Master's thesis

International Business Management

2014

Irina Pennanen

DATA GOVERNANCE

– Intelligent way of managing data

TURUN AMMATTIKORKEAKOULU
TURKU UNIVERSITY OF APPLIED SCIENCES

MASTER'S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

International Business Management

2014 | 63

Instructors: Laura Heinonen, Matti Kuikka

Irina Pennanen

DATA GOVERNANCE

– Intelligent way of managing data

Today's business is run by data. Still there are many companies that don't think data as a business critical asset. How have we ended up to this situation? World is changing all the time and companies should change too. What are the benefits of governing data well? How should it be done in organizations?

This thesis is trying to point out the background for this problem, find out what companies can do to make situation better and what are the advantages to business that can be achieved with better data management.

KEYWORDS:

Data, information, data governance, data management

OPINNÄYTETYÖ (YAMK) | TIIVISTELMÄ

TURUN AMMATTIKORKEAKOULU

International Business Management

2014 | 63

Ohjaajat: Laura Heinonen, Matti Kuikka

Irina Pennanen

DATA GOVERNANCE

– Älykäs tapa hallita tietoa

Yritystoiminta perustuu tietoon, joka on nykyään lähes aina sähköisessä, tietokoneiden ymmärtämässä ja tulkitsemassa muodossa. Koska tietokoneet eivät ajattele kuten ihmiset, on tiedolle määritettävä muoto ja opetettava tietokone ymmärtämään sitä. Mikäli muoto vaihtelee sovitusta, tietokoneet ja –järjestelmät eivät toimi.

Yrityksen panostavat koneisiin, rakennuksiin, henkilökuntaan, mutta panostus tiedon laatuun on suhteellisen vähäistä. Harvassa yrityksessä ymmärretään tiedon laadun merkitys yrityksen toiminnalle ja vielä harvemmassa on sitä varten organisaatiossa roolit ja määritykset ja tuloksia seurataan.

ASIASANAT:

Tieto, informaatio, tiedonhallinta, tiedon hallitseminen, tiedon laatu

PICTURES

FIGURES

TABLE

LIST OF ABBREVIATIONS (OR) SYMBOLS

OED	Oxford English Dictionary
SOA	Service Oriented Architecture
SAP	client/server enterprise application software
ERP	Enterprise Resource Planning –system
PDM	Product Data Management –system
DQM	Data Quality Management

1 INTRODUCTION – OBJECTIVES OF THE THESIS

Companies are investing money to new factories and machines, to product development, systems and personnel. But there are only few companies in Finland investing to the management of data, although companies are run by data.

Why companies should invest to data management? What are the advantages a company can achieve if doing that? Management is not usually thinking data as a business asset that it actually is. Especially in smokestack industry data is seen like an enemy: something weird that management can't comprehend - complicated IT stuff. It was handled in IT departments for a long time, but today IT tasks are centralized and more often outsourced and the data management is transferred under the business organization (where it actually belongs).

Data problems come up usually when company is buying a new system and implementing that. Good example is ERP (Enterprise Resource Planning) system implementation: prices of ERP's are extremely high, and you can only buy the system from the store. But the system is run by the data, and if the data is wrong, it seems to the management that system is not working.

I have faced these problems in my career in Metso Automation. I have been involved in several system projects and ERP roll-outs. All the time the problem is the same: data quality or actually lack of data quality. In a global company like Metso Automation it is extremely hard to find out the data owners: who is responsible of certain data content. There are a huge amount of users who create data to the systems and all the rules and roles should be clear to everyone. Metso Automation has grown a lot during the years, mostly through acquisitions. Several companies have combined as a one company and that is the moment when the problems start with data management. Suddenly there are hundreds of users creating data instead of tens.

From the system point of view there is of course user rights that defines when someone is allowed to only view data in the system or also editing it. But because there are hundreds of editors needed, company should create a governance model for managing the data. Typically there are a lot of fields in ERP which are mandatory to fill in. For example the acquisition code: is certain product or part a purchased one or a manufactured one? User fills the mandatory data but system can't know if the information is correct or not.

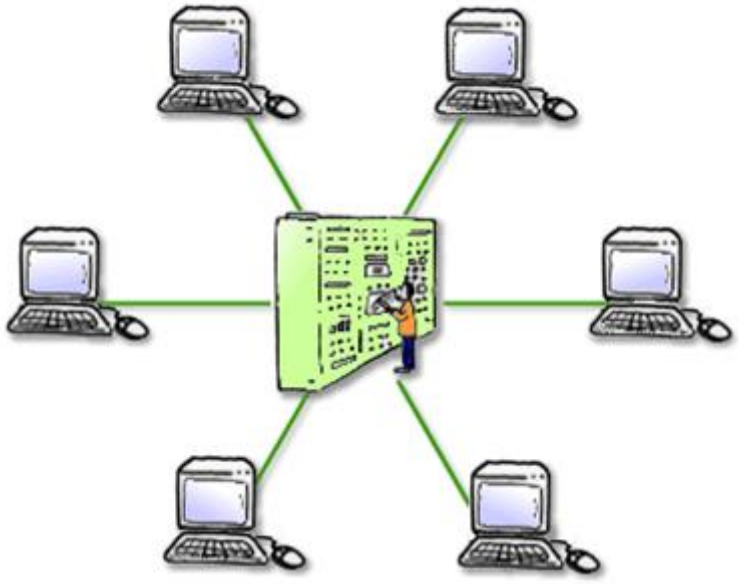
If you don't have clear rules how to fill the information, you can't trust it and system is not working the way people are expecting it to work. Will the rules and roles help then? When making data governance official it means that management can also monitor the data quality. If every editor has certain role and responsibility, it can be included to person's annual targets: like data correctness must be over 95 percent. When people have targets and it is measured, it motivates to do things right.

This thesis is trying to open up this issue: not from the IT point of view but from the business point of view. What are the advantages a company will achieve with governing data and investing to data management?

2 THE HISTORY AND THE FUTURE OF DATA, COMPUTERS AND INFORMATION

"Information is the oil of the 21st century, and analytics is the combustion engine," presented Mr. Peter Sondergaard, Senior Vice President at Gartner and Global head of Research, in ITxpo 2011 in Orlando. "Pursuing this strategically will create an unprecedented amount of information of enormous variety and complexity." (http://www.gartner.com/newsroom/id/1824919)

Industrialization started in the end of 1800's and by the mid of 1950's it was clear that automatization is next big thing. So computers started to do some things on behalf of humans. This required extremely stable data to "teach" the computer to do tasks. Because of the criticality of data only few people had access to it. This was called centralized data handling or mainframe usage, see Picture 1:



Picture 1. Use of central mainframe.

2.1 Mainframes

Mainframes were used before microcomputers became general. They were usually very expensive, powerful and operated by special software. Mainframes were typically used by large companies, public authorities and universities for their data handling tasks. Typical tasks done with mainframes:

- *File maintenance*: This is perhaps the most common use of mainframes. Maintaining records is a huge task for institutions. Records can contain information on sales, credit card status, payroll details, social security details, health records, stock inventory, etc. These either need to be accessed by different people in real-time (for instance a travel agent booking an airline ticket) or updated in batches (for instance warehouse stock levels at the end of each day). In such cases it is necessary to have the data stored centrally and with accessibility for those who need it. A lot of minicomputers are now capable of performing these tasks in medium-sized companies.

- *Simulations*: Many physical and engineering problems cannot be solved without the help of complex computer simulations. These require intensive mathematical work, and so take advantage of a mainframe's computational power. Examples include weather forecasting, or calculating the position of astronomical bodies with extreme accuracy. Many minicomputers or workstations are now used for this type of problem.

- *General purpose*: Many universities used a mainframe to act as a general purpose computing facility. Each user can then be given their own area on the mainframe to store files, and different departments can use its resources to perform different tasks, e.g. predicting bird populations in the Biology department and calculating metal stress in the Engineering Department. PCs are now used to perform many of these tasks. (http://labspace.open.ac.uk/mod/oucontent/view.php?id=426285&printable=1)

Companies used to have own data organizations that only handled data input and also served rest of the organization to have outputs of the data. This was the way to keep data quality in high level, 100 % valid. This was working perfect, but of course structures of organizations were quit heavy.

