

# DATA

All forward-thinking organisations know that by unleashing the data locked in their systems they have significant potential to advance their strategic goals. Reaching that full potential means a full transformation into a data-driven organisation. Techobscura can make that happen by guiding you through the many components to create a **leading-edge data strategy**.

**techobscura**  
trusted • independent • expert

# DATA

A whitepaper from Techobscura

<b>DATA STRATEGY .....</b>	<b>2</b>
<b>EVIDENCE-LED ORGANISATIONS .....</b>	<b>3</b>
<b>BUSINESS INTELLIGENCE .....</b>	<b>6</b>
<b>DATA GOVERNANCE.....</b>	<b>9</b>
<b>DATA-DRIVEN SERVICES.....</b>	<b>12</b>
<b>DATA SCIENCE.....</b>	<b>15</b>
<b>STAKEHOLDER DATA REPORTS (FACTBOOKS) .....</b>	<b>17</b>
<b>OPEN DATA .....</b>	<b>19</b>
<b>ABOUT TECHOBSCURA.....</b>	<b>22</b>

*... for complete initiation as an efficient citizen of one of the new great complex world wide states that are now developing, it is as necessary to be able to compute, to think in averages and maxima and minima, as it is now to be able to read and to write.*

**H.G. Wells, Mankind in the Making, 1903**

# DATA STRATEGY

All forward-thinking organisations know that by unleashing the data locked in their systems they have significant potential to advance their strategic goals. Reaching that full potential means a full transformation into a data-driven organisation. Before this happens there are generally two common questions posed:

1. What's our vision for how far we want this to go?
2. Where do we start?

To help answer those and provide an introduction to our more detailed notes, here are the four steps of a data strategy.

## Step 1: Embracing the cultural change

The cultural change needed is to become an **Evidence-Led Organisation**. This means asking for and relying on “data-as-evidence” for all decision making, starting at the top of the organisation and supporting that with the training of key staff to introduce data-literacy as a new skill throughout the organisation. This ensures that **Business Intelligence** becomes embedded as a resource used by staff at all levels of decision-making.

## Step 2: Identifying and managing your data

Understanding and documenting what data you have and how it is held, is vital if you want to move to any of the next steps. This can be as simple as drawing up a partial data asset inventory, through to the implementation of a full **Data Governance** framework.

## Step 3: Service innovation

You're now ready to innovate with the data, which you can jump straight into with the development of **Data-Driven Services** using data you already know you can get, or you can take a longer-term approach by building a **Data Science** team and using their outputs to drive your service development.

## Step 4: Customer, stakeholder and community engagement

The same promise that “data-as-evidence” will transform your organisation internally, can be made for your relationship with your customers, stakeholders and wider community. The simplest step is to produce regular **Stakeholder Data Reports (Factbooks)** that demonstrates the value that your organisation or industry provides by showcasing your data against a number of key comparators. The more advanced step is to publish your data as **Open Data** and take your stakeholder engagement to a new level.

## How we can help with Data Strategy

- Advise your board on the benefits of a data strategy.
- Work with you to introduce a complete data strategy.
- Help you to develop any missing elements to your current data strategy.
- Review your current data strategy and implementation.

# EVIDENCE-LED ORGANISATIONS

## Building an evidence-led culture

As with all cultural change, there has to be clear commitment from the top for the change to permeate the whole organisation. This means the Board and CEO always asking the question “what data do we have to back up this idea?” and repeatedly giving the instruction “go and get the data we need to inform this decision”. If the Board and CEO do this regularly then the rest of the organisation will change to think this way. More formally, this commitment is substantiated through the following:

- Including “evidence-based decision making” or equivalent sentiment as one of the key organisational values and ensuring that this value is followed.
- Pushing the key insights that the organisation has collected to all staff as part of an ongoing education programme on the evidence that drives the organisation.
- Training staff in the basics of data interpretation so that they can understand and interpret dashboards and figures.
- Broadening access to tools that allow people to explore and interpret data themselves and the raw data for those tools, beyond the traditional business intelligence team, to support power users.
- Using “data as evidence” in all the top-level organisational documents – the strategy, the annual report and board papers.
- Providing sufficient budget for the ongoing use of external survey companies and/or survey tools to ensure that the collection of new data becomes habitual.
- Designing instrumentation into all automated systems and adding it into key legacy systems.

