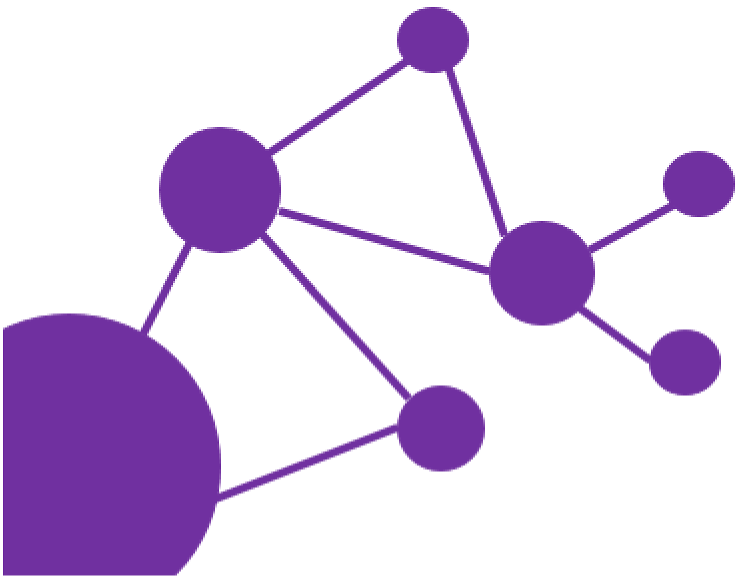
A Guide to Complexity of Database Systems

"The important thing is not to stop questioning.

Curiosity has its own reason for existing"

Albert Einstein



Research meets industry: Academic research report 2017

The Guide

In the turbulent fast moving field of database systems, complexity is found everywhere. The volume, variety and velocity of data is continually expanding as well as the accessibility and realization that businesses have a wealth of untapped data that can be democratised. Not only this, and changes in new technology, but also with the shift in business markets, organisational changes, knowledge required by operating staff and numerous stakeholders, adds to the complexity. Many of these complexities have been discussed in the Claremont and Beckman Reports (Agrawal *et al*. 2009; Abadi *et al*. 2016).

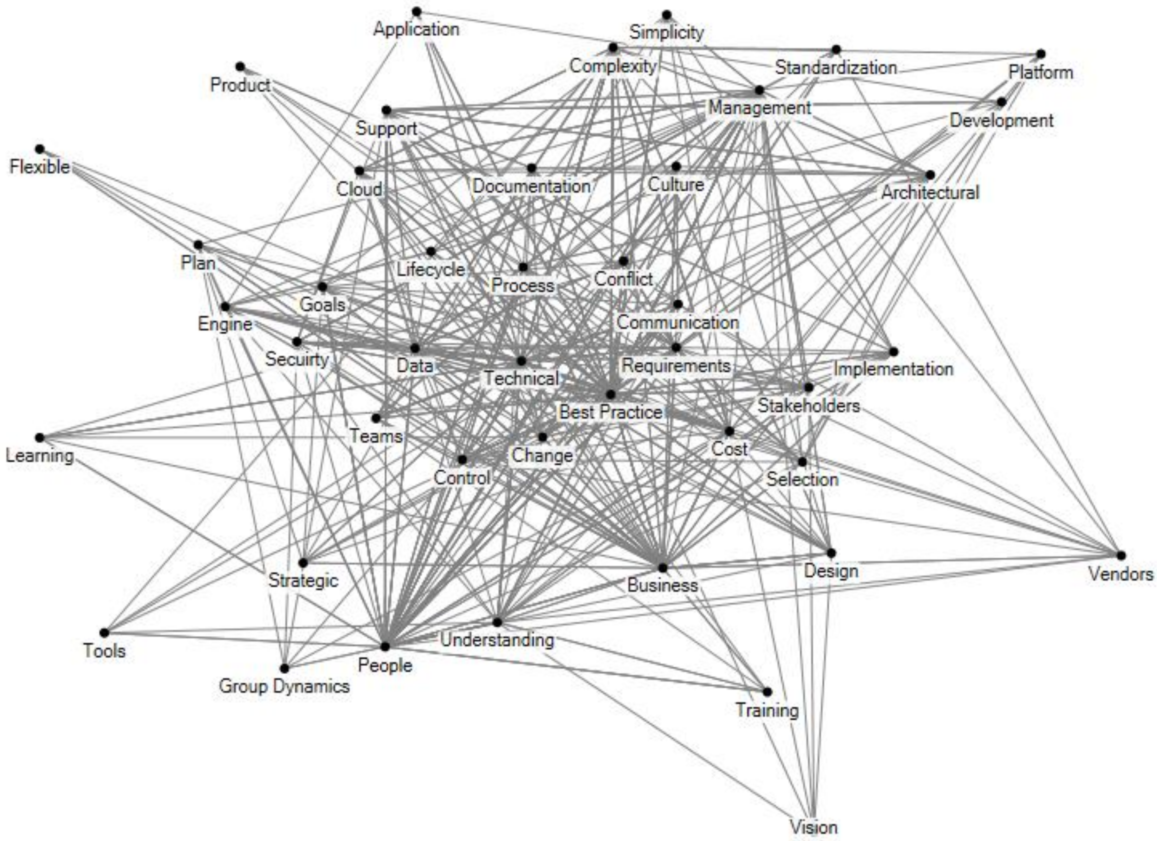
This guide takes a 360 degree view of the situation through a systems thinking lens, providing synthesis between the cross disciplinary fields. To be able to explain what complexity is shapes our understanding of the situation and a basic visualisation of this is shared through the use of a graph. The usage of graphs as visual representation are discussed with the presentation of the graph metrics leading to the CODEX, a blueprint for the management of database systems. The CODEX could enable transformation of the management of database systems so that actionable insight can be achieved.

