



Aalto University
School of Business

Business Intelligence

Lecturer: Assoc. Prof. Yong Liu (yong.liu@aalto.fi)

37E00500

Brief answers to the question: How can we develop BI in case there is limited top management support?

- 1) Approach business users and understand their business needs and focus on the business you can influence through your analytic results.
- 2) Data quality can be an issue, and you should focus on the data you have and trust. You need to live with data of poor quality since you have limited resources/support to address the data quality issue.
- 3) Timeliness of data analysis can be important, while the performance of analysis (e.g., R2, predictive accuracy) can be less important.
- 4) Find ways to measure your influence on the business and use it to justify your asking for more resources.
- 5) Keep your communication with the top managers.
- 6) If the company has no BI strategies, you should have your own!

Agenda

1. Data
2. Data Quality
3. Prevalence and Cost of Data Quality Problem
4. Root Causes of Data Quality Problem
5. Two case studies

Gartner's recommended solution

Assume you are running have a company

- Data quality project led by IT department
- The IT department wants to follow “12 Actions to Improve Data Quality” strategy offered by Gartner
- The first step of the strategy is to build data-quality dashboards... and ...

You can find the specific report by searching “12 actions to improve your DQ, gartner” at Google.

About 315 000 results (0,39 seconds)



Gartner

<https://www.gartner.com> > [newsroom](#) > [press-releases](#) ⋮

Gartner Identifies 12 Actions to Improve Data Quality

22 May 2023 — Focus on **the** Right Things to Set Strong Foundations · Apply Data Quality Accountability · Establish “Fit for Purpose” Data Quality · Integrate Data ...



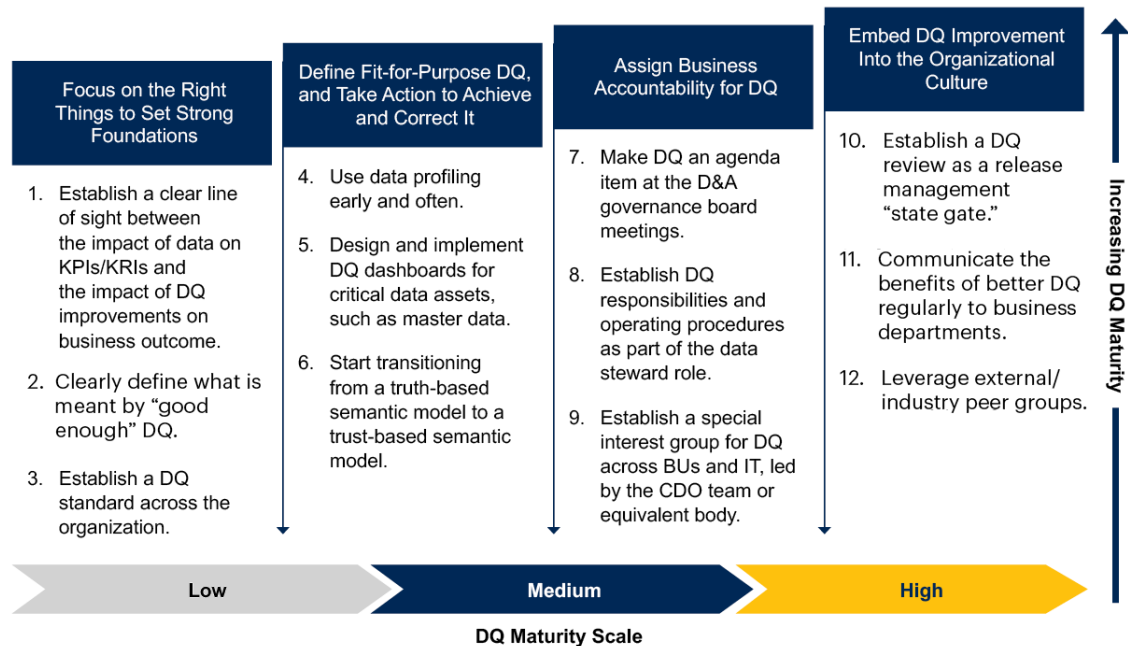
Gartner

<https://www.gartner.com> > [smarterwithgartner](#) > [how-t...](#) ⋮

12 Actions to Improve Your Data Quality

14 Jul 2021 — No. 1: Establish **how improved** data quality impacts business decisions · No. 2: Define what is a “good enough” standard of data · No. 3: Establish ...

12 Actions to Improve Your DQ



gartner.com

Source: Gartner
© 2021 Gartner, Inc. All rights reserved. PR_1363600

Gartner

8.4.2024

6

What is your opinion?

Following Gartner's approach, how likely will the company successfully improve its data quality?

Please offer your answer at: <https://presemio.aalto.fi/bic/>

Starts with a common business decision...

As the CEO of a large company with complex IT systems, you have heard the BI team report several times that data quality is becoming a key barrier to your company's continuous development. You concur, and finally, the company decides to invest in enhancing its data quality.

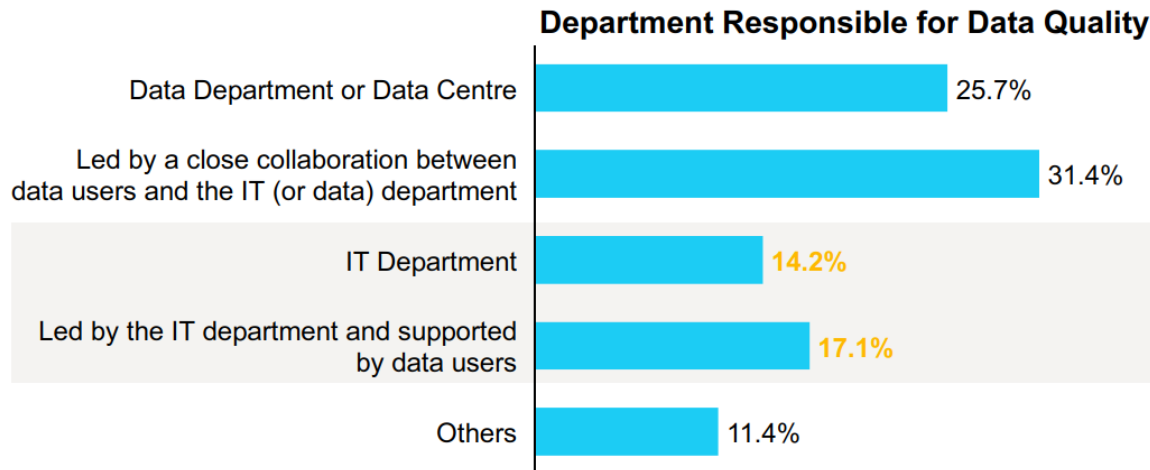
As the CEO of a company, who will you assign to lead a data quality project?

Please offer your answer at: <https://presemo.aalto.fi/bic/>

Responsible Departments for Data Quality

Among participant companies, there are multiple IT Department-led data quality projects.

Data users should play an important role in enterprise data quality projects. **31.4%** of surveyed companies place data users in a leading role in data quality projects. **25.7%** of companies have data departments or data centers responsible for data quality. **Notice that different reports indicated that data quality led by the IT department alone has a relatively high chance of failure.**



After class reading

- **Redman, 2013. Data's Credibility Problem, Harvard Business Review**
- **Desai et al. 2022. A better way to put your data to work, Harvard Business Review.**

Section 1: Data



Why Data is Important for Business?