## Aligning staff around key evidence

If the staff of an organisation are to fully align behind the strategy, then they need to see the evidence behind that strategy so that they can make the same intellectual connection between the evidence and the strategy that the organisational leaders have.

Some of this can be achieved by opening up access to corporate dashboards, reports and other data insights, which are often locked up within personal or department silos. However, key insights need to be pushed to staff as part of a company-wide programme to ensure 100% coverage.

## Training staff

In 1903 the famous author HG Wells wrote the following, which even now over a hundred years later seems ground-breaking:

*... for complete initiation as an efficient citizen of one of the new great complex world wide states that are now developing, it is as necessary to be able to compute, to think in averages and maxima and minima, as it is now to be able to read and to write.*

*Mankind in the Making, <http://www.gutenberg.org/files/7058/7058-h/7058-h.htm>*

All office workers are now effective users of word-processing software, but that was not always the case. Only thirty years ago there were still many organisations with pools of

secretaries or typists who wrote up documents from dictation tape. It did not take long for the benefits of everyone having access to a word-processor to become the standard, and the same approach needs to be taken to data-literacy. The quicker that we can put data into the hands of most staff and help them to become data-literate, then the quicker the benefits will flow through.

Achieving a data-literate workforce requires training in such basics as:

- calculating representative percentages (what is counted, what population, what values can be aggregated, etc)
- understanding margins of error and statistical significance
- judging correlation and causation
- recognising common forms of cognitive bias

## Surveys

The core mechanism for evidence gathering about people, their motivations, how they make decisions, what they are influenced by, and related questions, should be through surveys – surveys of your customers or those who use your services, surveys of potential customers, surveys of staff, surveys of the public and so on. These are the best ways to provide direct and unfiltered evidential data but getting the best out of surveys requires expertise.

Most people know what they want a survey to ask and can make a reasonable attempt at writing the questions without confusing, leading or disengaging the reader. However, they soon find out that understanding what the results actually mean is far more complex than it appeared at the outset and there's a good chance the questions weren't asked in quite the right way needed to join up the dots. Quite often people get knowledge from surveys, but not **actionable insights** and the survey becomes a wasted opportunity.

To get the most from a survey we start with the definitive results we want the survey to deliver and work backwards from there to create questions that will provide an 'evidence chain', where each step in the reasoning to the result links to the next. This ensures that when your results say "customer/stakeholder segment A prefers Y over X" you are not left asking "but would they prefer Z if it were offered to them?" or "what do we need to change with X to make that preferred?". That way, the results are actionable, not just interesting.

## Instrumentation

All IT systems collect and log data, but this is nearly always for operational and not strategic reasons. The common problems that come from this are:

- Not collecting the full data to support strategic analysis
- Difficulty in viewing the data as at a specific point in time in the past, and a related difficulty in generating historical trend data

Instrumentation of systems for strategic reasons has a number of components:

1. Specifying additional data points in addition to the technical logging to provide a fuller picture.
2. Creating point in time snapshots of data to enable historical analysis.
3. Building hooks for automated survey distribution and data collection.

## **How we can help you become an Evidence-Led Organisation**

- Advising your board on the benefits of this approach.
- Planning the cultural change to become an evidence-led organisation.
- Training or advising on training and tools for staff.
- Implementing a data governance framework.
- Designing surveys.
- Analysing survey results and produce a full report of actionable insights and other key information.
- Reviewing the results already produced from a survey to check they stack up and see what other insights can be inferred.

# BUSINESS INTELLIGENCE

Business Intelligence (BI) has been around in one form or another for many years and has developed from printed reports or a weekly statistics email to interactive dashboards.

Despite the availability of many different BI tool and the creation of internal BI team, this area remains complex, with many organisations still dissatisfied with their BI and many BI teams struggle to adopt a vision for business intelligence that gets strong internal support.

## Business intelligence strategy

Delivering effective BI means lifting above the operational view and taking a more strategic approach.

