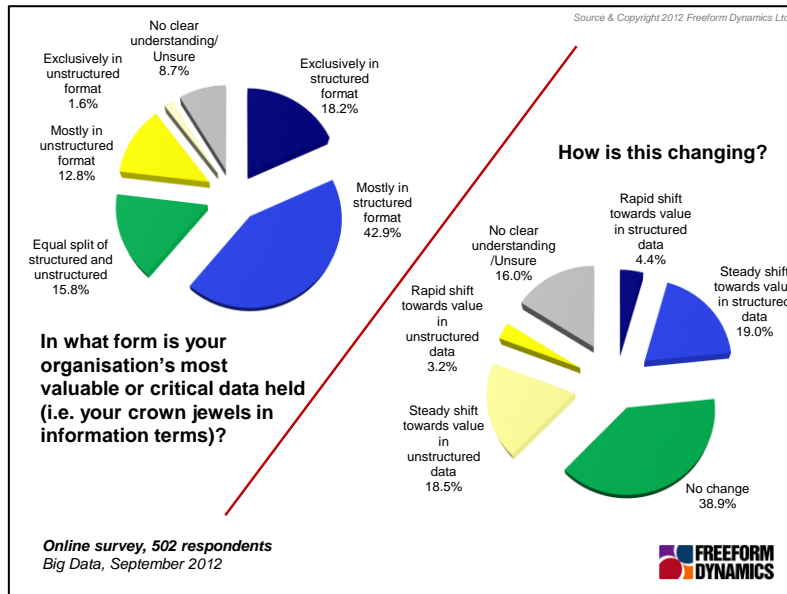


Figure 2

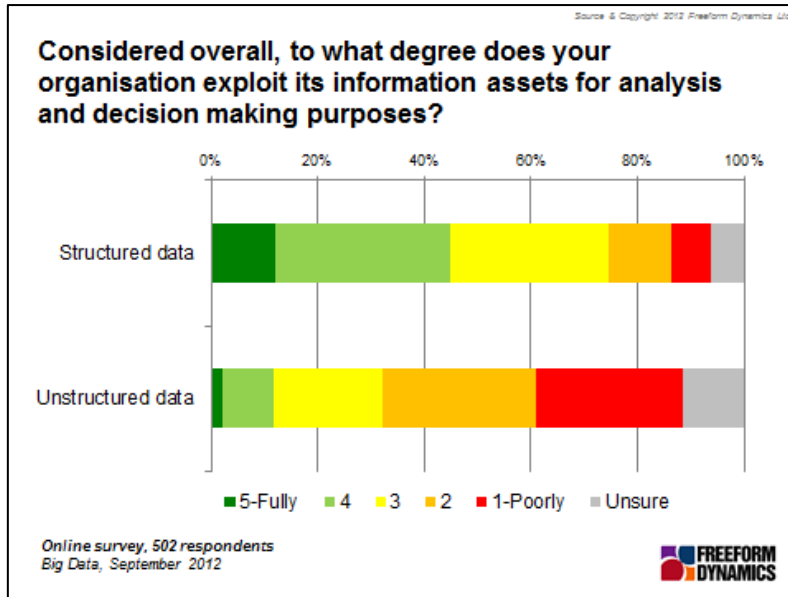
That said, it is worth noting that nearly one in ten of the respondents indicate that they are not sure how such data is held. It is easy to understand why this should be so when the question of data storage architecture has never been an area of intense study for most IT professionals and has never even been thought about by the vast majority of business users.

When asked how things are changing, the research indicates that roughly equal numbers of respondents expect there to be a shift towards either structured or unstructured sources as the repositories for business critical data. That said, the most frequently offered opinion is that there is likely to be little change evident in the near future. This illustrates that Big Data will not be sweeping away the very strong position that relational databases and other established solutions hold in the near term.

With so much data out there just how well are organisations exploiting what they already store?

