



JD Software Pvt Ltd

# dealing with DATA

*Data has to have multiple dimensions to be useful. Time and place are the main two dimensions that need to be considered - but these have complex needs and also need to have additional context to provide the maximum value to your organisation.*



*Data is worthless until it has had some level of analysis placed against it. For an organisation, data has historically been used to understand where it has been, where it is now and where it will be going in the future. This uses the time dimension of data- but there is a lot more that should be considered.*

*The place dimension can add immense power to data analysis - particularly when combined with other contextual data. This short white paper takes a look at the possibilities of using multi-dimensional data.*

## **A short history of DATA**

The actual capability of any organisation or individual to operate successfully is based around having enough information available on which to make the right decisions. However, this information has to be based on data - and the data has to be accurate and complete enough to enable information to be gleaned from it.

There is a basic 'Strategic Insight Pyramid' that is used to show how data has to be acted on to increase its value (see Figure 1). This shows how data needs to be dealt with to become information, from which analysis can create knowledge. Only from Knowledge can decisions be made - which then leads to strategic insights that can be used within the organisation.

That strategic insight is where the true value to the organisation lies - this is how intellectual property gets created and how the competitive actions, services and products that keep an organisation ahead of its competitors are progressed. Therefore, the main aim of any organisation has to be figuring out the best way to deal with its data needs to ensure that maximum knowledge can be extracted from the data. The main problems with modern data are that there are far too much of it, and that it has no value in its raw form.

Let's take the second point first - how come that data has no value in its raw form? In essence, the data that will be used in any analysis will be in electronic form and will be stored on a set of disk drives. Without an understanding of how that data is stored and the format

that it is in, there is no value to it - it is just data - some bits on the storage. A lot of data within an organisation has been held in formal databases, created by enterprise applications such as Customer Relationship Management (CRM) and Enterprise Resources Planning (ERP) systems. This structured data is easier to deal with than its less structured cousin: the data held within documents, emails, web searches, images, video, voice, etc. Business Intelligence (BI) systems find structured data easier to deal with - and are now having to deal with this growth in less structured data.

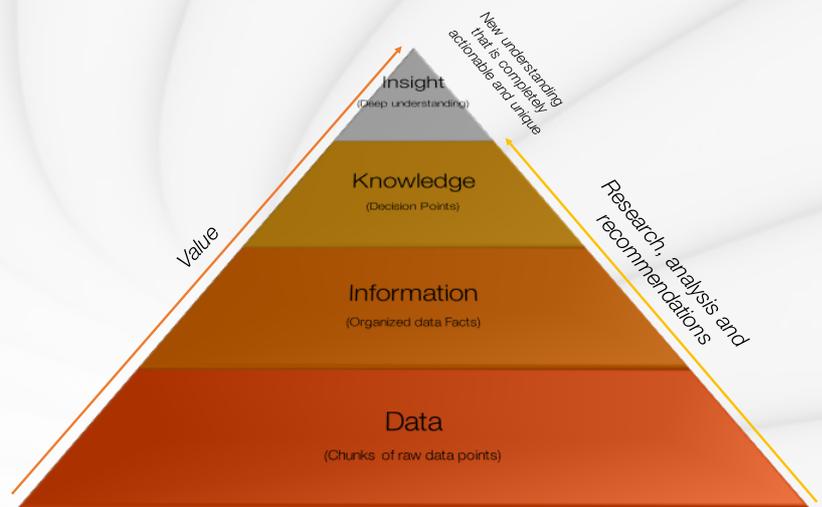
And herein lies the first problem - along with high growth rates in structured data, there is explosive growth in less structured data. Gaining insights into the less structured data needs additional efforts - but it can be well worth it. This 'Big Data' issue requires new approaches to how data is collected, filtered, analysed and reported. It requires a

### **Data has no value in itself**

**"In essence, the data that will be used in any analysis will be in electronic form, and will be stored on a set of disk drives. Without an understanding of how that data is stored and the format that it is in, there is no value to it - it is just data."**



Figure 1: The Strategic Insight Pyramid



different mindset around the sourcing of data that will help breach the gap between data and information. It requires different BI systems that can aggregate and understand the relationships between the different data sources and types.