What is Complexity

Complexity science, as defined by Johnson (2009, p.3), is the study of the phenomena which emerge from a collection of interacting objects. Throughout history and science, complexity has been discussed as a self-organizing network of interconnected and interdependent components that form a living system. Complexity exists within all stages in the database systems from inception. The interaction within the DBMS and external factors form the holistic Database System that makes this a complex environment.

Capra & Luisi (p.81) also argue the study of complexity is through qualitative analysis mapping relationships through visual patterns and the Beckman Report (Abadi *et al*. 2016) highlighted that visual analytics was essential to cope with large volumes of data in the database.

Complex interactions for the purpose of the research are three or more interactions that are linked with a single task or component. Analysis of the qualitative data and the systems mapping showed the complexity of database systems depicted in this undirected graph.



System Thinking

Systems thinking was used to advance understanding of the operation of database systems. Navigating through the ubiquitous database system incorporating management best practices, new technology, and information systems led to the adoption of a systems thinking approach to examine the interconnected components of information systems. Systems thinking has only infrequently been applied to the management of database systems. A system has been defined by a number of people and the definitions by Ackoff, Checkland and Senge set the scene for the perspective used.

Ackoff - A system
"A system is a set of two or more elements that satisfies the following three conditions.
(1) The behaviour of each element has an effect on the behaviour of the whole.
(2) The behaviour of the elements and their effects on the whole are interdependent. This condition implies that the way each element behaves and the way it affects the whole depends on how at least one other element behaves.
(3) However subgroups of the elements are formed, each has an effect on the behaviour of the whole and none has an independent effect on it."
(Ackoff 1981a, p.15)

- **Checkland - A system**
  "A model of a whole entity; when applied to human activity, the model is characterised fundamentally in terms of hierarchical structure, emergent properties, communication, and control. An observer may choose to relate this model to real-world activity. When application to natural or man-made entities, the crucial characteristic is the emergent properties of the whole." (Checkland 1999, pp.317–318)
- **Senge - Systems thinking**
  "The discipline for seeing wholes. It is a framework for seeing interrelationships rather than things, for seeing patterns of change rather than static "snapshots" [...] Today we need systems thinking more than ever because we are becoming overwhelmed by complexity. Perhaps for the first time in history, human kind has the capacity to create far more information than anyone can absorb, to foster far greater interdependency than anyone can manage, and to accelerate change far faster than anyone's ability to keep pace […] Systems thinking is a discipline for seeing the "structures" that underlie complex situations" (Senge 1990, pp.68–9)

What are Graphs?

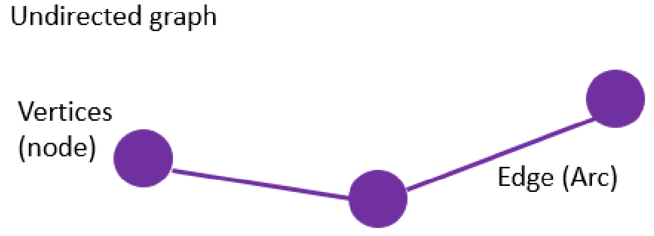
Graphs can provide visual representation of the ubiquitous complexity. The mathematician Euler began the study of graphs with his study of the 'Bridges of Königsberg' in the 18th century. The classic problem was based on seven bridges and two islands and how to cross each bridge only once from one starting location. Today graphs are used in social media to find who knows who, to find patterns of purchases through recommendations, detect fraudulent activity, understand network connectivity, and by haulage companies to calculate economical fuel routes.

