What have we found out about Data Maturity so far?

Sian Basker and Madeleine Spinks, Data Orchard, 16th May 2016

Over the past couple of months we've been finding out what existing work has been done on Data Maturity and what we can learn from it. We've been trawling the internet, reading the literature, and interviewing some leading people with relevant knowledge and expertise (see Appendices 1 and 2 for details).

1. Data Maturity Frameworks

The concept of data maturity is relatively new and seems to be most widely used and understood in the data science community. Whilst there's no listing for the term on Wikipedia, we found around 40-50 different models/frameworks and related theories. Indeed there are reported to be hundreds. Many of these focused on a particular industry or aspect/s of data: information maturity, analytics, business intelligence, data governance, open data, data warehousing , IT architecture, or big data. Various examples are listed in Appendix 4.

The earliest references we found to 'data maturity' were around 2005-2007 when both Gartner and IBM developed models for data quality and data analytics maturity. IBM built its first Analytics Quotient model in 2010 offering a quiz which identifies businesses as 'novice, builder, leader or master'. Many other models emerged around 2012-13 when 'big data' first gained high media profile. As knowledge and understanding has advanced more sophisticated/ updated versions have appeared.

Many data maturity models have been created by specialist vendors and consultancies as a means for selling products and services. They offer varying levels of simplicity/ technical detail to enable potential clients to understand where they are and where they might be going. Published examples include: an updated 2014 IBM model, Cardinal Path, Adobe Applied AI, Accenture . Typically they offer an organisational diagnostic and, in some cases, an assessment report e.g. 2014 The Data Warehouse Institute analytics maturity tool assesses five dimensions (organization, analytics, data management, infrastructure and governance). Much of the published literature and resources are in fairly high tech language largely focused around data architecture and data governance. These tend to be aimed at more technical audiences within large enterprise environments.

In essence what many models explain is the journey from looking at retrospective ad hoc data to explain the past, to a more continuous 'current/real-time' understanding of the here and now, a level of optimizing for efficiency and effectiveness, through to the ultimate state of predicting and creating the future. Some models use analogies to human development pre-natal/infant/child/teenager/adult/sage; or (with reference to machine power) crawling/walking/running/riding a bike ; others focus on practical processes, tasks or action.

Most models we found were conceived and applied within private sector markets where primary drivers have been efficiency, risk management, maximizing revenues, and competitive advantage. They tend to be aimed at large/very large enterprises where turnovers below $10M are regarded as small. The sectors where data capabilities are being most rapidly and innovatively advanced appear to be in competitive markets: retail, technology, energy, health, insurance, and increasingly, banking and finance.

The only commercial sector benchmarking research around analytics maturity we found was published by Accenture in the Netherlands in 2015. Based on the DELTA model by Tom Davenport of the International Institute of Analytics, it had surveyed 250 companies in 2012 and again 2015 using key indicators along the themes of: Data, Enterprise, Leadership, Targets and Analysts. It identifies a move from the earlier use of analytics for improved efficiency towards a more recent development of data to support new and improved ways of working and decision-making. It also shows data analytics maturity strongest in sales and marketing roles.

Data Analytics has become an integral component of the service offer amongst leading business consultancies e.g. PWC, Deloitte, KPMG. Whilst these primarily serve the private sector, they also count government and charity organisations amongst their clients.

The public sector also has its data maturity story. Whilst efficiency, security, and risk management were early drivers; increased transparency and public accountability are also key. The Environment Agency Data Maturity Model started in 2011 and now uses this alongside the 2015 Open Data Institute (ODI) model. The latter has been used to assess whether departments are ready to open data up e.g. in November 2015 it published a scored DEFRA assessment against five themes. The primary departments involved in supporting this work appear to be: The Cabinet Office, DEFRA and BIS. Nesta is also currently undertaking research on data in the public sector.