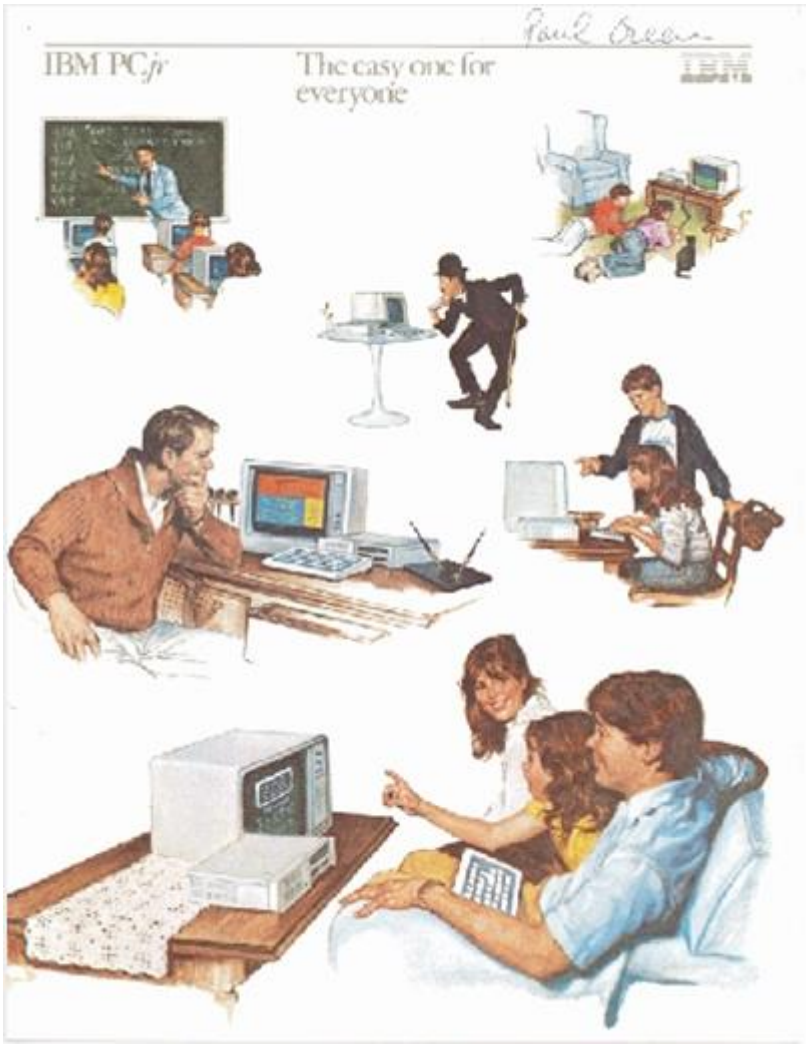
2.2 Personal computers

Personal computers (PC's) became more general in 1980's and centralized model of data handling were surrendered. Each and every one had their own PC and availability to update data according the user rights. Unfortunately this has led us to the situation where company's business critical information is updated by anyone, without common rules and follow-up.

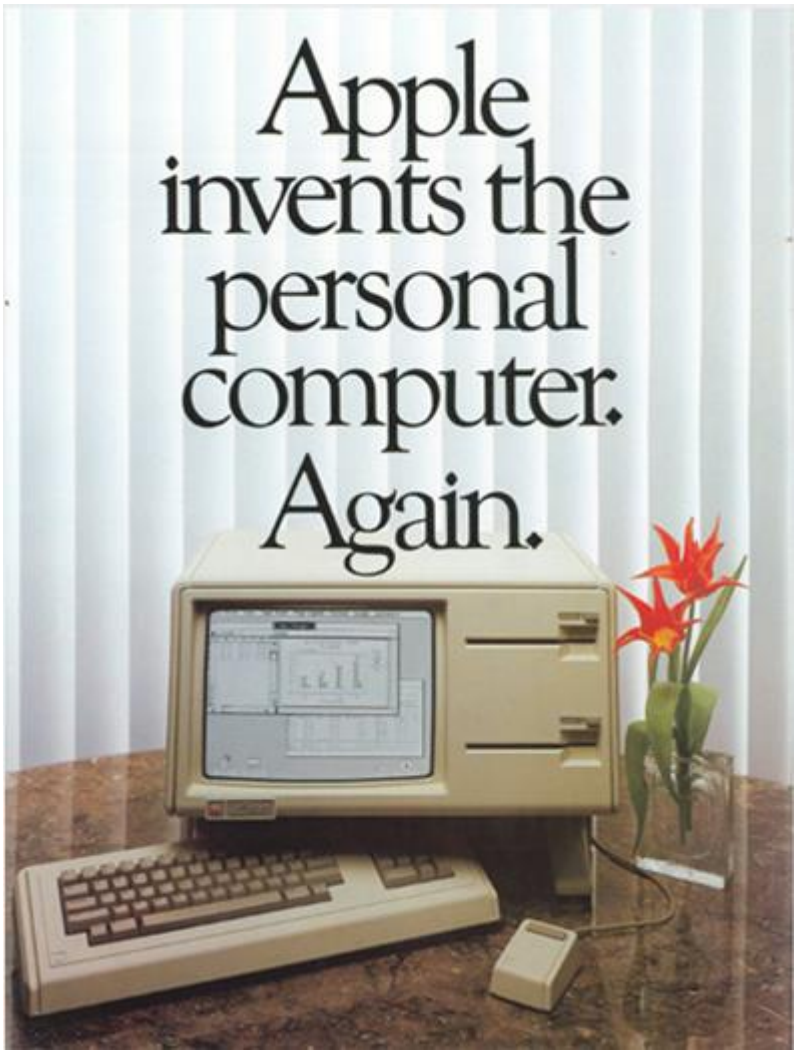
The term "personal computer" has been applied to a wide variety of machines (often in hindsight) where an individual user would have direct control of the entire computer (e.g., the LGP-30, Bendy G-15, and others). (http://www.computerhistory.org/brochures/categories.php?category=thm-42b97f98dbaf2)

However the true personal computer (as we know it today, that is a mass market item found in both home and office settings) had to await the development of the integrated circuit CPU in the form of the microprocessor and, more directly, its appearance in machines such as the Radio Shack TRS-80, the Apple II, and Commodore in 1977. When IBM introduced its Personal Computer (PC) in 1981, a slow shift in perception began in which the personal computer changed from being viewed as a toy to a business tool. Today most personal computers have much greater computational power than even the most powerful mainframes of only a few decades earlier, Pictures 2 and 3.

(http://www.computerhistory.org/brochures/categories.php?category=thm-42b97f98dbaf2)



Picture 2. PCjr, The easy one for everyone.

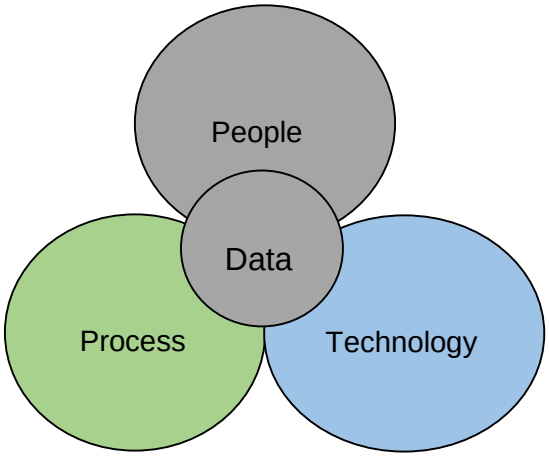


Picture 3. Apple invents the personal computer. Again.

In a small company this worked fine: people were sitting next to each other and if the data creator wrote something wrong the data user asked about it and the data was fixed. But then companies started to do acquisitions and manufacturing in other locations, meaning that companies became global. That was the point when the data quality turned out its significance.