At a highly general level, organisations have moved over the past 30 years from having very little data being held electronically, through having proportion of data held in relational databases and a little of their less structured data held in enterprise content management systems, to a point where a vast majority of an organisation's own data is being held electronically. What few organisations now use is data from outside their organisations. It is increasingly difficult to draw a line around an organisation and say "this is us". Organisations are dependent on suppliers (and their suppliers), as well as customers (and in some cases, their customers). There is a wealth of data held along this value chain - but also even more, possibly useful, data outside of the constraints of the chain.

For example, many organisations will be making use of data from Dun and Bradstreet or LexisNexis. Some will be making use of consumer profiling data sets such as Equifax or Experian. However, there are far more data sets available in the market - some are free, others available at a small charge. For example, gaining access to address verification databases makes not only customer relationship management (CRM) systems more effective, it can also help in fraud detection when different datasets can be combined and compared to ensure that a person's name and address do verify along with other available data.

How about gaining access to other data sets, such as geographic data on the positioning of certain features? Would knowledge of local transport (such as bus stations and train stations) help in deciding where to build a new retail outlet? Would knowledge of where main utilities pipe and cable runs are make a difference to where you run your data cables to your new offices? How about using the time dimensions as well? Is the prevalence of flooding in a certain area increasing over time - and what impact does this have on your plans to build in the area? Bringing in such data sets from organisations who have the domain expertise and capability to maintain the data as current as possible can not only save your organisation time and effort, but can also add direct value to how decisions are made and so impact the bottom line positively.

This short paper looks at how an organisation should look at its data assets and ensure that it gains as much value from them as possible. The paper should be of interest to anyone involved in ensuring that their organisation has sufficient information at its disposal to make the right decisions at the right time.



## Just what is information?

So, data in itself is worthless. Information is what is uncovered once the data is understood. For example, knowing that the data is held as a relational database, it is easy to understand the rows and columns that are hidden within the data. However, this has little value on its own - a row that is 'Inedible Foods Ltd' '250000' '03032015' 'ABC23456' '256' hardly enlightening. Once the database structure is understood, then it may come to light that our customer, Inedible Foods Ltd, has a credit limit of \$250,000 with our organisation, and that its last interaction with the organisation was on the third of March 2015, where it ordered 256 of stock number ABC23456.

Great - we now have some understanding of what is happening - but this is still not knowledge. All we have is one data point - we cannot see if Inedible Foods Ltd is going to buy more of ABC23456 next month; we have no idea where Inedible Foods Ltd is; we do not know whether it has a good credit history - we have no frame of reference to add any value to the information. This has been the problem with many data analysis systems - they can only analyse what is directly available, and tend to either give a historical snapshot view of what has happened or of what is happening now. There is a need to be able to access enough data sources to be able to provide better analysis of what is happening now, and what is likely to happen in the future.

Data has to be suitably aggregated, filtered and analysed so that it can be presented to the right person at the right time so that effective decisions can be made to create new strategic insights.

## Gaining knowledge from information

It is increasingly unlikely that all required data will be held in one place. In the above example, the data is likely to have been pulled from an order system, which may not be linked to the CRM or other systems. By pulling multiple different databases together, more context can be pulled out.

We may find out that the organisation has been dealing with Inedible Foods Ltd for 20 years, that it orders every quarter and that it has never missed a payment. We may find a list of people within Inedible Foods Ltd that the organisation has dealt with. Context is now being applied to the data - and we are gaining greater knowledge by being able to apply time lines to what we know, allowing us to carry out a level of prediction - for example, we can be pretty sure that Inedible Foods Ltd will place an order within the next quarter - and based on buying patterns, what this might be. We are better positioned to know who to contact within Inedible Foods Ltd based on who ordered last time.