To date we've found two models relevant to the social sector, both from the US. The first is Educause which has benchmarked analytics in hundreds of higher education organisations (many of them charities) since 2012. It offers some great resources and useful insights into how organisations are progressing and not progressing. The second is data maturity model developed for social impact. Published in April 2016 by the Centre for Data Science and Public Policy, University of Chicago, it offers a framework around two aspects: data and technology; and organization readiness (Appendix 6). They plan to collect data from non-profit and government organisations and use this to benchmark.

There's general agreement that whatever the sector, very few organisations are operating at the very advanced levels. Indeed most are at the early stages.

Much of the discourse and literature suggests the journey to maturity is fairly long and challenging, though there's evidence to show it's worthwhile in the end. The most recent well-rounded and applicable resource we found was a book called "Creating a Data-Driven organisation: practical advice from the trenches", by Carl Anderson in 2015. Experts suggest it can take five years or more to develop, implement and reap the rewards of becoming a data mature and data driven organization. However, there seems to be no single 'truth' about data maturity. Those we spoke to had very different perspectives.

"I don't see any use of formal rubrics inside of most companies. There's a wide range of sophistication around how people document data, provide access, and think about data governance, but I don't see a standard way of thinking about this. This is likely due to the different regulatory standards and norms in each industry, and each company has their own point-of-view around those norms. We work with folks in insurance, banking, and legal, and they are all very different." Hilary Mason, Data Scientist, Fast Forward Labs.

Whilst many of the maturity models are fairly simple, many recognize the complexity and interrelationships with other key aspects of organization development. Notably: leadership, business planning and strategy, culture, as well as the policy, security, data governance and underlying infrastructure digital tools and systems. There's no evidence to suggest how widely used the various maturity models are nor how useful. Indeed some commentators suggest they're not worth doing at all. However comparing with peers and market leaders through benchmarking appears to be a popular approach to raising aspirations and understanding stages of development.

2. Do the existing frameworks concur with DataKind's theoretical 5 stages[1]?

DataKind's theoretical model in Appendix 3 suggests a pathway of: Nascent, Explanatory, Exploratory, Developing, Mastering. Many of the other frameworks we identified had 3, 4 or 5 stages that are not dissimilar in principal. However the detail and language does vary significantly and will need further research and development. The typical model in the private sector is unlikely to resonate with the vast majority of charities and social enterprises. However many of the larger ones may already be operating at that level to some extent i.e. the 1.2% of charities and 6% of social enterprises with turnovers of £5m+.

3. What enables organisations to become more data driven

The seven key factors that appear most influential and effective in enabling organisations to grow and develop in their data maturity are:

---

[1] DKUK draft model based on one from Terradata via Duncan Ross, Chair of DataKind UK, See Appendix 3.

- Data people at the heart/centre of the organisation, adjacent to leadership team
- Data recognized and valued as a key asset with and data culture established as a collective effort i.e. data is a team sport, not the responsibility of just one data person, data becomes intrinsic skill and asset for every team in the organization.
- Data must be accessible to many in the organisation. Therefore people need to be able to query, join/relate and share the data across the organisation.
- Quality Data: the organisation must be collecting the right data, relevant to the question at hand, and be able to trust it with confidence.
- Skills: people with the right skills to steward and query the data, asking the right questions.
- Time to absorb, discuss and challenge using data.
- Forward looking: moving from reporting on the past (what happened?) to the present (what's happening now) to the future (extrapolation, modeling, recommended action, prediction/simulation).

According to Anderson[2], the top barriers stopping organisations making effective us of data are:

- Lack of understanding on how to use analytics to improve what they do
- Lack of management capacity (competing priorities)
- Lack of internal skills
- Existing culture doesn’t encourage sharing.

(For full list of barriers: see chart on p. 16 of Anderson, 2015).

What next?