### 1. Ensure basic data literacy

Organisations will benefit from a strategic decision to coach, train and develop as much of the staff as possible, so that they can become data-literate and contribute to the overall business intelligence ecosystem within the organisation. To understand why this is a critical, see our notes on ***Evidence-Led Organisations***.

### 2. Support power users

There have always been staff with a high-degree of data literacy who have used basic tools such as spreadsheet software to enhance their work. An organisation can derive significant benefit from providing these people personal BI tools and access to data so that they can explore and analyse data themselves and so generate useful BI.

This is often the source of internal friction as the tools that BI teams use are generally very expensive and so not suitable for a widespread rollout. A number of vendors of high-end commercial tools are recognising this and have begun to provide free or low-cost personal tools to fill this gap. To be clear, this means dashboard/visualisation creation tools, not viewing tools which are much cheaper but do not give people the ability to create their own BI.

### 3. Transform the role of the BI team

Adopting this strategy is significantly easier if the BI team transforms to take on the following responsibilities:

- A centre of excellence for BI skills, mentoring and training staff to help them become data-literate and proficient in producing their own business intelligence.
- Support centre for the personal BI tools used by staff.
- Publishing data catalogues and making data accessible to the rest of the organisation.
- Engaging across the organisation to identify what datasets people would use and then procuring and publishing those.
- Opening up central dashboard or reporting systems so that all staff can self-publish the BI they generate.

A BI team is often well suited to take on key aspects of the data governance function, which is responsible for setting the overall framework for the management of data.

## Engaging BI

### Data visualisation

Data visualisation allows people to gain insight from data quicker than looking at the raw data. In less than ten years advanced data visualisation has gone from niche to mainstream and there are now many tools that support a wide range of useful and innovative visualisation types. Rather than being seen as a “nice to have”, data visualisation should be considered a key output of business intelligence.

### Dashboards

Dashboards bring together a set of visualisations to give a broad view of a subject and are increasingly important as this reflects the broadness of strategic goals and targets. Interactive dashboards have the most impact as they allow people to conduct limited exploration of the data and produce greater insight. Interactive dashboards should not be seen as a “nice to have” but a key output of business intelligence.

### Managed discussion

Business intelligence can be made much more valuable by enabling managed discussion around specific data points or charts. For example, if someone is able to annotate a chart and write “does anyone know why this is trending down?” then the subsequent discussion will normally generate insight that has so far been missed, or only shared by a few people.

## Data repositories

Business intelligence relies on good access to data, which is generally held in a variety of systems throughout the organisation. The choice of mechanisms and tools for providing that access is a complex one, with multiple factors to consider.

### Problems

There are some common problems the BI practitioners face when accessing data:

- Access. Some systems can be hard to query and extract data from.
- Structure. The way the data is held in some systems makes it difficult to process or use.
- Efficiency. Even simple queries can sometimes require the extraction of huge volumes of data from a system, which is inefficient if the same data is repeatedly extracted.
- Interoperability. Different systems refer to the same data in different ways making it hard to create queries that combine data across systems.
- Performance. Directly or repeatedly querying a source system may add sufficient load to impact its operational performance.

### Data warehouses

To address these problems, Data Warehouse products appeared. These are highly structured data stores where the data that is ingested is shaped and formatted into a consistent form, solving many of the problems above. Unfortunately, data warehouses have their own problems, namely that it is too much work to structure the data from each new system and so data warehouses rarely contain all of the data that people need access to. The risk is that a data warehouse only contains the data that is easy to work with and so cannot support much of the analysis people want.



## Data lakes

The solution was a new breed of Data Lake products, which take the data from systems “as is”, allowing far more data to be ingested, and then create the structure and interoperability at the time of analysis. However, data lakes also have their problems, including lots of data being ingested but never used and slow processing as the data is structured for each query. The risk is that a data lake becomes a data swamp, full of data that nobody ever queries because it is too hard.

## Streaming, real-time, near real-time and batch processing

Another complication with data repositories is what forms of processing do they support:

- **Streaming.** The data is inspected as it is generated, generally to identify issues or anomalies that are then flagged for later analysis. Once processed the data is either discarded or a short rolling window is kept, which is sometimes described by the labels “real-time” and “near real-time”
- **Batch.** This is scheduled or on-demand processing of a full dataset. The requirement may still be for a fast response to any query.