2.3 Business dimensions

Companies have been measuring their business along three dimensions for many years: people, process and technology. Strout and Eisenhauer have strong opinion that there is a fourth dimension: data. They believe that data is like blood in the body: its existence is absolutely necessary to live. By investing heavily in people, process and technology companies were ensuring their operational efficiencies and gaining insights to their business and maintaining better control of their business. Thus, data is to operations as blood is to our body, see picture 4. (Strout&Eisenhauer, 2011, 3)



Picture 4. The fourth dimension.

We've all heard it for years: "Garbage in, garbage out." Unfortunately only few businesses have put in place the appropriate processes to ensure high data quality on a consistent basis.

Data is a very valuable asset of any business. It is owned by business and it should be treated as such. This means that it must be protected, guarded, managed and governed in such a way that it retains or increases in value. Companies spend a lot of money for doing

it other business assets like vehicles, plants, brands, copyrights and patents. John Eisenhauer has been asking the very proactive question in Data Governance Society meetings: “If data has an economic value, why shouldn’t we put data on the balance sheet?” Exactly, should we? We spend a lot of money for creating data, maintaining it and protecting it. Do we get the most value from it? (Strout&Eisenhauer, 2011, 4)

3 DATA AND INFORMATION

Information versus data… Information is something that gives you the overall description of the area of interest. One example can be time tables: you can have the data of departing time: 15.00PM. Okay, that is valid data, but it does not create any information to you. To get the information you need all pieces of data:

1. the type of vehicle (train)
2. the place of depart (Helsinki, Pasila)
3. the date of depart (1.1.2014).

When you add here the fact that you know what is a train and where Helsinki and Pasila are located and how to get there, you have the needed knowledge to use that train.

So you may say that information is something you can use to do things. And to get it, you need lot of data that is valid and only together this data creates useful information for you.

Peter R. Benson, Project Leader for ISO 8000, the International Standard for Data Quality and the Founding and Executive Director of the Electronic Commerce Code Management Association (ECCMA) has defined data and information:

"*The terms (words) data and information are often used interchangeably as synonyms. Understanding that they are in fact different concepts is important to understanding data quality and data governance. You cannot copyright information, only data. Until a song or a performance is translated into fixed form as data, it cannot be protected under the laws of copyright. The law of copyright is important to data in many ways, specifically the concept of a joint work where the work of more than one author is included in a work. Data is rarely the work of a single author and tracking what is and is not a joint work can be challenging.* (Benson 2012, 23)

Thomas Stearns Eliot (1888-1965) (http://en.wikipedia.org/wiki/T._S._Eliot) wrote a famous poem “The Rock” that Benson is referring in his book Managing Blind:

> “Where is the wisdom we have lost in knowledge?
>
> Where is the knowledge we have lost in information?”

It is common to see the relationship between data and information represented as a pyramid with data as the base, rising through information and knowledge, to wisdom as the apex:

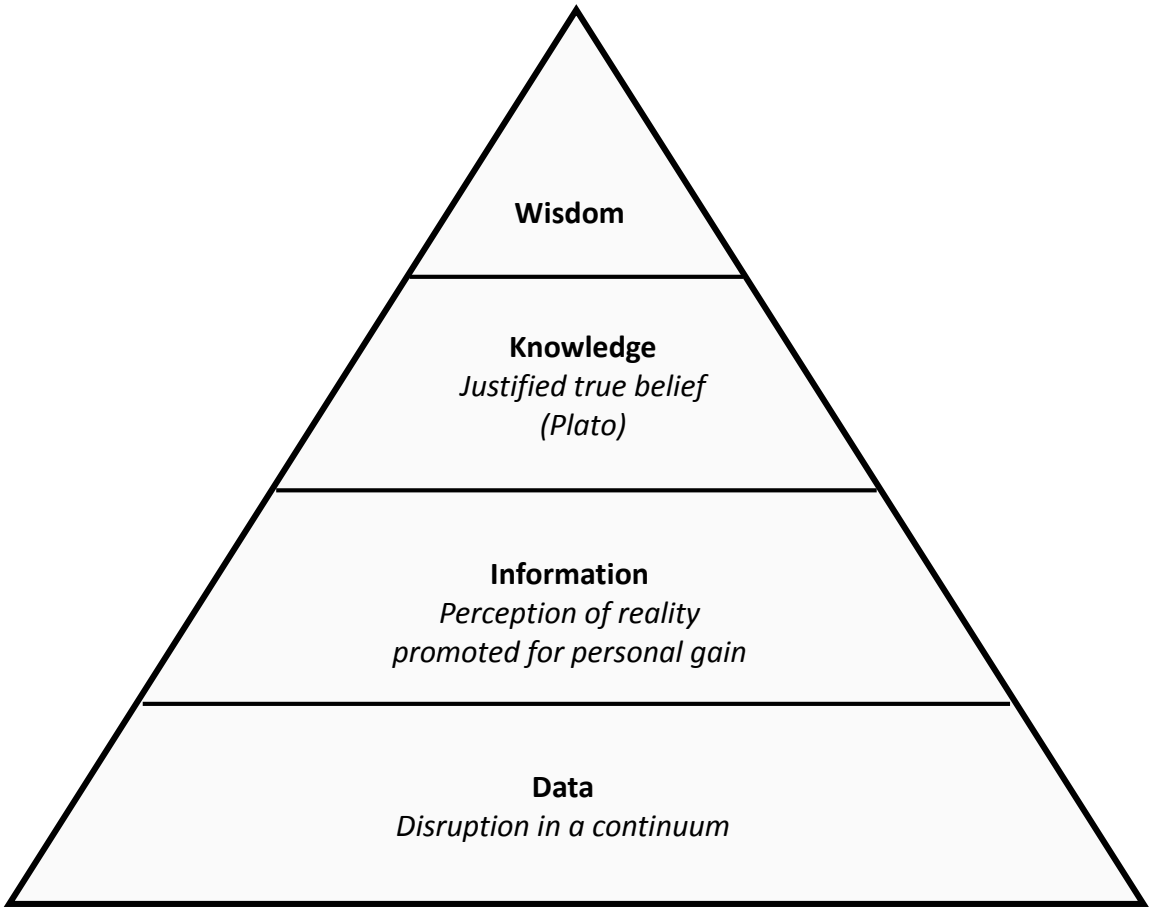


Figure 1. The The Rock (Benson 2012).

Of course data was conspicuous by its absence in T.S. Eliot's poem but it is not hard to understand the omission given that the first "freely programmable computer," the Z1 Computer, was only invented two years later in 1936 by Konrad Zuse. It was not until the early 1950s, 25 years later that we saw the first commercial computers.

Because of the close relationship between data and information it is actually very challenging to find good definitions. According to the international standard for writing definitions, ISO 704:2009, a good way to test a definition is to substitute it for the term in a sentence. That is why definitions are written as fragments and not sentences without a preposition or with a capital at the beginning and a period at the end.
A circular definition is one where when you substitute the term for its definition you end up with the term used to define itself. (Benson 2012, 24)

Benson has also said that "data is what data is and information is what you make of it". The characteristic that determine the quality of data will be inherent to the data itself, while the characteristics that determine the quality of information require a third party perspective or opinion. (Benson 2012, 29)

3.1 What is data?

"Data is the raw material for what can become information" states Steven Strout and John Eisenhauer in their book "The elephant in the room: data". (Strout&Eisenhauer, 2011, 7)

ISO 8000-2:2012(E) standard defines 'data':

> symbolic representation of something that depends, in part, on its metadata for its meaning and other ISO/IEC standard 2382-1:1993, definition 01.01.02:
>
> re-interpretable representation of information in a formalized manner suitable for communication, interpretation, or processing

Oxford dictionary describes word ‘data’:

- facts and statistics collected together for reference or analysis:

  there is very little data available

- the quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.
- Philosophy things known or assumed as facts, making the basis of reasoning or calculation.