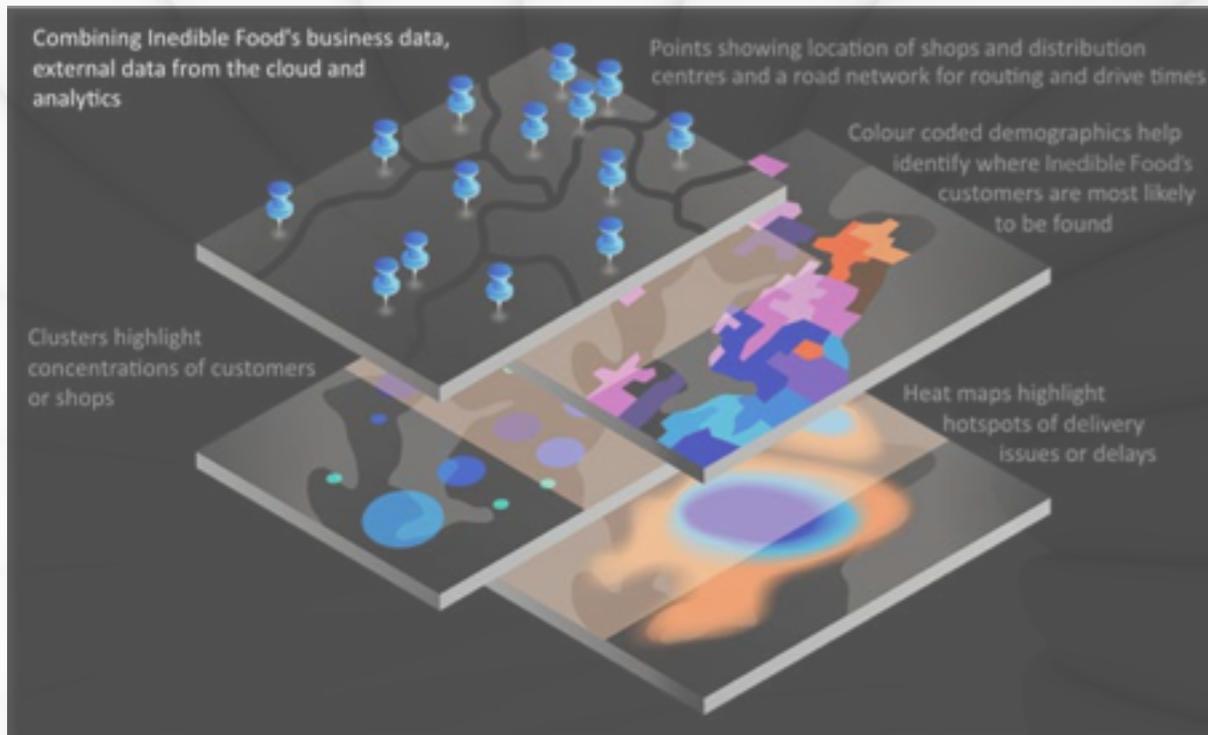
This enables a set of decisions to be made - what should be made ready to fulfil the order; when to contact Inedible Foods Ltd; who to contact within the organisation; send out an invoice and so on. We can start to put in place schedules for logistics. Value is now being added directly to the organisation - but it is still relatively low-level value.

## Adding to the knowledge

Now let's consider how other data dimensions can add to the value of the knowledge already obtained. Let's assume that the address for Inedible Foods Ltd within the master record is '1 Inedible Foods Way, Sometown, Anywhere, AA1 3AA'. We now have a geographic fix for the office. However, we know that Inedible Foods, Inc is a retailer, and that it has multiple shops around the country. Is this valuable to us? It could be - it could well be that Inedible Foods Ltd has multiple distribution centers. Is there value in delivering directly to each distribution center, rather than to Inedible Foods Ltd's main central distribution centre? Will Inedible Foods Ltd pay extra for that? How



about direct delivery to each outlet - will that enable better 'just in time' (JIT) inventory management for both our organisation and Inedible Foods Ltd? That could have great value for both organisations - and just requires a little



knowledge of where everything is and suitable mapping software to compute optimal vehicle routing (see Figure 2).

We do, however, need to be able to easily visualise this. It is not that easy for an individual to look at a list of addresses and figure out the best way to arrange logistics to visit the addresses in turn. Layering the data directly onto a map makes this much easier for the individual to understand. The data already has geographic context within it - each address has its own postcode. It therefore becomes easy for the right analytics systems to take that data and layer it over a map - and then do carry out further analysis and advise on the best route for trucks to take to minimise the time and fuel involved in getting to each destination.

This can then be combined with other datasets, such as inventory, to ensure that outbound loads are maximised. By analysing customer needs, it also makes it possible to offer them reverse logistics - do they need something transporting from their site to a point along your vehicle's return trip? If so, a charge can be made for this that offsets the cost of outbound logistics, while adding extra value to the customer relationship. With inventory, being able to add a dimension of place to each item can be valuable - for example, we may not have what is required to fulfil an order in the warehouse we have direct responsibility for, but is the item in any other warehouse worldwide, and how long and at what cost would it take to get the item either to our warehouse or directly to the customer? With 'big data' coming through, do we have access to live inventory information, such as where a specific pallet or item is in its journey? If so, can we make this available to our customers so that they can self-track items and so cut down on enquiries to our help desk?