Heikki Lempinen
CEO at Reima

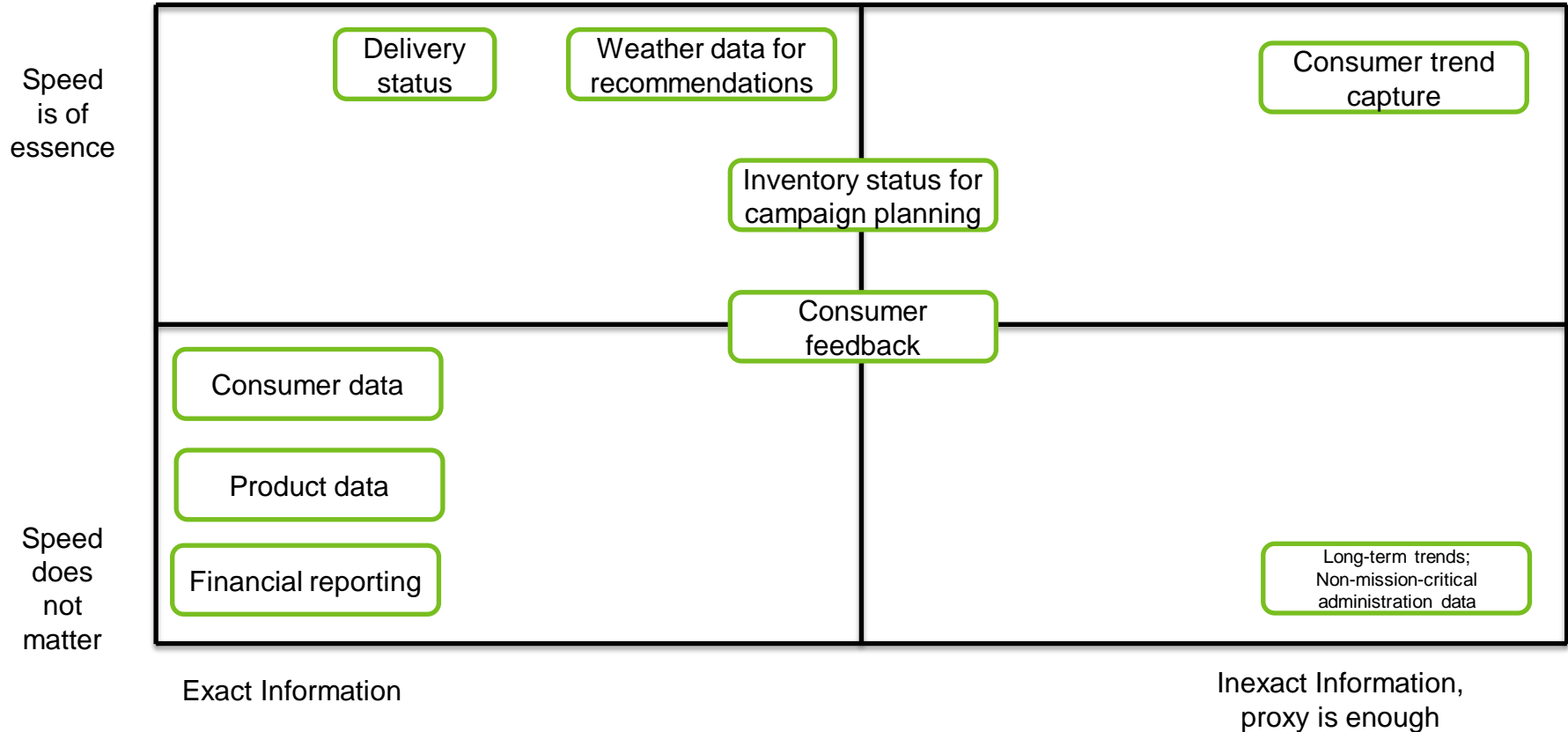
Data View

- 1. If you cannot measure it, you cannot manage it**
- 2. If you cannot measure it, it does not exist.**
- 3. Move fast – the right data at the right time.**
- 4. Data: Speed & Precise**

My addition: If data measures something unknown or not understandable, it is a cost!

Speed and Precision of Data

Example

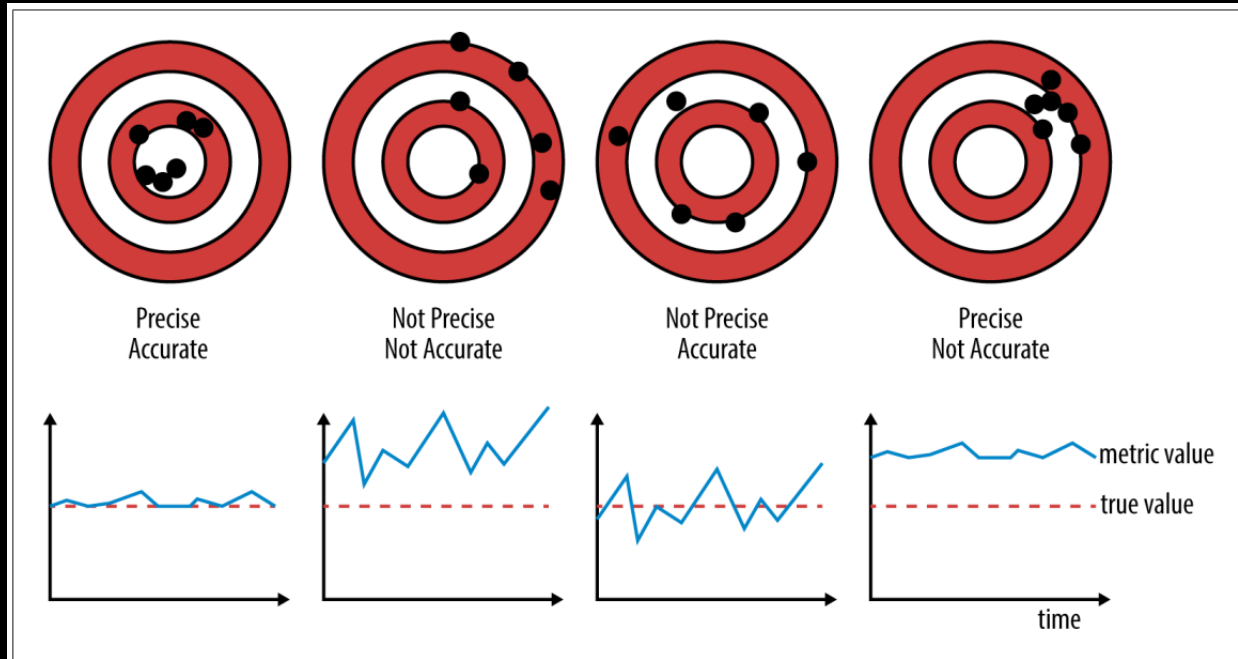


8.4.2024

13

Source: adapted from a Heikki's presentation

Data measure reality, but not necessarily in line with reality.



Speed: Expectation and reality

How long would it take for the top 543 Finnish companies in revenue to calculate their revenue for the last month?

Please offer your answer at: <https://presememo.aalto.fi/bic/>

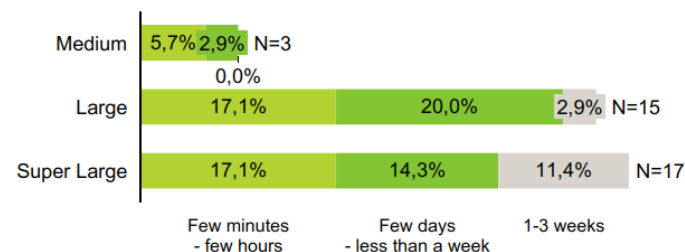
Analytics for Basic Tasks

The time needed to calculate last month's revenue is used to show the data infrastructure performance.