## Where to from here

Many of the problems identified above are being solved as applications mature, making it easier to work directly with source systems and thereby enabling hybrid products that both query source systems directly and pull data into a repository as needed. These developments are largely coming from the addition of REST APIs as a means of extracting data from systems:

- Access. Increasing use of APIs.
- Structure. As part of the adoption of REST, the data provided is getting simpler and with machine readable structures.
- Efficiency. Through provision of an API call that allows only data that has been added or changed to be retrieved.
- Interoperability. Again, the adoption of REST is helping to create common structures and so enhance interoperability.

## Internally developed systems

It should be clear that internally developed systems need to have REST APIs for access to their data if they are to be usable by the broadest possible range of BI and analytics tools.

### How we can help you with Business Intelligence

- Developing a data-literacy training programme for your team
- Advising on business intelligence tools and online services
- Advising on the architecture of a data repository and how systems feed into it
- Review an existing business intelligence team or strategy and recommend improvements

# DATA GOVERNANCE

Data is an increasingly important strategic asset for an organisation and like all important assets it should be managed within a formal framework. This reduces risk, increases opportunities to add value, and gives a clear picture of an otherwise hidden investment. This management of data as assets is called Data Governance and is made up of the following components:

## Data Asset Inventory

This is a simple register of all datasets that match any of the following general indicators that those datasets are an asset to the organisation:

- The dataset is used for decision making.
- The dataset is held in an application.
- The dataset is regularly updated.
- Loss of the dataset would be problematic.

The details of each dataset held in the register (the metadata) are similar to those held for any other asset and include:

- System of record (where the data is held).
- Owner
- Maintainer(s)
- Keyword(s)
- Theme (see Taxonomy below)
- Format
- Size
- Update frequency
- Language
- Classification (see Classification below)
- Details of any Personally Identifiable Information
- Details of any publication

Creating an initial data asset inventory is often the hardest part of data governance as the whole concept is new to most people and so the process requires a set of interviews with data owners that is part investigation and part training.

## Responsibilities and training

In the same way that managers in an organisation have specific responsibilities for the physical assets in their care, such as laptops or even desks, managers will need a specific set of responsibilities around data assets. These include:

- How to identify a data asset and the owner of that asset
- Ensuring the data is updated and meets any defined quality standards
- Updating and maintaining the data asset inventory
- Correct classification and description of the data
- Publication (where a decision is made to publish)

As this is a new area for most managers, the introduction of these responsibilities is normally accompanied by a training programme to introduce the concepts and explain the details.

## Standards

Unlike the financial industry which has long-established industry bodies that develop standards that are used ubiquitously throughout the world, the data industry is still in its infancy and so each organisation or group of organisations must develop and maintain its own set of standards. These standards cover the following:

1. **Metadata.** This is both the list of details that are recorded for each data and the format by which metadata is published and shared. There are a number of published government and industry standards in this area, but some are quite complex to use and as a result there is no single standard, with organisations creating their own based around a published standard.
2. **Classification.** This is the security level of the data and the process for determining that. A classification is often a set of levels of what needs to be done for the data to be shared, such as:
  - a. Shareable as is
  - b. Shareable after automated redaction
  - c. Shareable after manual redaction
  - d. Shareable only in aggregated form
  - e. Shareable only with access control
  - f. Not shareable
3. **Taxonomy.** A limited set of subjects that are relevant to the organisation and which can be used to categorise the datasets. This is normally part of the information management strategy of an organisation.
4. **Conceptual Data Model** (optional). It's likely that your IT team has some parts of this from IT projects that restructured your data. A full model for your entire organisation is exceptionally valuable in resolving long-standing data problems, planning changes, and understanding the potential of your data.
5. **Formats** (optional). Some organisations choose to be quite rigid about the formats that data can be held in, in order to ensure interoperability, control the costs in managing data, prevent being locked into certain tools or being locked out of data.
6. **Licensing** (optional). When data is published it should be licensed (or explicitly unlicensed) and that decision is best made on an organisation-wide basis.