Now the true value of data is being made visible. We have passed beyond just being there as information; we have added value to it so that it becomes knowledge that business nations can be taken against. The data has reached a position where it is not only helping our organisation take informed decisions, but it is also giving us capability to help our customers and suppliers - making them more likely to maintain high value relationships with us.

### **The application of context**

So far, we have considered the direct analysis of available data on a single entity - in this case, Inedible Foods Ltd. However, to become a truly intelligent organisation based on analysis of all available information, we need to extend our reach and add further context. For example, Inedible Foods Ltd is player within a specific retail market. By bringing in information about others in the same market, we may be able to both create additional value for ourselves and for Inedible Foods Ltd.

Maybe we can see that two of Inedible Foods Ltd's competitors are merging, and that this will result in a major overlap in the new entity's stores. By sitting down with Inedible Foods Ltd, maybe we can help them to understand where to put in more investment where the new entity's outlets may be closed down and to rethink plans where existing outlets are likely to be increased in size. Maybe we can work with Inedible Foods to come up with an optimised delivery system where we deliver to major shops and they take some of that to smaller shops themselves, and we charge them less for the delivery.

We may also be tracking the market and find that there are increasing problems in the market that Inedible Foods Ltd operates in. Maybe we should be looking at tightening our credit limits on them - or maybe not, if the information only applies to geographical areas where Inedible Foods Ltd does not operate. It is pretty obvious that there are three main aspects to ensuring that an organisation can gain optimal value from its data assets - having access to as much real and correct data as possible; being able to understand the context of data with regards to place. This combination enables fast analysis of data to provide the information to a visualisation engine in the correct manner so that knowledge can be made visible to a person, enabling timely strategic decisions to be made - so completing the Strategic Insight Pyramid.

### **Increasing the use of Open Data**

An aspect of data availability has been growing lately is the availability of 'Open Data' sets. In the Europe, the Government has been making great strides in publishing own data to anyone who wants to use it at no cost. These Open Data sets are available either as downloadable data sets (generally Excel spreadsheets) or via application programming interfaces (APIs).

As an example of such data sets, the Environment Agency makes available data on flood events. Using such data and projecting it onto a map helps decisions be made on where to build a new office or retail outlet - an area which be seen to flood regularly may not be the best place to position a strategic building. Similarly, the Land Registry provides data on land property transactions, price paid for properties and data to support the EU's INSPIRE polygon land parcel data. These data sets can be combined with some of the Ordnance Survey's data sets around building boundaries (such as OS VectorMap District, an open data set showing aggregate property boundaries based on the per-building boundary commercial Mastermap data product) showing, for example, how different areas and different building plot sizes are trending at a price level. Paid-for data from the Land Registry can provide the title details to a property - using this data to identify any land usage constraints (e.g. is this land adjacent to the building plot we are



considering buying contracted under a farming only covenant?) and can ensure that decision are made that have greater value over the longer period.

Other data sets that the government makes available include the open register based on the electoral register, the official government contract request for proposal database, the Companies' House information service, the Ordnance Survey maps service, the Highways Agency's live road traffic information service - the list goes on to currently around 27,000 different datasets that can be accessed through the National Information Infrastructure. A list of available data sets can be found at <http://data.gov.uk/data/search>. Similarly, local councils, police forces and other public bodies are also making available data that could be useful in adding to the capability to extract knowledge for an organisation to act on.

Much of this data is formatted around geospatial data - the basis point for the data will often be a point in a two -or three-dimensional space. The data sets tend to use standardised latitude/longitude or, as with OS data in the UK, British National Grid coordinates (plus altitudes for 3D). This data can be directly ingested and used by suitable business intelligence and geographic information systems (GIS) to conduct advanced spatial analysis and add value to other data. by using the spatial data as reference data, different data sets can be very easily normalised, making data aggregation far easier.