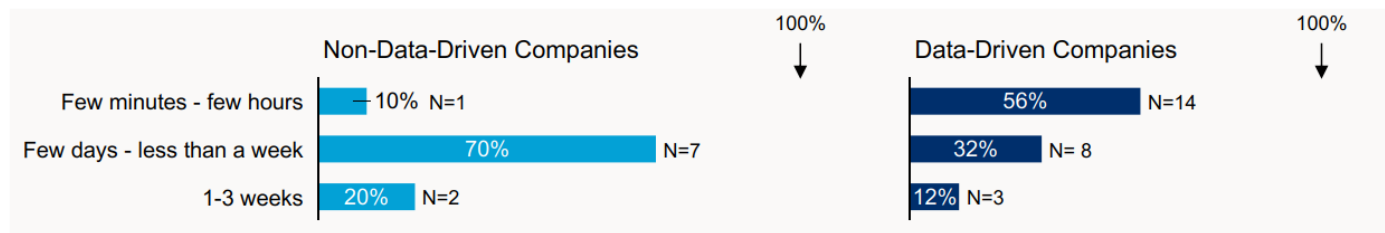
42.9% of the surveyed companies can calculate last month's revenue within a few minutes to a few hours, while another 42.9% would need from a few days to less than a week. However, 14.2% of companies need 1-3 weeks to calculate the revenue.

Companies identifying themselves as data-driven can calculate their revenue faster, i.e., in less than a few hours (56%) than those identifying themselves as non-data-driven (10%).

Company Size and Revenue Calculation Time

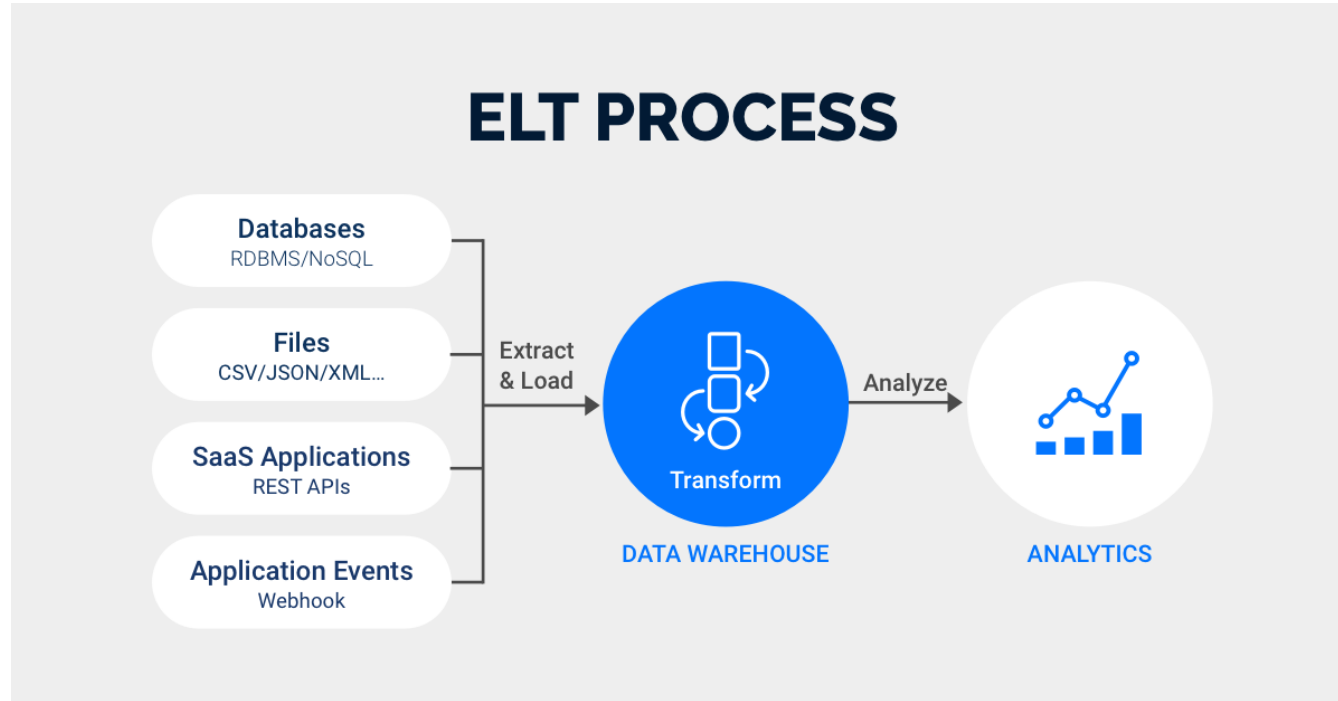


Time Needed to Calculate Last Month's Revenue: Data-Driven vs. Non-Data-Driven Companies



Extract, Transform, Load (ETL)

ETL is the process of combining data from multiple sources into a large, central repository called a data warehouse. ETL differs from ELT (Extract, Load, Transform).



Recap

- 1. If you cannot measure it, you cannot manage it**
- 2. If you cannot measure it, it does not exist.**
- 3. It is a cost if data measures something unknown or not understandable!**
- 4. Data accuracy is significant, though it may not always be paramount.**



Aalto University
School of Business

Section 2: Data Quality



Data Quality

CustomerID	GlobalID	CustomerSegment	CustomerStatus	Country	CustomerNAME	City	AddressLine1
1000	55001000	IM-S	A	GB	Jane Doe	London	West Road 3
1001	55001001	IM-M	A	US	Will Arnold	NY	A
1002	55001002	BM-S	A	GB	Sales	London	London Road 56
1003	55001003	BM-L	I	DE	Manfred Geld	GB	
1004	55001004	IM-S	A	PL	Anna Nowak	Gliwice	Lutnicza 1
1005	55001005	IM-M	A	AU	Pawel Kowalski	Linz	Dworcowa 3
1006	55001006	IM-L	A	AU	Roland Linde	Wien	Berggasse 5
1007	55001007	IM-S	A	DE	Alex Haller	Stuttgart	Neugasse 3
1008	55001008	BM-M	A	DE	Thomas Bond	Munchen	Axelplatz 4
1008	55001009	BM-M	I	DE	Thomas Bond	Stuttgart	Altgasse 9
1010	55001010	IM-S	A	CH	Long Road 23	Zug	Limmatquai 10
1011	55001011	IM-M		DE	Joseph Kaller	Frankfurt	Erdingerplatz 3
1012	55001012	IM-S		PL	Alina Schwarze	Warszawa	Rondo 3
1013	55001013	IM-S	A	CH	Alina Kowalska	Zurich	
1011	55001011	IM-M	A	US	Irena Nueller	4	
1015	55001015	BM-M	I	GB	John Smith	3a	Old Rolad 666

CustomerID	GlobalID	CustomerSegment	CustomerStatus	Country	CustomerNAME	City	AddressLine1
1000	55001000	IM-S	A	GB	Jane Doe	London	West Road 3
1001	55001001	IM-M	A	US	Will Arnold	NY	A
1002	55001002	BM-S	A	GB	Sales	London	London Road 56
1003	55001003	BM-L	I	DE	Manfred Geld	GB	
1004	55001004	IM-S	A	PL	Anna Nowak	Gliwice	Lutnicza 1
1005	55001005	IM-M	A	AU	Pawel Kowalski	Linz	Dworcowa 3
1006	55001006	IM-L	A	AU	Roland Linde	Wien	Berggasse 5
1007	55001007	IM-S	A	DE	Alex Haller	Stuttgart	Neugasse 3
1008	55001008	BM-M	A	DE	Thomas Bond	Munchen	Axelplatz 4
1008	55001009	BM-M	I	DE	Thomas Bond	Stuttgart	Altgasse 9
1010	55001010	IM-S	A	CH	Long Road 23	Zug	Limmatquai 10
1011	55001011	IM-M		DE	Joseph Kaller	Frankfurt	Erdingerplatz 3
1012	55001012	IM-S		PL	Alina Schwarze	Warszawa	Rondo 3
1013	55001013	IM-S	A	CH	Alina Kowalska	Zurich	
1011	55001011	IM-M	A	US	Irena Nueller	4	
1015	55001015	BM-M	I	GB	John Smith	3a	Old Rolad 666



Integrity



Duplication



Missing value



Accuracy

Facets of Data Quality

Accessible: the data is accessible by appropriate tooling.