This research will be used by Data Orchard and DataKind UK as part of the Data Evolution project to explore social sector data maturity. The findings will be shared as a blog post and feedback and comments will be welcome. Discussions about this research will also be held with social sector organisations at various events and workshops during 2016. More specifically the research will help shape our questions with charities and social enterprises as part of a national survey.

Contact

Sian Basker sian@dataorchard.co.uk ; Madeleine Spinks mads@dataorchard.co.uk; Emma Prest emma@datakind.org.uk www.dataevolution.co.uk

[2] Anderson C.,"Creating a Data-Driven organisation: practical advice from the trenches", 2015

Appendix 1 Key Sources on Data Maturity

Accenture, Analytics Maturity Assessment, Netherlands 2015

Anderson C.,"Creating a Data-Driven organisation: practical advice from the trenches", 2015

Booz, Allan & Hamilton, "The Field Guide to Data Science", 2015 (2nd Edition).
(Lit. Review list & Data Science Maturity Model on page 35)

Davenport T., Assessing your analytical and big data capabilities, Wall St Journal, July 2014

Eckerson W., The Data Warehousing Institute, Business Intelligence Maturity Model

Fisher D., "Data Analytics Maturity Model", 2014

Howard J., Review of the INFORMS analytical model, 2014

Howson C., TScore Overview for BI and Analytics, Gartner 2015

Marsh M., "Review of skills and leadership in the VCS sector" (section on data-informed social change), 2013

Mason H., video and e-mail comms, 2016.

McSweeney A., "Review of Data Management Maturity Models" 2013

Parenteau P., Sallam R., Howson C., Tapadinhas J., Schlegel C., Oestreich T., Magic Quadrant for Business Intelligence and Analytics Platforms, Feb 2016

Polynumeral Blog, the number one question CEOs ask about data 2016

Patil D.J., Mason H., "Data-driven - Creating a Data Culture", 2015

Sedar J., Data Science maturity model blog, March 2016

Soares S., "The IBM data governance Unified Process", Sept 2010

Yanosky R., Arroway P., The Analytics Landscape in Higher Education, Educause, Oct 2015

Appendix 2 People we interviewed

Jake Porway, Founder and Executive Director at DataKind, New York

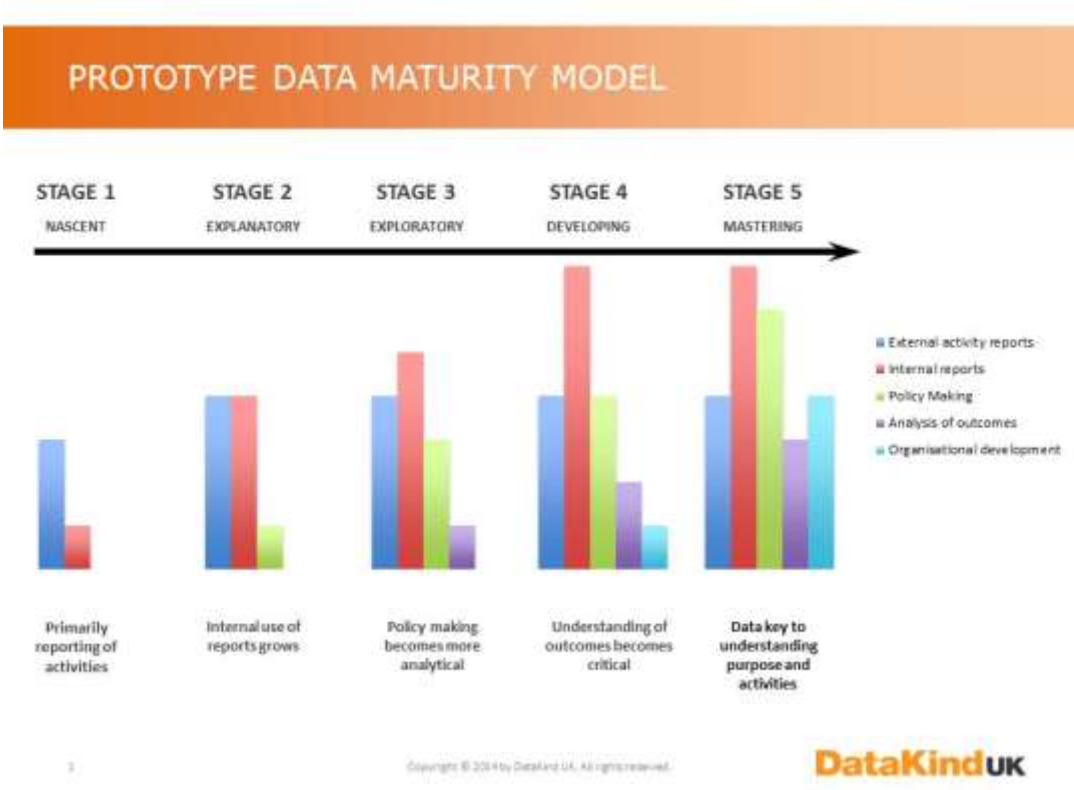
Duncan Ross, Data and Analytics Director, TES Global (Founder/Chair DataKind UK)