(http://www.oxforddictionaries.com/definition/english/data?q=data)

Peter R. Benson uses the following definitions for a word data:

> **Data:** application processable representation of
>
> or
>
> **Data**: elements into which information is transformed so that it can be stored or moved

Data is by its very nature, a historical record. Data is a presentation of entities and events. Entities are individuals, organizations, locations, goods, services, processes, procedures, rules and regulations as you can see for the figure 2:

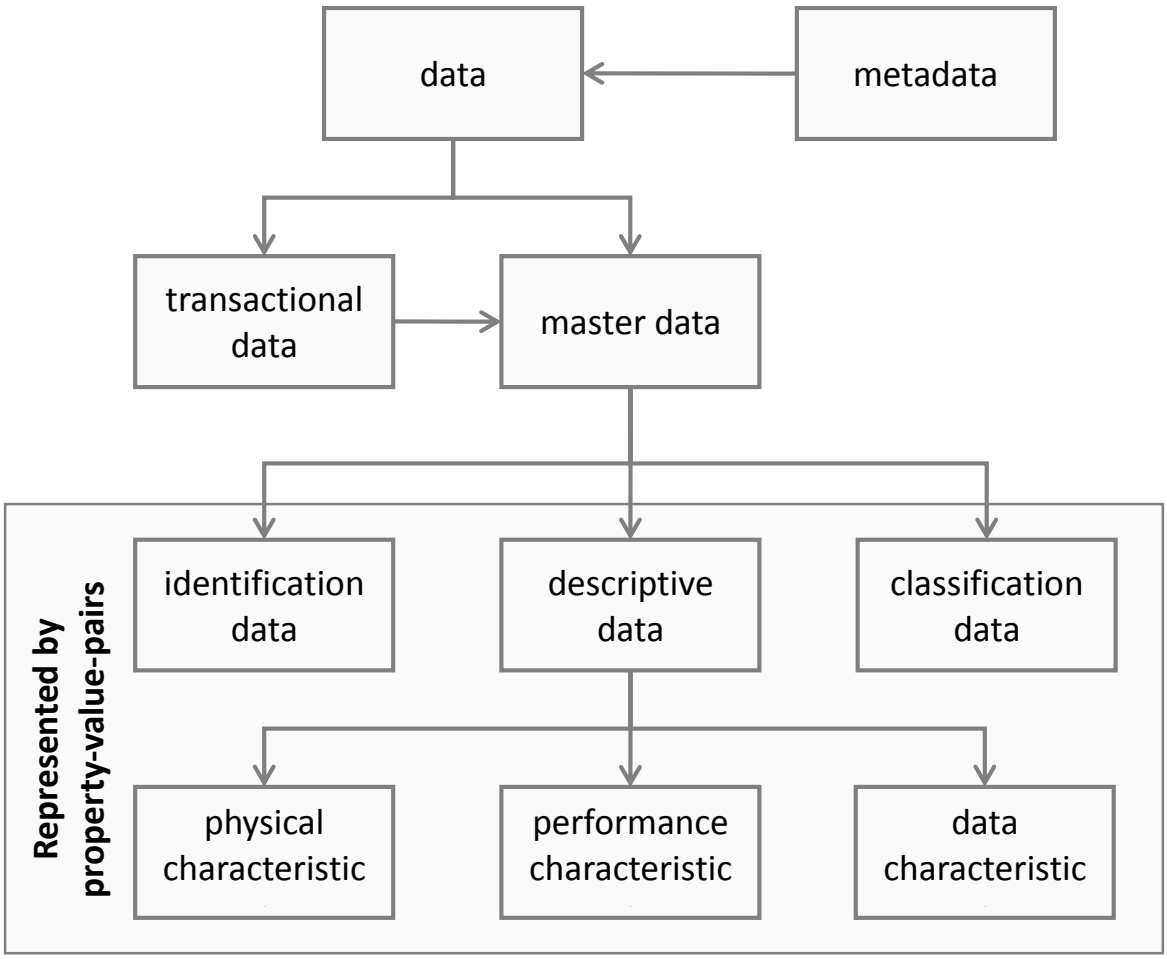


Figure 2. A simplified taxonomy of data (Benson, 2012).

T.S. Eliot's poem, "where is wisdom we have lost in knowledge and where is the knowledge we have lost the information," the answer is simple – it is in the data! Data is what we use to transfer wisdom, knowledge and information. So data quality and governance are absolutely critical. (Benson 2012, 27).

Luciano Floridi has defined data in his theoretical article "Semantic Conceptions of Information" very philosophically:

The Diaphoric Definition of Data (DDD):

A datum is a putative fact regarding some difference or lack of uniformity within some context.

Depending on philosophical inclinations, DDD can be applied at three levels:

1. data as diaphora de re, that is, as lacks of uniformity in the real world out there. There is no specific name for such "data in the wild". A possible suggestion is to refer to them as dedomena ("data" in Greek; note that our word "data" comes from the Latin translation of a work by Euclid entitled Dedomena). Dedomena are not to be confused with environmental data. They are pure data or proto-epistemic data, that is, data before they are epistemically interpreted. As "fractures in the fabric of being" they can only be posited as an external anchor of our information, for dedomena are never accessed or elaborated independently of a level of abstraction. They can be reconstructed as ontological requirements: they are not epistemically experienced but their presence is empirically inferred from (and required by) experience. Of course, no example can be provided, but dedomena are whatever lack of uniformity in the world is the source of (what looks to information systems like us as) as data, e.g., a red light against a dark background. Note that the point here is not to argue for the existence of such pure data in the wild, but to provide a distinction that will help to clarify why some philosophers have been able to accept the thesis that there can be no information without data representation while rejecting the thesis that information requires physical implementation;

2. data as diaphora de signo, that is, lacks of uniformity between (the perception of) at least two physical states, such as a higher or lower charge in a battery, a variable electrical signal in a telephone conversation, or the dot and the line in the Morse alphabet; and

3. data as diaphora de dicto, that is, lacks of uniformity between two symbols, for example the letters A and B in the Latin alphabet. (Floridi 2013)

3.2. What is information?

Strout and Eisenhauer wrote that “data is the raw material for information”. (Strout&Eisenhauer, 2011, 7). So that means that information is then the “end product”, something that has been manufactured from data.

Oxford dictionary defines **information**:

1. facts provided or learned about something or someone:

   a vital piece of information

   - *[count noun] Law a charge lodged with a magistrates' court: the tenant may lay an information against his landlord*

2. what is conveyed or represented by a particular arrangement or sequence of things:

   genetically transmitted information

   - *computing* data as processed, stored, or transmitted by a computer.
   - (in information theory) a mathematical quantity expressing the probability of occurrence of a particular sequence of symbols, impulses, etc., as against that of alternative sequences.

(http://www.oxforddictionaries.com/definition/english/information?q=information)

ISO 9000:2005, definition 3.7.1 explains information very simple way:

> Information is meaningful data.

ISO/IEC 2382-1:1993, definition 01.01.01 describes information:

> Knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts that within a certain context has a particular meaning.

Luciano Floridi has also defined information in his theoretical article “Semantic Conceptions of Information”:

“It is common to think of information as consisting of data. It certainly helps, if only to a limited extent. For, unfortunately, the nature of data is not well-understood philosophically either, despite the fact that some important past debates — such as the one on the given and the one on sense data — have provided at least some initial insights. There still remains the advantage, however, that the concept of data is less rich, obscure and slippery than that of information, and hence easier to handle. So a data-based definition of information seems to be a good starting point.

Over the last three decades, several analyses in Information Science, in Information Systems Theory, Methodology, Analysis and Design, in Information (Systems) Management, in Database Design and in Decision Theory have adopted a General Definition of Information (GDI) in terms of data + meaning. (GDI has become an operational standard, especially in fields that treat data and information as reified entities (consider, for example, the now common expressions “data mining” and “information management”). Recently, GDI has begun to influence the philosophy of computing and information.

A clear way of formulating GDI is as a tripartite definition:

The General Definition of Information (GDI):

σ is an instance of information, understood as semantic content, if and only if:

(GDI.1) σ consists of one or more data;

(GDI.2) the data in σ are well-formed;

(GDI.3) the well-formed data in σ are meaningful.

GDI requires a definition of data. According to (GDI.1), data are the stuff of which information is made. We shall see that things can soon get more complicated. (Floridi 2013)

So as a conclusion we can say that information is pieces of data collected together and interpreted by human being. In companies there are a lot of data: cost centers, accounts, routings, machines, items, documents, products, parts, persons, salaries. To get the facts how company is doing, you need not just collect, but connect all this data together to get the information needed. This makes all data critical: if you have exactly correct time for depart but you are in a bus station instead of a train station that data is worth of nothing.