Accurate: mistyped date of birth, out-of-date customer address.

Coherent: Data can be combined with other relevant data in an accurate manner. For instance, a sales order should be able to be tied to a customer, one or more products in the order, a billing and/or shipping address, and possibly payment information

Complete: no missing data. For instance, a missing first name in a customer record.

Consistent: Data is in agreement. For example, an email address for a particular customer in one data source matches the email address for the same customer in another data source. When there are conflicts, one source should be considered the master source or they are both unused until the source of disagreement is understood and fixed.

Defined: Individual data fields each have a well-defined, unambiguous meaning. Well-named fields accompanied by a data dictionary aid quality.

Relevant: Relevant data should be saved at the same table or database.

Timely: There is a short or reasonable duration between the collection of the data and the availability or release to analysts. In practice, this means that the data arrives in time for analysts to complete the analysis before it is due. It is possible that a major corporation whose data warehouse had a latency of one month.

Take away

- **There are many data quality measures, each describing different facts about data quality issues. However, they are not the root causes.**
- **Understanding data quality measures can be important, such a problem cannot solve the data quality problem.**



Aalto University
School of Business

Section 3: Prevalence and Cost of Data Quality Problem



The High Stakes of Wrong Pricing in E-Commerce

FIFA 23

- Pre-order of FIFA 23 on the Epic Games Store
- A decimal point error
- The game was listed for less than \$1 in India (the intended price is \$60)
- Massive orders from around the globe
- EA Sports honored all sales

Zappos

- Mistakenly capped the maximum price of all their items at \$49.95 in 2010
- Losses exceeding \$1.6 million
- Zappos honored all sales

Best Buy

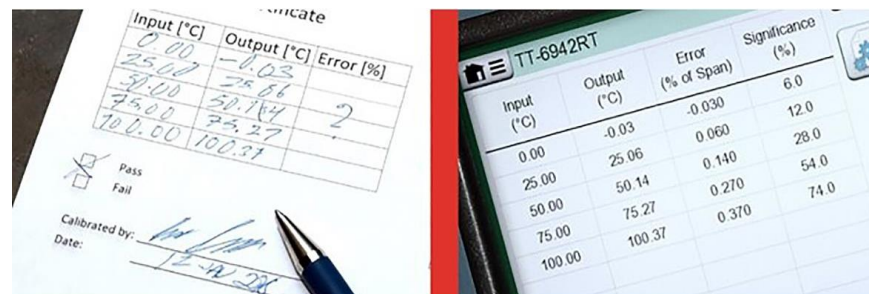
- Pricing error in 2015
- \$200 gift card being sold for \$15
- The pricing error went viral.
- Best Buy decided to not honor the sales.
- Significant customer backlash and negative publicity

Data Quality Problem In Design

Something as simple as a misplaced decimal point can cost you thousands, millions, or even billions of dollars. Just look at the \$2.7 billion Spanish S-80 submarine program. One misplaced decimal point led to submarines 70 tons heavier than planned, and engineers feared they would be incapable of resurfacing once submerged. Spain had to spend \$14 million to have their weight reduced.



Manual Data Entry Error



In 2022, many Finnish retailers (e.g., XXL, Intersport) offered a chance for customers to register as a member by providing their e.g., the number of phones.

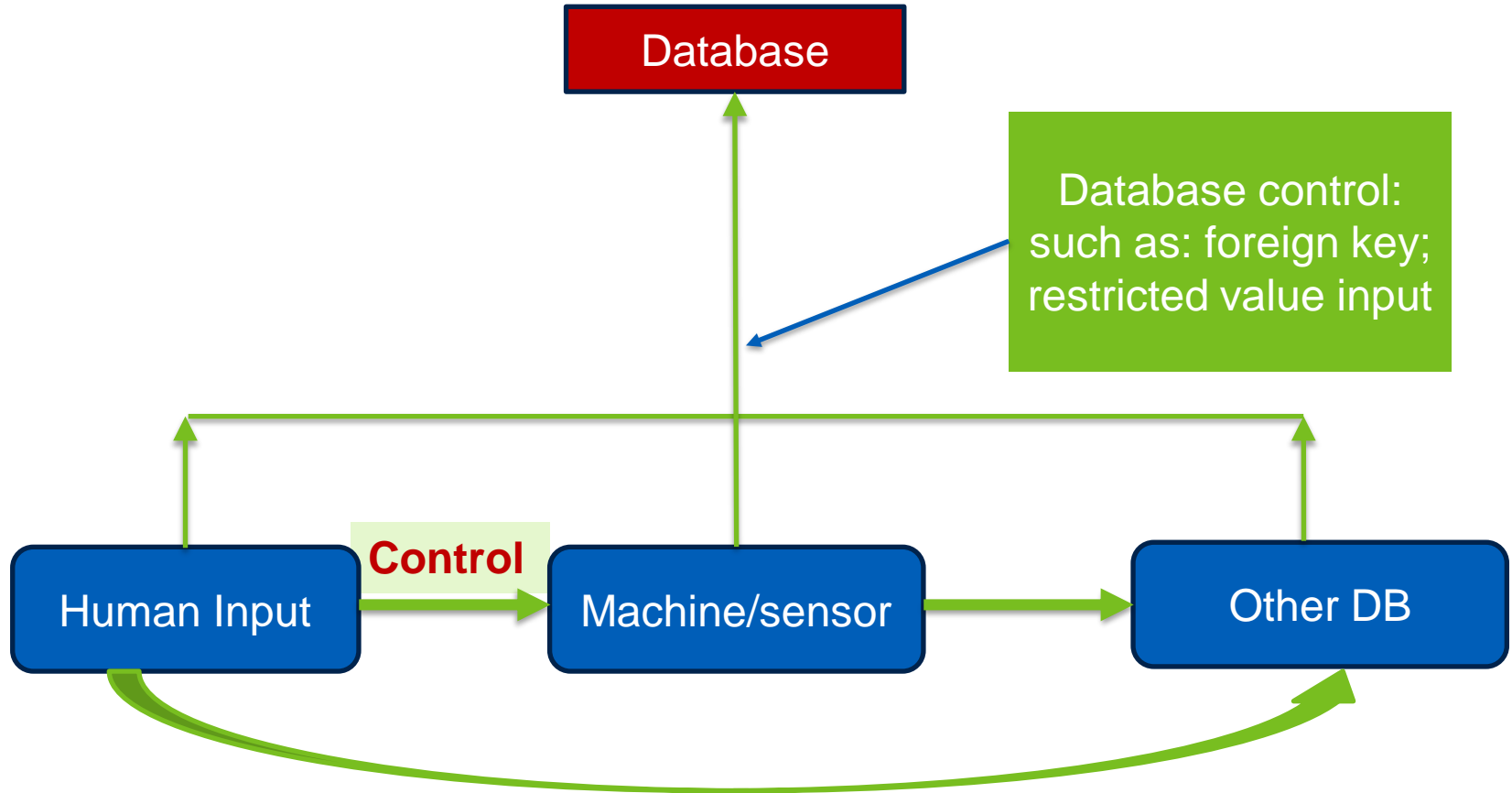
Based on your intuition, what is manual data entry's general typical error rate?

Please offer your answer at: <https://presemo.aalto.fi/bic/>

more problematic than others, but it's practically a guarantee that manual data entry mistakes will occur. A general rule of thumb is that error rates usually hover around 1%. One study involving laboratory tests found an error rate of almost 4%, with over 14% of errors containing significant and potentially dangerous discrepancies. Calculations or large data tables can compound mistakes.