Graph theory can look at complete networks where relationships may or may not be reciprocal (symmetric or asymmetric) using matrixes to record this. Different kinds of graphs are explored to investigate complex data patterns (Lenharth *et al*. 2016).

Graph theory algorithms can be useful when the data is sparse and the user needs to predict properties. This research problem to predict as accurately as possible what the list of complex components are for each change or new system setup, is a data science problem and needs to follow the data science process. Finding the best way forward could be through either a deterministic model or a probabilistic model. A deterministic model could describe exact outcomes from an experiment. In deterministic models every event has a cause. A probabilistic model could give a distribution of outcomes and the likeliness of an outcome occurring. The probabilistic models do not have all the information for a specific event. Probabilistic graphical models can be used with machine learning.

A graph consists of vertices and edges to show relationships between the objects. Vertices (nodes) are entities such as 'learning' and 'control'. Edges (arc) describe the relationship between vertices. They can be labelled relationship and have directions.

There are two types of graphs directed and undirected.

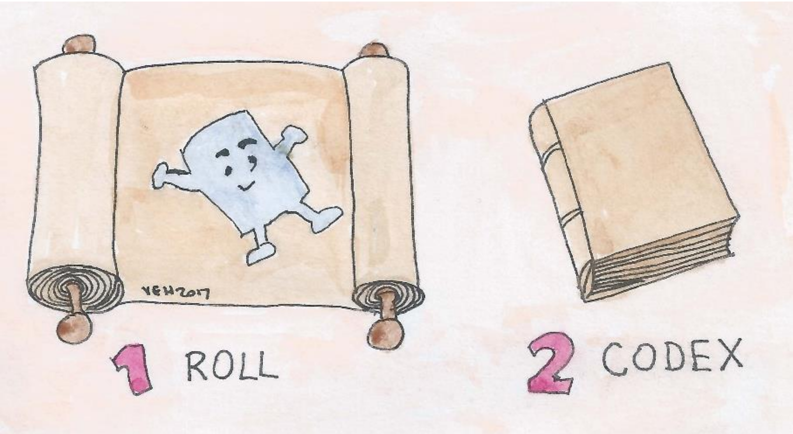


For ease of display, the global network metrics for the graph in this paper shows the complexity of management of database systems in an undirected graph.

A few of the graph metrics are shown in the table below.

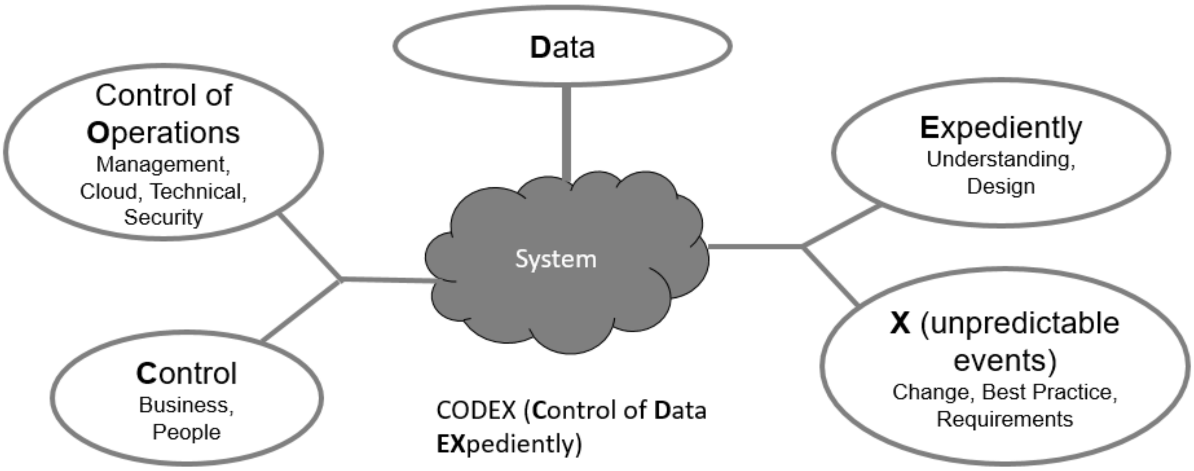
Graph Metric		Value
Graph Type	Directed or undirected	Undirected
Layout	The graph was laid out using the Harel-Koren Fast Multiscale layout algorithm.	
Vertices	The number of nodes in the graph	44
Unique Edges	The number of edges that do not have duplicates.	161
Edges With Duplicates	The number of edges that have duplicates, incident to the same two vertices.	459
Total Edges	The number of edges in the graph. This is the sum of Unique Edges and Edges With Duplicates.	620