Jonathan Sedar, Consulting Data Scientist at Applied AI Ltd

Shyann Seet, Independent Data & Analytics Advisor

Hilary Mason, Data Scientist, Fast Forward Labs. (comment via E-mail).

Appendix 3 DataKind's prototype data maturity model



Appendix 4 A selection of the models and frameworks we found

Note: Due to copyright restrictions readers are directed to the online source rather than reproducing diagrams and images here.

Applied AI Data Science Maturity Model
GartnerMaster Data Management Maturity Model Gartner Business Intelligence and Analytics Model
The Data Management Maturity (DMM) (see CMMI institute)
Steven Mills, Chief Data Scientist, Booz Allen
Educause 2012
Gapbridge Analytics
Infofarm Slideshare slide 26
Dan Fisher, Data Analytics Maturity Model, 2014.
Jay Zaidi, 2015(AlyData): Data Management Maturity Model (DMM) developed by the Software Engineering Institute at Carnegie Mellon University.
The Data Warehouse Maturity Model Business Intelligence Maturity Model
Big Data Maturity Model (2012), comes from an IT perspective. Advanced version has more detailed emphasis on value creation, risk management, compliance, competency, architecture, policy, security, organization, audit.
Comparative view by A McSweeney. McSweeney A., "Review of Data Management Maturity Models" 2013

Appendix 5 Open Data Maturity Model

The Open Data Institute launched the first edition of its The Open Data Maturity Model in March 2015. They've built a 'Map your pathway' App which offers to help assess where you're at, set goals and track progress towards them. Benefits it promises include: Discover your organisation's strengths and weaknesses; Identify areas of improvement to optimise progression; receive practical recommendations to help achieve your goal.