## Models for governance

There are multiple models for Data Governance and an organisation should choose that which suits it best. These models include:

- A cross-organisation team headed by a senior executive. The main advantage of this model is that it allows multiple departments and key people to be involved in the development of this new function and the identification of training needs.
- Within the IT department. This is a natural choice where the IT department already has a strong focus on formal information management and a strong service ethos.
- Within the product management team. If the primary reason for embarking on data governance is for the creation of data-driven services then the product team is sometimes the best place to start up data governance, even if it is limited to the datasets they initially manage under this framework.

Finally, it should be noted that whatever model is chosen, ongoing engagement both within the organisation and with the customer/stakeholder community is vital to ensure that the data governance meets their needs. This engagement should be formalised as part of the responsibilities of the chosen model.

## **How we can help with Data Governance**

- Develop a detailed long term data strategy covering governance, data services, data science and open data.
- Introduce a full or partial data governance framework.
- Generate a comprehensive data asset inventory, data classification standard, metadata strategy and data governance roles.
- Develop staff support resources and deliver the cultural change needed to support formalised data governance.
- Specify measurements and metrics to ensure the ongoing compliance with the data governance framework and minimise issues from non-compliance.

# DATA-DRIVEN SERVICES

Data can be sold as is, turned into data-enabled products and become the foundation for consultancy services. Choosing the right product strategy can be complex and then implementation introduces a range of considerations - how to price correctly for varying customer types, how to brand and sell the product and how to manage the data to meet customer expectations.

At techobscura we have developed a multi-level framework for commercialisation that addresses all of these issues and takes an organisation from the simple to the more ambitious products and services.

## Benefits

Data-driven services are an exceptionally powerful add-on to an existing service portfolio as they provide three otherwise elusive benefits:

### 1. Immediate value

Data-driven services that are designed around *actionable insight* will give customers immediate value. Actionable insight comes in a number of forms:

- Problem identification. These are insights where the customer is told exactly what is wrong and, if possible, how to fix it.
- Anomaly detection. These are insights where anomalous behaviour is detected which the customer should investigate as it may mean a serious issue.
- Predictive recommendation. Using predictive analytics, a recommendation can be given to take a specific action based on the likelihood of something happening.

### 2. Long-term potential

We are still at the very early stages of exploring and unleashing the potential of data, as evidenced by the rate of growth of tools, techniques, datasets, standards and services. While organisations have for a long time used pre-packaged external data sources, the trend is now towards building internal data analytics (business intelligence) capabilities as organisations recognise that expertise is necessary to unlock value of the data they hold.

As an organisation builds its internal capabilities around data analytics and data-driven service design then it creates long-term potential for new services and new value. See more in our notes on **Business Intelligence** and **Data Science**.

### 3. Deepened engagement

With the additional value they are getting, customers are already more interested in you as a service provider. When they know that you are committed to a path of developing new data-driven services then they are even more interested because they also see the long-term potential and if you're going to be doing some of the heavy lifting for them then they want to be part of that.

## Key design considerations

When designing data-driven services, the following key design considerations should be thought through for each service:

### 1. Strategic goals

The key goal of any data-driven service is either to increase the knowledge of the customer or to change their behaviour. Behavioural changes include:

- Better infrastructure hygiene. Customers do more to reduce the cyber abuse perpetrated or enabled by their customers.
- Reduced support costs. Fewer support calls as customers use self-diagnosis and other self-service tools.
- Better data quality. Customers get better at keeping their data up to date and get better quality data from their customers.
- Evidence-based. The more informed a customer is then hopefully the more evidence-based their decisions will be and the more successful the outcome of those decisions.

### 2. Actionability

As noted above, the more immediate the action that can be taken then the more instantly recognisable the value of the service is. A service that requires a customer to combine the data provided with their own data and then analyse it may provide more value at the end, but it will inevitably have a slower take-up and may never fully get off the ground.

### 3. Basic competitive advantage

The competitive advantage of the service can come from a number of reasons:

- Unique data that nobody else can provide.
- Data that customers are keen to use but it is too expensive for them to purchase individually.
- Data that is available elsewhere but through an unwieldy interface or hidden behind multiple obstacles.
- Data that is of a better quality, generally because more work has gone into cleaning and testing it.