4 DATA QUALITY

How to define data quality? Quality is probably one of the most misunderstood concepts. We all know what quality is, yet we cannot really define it. (Benson 2012, 36)

ISO 9000, the standard that contains the terminology for the 9000 series of standards contains a definition for the term quality:

> "degree to which a set of inherent characteristics fulfills requirements"

This introduces two very important concepts; first the "degree to which", something we can measure and secondly, the most important "fulfills requirements."

Quality is about "fulfilling requirements." We cannot measure data quality unless we can specify the requirements for data.

The OED defines "requirement" as: a thing that is needed or wanted or a thing that is compulsory; a necessary condition. The definition in ISO 9000 simply adds that the need or expectation must be "stated" and this is important.

ISO 9001 is the standard that contains the clauses that you must comply with if you wish to be ISO 9001 compliant. The difference between ISO 9001 and ISO 8000 is how the requirement is "stated" and how compliance is measured. To be compliant with ISO 9001 you must have "documented" the requirements. ISO 8000 takes this one step further. ISO 8000 mandates that requirements must be "stated" in a computer processable form. This is critical to ISO 8000; after all it is a standard about computer processable data, so it makes sense that the compliance with requirements must be capable of being performed by a computer and not by a person with a clip board. (Benson 2012, 38-39)

Oxford English Dictionaries (http://www.oxforddictionaries.com/definition/english/quality?q=quality) defines ***quality*:**

noun (plural **qualities**)

- *[mass noun]* the standard of something as measured against other things of a similar kind; the degree of excellence of something: *an improvement in product quality [count noun]:people today enjoy a better quality of life*
- general excellence of standard or level: *a masterpiece for connoisseurs of quality*
- (usually **qualities**) *British* short for quality paper.
- *archaic* high social standing: *commanding the admiration of people of quality*
- *[treated as plural] archaic* people of high social standing: *he's dazed at being called on to speak before quality*
- distinctive attribute or characteristic possessed by someone or something: *he shows strong leadership qualities the plant's aphrodisiac qualities*
- *Phonetics* the distinguishing characteristic or characteristics of a speech sound.
- *Astrology* any of three properties (cardinal, fixed, or mutable), representing types of movement, that a zodiacal sign can possess.

adjective

- *informal*
- *of good quality; excellent: he's a quality player*

Data and information quality are now widely recognized problems in companies large and small, ranging from manufacturing and processing, to finance and health care. Incomplete or duplicate records, poor quality descriptions and inaccurate information cause inefficient allocation and use of resources. This can add up to a 20% increase to direct and indirect costs. Poor quality data is a barrier to effective marketing and the leading cause of

transparency issues that drive up the cost of regulatory compliance. (http://www.eccma.org/iso8000/iso8000home.php)

Already Albert Einstein (1879-1955) (http://einstein.biz/biography.php) understood the problem when he defined insanity as “doing the same thing over and over again and expecting different results.” Solving data quality and governance issues requires change, and changing a stable system requires effort and effort requires motivation. Individuals are motivated by greed and fear. In order to address the issues of data quality and governance we must be clear about our motives, our goals and our objectives. We must also find ways to translate these in ways that will motivate others. (Benson 2012, 5).

From a quality perspective, only two moments matter in a piece of data’s lifetime: the moment it is created and the moment it is used. The quality of data is fixed at the moment of creation. But we don’t actually judge that quality until the moment of use. If the quality is deemed to be poor, people typically react by working around the data or correcting errors themselves. But improving data quality isn’t about heroically fixing someone else’s bad data. It is about getting the creators of data to partner with the users — their “customers” — so that they can identify the root causes of errors and come up with ways to improve quality going forward. (http://hbr.org/2013/12/datas-credibility-problem/ar/1)

The good news is that a little communication goes a very long way. Time and time again, in meetings with data creators and data users, I’ve heard “We didn’t know that anyone used that data set, so we didn’t spend much time on it. Now that we know it’s important, we’ll work hard to get you exactly what you need.” Making sure that creators know how data will be used is one of the easiest and most effective ways of improving quality.

Even better news is that addressing the vast majority of data quality issues does not require big investments in new technologies or process reengineering. To be sure, disciplined measurement, automated controls, and methodologies like Six Sigma are helpful, particularly on more sophisticated problems, but the decisive first step is simply getting users and creators of data to talk to each other. (http://hbr.org/2013/12/datas-credibility-problem/ar/2)

Once a company realizes that its data quality is below par, its first reaction is typically to launch a massive effort to clean up the existing bad data. A better approach is to focus on improving the way new data are created, by identifying and eliminating the root causes of error. Once that work has been accomplished, limited cleanups may be required, but ongoing cleanup will not. (http://hbr.org/2013/12/datas-credibility-problem/ar/2)

Very often, data creators are not linked organizationally to data users. Finance creates data about performance against quarterly goals, for example, without considering how Sales will want to use them or Customer Service analyzes complaints but fails to look for patterns that would be important to product managers.

When quality problems become pervasive or severe, the organizational response is often to task the IT department with fixing them, usually by creating a special unit in the group to spearhead the initiative. This may seem logical, since IT is a function that spans all silos. But IT departments typically have little success leading data quality programs. That's because data quality is fixed at the moment of creation. With rare exceptions, that moment does not occur in IT. To address problems, IT people can talk to creators and users, but they can't change the offending business processes. All they can do is find and correct errors, which, as we've seen, is not a long-term solution. (http://hbr.org/2013/12/datas-credibility-problem/ar/3)

Smart companies place responsibility for data quality not with IT but with data creators and their internal data customers. For most companies, the real barriers to improving data quality are that some managers refuse to admit their data aren't good enough, and others simply don't know how to fix poor-quality data. The first bit of progress occurs when a manager somewhere in the organization (possibly a senior executive, but more often someone in the middle) gets fed up and decides that "there has to be a better way." The manager launches a data program and, if the prescriptions noted here are followed, usually gets good results. (http://hbr.org/2013/12/datas-credibility-problem/ar/3)

4.1 ISO 8000 – The international standard for data quality

A standard is a document that provides requirements, specifications, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose. (http://www.iso.org/iso/home/standards.htm)

Standard are developed by committees and they represent negotiated compromises between domain experts. Standards therefore represent the consensus of domain experts. Under the ISO procedure a standard must go through a series of ballots. An important part of the ISO process is ballot comment resolution. A committee must address and answer all the comments raised by member of countries through their national Technical Advisors Group (TAG). In the end, a two thirds majority is required to publish an international standard. (Benson 2012, 29)

What are the benefits of using standards? Standards ensures that products and services are safe, reliable and of good quality. For business, they are strategic tools that reduce costs by minimizing waste and errors, and increasing productivity. They help companies to access new markets, level the playing field for developing countries and facilitate free and fair global trade. (http://www.iso.org/iso/home/standards.htm)

International standards bring technological, economic and societal benefits. They help to harmonize technical specifications of products and services making industry more efficient and breaking down barriers to international trade. Conformity to International Standards helps reassure consumers that products are safe, efficient and good for the environment. (http://www.iso.org/iso/home/standards/benefitsofstandards.htm)

In ISO 8000 development the committee identified five characteristics of data that determine data quality. These characteristics are:

- syntax
- semantic encoding
- meets requirements
- provenance
- accuracy
- completeness

(Benson 2012, 28-29)

In ISO 8000-1:2011(E) standard defines:

- data quality involves data being fit for purpose, i.e. decision it is used in
- data quality involves having the right data in the right place at the right time
- data quality involves meeting agreed customer data requirements
- data quality involves preventing the recurrence of data defects by improving processes to prevent repetition and eliminate waste. (ISO 8000-1:2011(E), 6 Principles of data quality)

4.2 Data quality management

ISO 8000-150 Data quality – Master data: Quality management framework describes:

“The ability to create, collect, store, maintain, transfer, process and present data to support business processes in a timely and cost both an understanding of the characteristics of the data that determine its quality, and an ability to measure, manage and report on data quality.” (ISO 8000-150:2011(E), 6) effective manner requires

Enterprises need Data Quality Management (DQM) to respond to strategic and operational challenges demanding high-quality corporate data. Hitherto, companies have mostly assigned accountabilities for DQM to Information Technology (IT) departments. They have thereby neglected the organizational issues critical to successful DQM. With data governance, however, companies may implement corporate-wide accountabilities for DQM that encompass professionals from business and IT departments. (Weber et al. 2009)

To manage the master data quality successfully, organizations shall keep the following fundamental principles:

- Involvement of people: people at all levels who have roles for data quality management are involved to improve data quality of an organization. Although data processing of end users with lower-level role has the most direct effect on data quality, intervention or control of data administrators with middle-level role is required to implement and settle down processes for data quality improvement in the organization. In addition, involvement of managers who are in charge of organization-wide data quality with high-level role is inevitable to change and optimize roles, authority, and processes of the organization.

