

Sourcing Data Werewolves on LinkedIn

Alison Holt



LONGITUDE 174
LIMITED

Proud Sponsor of the
New Zealand Symphony Orchestra

Taking Standards to the World

Thanks for the introduction – I’m honored to be here! It’s great to be part of such a prestigious conference.

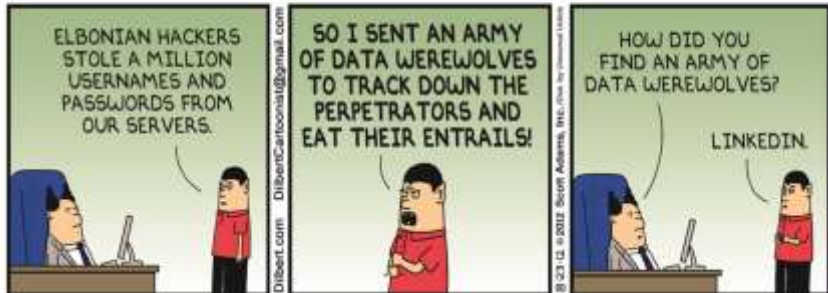
Let’s get started.....

We’re here today because we understand the value of data and, even more importantly, the power of data to transform lives and businesses – to find cures for cancer, predict public transport requirements, build intelligent maintenance services for remote locations and generally match customers to our products & services.

But there are risks associated with collecting and storing data.

Today we’re going to look at strategy and governance frameworks that will help us address these risks without having to resort to hiring data werewolves to eat the entrails of the dastardly cyber criminals who threaten to steal our data.

We all know of data disasters



DILBERT © 2012 Scott Adams. Used By permission of ANDREWS MCMEELE SYNDICATION. All rights reserved.

LONGITUDE 174°

We all know of data disasters.

LinkedIn was hacked on 5th June 2012 and by 6th June 2012 6.5M decrypted user names and passwords were available for sale online. It was initially thought that the user account information had been stolen by Russian Cyber Criminals. In May 2016 another 100M user accounts became available, and these were thought to be taken as part of the initial attack in 2012.

LinkedIn was criticised for using hashed but not salted passwords that were easy to break open. However LinkedIn customers chose poor passwords. Out of the original 6.5M user accounts, 1M passwords were made up roughly as follows – 750k – 123456, 170k – linkedin, 140k – password

Organisations have a responsibility to protect customer data through better anonymisation and encryption techniques. However maybe there is a customer responsibility too or at least a responsibility to educate customers?

Data comes with Value, Risk & Constraints



© iDreamstime.com

LONGITUDE 3M[®]

In Nov 2015, Mattel announced "Hello Barbie" – a Siri like Barbie doll that could have, and could record conversations with your child. Mattel outsourced the development of the software, and the software partner outsourced the hosting of the data. Each outsource partner set an expectation that if their company was bought, the data could be sold with the company. Any of you in the room with little children will cringe to think of the sort of things your child might share with its dolly. Years ago, we were considering fostering. Social Services called the house and my daughter, then 8 years old answered the phone. "Mum's fast asleep and I've no idea where Dad is." I dread to think the sort of things she might have shared with a doll when she was feeling less reserved.

And what if the Siri like function is hacked, and your child starts learning an altogether different and more troubling vocabulary?

Meanwhile, the only risk envisaged by Mattel – was that the Internet could be down and your child could be disappointed, but at least your child could still play with the doll as an old fashioned non-responsive doll.

Since 2015 we have seen several cases of hacks into data collected through

children's intelligent toys.

Is this an
option?



© Dreamstime.com

LONGITUDE 3M[®]

Is the data business too risky then?

Can we just avoid developing services and products that require data collection and sharing?



(4 clicks) No we can't – because business has changed – thanks to the incredible evolution of ICT systems as predicted by Moore, Kryder, Gilder and Metcalfe.

We can crunch data, store data and move data as never before and its getting faster, cheaper and easier all the time to set up data driven products and services.
XXXXX

Moore, Kryder, Gilder and Metcalfe foresaw this incredible evolution of ICT systems. Moore (1965) suggested that, “processor speeds or overall processing power for computers will double every two years”. Kryder’s correlate to “Moore’s Law” predicted a similar expected increase in the density and capability of hard drive storage media over time (Walter 2005). Gilder (2000) calculated that the total bandwidth of communication systems triples every 12 months, and Metcalfe added that the value of a network is proportional to the square of the number of nodes, so as a network grows, the value of being connected to it grows exponentially, whilst the cost per user remains the same or reduces (Shapiro and Varian 1999). In summary, our ability to process and store data is getting easier and cheaper. Whilst the cost of sharing data is reducing, the value of sharing and connecting users together is increasing significantly.

**...and now
we're all in
the data
business.**

- Bicycle accessories
- Dairy farming
- Elevator maintenance
- Toy manufacturer
- Taxi cabs
- Accommodation
- Pizza maker
- Recruitment
- Real Estate
- Government

Taking Standards to the World

LONGITUDE 3M[®]

It no longer matters what you do or who you are – you are in the data business...or at least your peers and competitors are.