The CODEX



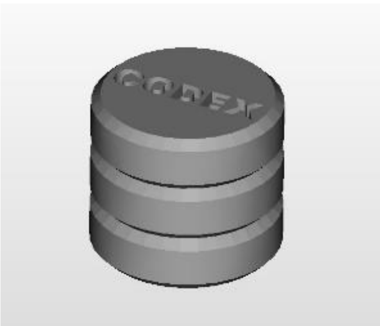
The CODEX is a blueprint for database management. The acronym CODEX has been selected by analogy with the revolutionary introduction of the Codex (Netz & Noel 2007, pp.69–85) in the first century AD which changed the storage medium from a roll to a Codex (book format). This brought challenges migrating the data, but significant benefits of increased speed of data access, reference and durability (of the parchment). Not all texts were migrated from rolls to Codex and those that were not migrated became defunct. Text case was changed from capitals to lowercase and minuscule copies made, resulting in further change; original majuscule manuscripts have not survived. This scholarly activity led to a revival in reading classic documents and a development of a centre of culture.

The CODEX is an output from the research, based on interpretation of the data



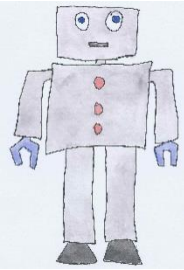
The database system CODEX

This suggested CODEX (Control of Data EXpediently) is a pattern to help improvement in the management of database systems. An important part of the CODEX is that over time the number of components and the complex interactions of components are likely to change. Continuous feedback is required from all elements of the system.



The introduction of a CODEX to help in the management of database systems consists of multiple inputs and was based, in the core, on best practice and complex component interactions. The CODEX is the usage of knowledge through control, operations, data, expediency and diverse environment variable factors.

Transformation of database systems



It is possible to transform and enhance the management of database systems. It is not enough to think of the transformation of the management of database systems as complete once the databases have been migrated to the cloud as DBaaS. The delivery of autonomously managed database estates are only the foundation stones. The data management estates are the walls, the infrastructure, and the people with data being the contents. This is finally covered with a roof including predictive analytics and artificial intelligence (AI) which enables actionable insights for organizations to make a home.

The principal findings of the research were:

- The plurality of sources of materials, technology, skills, teams of people, locations, and the related interconnectedness means that what were once straightforward management techniques need careful control.
- Constant, rapid change is one of the most difficult challenges in the management of database systems.
- Complexity makes it difficult to define and use best practices for the management of database systems.
- The satisfactory control of all components is a sociotechnical problem i.e. the people aspect is as important as the technology aspect.

Thus innovation based on knowledge, gained through the research, leading the way forward to artificial intelligence, will transform the management of database systems.

Methodology

The research consisted of a mixed method approach with a sequential explanatory design. A quantitative survey with participants from around the world was followed by a set of qualitative focus groups.

References

1. Abadi, D. et al., 2016. The Beckman Report on Database Research. *Communications of the ACM*, 59(2).
2. Ackoff, R.L., 1981a. *Creating the Corporate Future*, New York: John Wiley & Sons
3. Agrawal et al., 2009. The Claremont Report on Database Research. *Communications of the ACM*, 52(6), pp.56–65.
4. Capra, F. & Luisi, P.L., 2014. *The Systems View of Life*, Cambridge: Cambridge University Press.
5. Checkland, P.B., 1999. *Systems thinking, systems practice* Reprint 19., Chichester: John Wiley & Sons.
6. Johnson, N., 2009. Simply Complexity: a clear guide to complexity theory, Oxford: Oneworld Publications.

7. Netz, R. & Noel, W., 2007. The Archimedes Codex, London: The Orion Publishing Group Ltd.
8. Lenharth, B.Y.A., Nguyen, D. & Pingali, K., 2016. Parallel Graph Analytics. Communications of the ACM, 59(5).
9. Prell, C., 2012. Becoming familiar with social networks. In Social Network Analysis: history, theory & methodology. London: SAGE Publications Ltd, pp. 7–18.
10. Senge, P.M., 1990. *The Fifth Discipline: The Art and Practice of The Learning Organization* Reprint., Chatham: Mackays of Chatham.

Dr Victoria Holt

Database Researcher

T: @victoria_holt

L: https://www.linkedin.com/in/victoriaholt/