- Process approach: data-centric measurement and correction is not enough to improve data quality of the whole organization. Desired data quality is achieved more efficiently when activities and related resources for data quality are managed by processes.

- Continual improvement: data quality is improved continuously through the processes of data processing, data quality measurement and data error correction. However, with these processes only, identical data errors that occur repeatedly cannot be prevented. Recurrence of data errors
  can be prevented when the processes to analyze, trace and improve root causes which hinder data quality goes with these processes. For this, management processes concerned with data architecture/schema, data stewardship and data flow shall also be supported. In addition,

organizations shall improve not only processes for data quality management but also business processes where data are directly operated.

- Master data exchange: all processes to manage master data quality comply with requirements that can be checked by computer for the exchange, between organizations and systems, of master data that consists of characteristic data.

Master data quality management framework is built for three processes and three roles as can be seen in figure 3:

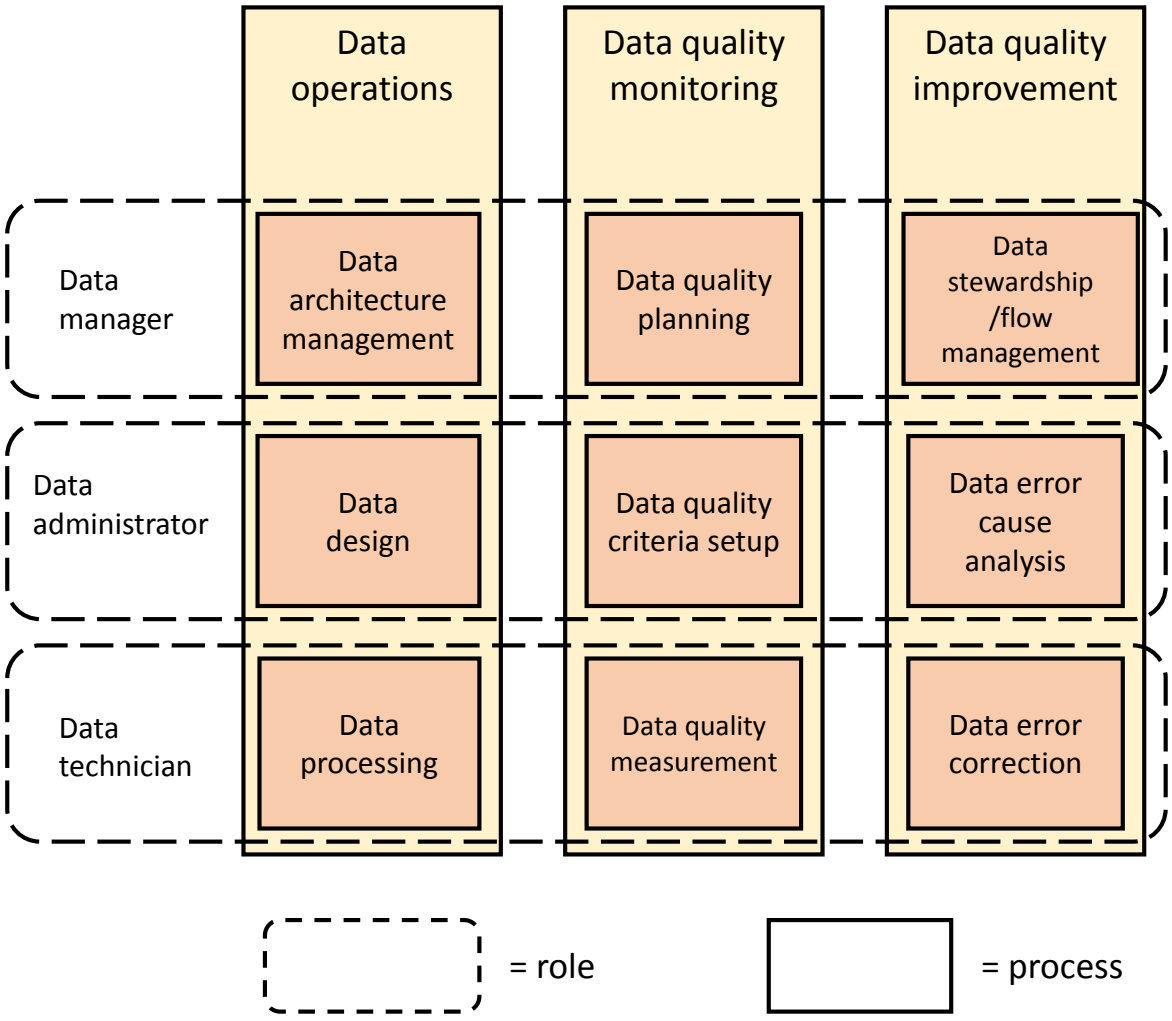


Figure 3. Master data quality management framework. (ISO 8000)

4.2.1 Three data quality management processes

The three top-level processes in the framework shall be:

- data operations
- data quality monitoring
- data quality improvement

The data operations process identifies factors that affect data quality and ensures data is available at the right place in a timely manner. This top-level process shall consist of the following processes:

- data architecture management; the process that manages organization-wide data architecture from the integrated perspective to use data in distributed information systems with consistency and therefore ensure data quality
- data design; the process that designs data schema, and implements a database to make data users apply data without mistake and ensure data quality
- data processing; the process that creates, searches, updates, deletes data in accordance with guidelines of data operations

The data quality monitoring process identifies data errors through a systematic approach. This top level process shall consist of the following processes:

- data quality planning; the process that sets up objectives of data quality in alignment with the strategies of an organization, identifies factors to be managed, and performs actions in order to accomplish objectives. This process also includes assurance of data quality and adjustment of objectives on the back of assurance results
- data quality criteria setup; the process that sets criteria that include characteristics of data, target data, and methods to measure
- data quality measurement; the process that measures target data with the criteria set in the process of data quality criteria setup on a real time basis or periodically

The data quality improvement process corrects data errors detected and eliminates root causes of the data errors by tracing and identifying them. In order to support the top-level process effectively, adjustment of data stewardship in accordance with data flows tracing is required. This process has the function of process improvement not only data quality improvement. Processes for data management are improved at the data administrator level while business processes at the data manager level. This top-level process shall consist of the following processes:

- data stewardship and flow management; the process that analyses data operations and data flows among businesses or organizations, identifies responsible parties and their data operation systems which influence data quality, and manages the stewardship of data operations
- data error cause analysis; the process that analyses root causes of data errors and prevents a recurrence of the same errors fundamentally
- data error correction; the process that corrects the data that turns out erroneous

4.2.2 Three data quality management roles

The three roles in the framework are responsible for performing the processes in the framework. These roles shall be:

- data manager
- data administrator
- data technician

The data manager shall perform the following processes within the framework:

- data architecture management
- data quality planning
- data stewardship and flow management

The data manager performs the role that directs a guideline for master data quality management in compliance with objectives of an organization, manages factors that impact data quality at an organization level, and establishes the plans for performing data quality activities in the organization. Along with each major top-level process, the data manager maintains data consistency in individual information systems through the organization-wide data architecture management, and analyzes factors that affect data quality in data quality planning. In addition, the data manager takes a role of granting data administrator's authority to trace and correct data over the information systems or organization.