Your ability to collect and share data will play a large part in your success as an organisation.

We've seen pizza companies introduce tracking so that you can track your pizza as it is being delivered.

We've seen bicycle accessories that mash up your cycle ride with your personal fitness, weather & map information.

How can we
navigate
through the
risks &
constraints?



© Getty Images

LONGITUDE 134°

So how can we navigate through the risks and constraints?

This is Darius I of Persia – a surprise appointment of a leader of the Persian Empire. Out of the short listed candidates his horse was the first to neigh at sunrise, but that was a bit of a trick. Darius's trusty servant had rubbed a cloth against the mare that Darius's horse was fond of, and as the sun rose, the servant held the cloth underneath Darius's horse's nose and it neighed.

Despite the alleged trickery in the selection process, Darius was very successful – he took a shrinking empire and built a stable and then expanding empire. He understood the value of data and information and the good governance of both.

He put in place – satraps – an early civil service that collected data to inform his decisions and distributed information.

He built infrastructure, and introduced a uniform currency and language.

He reformed his defence force by picking battles carefully and training his troops.

And by the end of his reign his empire was home to over 44% of the world's population



What can success look like?

An example to demonstrate what can be achieved.



Before we look at applying the strategy building and governance lessons from Darius and the successful empire builders – let's take a quick look at what data value and success can look like.

Protecting Premature Babies



© iStockphoto.com

LONGITUDE 3M[®]

There have been several big data projects around the world to address the survival rate of premature babies – looking for patterns around infection and survivability – and these studies include a case study that can be found on the SAS web site.

By applying early intervention to the early warning signs of infection, survival rates have increased noticeably.

In one study, researchers observed a pattern from the collection of large data sets where the temperature of the prem baby would spike one day, return to normal the following day, and spike again on the third day. By then, alas, it was too late to give medication to arrest the infection and the baby subsequently died. By giving medication on the first spike, many little lives were saved.

Predicting Anti Microbial Resistance (AMR)



LONGITUDE 174°

The serious implications of the worldwide threat of rising antibiotic resistance are becoming increasingly clear, along with its link to the misuse of antibiotics^{2,3,4}. There is a need to share generated antibiotic surveillance data externally to guide appropriate antibiotic prescription and to set up new breakpoints for antibiotics.

I'm working on a project with leading pharmaceutical organisations, interested academic and research organisations, Google Web Services and the Open Data Institute in the UK; that will focus on making industry sponsored antibiotic surveillance data openly published. We will agree common standards and develop a web based platform for making AMR surveillance data available so that learnings can be derived which will benefit clinicians, global organisations such as the United Nations, and wider society. My aim is for us to collect sufficient data to be able to produce a real time map of resistance, guidance for all stakeholders associated with prescribing, and eventually successfully predict patterns of resistance. We started off in April with data from one nation, as of today we have data from 19 nations and tonight I'm hoping that number will have increased with another large pharmaceutical joining the project.

Commitment

*“Continue to **share** the surveillance data we generate with public health bodies and healthcare professionals, and work with them to **improve understanding** of resistance trends, inform appropriate antibiotic and vaccine use and, over time, thereby help increase surveillance capabilities globally.”*

LONGITUDE 374°

The commitment from my pharmaceutical associates is as follows.



Building a successful data strategy

Develop a strategy with associated policies and processes to ensure that the strategy can be delivered safely.

LONGITUDE 3M[®]

How can we address the risks and constraints associated with data sharing and avoid the issues faced by LinkedIn, Mattel, and many other organisations around the world?

Specifically – how can we avoid that awkward and embarrassing situation where we need to call in the data werewolves?

Firstly we need a data strategy - stand back – hands off the data and take a good long look at what you're trying to achieve.

What's your data strategy?

- How much do you want to be in the “business of data”?
 - What data should you be collecting?
 - Should you be buying data?
 - Should you be selling data?
- If you're accountable for data and its use in your organization.
 - Do you have the right strategy, policies, controls and monitoring in place?
- The value of data is in the value of the decisions made using that data.
 - Are you maximizing your investment in data?
 - Are you maximizing the potential of data?
- If you're accountable for decisions made in your organization
 - Are the decisions in your organization made on good data?
 - Who has access to this good data?
 - All your staff? Your customers? Your competitors? Criminals?
 - Decision making is becoming increasingly automated
 - How are you accounting for biases, discrimination, privacy?
 - Do the decisions made match the level of responsibility assigned?

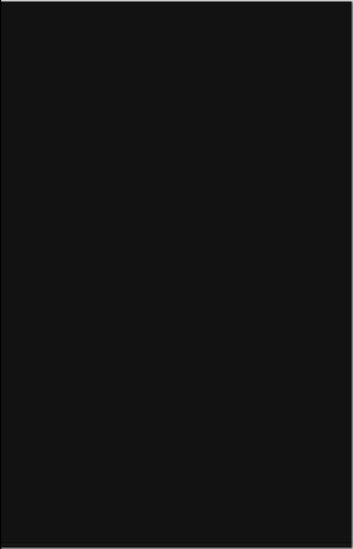
(4 clicks) Here's some questions to get you started.