Appendix 6 Social Impact Data Maturity Model, University of Chicago

Center for Data Science & Public Policy
THE UNIVERSITY OF CHICAGO

Data Maturity Framework
Data and Tech Readiness Scorecard

Category	Area	Lagging	Basic	Advanced	Leading
How is Data Stored	Accessibility	Only accessible within the application where it is collected	Can be accessible outside the application but proprietary format, requiring specialized analysis software	All machine readable in standard open format (CSV, JSON, XML, database)	All machine readable in standard open format and available through an API
	Storage	Paper	PDFs or Images	Text Files	Databases
	Integration	Data sits in the source systems	Data is exported occasionally and integrated in ad hoc manner	Central data warehouse - realtime aggregation and linking (Automatic)	External data also integrated
What is Collected?	Relevance and Sufficiency	The data you are collecting on subjects of interest is irrelevant to the problem you want to solve: ie you want to do predict which students need extra support to graduate on-time but don't have data on graduation outcomes	Some of the data you have is relevant, but it is insufficient because key fields are missing, ie no data on academic behavior or attendance history, etc.	You have data that is helpful and relevant for solving the problem but not sufficient to solve it well. ie you have yearly academic and demographic information but are missing extra-curricular activities, or interventions they were targeted with	You have all the relevant data about all the entities being analyzed and it's sufficient to solve the problem you are tackling
	Quality	Missing rows (people/address level entities missing in the data)	Missing columns (variables missing)	No missing data but errors in data collection such as typos	No missing data and no errors in data collection
	Collection Frequency	Once and never again	yearly	frequently	realtime
	Granularity	City level aggregates	Zipcode/Block level aggregates	Individual level (person or address) level data	Incident/Event level data
	History	No History Kept - old data is deleted	Historical data is stored but updates overwrite existing data	Historical data is stored and new data gets appended with timestamp, preserving old values	All history is kept and new data schema gets mapped to old schema so older data can be used
Other	Privacy	No privacy policy in place	no PII can be used for anything	ad-hoc approval process in place that allows selected PII data to be used for selected/approved projects	Software defined/controlled privacy protection that allows analytics to be done while preserving privacy based on predefined policies
	Documentation	no digital documentation or metadata: data exists but field descriptions or coded variables are not documented	data dictionary exists (variables and categories defined)	data dictionary plus full metadata available (including conditions under which the data were captured)	data dictionary plus full metadata available including collection assumptions, what's not collected, and potential biases

Center for Data Science & Public Policy
THE UNIVERSITY OF CHICAGO

Data Maturity Framework
Organizational Readiness Scorecard

Area	Lagging	Basic	Advanced	Leading
Staff Buy In	Staff at the organization have some idea that data exists but doesn't understand it is important	There are a few individuals who deeply understand the data available and what can be done with it	Organization has a clear idea of how data can be used to drive business decisions beyond justification of funding	Organization has a culture of data within the organization and demands data to justify all programmatic decisions
Data Collector Buy In	On the ground staff provide data seldomly, sporadically, or incompletely because they are required to but it is seen as a hindrance to their "real job"	On the ground staff regularly provide data because they are required to	On the ground staff provide data on a regular basis and eventually get actionable insights in return	On the ground staff provide data in real time and make decisions based on the data and insights available to them, and offer suggestions on what is collected/what information they could use to improve their job effectiveness
Leadership Buy In	Leaders at this level fundamentally don't know how data can help advance the organization's mission.	Leadership wants to use data but don't have a clear path forward to use data	Leadership has a clear idea of how data can be used to drive business decisions beyond justification of funding	Leadership builds a culture of data within the organization and demands data to justify all programmatic decisions
People Resources	Individual stakeholders maintain siloed data sets	The organization knows how data can help, what data they need, and are able to access it, but lack the in-house data skills, tools, or infrastructure to be able to turn data into meaningful insights that affect human action.	Organizations know how data can help, what data they need, and are able to access it, but lack either the infrastructure or the people to be able to turn data into meaningful insights that affect human action.	The organization has dedicated staff who own data storage AND data content owners who own the cleaning and rigor of the data
Data Use Policy	No policies exist around use, transfer, and sharing of data	Organization has policies in place for the use, transfer, and sharing of data but it does not cover all data that exists within the organization	Organization has policies in place for the use, transfer, and sharing of data internally	Organization has policies in place for the use, transfer, and sharing of data internally and externally
Intervenor Buy In	No partnerships exist	Partnerships exist but data is not shared	Partnerships exist and have policies and technology in place to share data occasionally or through a manual process	Partnerships exist and have policies and technology in place to share data in real-time
Funder Buy In	Funders do not require data other than vanity metrics	Funders ask for key performance metrics	Funders ask for key performance metrics and provide funding for data infrastructure and maintenance	Funders require data driven decision making and provide funding for data infrastructure, maintenance, and usage