Harvey, Everett, Parmar, & on behalf of the CHART Steering Committee, 1994): Estimates put the error rate around 0.20% (Reynolds-Haertle & McBride, 1992) or between 0.04% and 0.67%, depending upon the type of data (Paulsen, Overgaard, & Lauritsen, 2012). However, in research contexts, data-entry errors are more common. Error rates typically range from 0.55% to 3.6% (Barchard & Pace, 2011; Bateman, Lindquist, Whitehouse, & Gonzalez, 2013; Buchele, Och, Bolte, & Weiland, 2005; Kozak, Krzanowski, Cichocka, & Hartley, 2015; Walther et al., 2011), although

Sources of Data in Database

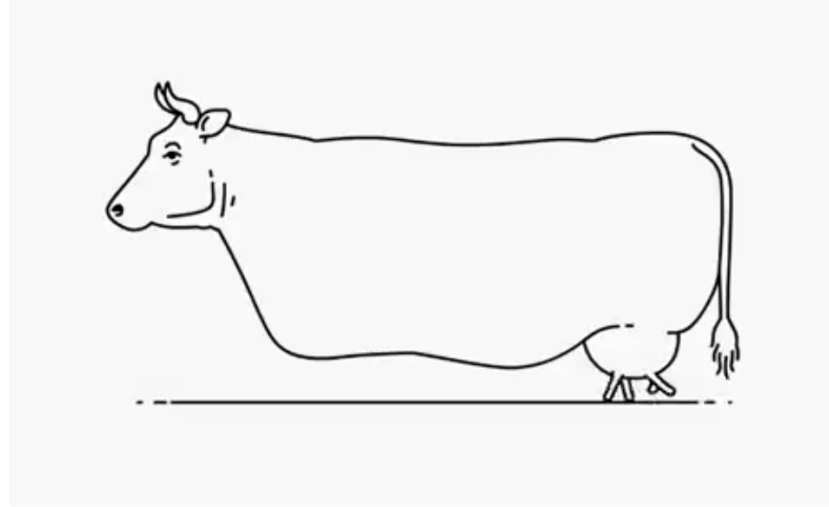


Can the company still do business and profit despite poor data quality?



Faint Knowledge of Data / System

First rule of programming; if it is work, don't touch it



An e-commerce site



This Photo by Unknown Author is licensed under CC BY-NC

A forest-related company



https://c.pxhere.com/photos/20/31/job_work_logging_timber_wood_lumber_worker_loading-779448.jpg

A public health website



24

30

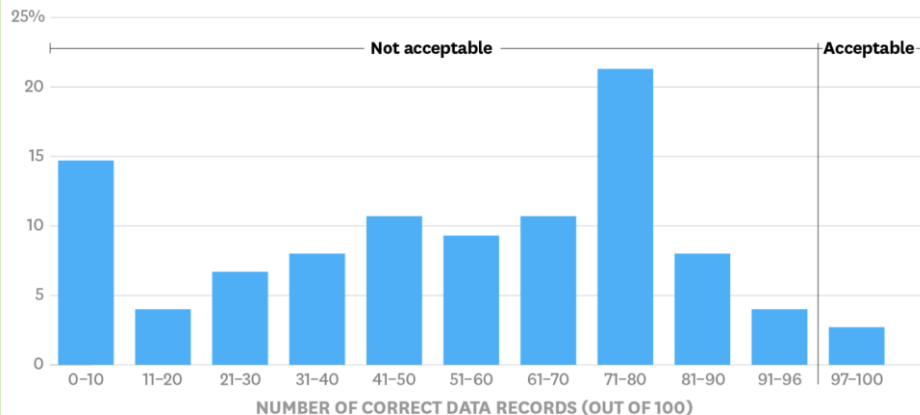
<https://9gag.com/gag/a9n1LLj>

Harvard Business Review: Only 3% of Companies' Data Meets Basic Quality Standards

Data Quality Is in Worse Shape Than Most Managers Realize

In a study involving 75 executives, only 3% found that their departments fell within the minimum acceptable range of 97 or more correct data records out of 100.

PERCENTAGE OF DEPARTMENTS



SOURCE TADHG NAGLE ET AL.

© HBR.ORG

The method is widely applicable and relatively simple: We instruct managers to assemble 10-15 critical data attributes for the last 100 units of work completed by their departments — essentially 100 data records. Managers and their teams work through each record, marking obvious errors. They then count up the total of error-free records. This number, which can range from 0 to 100, represents the percent of data created correctly — their Data Quality (DQ) Score.

On average, 47% of newly-created data records have at least one critical (e.g., work-impacting) error.

Data Dictionary & Why it is important?

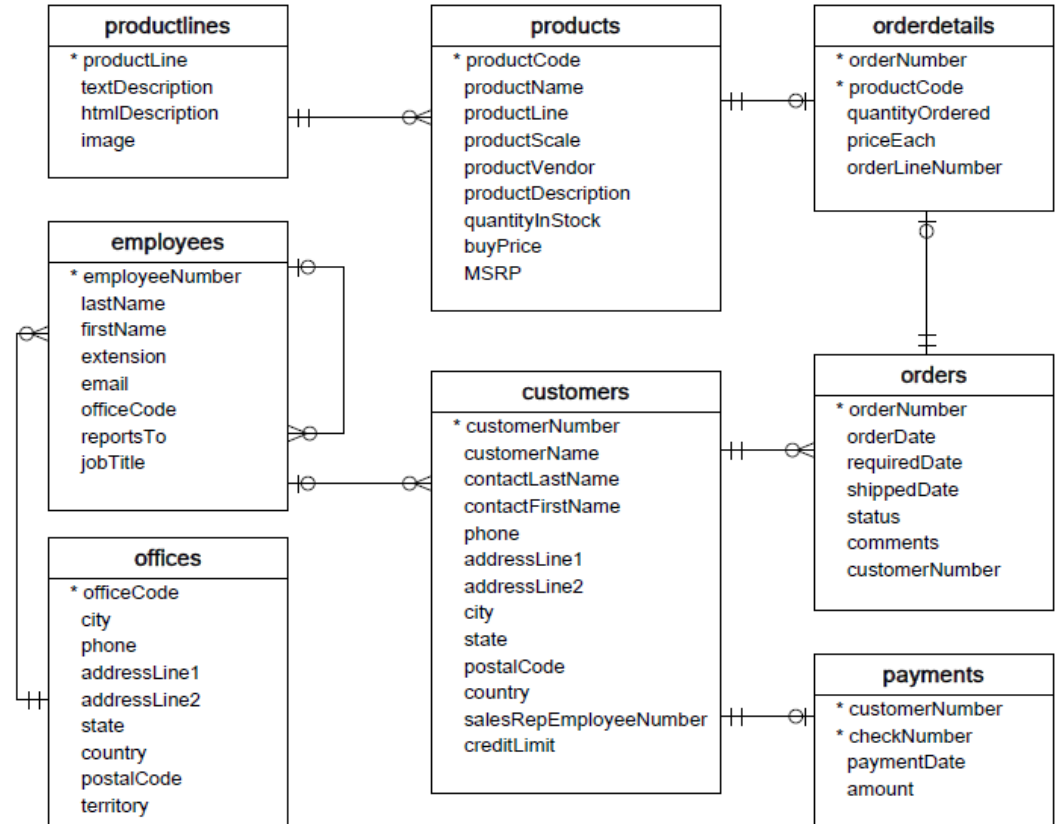
Data dictionary is an inventory of data elements in a database or data model with detailed description of its format, relationships, meaning, source and usage.