5 DATA GOVERNANCE

Data management is defined as the act of managing data, while data governance is concerned with managing the rules for managing data, the “authority” over the data. The concept of data governance is still new so what is included and what is excluded varies from one company to another.

It is important to remember that governance is first and foremost about people. Authority implies a chain of command and the delegation of duties and responsibilities.

The meaning of the word ‘governance’ according to Institute of Governance:

“The need for governance exists anytime a group of people come together to accomplish an end. Though the governance literature proposes several definitions, most rest on three dimensions: authority, decision-making, and accountability. Shortly:

Governance determines who has power, who makes decisions, how other players make their voice heard and how account is rendered.” (http://iog.ca/defining-governance/)

In article “Designing Data Governance” Khatri & Brown are defining governance:

“Governance refers to what decisions must be made to ensure effective management and use of IT (decision domains) and who makes the decisions
(locus of accountability for decision making).” (Khatri&Brown, 148. 2010)

Governance in general “refers to the way the organization goes about ensuring
that strategies are set, monitored, and achieved” (Rau 2004, 35 in Weber et al. 2009).

In successful data governance there are several decision domains that should be defined as seen is figure 4 below:

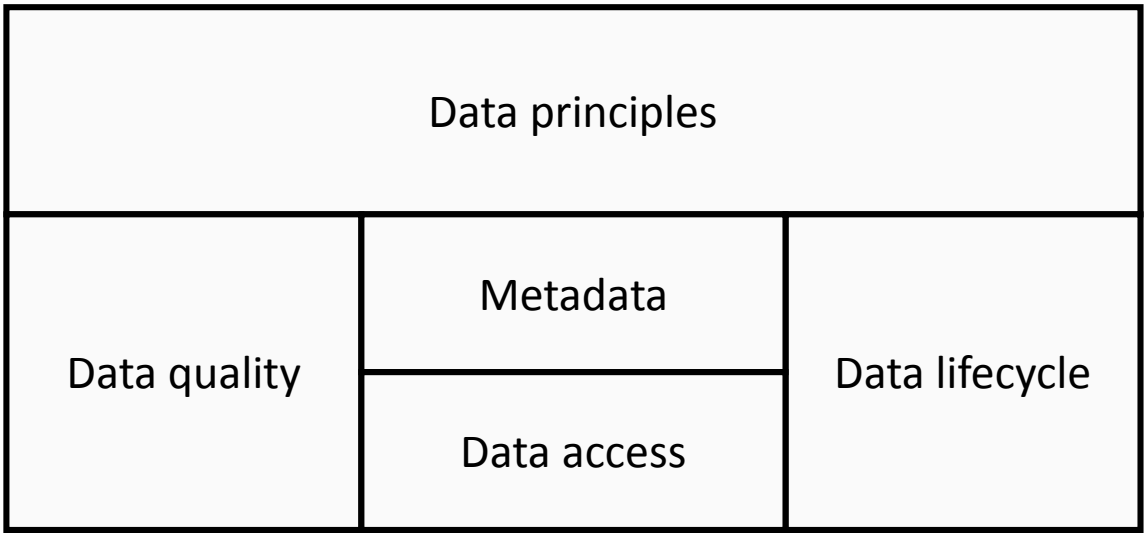


Figure 4. Decision domain for data governance. (Khatri&Brown, 2010)

If summarizing governance and data definitions it makes sense that a company needs to take care of the data quality and it is possible only if there is a governing model in use.

One of the primary functions of a data governance program is to identify the sources and the applications of data. This can be a very challenging mission. It is not uncommon for businesses to be buying data from many sources only to find that not only are the sources duplicative, but worse they are conflicting. Defining the rules for resolving these conflicts is an important part of data governance. (Benson 2012, 66)

By answering to questions: who, what, when, where, why and how it will be quite clear what Data Governance means.

Who?

- Data Governance is of concern to any individual or group who has an interest in how data is created, collected, processed and manipulated, stored, made available for use, or retired. We call such people Data Stakeholders

What?

- Data Governance means "the exercise of decision-making and authority for data-related matters."

- More specifically, Data Governance is "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods."

When?

- Organizations need to move from information governance to formal Data Governance when one of four situations occur:

    - The organization gets so large that traditional management isn't able to address data-related cross-functional activities
    - The organization's data systems get so complicated that traditional management isn't able to address data-related cross-functional activities.
    - The organization's Data Architects, SOA teams, or other horizontally-focused groups need the support of a cross-functional program that takes an enterprise (rather than siloed) view of data concerns and choices.
    - Regulation, compliance, or contractual requirements call for formal Data Governance.

Where?

- Data Governance can be placed within Business Operations, IT, Compliance/Privacy, or Data Management organizational structures. What's important is that they received appropriate levels of leadership support and appropriate levels of involvement from Data Stakeholder groups.

Why?

- Data Governance Frameworks help us organize how we think and communicate about complicated or ambiguous concepts. The use of a formal framework can help Data

Stakeholders from Business, IT, Data Management, Compliance, and other disciplines come together to achieve clarity of thought and purpose.

The use of a framework can help management and staff to make good decisions - decisions that stick. It can help them reach consensus on how to "decide how to decide." That way, they can more efficiently create rules, ensure that the rules are being followed, and to deal with noncompliance, ambiguities, and issues.

How?

- Organization decides what's important to them - what their program will focus on. Then they agree on a value statement for their efforts. This will help establish scope and to establish goals, success measures, and metrics.
- Next, develop a roadmap for the efforts, and use this to acquire the support of stakeholders. Once achieved, design a program, deploy the program, go about the processes involved in governing data, and perform the processes involved in monitoring, measuring, and reporting status of the data, program, and projects.

(http://www.datagovernance.com/adg_data_governance_basics.html)

5.1 Benefits of good quality data

Managers should make the right decision every time they make a decision. This brings with it the need for correct and adequate information. A decision should be based on real knowledge, which is a holistic point of view and not on “educated guesses”, intuitive feelings or limited information that is only looked at from some point of view. (Reunanen 2013)

Bernand Marr, Founder and CEO, Advanced Performance Institute, has introduced an intelligent company model for management: how to make good decisions based on agreed requirements, collected and qualified data and analyzes of the data. Figure 5 Intelligent company model (Marr 2013):