### 4. Aggregation, comparability and gamification

A more sophisticated competitive advantage is possible when a customer's data is aggregated with that of multiple other customers. This creates a significant potential for comparative insight and that in turn can be gamified to drive behavioural changes even more. For example, if a customer is told that they are #1 worst for data quality out of 7000 customers then they have a strong incentive to correct that position.

### 5. Community platform

Data-driven services are often thought of as one-way services provided by the organisation to their customers. Another approach is where the service revolves around a community of data contributors who get the benefits of the service and the organisation gets the benefit of their data. For example, community surveys where only contributors get the final report. A hybrid approach also exists where customers submit their data to be combined and analysed against an otherwise unavailable dataset to produce actionable insights.

## 6. Insight into the next level down

An organisation that services intermediate customers who in turn have customers of their own can add some extra insight if they have visibility of that next level of customer. For example, they can:

- classify those next level customers looking across their entire intermediate customer base and so help the intermediate customers understand how their makeup of next level customers compares to the overall base.
- spot a problematic next level customer who takes service from a number of intermediate customers.
- understand the characteristics of next level customers that make them good or bad customers and help their intermediate customers recognise these.

### How we can help with Data-Driven Services

- Develop a data services/commercialisation strategy
- Design data products
- Build a customer strategy for data products
- Specify metrics to measure product usage and engagement
- Provide an independent review of existing data services/commercialisation

# DATA SCIENCE

When used effectively, data science can provide incomparable insight and evidence to drive business objectives. It's increasingly common to see organisations turning the output of their data science team into products, reports and actionable insights that enhance their customer or stakeholder engagement.

## Why data science?

There are a number of areas of research where data science delivers results that have otherwise only been possible with extensive manual processing, if they have been possible at all. These are all basically forms of pattern matching, including:

### 1. Classification

Classification is as simple as it sounds – working through a list of entities and classifying each one into one of a set of classes. The power of classification is in the huge range of problems this can be applied to. For example, classifying your customers by their standard industry code just from looking at their website.

### 2. Predictive analytics

Predictive analytics uses a broad set of inputs to calculate the probability of certain events happening and can also be applied to a huge range of problems. For example, predicting which of a range of markets to advertise in will deliver the best result.

### 3. Sentiment analysis

By using natural language processing it is possible to determine the sentiment of a message. For example, a customer service operative could be alerted that the customer they are about to talk to on a live chat is likely to be frustrated, based on their initial message.

### 4 Fraud detection

Similar to classification and predictive analytics, data science is increasingly being used as an effective mechanism for the early detection of fraud from what appears to be only a small amount of data.

## Building the team

Data science is well on its way to being a mainstream occupation but the whole field has a long way further to develop. For that reason, ongoing professional development and training are critical for the team. A data scientist who is unable to learn new things will soon find their knowledge out of date.

## Strategy for the team

It is vital to remember that you are building a science team, which is very different from a project team and so needs a different kind of strategy. The key difference is that results cannot be guaranteed as the team will be breaking new ground and not doing things that have been tried and tested before. This is the nature of research, which



cannot be planned in the same way as a project. It is not uncommon for a data science team to experience repeated failure, or to have to abandon a particular line of research and start somewhere else.

Another area of difference from a standard business process is that it is generally not possible to validate ideas before implementation. Data science is by its very nature a lengthy and complex process of validation.

It should also be noted that the data science team might not be the best people to see the value in their results as that often requires a significant depth of business knowledge.

A strategy for a data science team should reflect this uncertainty:

- Don't set precise goals and targets for the team or expect validation early on, but give them the freedom to explore widely.
- Accept that regular failure will be part of the process.
- Ensure extensive engagement with people outside of the team, at all stages of the research, to help identify the value and possibility of the work being undertaken.

## Tools and techniques

The tools and techniques of data science are evolving rapidly, including both the commercial tools and open source tools. As with many industries, the first tools are often open source, having been developed in academia, and then commercial tools develop as the market becomes clearer. Data science is still at the early stages and so open source tools dominate, while commercial tools can be expensive and niche. When new algorithms or techniques appear, it is often in the form of a new open source library or tool that is not yet polished.

A successful data science team gets to select their own tools. While most of the open source tools will be of high quality, some may not have the support or quality an enterprise expects, but that should not become an obstacle.