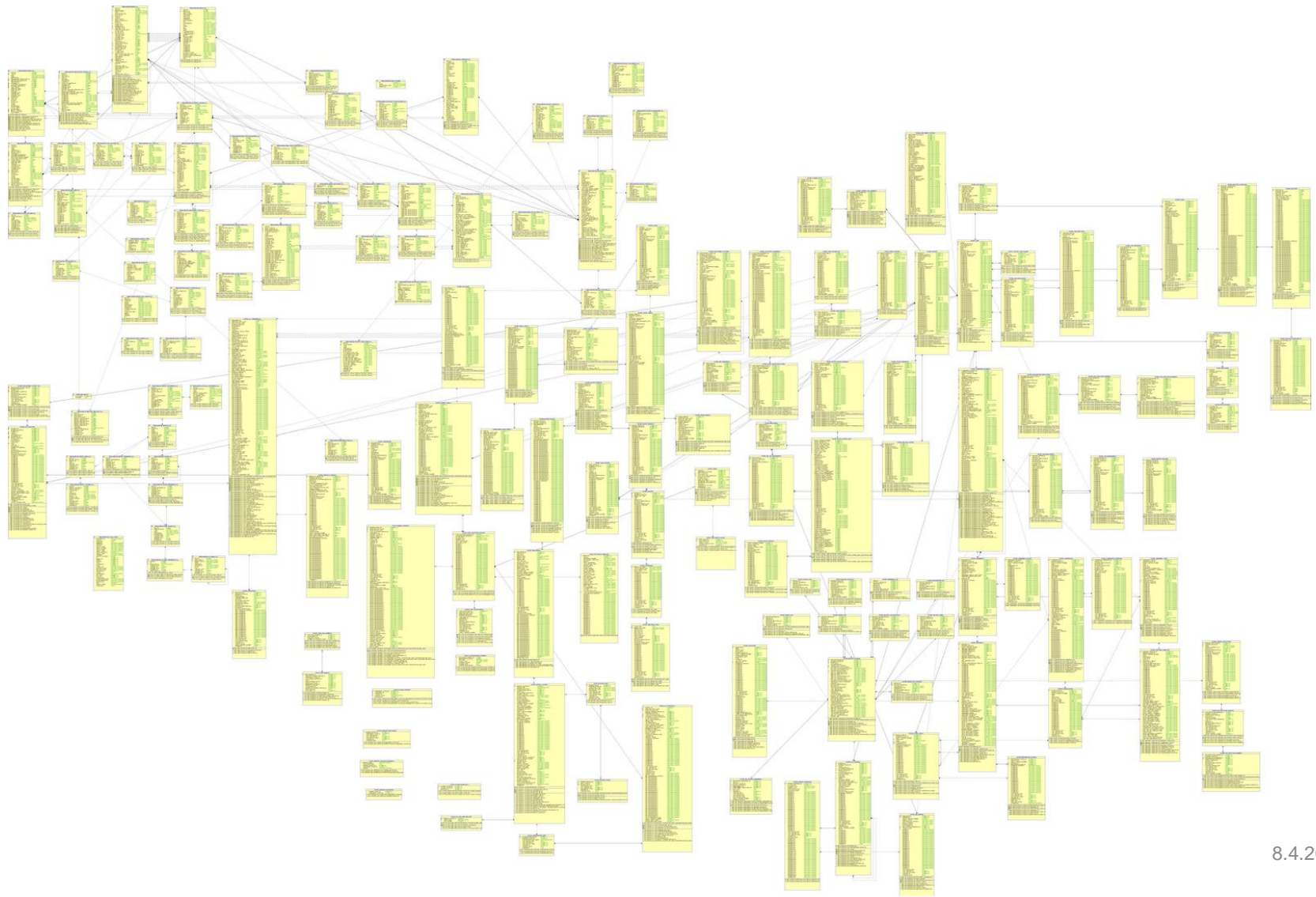
DATA					DATA DICTIONARY (METADATA)		
employee_id	first_name	last_name	nin	dept_id	Column	Data Type	Description
44	Simon	Martinez	HH 45 09 73 D	1	employee_id	int	Primary key of a table
45	Thomas	Goldstein	SA 75 35 42 B	2	first_name	nvarchar(50)	Employee first name
46	Eugene	Comelsen	NE 22 63 82	2	last_name	nvarchar(50)	Employee last name
47	Andrew	Petculescu	XY 29 87 61 A	1	nin	nvarchar(15)	National Identification Number
48	Ruth	Stadick	MA 12 89 36 A	15	position	nvarchar(50)	Current postion title, e.g. Secretary
49	Barry	Scardelis	AT 20 73 18	2	dept_id	int	Employee depamtnet. Ref: Departmetns
50	Sidney	Hunter	HW 12 94 21 C	6	gender	char(1)	M = Male, F = Female, Null = unknown
51	Jeffrey	Evans	LX 13 26 39 B	6	employment_start_date	date	Start date of employment in organization.
52	Doris	Bemdt	YA 49 88 11 A	3	employment_end_date	date	Employment end date.
53	Diane	Eaton	BE 08 74 68 A	1			

Metadata / Data dictionary

- Many data quality problems are rooted in metadata.
- A good working definition of "metadata" is "data about data"
 - ❑ for instance, units of measure.
- An example is Aera Energy.
 - ❑ As CEO Gaurdie Banister indicated, "High-quality metadata makes everything we do easier, from internal communications to planning new applications to making better decisions."
 - ❑ One firm that has done an exceptional job with its metadata is Aera Energy. It identified **53 common business terms**, such as "contract" and "customer," and then brought people from across the organization together to hammer out definitions of those terms, which serve as the core of its metadata.

Is Data Dictionary Enough?





Error-free data does not necessarily good data!!

Coveo found that the average employee spends 3.6 hours daily searching for information — an increase of one hour more from last year's report. IT employees spend half their day (4.2 hours) looking for relevant information (VentureBeat, 2022).

“employees spend 1.8 hours every day—9.3 hours per week, on average—searching and gathering information. Put another way, businesses hire 5 employees but only 4 show up to work; the fifth is off searching for answers, but not contributing any value.”

A day in the life of Consultant

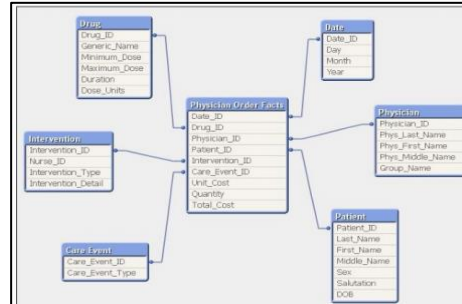


Tuukka Pesonen



Jesse Rajala

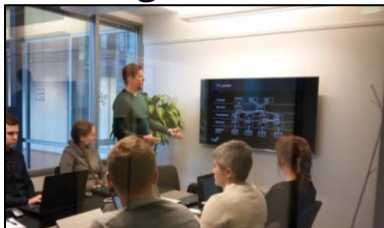
Data transformation and modelling ~ 70 %