**Data value is still ‘underexploited’**

The Reg readers who responded to the survey tell us that only rarely is the data their organisations hold exploited as well as it could be. Even a cursory glance at Figure 3 shows very clearly that only around one in eight organisations fully exploit their structured data whilst an even smaller fraction does so for unstructured data.



**Figure 3**

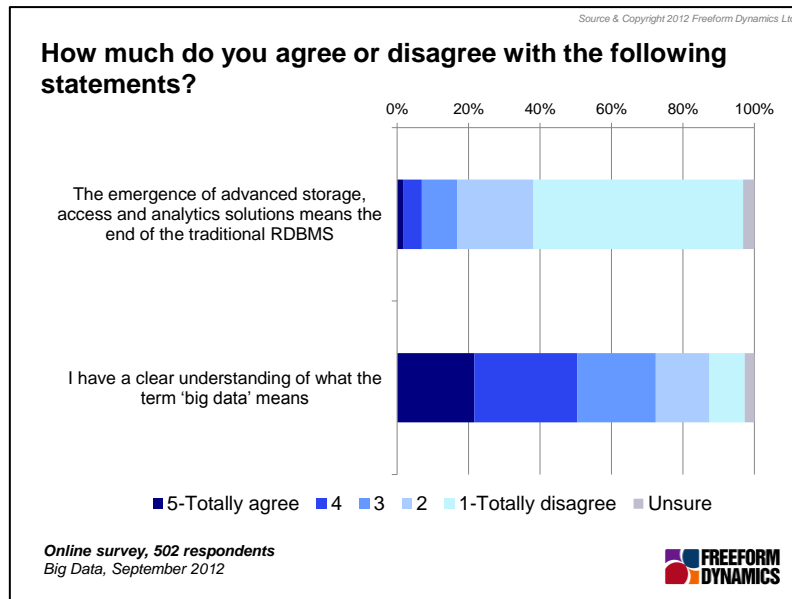
There is also a wide spread acknowledgement by many that they do a poor job in making full use of both types of data, with the effective exploitation of unstructured sources at remarkably low levels. Taken together, these results show that there is obvious potential for both ‘traditional’ business intelligence / data analytics tools and Big Data solutions to be utilised for the benefit of very many organisations.

Indeed, the fact that very large numbers of respondents indicate they could do better in terms of exploiting the information they hold can be interpreted as both a recognition of existing business pressures to get more value from their data as well as an acknowledgement that until now BI has been seen in many organisations as the domain of a few specialist analysts rather than being something that could help steer the company in its daily operations.

But if so few organisations are doing a good job generating value from their information assets, what does this tell us about the likely evolution of how data is held and processed?

**Relational Databases are here to stay**

Some writers would have you believe that the advent of ‘Big Data’ is a revolution opening the way into a new era of information processing that holds the potential to supersede everything that organisations do today to analyse data and turn it into information. A few even go so far as to claim that the age of structured database systems is over , and with it will come its passing as the primary repository for business critical data and the platform on which key business analytics are run. The results in Figure 4 show that this is not the case.



**Figure 4**

Even amongst readers of the Register, a population likely to contain no small number of early adopters of revolutionary systems, it is clear that the end of RDBMS solutions is anything but close. This should come as no surprise to the vendors marketing Big Data solutions as the prevalence of RDBMS systems is extremely wide and it is well known that making significant modifications to business critical systems is neither quick nor easy.

But the second point on the charts is one that the proponents of Big Data systems need to tackle very quickly, unless the visibility that the term has generated thus far be lost in a sea of noise. When fewer than half of our Register reading respondents state they have a clear understanding of what the term 'Big Data' actually means, it is fair to assume that an even smaller number in the wider world of IT and business have a good idea of what it is.

Even more importantly, it also means that the number of organisations that are ready to make use of Big Data to improve their lot is not likely to be large and that some who could potentially benefit from such systems are not yet ready to do so. In fact, the most likely scenario is that RDBMS and other 'traditional' business analysis and information storage platforms will run alongside new 'Big Data' platforms as they develop into solutions that can be widely deployed from the niche status they hold today.