As an example of this, consider an insurance company. It receives two claims against paint-splattered carpets from two different houses. Let's say that one claim is from 256, High Street, the other from 12, Market Place. Obviously, these two properties have to be different, so there is little chance of fraud at a multiple claim level. However, by applying geospatial data to the houses, it appears that the two properties back on toe acc other. This could indicate that discussions have been going on between the two claimants - and could be worth investigating to see whether this was just a means of getting new carpets from two people talking to each other. Other organisations are also making data available - the Open Data Institute (ODI) is driving acceptance and use of Open Data, with commercial companies such as Esri UK, Thomson Reuters, NTT Data, Amey and Adobe all members. More can be found on the ODI at <http://theodi.org>.

## **Advancing the use of information**

The increasing power IT systems allied with the improvements in capabilities of analysis systems is leading to many more possibilities in how the Strategic Insights pyramid can add value to the business.

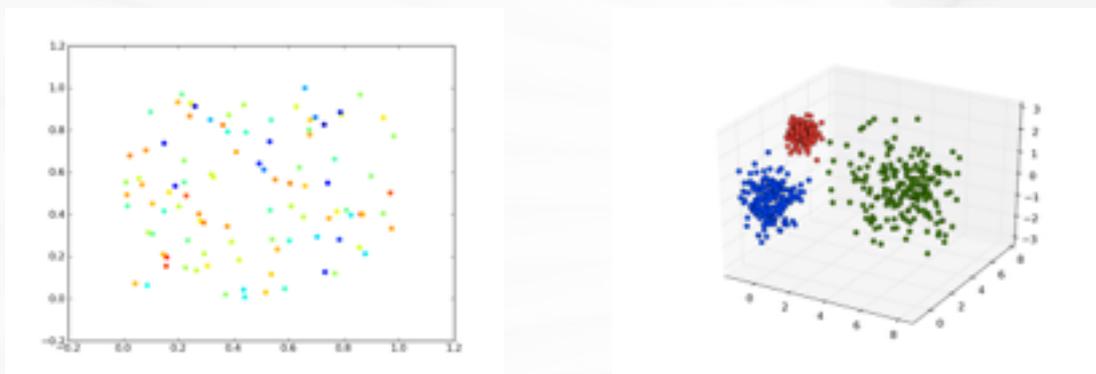
Increasingly, data sets are being loaded up directly into computer memory, enabling real-time analysis to take place. Even the use of flash-based storage arrays can make the analysis time of large data sets shrink to a point where more continuous analysis can be carried out. Big Data is enabling different types of data to be brought into the reach of analytical engines, and is providing more data that can be used to provide contextual analysis. Analytics engines are providing more ways of analysing different types of data along with different ways of visualising the results. Let's take as an example the use of multi-dimensional spatial data - there will be a further report looking purely at this subject, so we will keep it simple at this stage.

An oil rig is a complex and dangerous environment, with over 200 people working on it. The Oseberg C rig off the Norwegian coast is operated by Statoil, and is a combined production, drilling and accommodation rig. Note the complexity - what is essentially a fully contained five storey hotel at the rear of the structure, with helicopter landing site on top, along with multiple levels of working platforms around the drilling rig itself. The platform is currently being

extended to house more workers. Now assume that there is a problem with the rig. An evacuation is required - just where is everyone on this rig?

Oil companies have well-prepared and comprehensive procedures for when a problem happens on a rig - everyone knows which muster points they need to go to, and how to prepare themselves. However, procedures do not always work out - for example, a set of stairs that are meant to take people from one area to a specific muster point may not be usable. A person could be injured or otherwise disabled and unable to make their way to the muster point. Therefore, rescuers are put at distinct danger - if they cannot account for each and every person on the rig, then they will need to search the whole rig in order to find any missing people.

If each worker carried a wearable device that could plot their position, it would make life easier. However, if these



were just standard 2-dimensional positioning devices, all that rescuers would be able to see is something along the lines of a flat plat of dots - even though this could be a realtime plotting of the individual's moving around. The 2-dimensional view will show strong clusters of people on the rig itself, and what could be people on the helicopter landing pad - making rescue that much harder. However, once 3-dimensional capabilities are brought in to play, everything becomes clearer.

It becomes apparent that those clusters of people are not really clusters, as they are spread over the various levels of the rig. Likewise, those people that looked like they could be on the landing pad are actually spread through the accommodation quarters. Targeted information can therefore be sent to each person directing them to the spot where they can be best rescued from: anyone who cannot move to a rescue point can be easily pin-pointed and a rescue team sent directly to them.