Front end development ~ 20 %



Meetings ~ 10 %



Breakdown of working hours

09:50:05

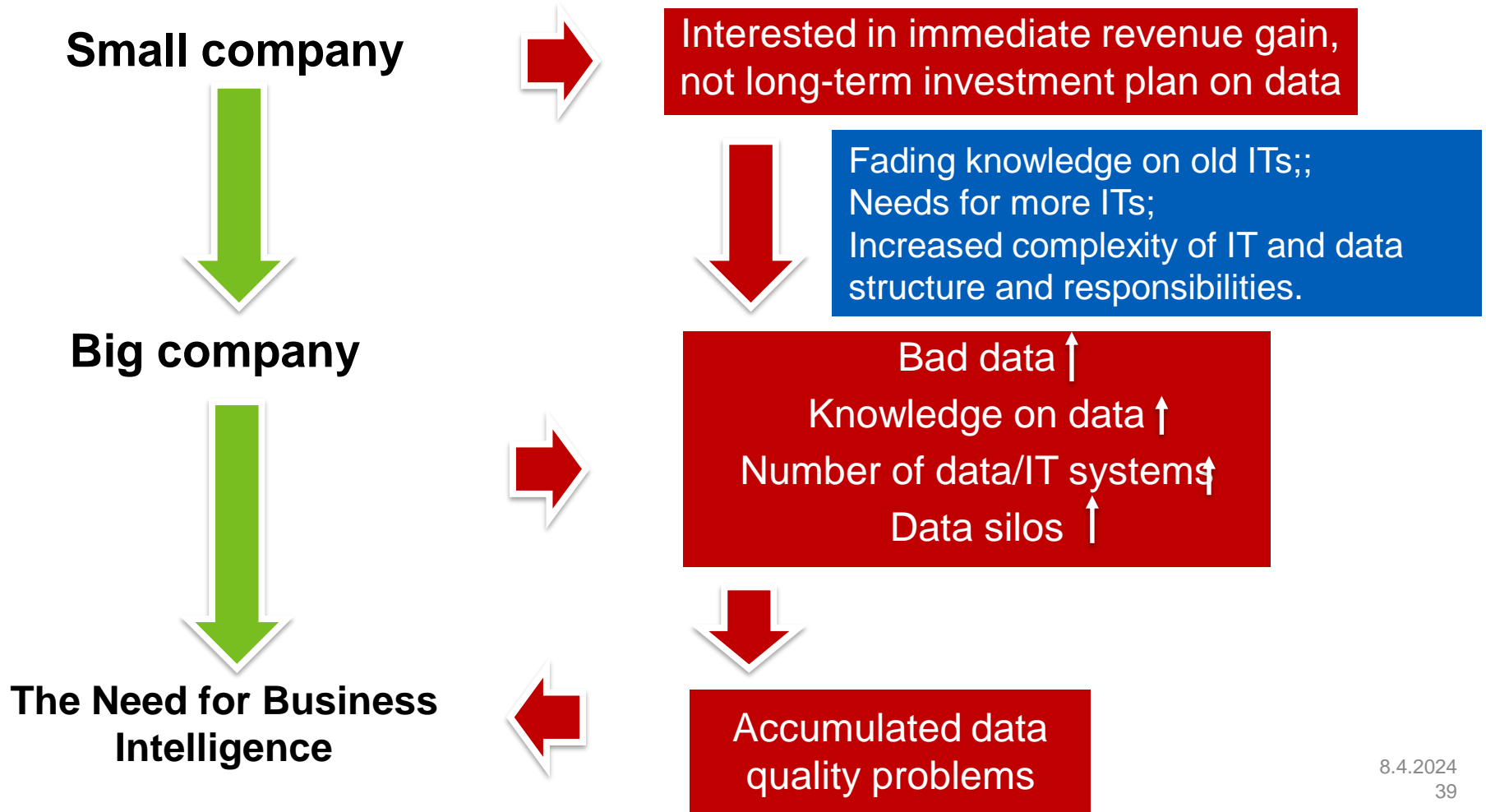


Aalto University
School of Business

Section 4: Root Causes of Data Quality Problem



Company & Business Intelligence Evolution

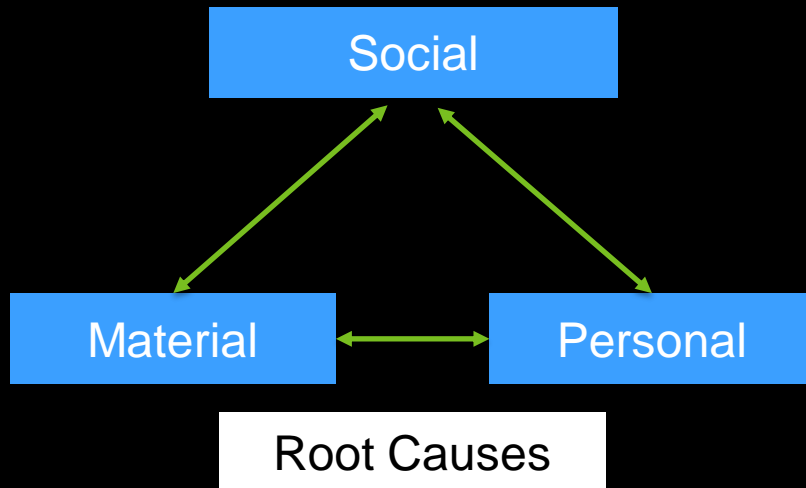




Illness

Fever & Cough

Root Causes of Data Quality Problem



Missing values

Integrity

...

Accuracy

Symptoms / Data Quality Problems

Recommended reading: Andrews et al. (2023). Root-cause analysis of data quality problems, *Journal of Business Analytics*, 5(1), p. 51-75.

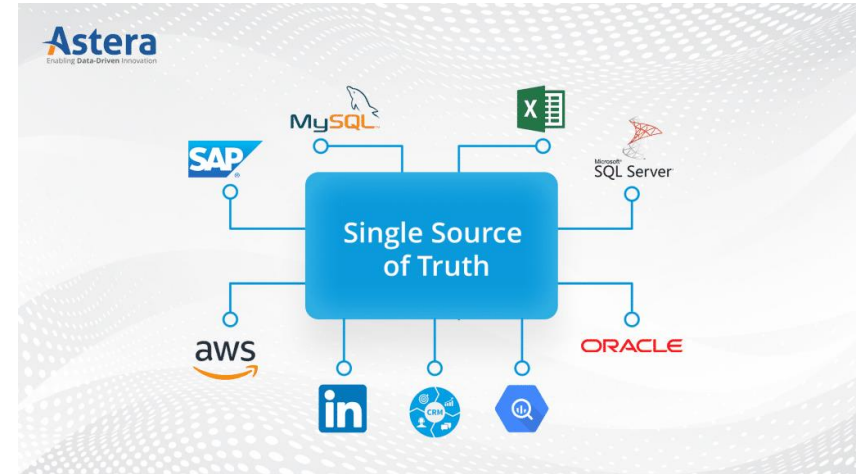
Accessibility, Speed and Management: Case 1

1. The BA received a simple request from the CEO to calculate the revenue of a large Finnish company.
2. The company has e.g. 3 data warehouses + 2 databases
3. Databases and warehouses have different owners/managers.
4. The BA has to communicate with different managers to get accounts to access their data. Of course, the BA's request has a very low priority for different managers.
5. It took probably 3 months for the BA to access all the needed data, albeit the calculation seems to be simple.

A possible solution: SSOT

Single Source of Truth (SSOT)

A single source of truth (SSOT) is the practice of aggregating the data from many systems within an organization to a single location. A SSOT is not a system, tool, or strategy, but rather a state of being for a company's data in that it can all be found via a single reference point.



While implementing a Single Source of Truth (SSOT) can yield numerous benefits, the process requires significant financial investment and allocation of resources!

Case 2: A new shipment location in ERP

A chain supermarket launched a new store in a Swedish city where an existing store had been operating for a long period. The system hosts product delivery information through the input of a shipment ID, such as 'ST0001', linked to a specific address.

Unfortunately, products were delivered to the old store (intended for the new store) because the shipment ID associated with that city was automatically filled in the system, and there was no distinction between the old and new stores. While the office manager has visibility only to the shipment ID (thought it was an ID for a city), lacking the address information, the delivery personnel have access to both the shipment ID and the corresponding wrong address.

Root Causes: System design + personal

Case 3: A B2B company

The company sells and delivers products to different companies. However, their ERP systems allow only one address for one company customer. This causes problems as one company customer may request a delivery to different locations, e.g., warehouse .

Root Causes: System design

Noticing the problem, the company adapted the system to allow multiple addresses for one company customer.

Later, the customer company moves its office. Sales managers find problems as one company is associated with three addresses, probably including a current invoice address, a delivery address, and an old invoice address.

Root Causes: System design + personal

Case 4: University Publication Records

1. The university needs publication information from staff, as it serves as a KPI of the university.
2. Filling in the publication reporting system is of very low priority for staff.
3. The university has made multiple attempts to persuade individuals to provide their publication information in time.