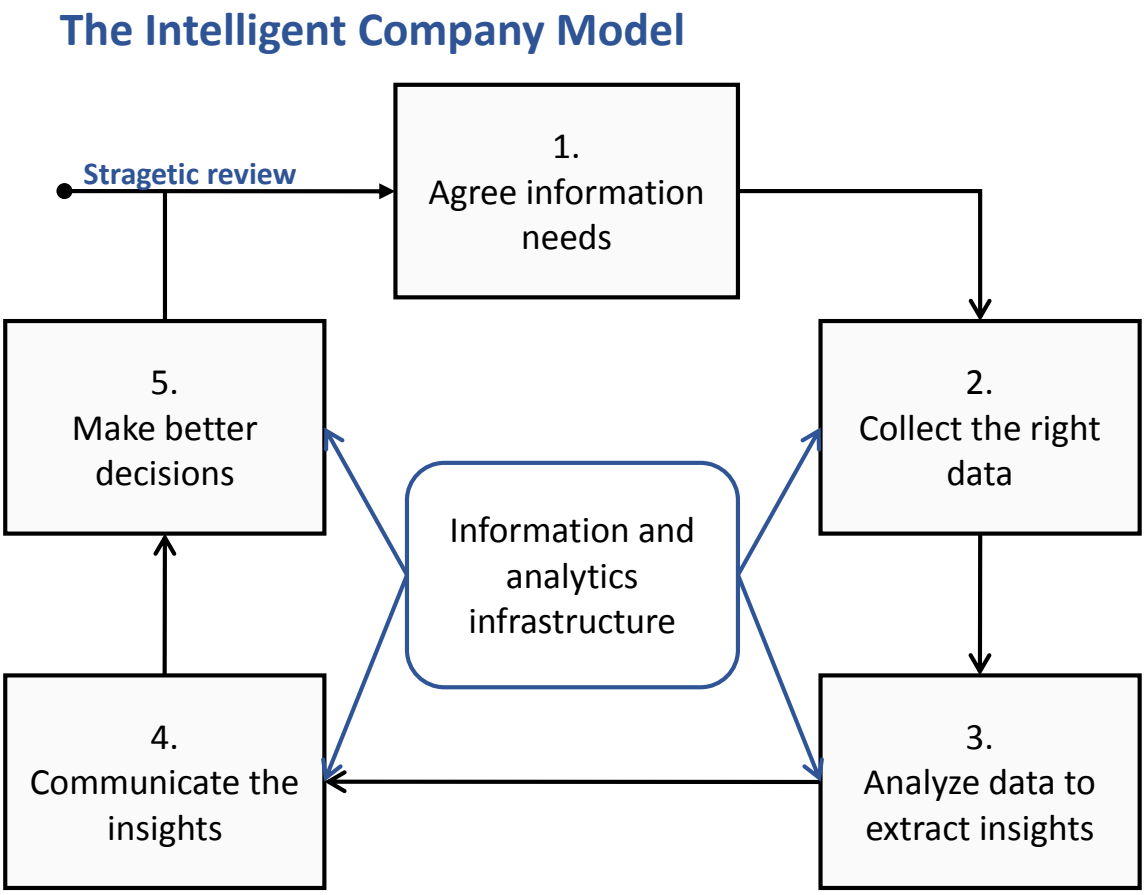


Figure 5. The Intelligent Company Model (Marr 2013).

Marr also makes seven conclusions about using the Intelligent Company Model:

1. Making better decisions is everyone's everyday job
2. Intelligent Strategy: Clear objectives, strategy maps and questions
3. Intelligent Data: Leveraging big and small data to answer your questions
4. Intelligent Insights : Analytics and experiments
5. Intelligent Communication: Balance numbers with narratives & visuals
6. Intelligent Decision Making: Court-style decision meetings
7. Create a culture of fact-based decision-making

(Delivering Business Insights From Analytics and Big Data, Marr 2013)

To be able to do the right decisions needed, managers should have the right data in use and available. When data are unreliable, managers quickly lose faith in them and fall back on their intuition to make decisions, steer their companies, and implement strategy. They are, for example, much more apt to reject important, counterintuitive implications that emerge from big data analyses. (http://hbr.org/2013/12/datas-credibility-problem/ar/1)

Fifty years after the expression “garbage in, garbage out” was coined, we still struggle with data quality. But I believe that fixing the problem is not as hard as many might think. The solution is not better technology: it’s better communication between the creators of data and the users of the data; a focus on looking forward; and, above all, a shift in responsibility for data quality away from IT folks, who don’t own the business processes that create the data, and into the hands of managers, who are highly invested in getting the data right. (http://hbr.org/2013/12/datas-credibility-problem/ar/1)

5.2 Meaning of qualified data in business decisions

International Standards are strategic tools and guidelines to help companies tackle some of the most demanding challenges of modern business. They ensure that business operations are as efficient as possible, increase productivity and help companies access new markets.

Benefits include:

- **Cost savings** - International Standards help optimize operations and therefore improve the bottom line
- **Enhanced customer satisfaction** - International Standards help improve quality, enhance customer satisfaction and increase sales
- **Access to new markets** - International Standards help prevent trade barriers and open up global markets
- **Increased market share** - International Standards help increase productivity and competitive advantage

- **Environmental benefits** - International Standards help reduce negative impacts on the environment

Businesses also benefit from taking part in the standard development process. (http://www.iso.org/iso/home/standards/benefitsofstandards.htm)

Levelling the playing field: How companies use data to create advantage is an Economist Intelligence Unit report, sponsored by SAP. The Economist Intelligence Unit conducted the survey and analysis and wrote the report.
(http://www.economistinsights.com/technology-innovation/analysis/levelling-playing-field, 10.12.2013)

The report's quantitative findings come from a survey of 602 senior executives, conducted in September 2010. The Economist Intelligence Unit's editorial team designed the survey. (http://fm.sap.com/images/kern/assets/sap_EIU_Levelling_The_Playing_Field.pdf, 10.12.2013)

Nearly all companies realize that the way to gain a competitive advantage is to obtain better data, interpret them quickly, and distribute them in easier-to-use formats. However, there are many obstacles to the effective use of data and few companies surmount them all—a fact that results in a lot of unused corporate data. Indeed, only 17% of companies use 75% or more of the data they collect.
How are companies using information to beat their rivals and create a more level playing field?

Below is a list of its major findings:

- Leading companies are keenly focused on data. Of the 38% of respondents who say their company performs ahead of its peers, 74% say that data are "extremely valuable" in achieving competitive advantage. The best corporate users of data devote substantial time to figuring out what sort of information they should track and

who within their companies needs it. They also invest in technology and training to make sure individual workers are able to capitalize on the data they have collected.

- Accuracy trumps detail. Accuracy and timeliness are the most important attributes of data, ahead of the amount of detail the data offers. This is because getting the basic insight—about a new prospect, a change in the price of some raw material, or an emerging problem at a manufacturing plant—is more important than being able to analyze every detail about it.

- Information supports competition in myriad ways. Seventy-seven percent of respondents say data make an important contribution to their customer support/customer relations efforts, and 71% say it helps them support their sales processes. Operations, cost management and product development are all aided by data as well. A less common benefit—cited by around half of all companies—is the contribution that business insights have made to helping executives strengthen awareness of a company's brand.

- Yet most companies remain awash in unused data. In fact, only 27% of respondents say their firms do a better job of using information than most of their competitors. A large amount of data sitting on a company's servers, unused, is not uncommon and can be a sign of a sub-optimal data strategy. In some cases, however, there are good reasons to hold on to older data. Financial service firms often need archived data as a defense against litigation; others may want data for future data-mining purposes.

- A top-down approach may stifle competitiveness. Companies sometimes end up unintentionally approaching data from a management perspective and ignoring its value to others lower down the hierarchy. The companies that find ways to "democratize" their data often gain an advantage. Indeed, 77% of the companies that aim their data initiatives at all employees, regardless of level, say they've found ways to make data extremely valuable to their business. Only 65% of companies where the data initiatives are intended primarily for managers agree.

REFERENCES

Benson, Peter R., Managing Blind, 2013

http://cacm.acm.org/magazines/2010/1/55771-designing-data-governance/

http://www.computerhistory.org/brochures/categories.php?category=thm-42b97f98dbaf2, 9.12.2013

http://www.datagovernance.com/adg_data_governance_basics.html, 10.12.2013

http://www.eccma.org/iso8000/iso8000home.php, 10.12.2013

http://www.economistinsights.com/technology-innovation/analysis/levelling-playing-field, 10.12.2013

http://einstein.biz/biography.php, 10.12.2013

Floridi, Luciano, "Semantic Conceptions of Information", The Stanford Encyclopedia of Philosophy (Spring 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2013/entries/information-semantic/>.

http://www.gartner.com/newsroom/id/1824919, 2.11.2013

http://hbr.org/2013/12/datas-credibility-problem/ar/1, 9.12.2013

http://iog.ca/defining-governance/, 25.3.2014

http://www.iso.org/iso/home/standards.htm, 10.12.2103

http://www.iso.org/iso/home/standards/benefitsofstandards.htm, 10.12.2103

http://labspace.open.ac.uk/mod/oucontent/view.php?id=426285&printable=1, 9.12.2013

Liker, Jeffrey K. The Toyota Way, 2010

Marr, Bernard, Intelligent Business Model, 2013

http://www.oxforddictionaries.com/definition/english/data?q=data, 10.12.2103

http://www.oxforddictionaries.com/definition/english/information?q=information, 10.12.2103

http://www.oxforddictionaries.com/definition/english/quality?q=quality, 10.12.2013

Reunanen, Tero, Leader's Conscious Experience Towards Time, 2013

http://fm.sap.com/images/kern/assets/sap_EIU_Levelling_The_Playing_Field.pdf, 10.12.2013

Strout, Steven; Eisenhauer, John. The Elephant in the Room: Data. What you need to know to best govern and manage your enterprise data. 2011