Finally, be prepared for some unusual requests for equipment or services. Big data requires big iron and it is not unusual for a data science team to request a thousand virtual servers, or a single machine with ten high-performance GPUs.

### How we can help with Data Science

- Advise you on the right data science strategy for your organisation.
- Help you build the team and the technology that supports them.
- Accelerate the delivery of insights and value by designing a research programme that will deliver highly innovative results.
- Review an existing data science function to ensure it is delivering best value.
- Analyse your algorithms and the results they deliver to ensure their integrity and effectiveness.

# STAKEHOLDER DATA REPORTS (FACTBOOKS)

Becoming a true data-driven organisation means using data as a key underpinning of your engagement with stakeholders and one of the simplest and most effective ways to do that is through a Stakeholder Data Report (also known as a Factbook). This can be a standalone publication, a section in an existing report, or a regularly updated web site. Making it a success means following a small set of common rules:

## **1. Small number of insights and explanatory analysis of each one**

Too much data loses readers, and with stakeholders that becomes even more complicated as their interests and data-literacy could vary widely. A good stakeholder report picks out a handful of key insights (between three and seven is ideal) and explains each one in an accessible but knowledgeable way. Accessible by using plain language, knowledgeable by covering important concepts such as sample size or margin of error.

## **2. Insights that connect with your stakeholders**

You need your stakeholders to feel a connection with the insights that you produce as that gives them a connection to you. Conversely, if you produce insights that have no connection to them then they will suspect that you do not understand them or are oblivious to them. Good examples of insights that make a connection are:

- Explain why the organisation has taken certain actions
- Identify a problem that stakeholder may be suffering from
- Predicts the likely future in a way the stakeholders can act on

## **3. Contextualise your data with industry, national or global data**

It is important to remember that stakeholders view your organisation as one of many and as part of a specific industry and in either a national or global context. That view should be addressed by contextualising your data with other data drawn from the broader industry, or your country or globally, to provide comparisons and trends. Examples of this include:

- Total number of employees or total value of all customers/stakeholders using data from a national statistics body.
- Change in global indicators, which your work contributes to such as the number of people lifted out of poverty.

## **4. Demonstrate the value of your organisation and its mission**

This is about finding measures of the things your organisation has done, as part of its mission, that have provided direct value to your stakeholders or to their missions.

For example, if your mission includes fighting cybercrime then include a statistic on how many bad sites you have taken down (your actions), and then use external data to estimate how much money lost to crime your action has saved and show that.

## How we can help with Stakeholder Data Reports

- Write the report for you
- Identify the key insights
- Find the appropriate external data to provide contextualisation
- Recommend statistics that will demonstrate the value of your organisation
- Measure and review the effectiveness of your existing data reports

# OPEN DATA

## Defining ‘open’

Open Data means data that is published in accordance with the principles set out in the Open Data Charter<sup>1</sup>, which speak for themselves, though in the private sector we would replace the word ‘Citizen’ with ‘Community’:

1. Open by Default
2. Timely and Comprehensive
3. Accessible and Usable
4. Comparable and Interoperable
5. For Improved Governance and Citizen Engagement
6. For Inclusive Development and Innovation

It’s important to understand just how ‘open’ this is. It means no sign-ups, no logins, no contracts and no restrictions on further sharing. Truly open.

## The case for open data

Publishing open data is a path that organisations take for more than just the benefits they get from it, they also take this path because it aligns with their mission of serving a community or simply because it is the right thing to do. In that context, the overall benefits of open data can be listed as:

- Informs the community and helps them make informed decisions.
- Builds a community around your data or deepens the engagement of your existing community.
- Finds the value in your data through crowdsourcing the exploration and creative use of the data.

One surprising outcome that every organisation that publishes open data soon discovers, is that internal staff switch from getting their data from an internal system and instead prefer to get it from the open data platform as if they were any other member of the community.

## Four levels of publication

Open data is still maturing, and different vendors and data publishers have taken different approaches to open data publication. The four levels below represent the full set (as we understand it today) but it is rare to see any vendor or data publisher fully addressing all four.