### **The challenge of integration**

As organisations seek, in the near future, to operate both traditional data / information management and analysis environments alongside new 'Big Data' entrants, there are other factors that will need to be addressed. Top of mind here is the requirement of how to integrate the two environments to allow both to be exploited without adversely impacting each other. Perhaps of even greater importance is the requirement to feed any results generated by Big Data solutions into the existing information management and business information / analytics environment so that the most effective use can be made of the insights gained.

For both existing BI systems and Big Data generated results to be exploited as widely and as rapidly, as possible, all systems must be designed to work within the existing landscape. It is essential that what could otherwise be independent islands of capability are built to work together with management tools and processes structured to mitigate potential fragmentation and disjoints.

Without such process and management, integration costs will be higher, results will be visible more slowly and overall exploitation of data resources will be impaired. Another decidedly non-trivial by-product a lack of operational integration could entail involves matters of data security and legislation / regulation. Unless processes are put in place it will be very easy for external restrictions on data usage to be overlooked, with potentially significant consequences.

## The People Element

Another major factor to bear in mind concerns the role that people play when data analytics and 'Big Data' are under consideration. It is not difficult to notice that there is today a considerable shortage of people able to generate good results from data stores and then understand what the numbers tell them in the context of the business. As one respondent pointed out:

*"I am a statistician doing data science, mostly on data sets under 500 GB. You are focusing on the Big Data side when there is not enough use being made of the smaller data sets due to lack of statisticians. The Big Data side is also seriously hampered by lack of statisticians, as there is lack of recognition that stats training is needed in Big Data as well as the hardware and the analytics software. A good knowledge of maths is needed to be able to learn the stats, so not everyone is suitable for training."*

As is often the case in IT, attention is usually centred firmly on the technological advances that technology can bring, with far lower consideration paid to how people need to interact with systems. In the case of BI and Big Data, magic software able to automatically analyse data in the context of business operations has yet to be created and the number of people with the numerical skills and business understanding to exploit analytics appears to be far too low. The value of skilled staff can be easily overlooked or taken for granted.

One potential way to minimise the requirement for highly specialist skills is the provision of 'templates' and models constructed to work in specific industries and scenarios, for example to help make sense of customer management in insurance, fraud management in banking, consumer behaviour in retail/CPG etc. Some vendors are actively talking about creating such tools, but it is still early days and the need to produce more data scientists is not likely to go away in the near future, if ever.

Interest in Big Data has been aroused, but there is also recognition that making better use of existing information to improve business operations must be addressed in many organisations, and that people have a pivotal role to play, but one that is perhaps not widely recognised.

## About Freeform Dynamics



Freeform Dynamics is a research and analysis firm. We track and report on the business impact of developments in the IT and communications sectors.

As part of this, we use an innovative research methodology to gather feedback directly from those involved in IT strategy, planning, procurement and implementation. Our output is therefore grounded in real-world practicality for use by mainstream business and IT professionals.

For further information or to subscribe to the Freeform Dynamics free research service, please visit [www.freeformdynamics.com](http://www.freeformdynamics.com) or contact us via [info@freeformdynamics.com](mailto:info@freeformdynamics.com).

### Terms of Use

This document is Copyright 2012 Freeform Dynamics Ltd. It may be freely duplicated and distributed in its entirety on an individual one to one basis, either electronically or in hard copy form. It may not, however, be disassembled or modified in any way as part of the duplication process. Hosting of the document for download on the Web and/or mass distribution of the document by any means is prohibited unless express permission is obtained from Freeform Dynamics Ltd.

This document is provided for your general information and use only. Neither Freeform Dynamics Ltd nor any third parties provide any warranty or guarantee as to the suitability of the information provided within it for any particular purpose.