You might not be accountable for the collection of data, data use and data-enabled decision making in your organisation. Hopefully your Board is across answering these questions.

What's your data strategy?

- How much do you want to be in the “business of data”?
 - What data should you be collecting?
 - Should you be buying data?
 - Should you be selling data?
- If you're accountable for data and its use in your organization.
 - Do you have the right strategy, policies, controls and monitoring in place?
- The value of data is in the value of the decisions made using that data.
 - **Are you maximizing your investment in data?**
 - **Are you maximizing the potential of data?**
- If you're accountable for decisions made in your organization
 - Are the decisions in your organization made on good data?
 - Who has access to this good data?
 - All your staff? Your customers? Your competitors? Criminals?
 - Decision making is becoming increasingly automated
 - How are you accounting for biases, discrimination, privacy?
 - Do the decisions made match the level of responsibility assigned?

I can't tell you what the answers will be (because they will vary across businesses, markets, risk appetites, etc). But I can suggest a way for you to get the answers you need, or to raise awareness amongst the decision makers in your organisation.



International guidance – here to help



I've spent the last 2-3 years developing a series of Governance of Data standards with my international colleagues for ISO (the International Standards Organisation).

The first of this series – standard 38505 Part 1 was published last month.

38505 – governance of data

All organizations use data, and the major proportion of this data is stored electronically across IT systems. With the advent of cloud computing, the realization of the potential of the "internet of things" and the increasing use of "big data" analytics, data is becoming easier to generate, gather, store and mine for useful information. This flood of data brings with it an urgent requirement and responsibility for governing bodies to ensure that valuable opportunities are leveraged and sensitive data is protected and secured.

ISO/IEC 38505-1



Taking Standards to the World

LONGITUDE 134°

38505 Part 1 sets out 6 principles for the governance of data and a governance model that shows how the management body interact with the governing body to deliver data enabled products and services in a safe and secure way. Most importantly the standard provides a data accountability map with 6 areas, as follows:

Collect, Store, Report, Decide, Dispose and Distribute

We believe that these areas represent the points where there needs to be a specific governance focus.

The standard is aimed at those responsible for writing data strategy, but will be of interest to anybody developing data enabled services or products.

38505-1 – governance of data

	Value	Risk	Constraints
Collect	<i>Data Business</i> 	<i>Data Risk Appetite</i> 	<i>Collection Policy</i> 
Store	<i>Allocate Resources</i> 	<i>Implement Security</i> 	<i>Ensure Conformance</i> 
Report	<i>Implement Tools</i> 	<i>Establish Interpretation Rules</i> 	<i>Aggregation Policy</i> 
Decide	<i>Establish Data Culture</i> 	<i>Decision Making Responsibilities</i> 	<i>Data Re-use And Learning</i> 
Distribute	<i>Distribution Strategy</i> 	<i>Implement Controls</i> 	<i>Distribution Rights</i> 
Dispose	<i>Disposal Policy</i> 	<i>Implement Processes</i> 	<i>Ensure Compliance</i> 

Taking Standards to the World

LONGITUDE 3M[®]

At each point of the data accountability map, we suggest 3 considerations:

- What is the value of this activity?
- What is the risk associated with this activity?
- What are the constraints – legislation, regulation or internal policy?

- Working through the 3 considerations of risk, value, and constraints for each area of the data accountability map will show where policy is required.

- For example – data collection – what is the value of this activity? What are we going to be using the data for? Are we collecting data of sufficient quality and quantity? Are we curating the data as we collect it? Does it have a shelf life? (Example of African pharmaceutical company & CSR credits from the UN Development Agency for sharing data in a charitable way)

- What are the risks associated with collecting data? What is our appetite for risk?

- What are the constraints? What legislation is in place that is applicable here? Does our data collection align with who we say we are as a company, and our own internal policies?

- Part 2 of the standard (which will be published later this year) will assist with the development of the policies that have been identified through the considerations matrix. Part 2 also includes case studies of organisations with good data governance in place. One of these organisations is the Chinese Securities Regulatory Commission that handles 4PB of high velocity, high value data per day. Another organisation runs a travel agency that suggests excursions based on analysis and matching of customer profiles – and does it very successfully.

Summary

- New technologies are unlocking great value from data at low cost
- No matter what industry you're in, you're in the business of data
- If you don't want to get "uberized" or have to call the data werewolves you need a sound data strategy
- The new Governance of Data (38505) standard can help



LONGITUDE 3M®

So – in summary – you can do great things with big data, but you want to protect against risks and meet the constraints put upon you.

You need a data strategy.

The recently published Governance of Data standard 38505 Part 1 can help you develop a data strategy.

I'd like to leave you with the following thought:

Keep the werewolves at bay – buy the standard today!