### 1. Catalogue

Many organisations choose to publish a comprehensive list of datasets as it is a key part of the implementation of data governance and the creation of a data asset inventory. Some organisations do not publish a comprehensive list, but only list those they are able to publish as open data. In very large organisations, such as governments, it is sometimes only possible to publish the catalogue with pointers to the datasets that are held within different divisions.

---

<sup>1</sup> <https://opendatacharter.net/principles/>

If done correctly, there is still a major technical component here as the catalogue should be available in machine-readable form as should the metadata for each dataset.

## **2. Data Publication**

This is publishing the data in an entirely open way. Just point your browser, Excel or other tool at a URL and download. Nothing else required.

Almost all open data publication platforms will make every dataset they hold available in any common data format: CSV, JSON, ODATA, SODA and more, enabling the data consumer to use whatever tool and format is best for them.

It should be noted that it can take a lot of work for an organisation to publish data if it means building multiple complex pipelines to extract, aggregate, anonymise and then push data to an open data platform. Some platforms aim to lessen this work by providing a wide range of harvesting tools, but with those tools come new security risks.

## **3. Interactive exploration**

Data consumers have different skill levels ranging from those that can build an interactive data visualisation in code, to those that find Excel too complex. In order to ensure as widespread use of data as possible, many open data platforms have built-in interactive data exploration tools that enable the user to sort, filter, aggregate and chart the data. These charts can then be saved, shared and embedded in other web pages.

It's easy to see how these features can build a community with people exploring the data, finding an insight, preserving it as a chart or table, and then sharing it for others to comment on and reply with their own observations.

## **4. Data Stories**

The final level of publication is using the data, often from multiple datasets, to tell a story through text and charts in a long-form publication. Because the data is open, those reading the story can download it themselves and verify any claimed insights or they can write their own story around the data.

With the embedding feature of interactive exploration, anyone can do this on their own site, but those stories are then isolated and hard to find. Open data platforms are now adding their own tools to support story creation and sharing in a central location alongside the data, thereby making them easier to find.

# **Open data platforms**

Most organisations choose to use an off-the-shelf platform to publish their open data rather than write their own, as that quickly enables the various levels of publication and all of the various features. The platforms range from the fully open source, to open source with proprietary add-ons to entirely proprietary. All are available as cloud platforms and the open sources platforms can also be installed in-house.

The market in these platforms is developing rapidly and they all have some significant differences from each other, which means that the choice of platform normally requires an assessment of the market against formal requirements or even a tender rather than simply picking one of the leaders.

## How we can help with Open Data

- Manage a full open data implementation programme.
- Advise on the best open data platform to meet your needs or conduct an RFP to test the market.
- Generate a comprehensive data asset inventory and associated standards.
- Introduce a full or partial data governance framework to support open data.
- Develop staff support resources and deliver the cultural change needed to embed open data.
- Specify measurements and metrics to demonstrate the ongoing value of open data.

# ABOUT TECHOBSCURA

## Our Mission

We help organisations evolve and secure their future, by adopting the following strategies:

- Strong corporate and personal integrity
- Professional and effective governance
- Evidence based decision making
- A data-literate and data-rich organisation, board and customer community
- Conforming to and contributing to best practice
- Regular independent assessment and review

In our view, organisations that adopt these strategies are more honest, resilient, engaged, successful and respected.

## Our Company

Techobscura is a boutique consultancy and services company, which helps transform organisations, markets and communities. Its customers include Internet NGOs, domain name registries and government regulators who need trusted, independent and expert advice.

## Our people

Techobscura was formed in New Zealand in 2018 by Jay Daley after a 30-year career in technology. His experience includes the CEO of the registry for the .nz top level domain, the interim CEO of .org, the CTO of .uk and the Head of ICT and e-government for Brighton and Hove City Council.

Jay has a technical background in software development, network management, general IT management and finally general management as a CEO. With that experience he has built networks, applications, products, services and world-class teams.

His skill set is as a global expert in domain names, a specialist in all things data and some aspects of the global Internet. In addition, Jay is highly experienced in Internet governance. Jay is a member of SSAC and a member of the board of Public Interest Registry which operates .org and a number of small TLDs.



*Techobscura Ltd is a New Zealand registered company, NZBN 9429046526465*

**techobscura**  
trusted • independent • expert