This addition of one extra dimension enables a major change in how data is analysed, presented and used. Many organisations are beginning to use more context around their existing data to uncover insights that are of value to how they carry out their operations. Consider how adding more context and dimensions to data can help in logistics, utilities, even the tracking of at-risk groups of people such as the very young or the infirm elderly. The opportunities are endless: those who fail to make the most of the extra data that is so easily available to them will all to be competitive in their markets.

# Structuring dimensional data analysis

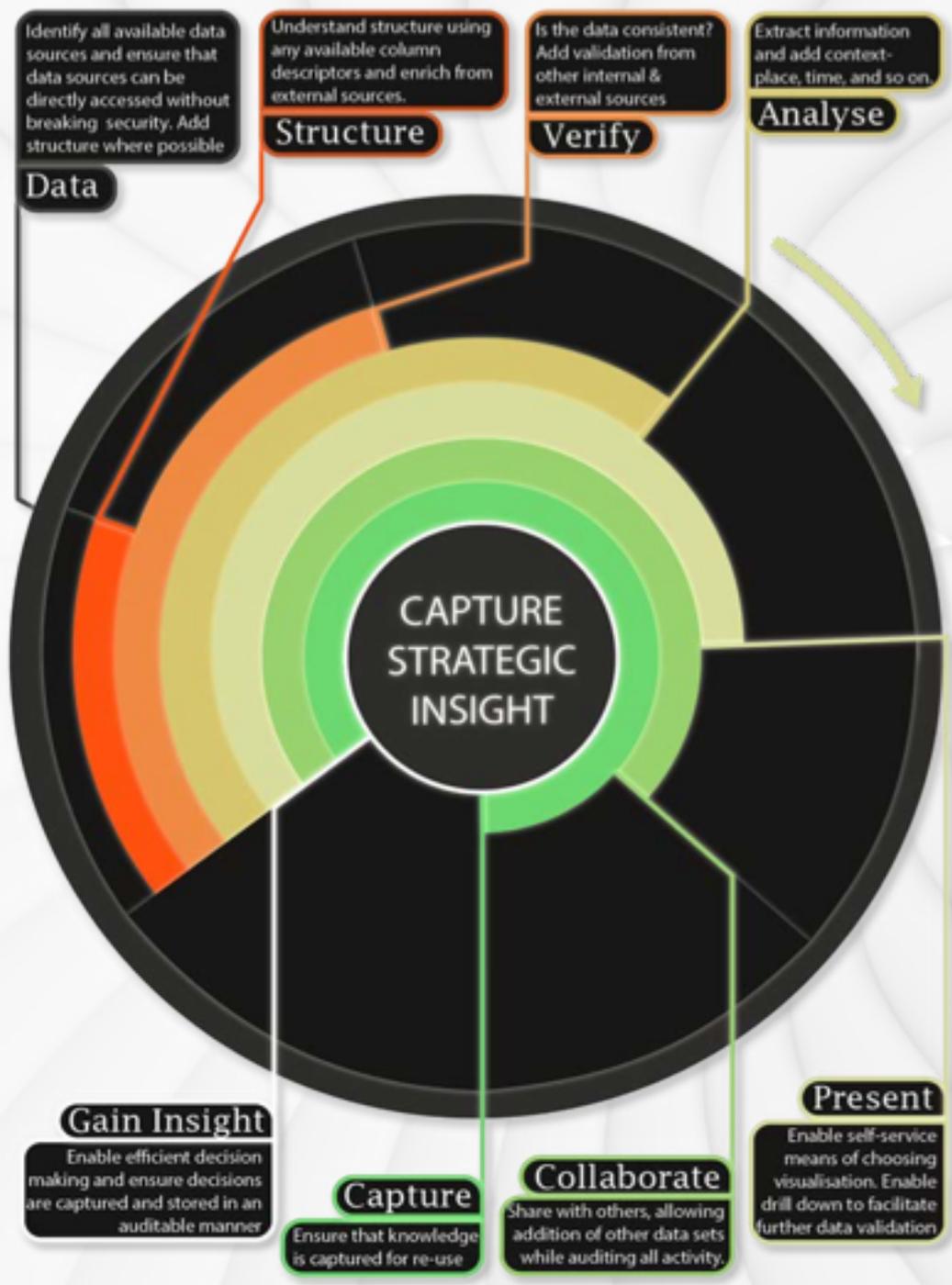
When looking towards a more contextually-oriented analysis platform, it is necessary to ensure that the right processes are in place to ensure that the maximum value is gained from the overall data creation/acquisition, analysis and reporting.

The diagram below outlines the tasks that JD SOFT advises be used to ensure that the strategic insights required are efficiently and effectively captured.

The process is made up of a set of tasks:

- Capture the data
- Structure
- Verify
- Analyse
- Present
- Collaborate
- Capture the Knowledge
- Gain the strategic insights

This then creates a 'virtuous circle', where new data adds to the existing knowledge within the organisation, and new strategic insights drive the identification of what extra data sets could be of use to add even further insights.





## Conclusion

Data has become fundamental to the general operations of organisations. The analysis and reporting of this data and the creation of strategic insights is what defines the successful organisations. The focus on true business value is moving away from the underlying hardware and software that create the data to the systems that can aggregate, analysis and present the data in a meaningful manner. Software as a service, external commercial data sets and Open Data mean that the basic platform is taken out of your organisation's hands: the key here is around how well you can negotiate service levels around data availability and cost.

All this is making data far more available - and that in itself is a problem. Data, on its own, is worthless-although the explosive growth in volumes of owned and available data make its storage and management cost high. Unless data is dealt with effectively, it is nothing but a cost to a business. Data has to be exploited in order to become valuable to the business. Yet trying to lower this cost by minimising data used to provide strategic insight is self-drafting: the decisions that are made will be based on a subset of the available data, and so will have a lower probability of being correct.

Any chosen system for dealing with data must be able to aggregate differing types of data, must also be able to deal with filtering the data into manageable volumes without losing any of the pertinent data. It then must be able to analyse the data in suitable timeframe, and must be able to present the information extracted in a way that helps people within the organisation to make better informed decisions.

The capability to bring in external datasets and apply them in context against other data is where the real value lies. The use of spatial data, whether this is at the macro level (geographic mapping) or the mic level (by building, function or item), provides data insights that can make an organisation more effective and more efficient, adding direct value to the organisation's bottom line. The use of multiple data dimensions enables far more value to be extracted from disparate data sets than any other method. The ability to normalise and layer data sets through the use of latitude/longitude/altitude, postcode and road number, or other broadly accepted spatial identifiers means that the majority of data that is of interest to an organisation can be made far more valuable. Trying to carry out business intelligence in a world of data without the context of time and place is becoming harder: as big data continues to be a major issue for organisation, now is the time to ensure that such context is understood and used effectively to provide the strategic insights required to be effective in your markets.

# about JDSOFT

JD SOFT is a global IT consulting, technology services and out-sourcing company, with more than 125 people serving clients in 5 continents. Combining unparalleled experience, comprehensive capabilities across all industries and business functions, and extensive research on the world's most successful companies, JD SOFT collaborates with clients to help them become high-performance businesses in the face of competition. To know more about JD SOFT please visit [www.jdsoft.in](http://www.jdsoft.in)

JD SOFT's Analytics CoE, is a research and analysis practice specialising in the business impact of information technology communications (ITC). With reach in 5 continents, JDSOFT provides in-depth insights into the views of the business data, buyers and influencers in large, mid-sized and small organisation. Its analyst team is made up of real world practitioners with first-hand experience of ITC delivery who continuously research and track the industry and its real usage in the markets.

Through scientific research perceptions, JDSOFT uncovers the real value hidden in the data that the business generates on daily basis - the personal and political aspects of an organisation's environment and the pressures of the need for demonstrable business value that they had to generate to their shareholders. This capability to uncover and report back on the end-user perceptions in the market enables JDSOFT to provide advice on the realities of the risks and its mitigation.

JDSOFT's research is always pragmatic, business oriented and conducted in the context of the bigger picture. Analytics has the ability to transform businesses and the processes that drive them. JDSOFT's mission is to help organisations improve their success rate in process enablement through better levels of understanding of their data and improve the competitiveness within the industry.

JDSOFT also works with global and local providers of data management tools apart from their own to help them in delivering on the promises that ITC holds for the business. JDSOFT client's include one of the big names in India and Africa in the retail, utilities, manufacturing and energy sectors.

Copyright © JD SOFT All rights reserved

JD SOFT, its logo and lead IT... are  
trademarks of JD SOFT