Result: outdated and incomplete publication information

Root Cause: Personal

Solution: ?

Case 5: A Bank System

The customer service specialist maintains the loan case status in the system ("open, awaiting internal review, approval pending, in progress, closed) throughout the life cycle of the case. However, there are instances where customer service specialists may overlook updating the status promptly.

Root Cause: Personal - Time and resource constraints pose challenges, as customer service specialists are engaged in handling more pressing and immediate tasks.

Personal

Digitalization is normally a good solution for data quality problems related to personals, but it may also lead to future data quality problems by increasing the number of systems in a company.

What is the difference between business intelligence and digitalization?

“Data’s Credibility Problem”

1. “When crude oil is thick, one of the major costs of working an oil field is stream-heating the crude in the ground to make the oil easier to pump.”
2. “To figure out how much stream is needed, field technicians point an infrared gun at the flow line, take a reading, and send the data to the reservoir engineer....”
3. But the flow line can get dirty, causing readings to be as much as 20 C lower than the true level.
4. Bad readings caused reservoir engineers to use more stream than necessary, causing additional operational expenses by tens of millions of dollars.



Solution: Connect Data Creator with Data Customers

The Problem

- Bad data
- Poor accessibility of data

How it happens

The people creating data have little understanding of others in the organization use that information. And when errors occur, the data users typically work around them or fix them as they go without addressing the root causes of the mistakes.

The Solution

The solution is not better technology; it's better communication between the creators of data and the user, a focus on looking forward, and the shifting of responsibility for data quality from IT folks to line managers, who are highly invested in getting the data right.

Strategies: existing data vs. new data

Like other companies, your company has a large amount of bad data. You have appointed an important role in improving the quality of data.

Will you focus on cleaning up existing bad data or improving how new data are created?

Please offer your answer at: <https://presemo.aalto.fi/bic/>

“Rather than launch a massive effort to clean up existing bad data, companies should focus on improving the way new data are created.”

– Thomas C. Redman

The nature of data quality problem

“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.”

Data Creators & Data Customers

Peter's Post

Thursday, November 21, 2013

Data's Credibility Problem

Print Logo

The Magazine - December 2013

by Thomas C. Redman

The quality of data is fixed at the moment of creation!
Why IT should not lead data quality project:

Put Responsibility for Data in the Hands of Line Managers

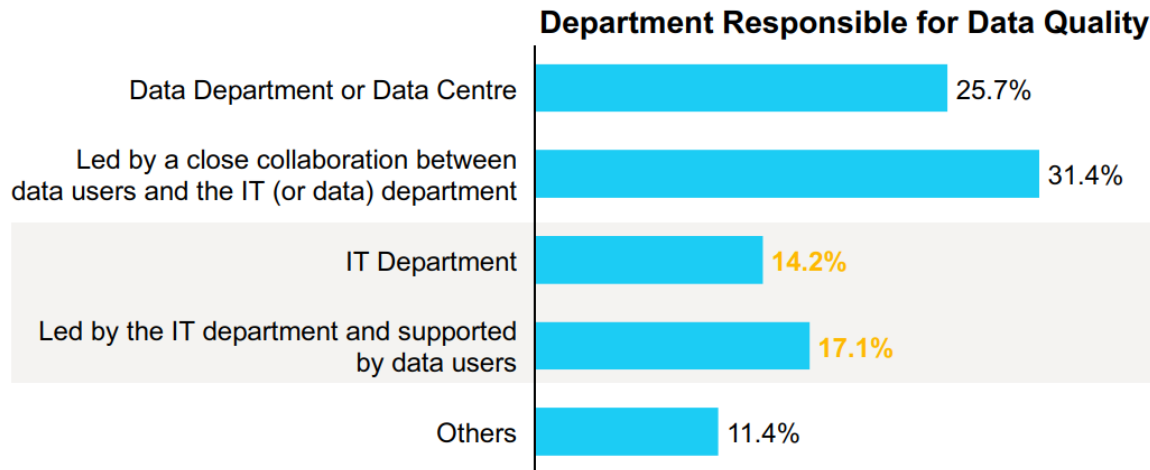
Very often, data creators are not linked organizationally to data users. Finance creates data about performance against quarterly goals, for example, without considering how Sales will want to use them or Customer Service analyzes complaints but fails to look for patterns that would be important to product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

Responsible Departments for Data Quality

Among participant companies, there are multiple IT Department-led data quality projects.

Data users should play an important role in enterprise data quality projects. **31.4%** of surveyed companies place data users in a leading role in data quality projects. **25.7%** of companies have data departments or data centers responsible for data quality. **Notice that different reports indicated that data quality led by the IT department alone has a relatively high chance of failure.**



Recap

- **Data quality is fixed at the moment of creation!**
- **“IT departments typically have little success leading data quality programs”.**
- Data creators are responsible for “garbage in,” and data consumers are responsible for “garbage out.”

After course reading

Five Reasons Your Data Governance Initiative Could Fail

<https://www.stibosystems.com/blog/five-reasons-your-data-governance-initiative-could-fail>

5 Reasons Data Quality Initiatives Fail

<https://www.precisely.com/blog/data-quality/reasons-data-quality-initiatives-fail>

Section 5: Two case studies



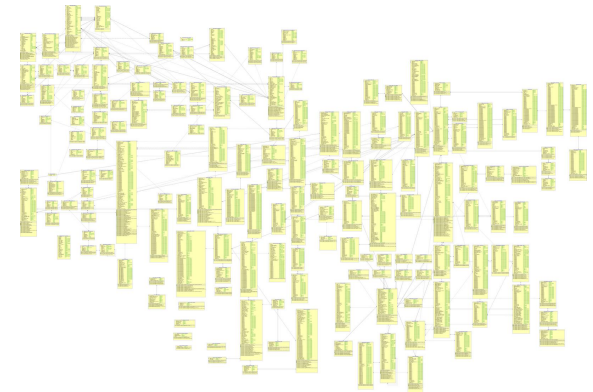
Case 1: Sanoma

Source: Mr. Kristian af Hällström, Business Intelligence Lead from Sanoma, 2023

Sanoma: Leading media brands in Finland with business covering the areas of news media, magazines, TV, and audio.

Problem: The old DW solution in 2017

- Table amount was huge due to long history
- Inactive and unused tables were not deleted.
- ETL duration: ~ 5 hours
- 1000+ tables



THE NEW DATA WAREHOUSE

Why we failed in previous DW project

- ✓ Due to challenges in data quality in MediaDW, the DW renewal project (data vault) started ~2015 and failed. Data quality was even worse and we were stuck with two data warehouses.
- ✓ The new DW (current) project started mid-2018, launched early 2019 and replaced two old data warehouses.
- ✓ The factors that causes the failure in data vault project
 - The project was led by technology
 - Data engineers didn't understand data nor business
 - Data structure and models did not meet the business needs, data model was too complex
 - Too many outsourced resources and consultants, several personnel changes

Key factors in the success of the DW project

- ✓ Keep it simple!
- ✓ Best practices
 - Naming convention, source keys, surrogate keys, data processing, testing processes, peer review, production deployment process, etc.
- ✓ Good cooperation between tech and business / data engineers and business analysts
 - Regular status meetings
- ✓ Specifications are made from business needs
- ✓ Ensure data quality and automate data quality processes

ERP Implementation Success Ratio

To investigate this question, the data was further analyzed and compared for the success rate in each case (Figure 11). Assuming 100-75 % as success and 50% & below as failure the percentage of success for the implementation where the project was lead by the business managers was 36.5% while for the projects lead by IT manager the success rate was 21.62%. While failure rates were 6.75% and 16.2% respectively. The success to failure ratio was estimated to be 5.4 for the business managers' case as compared to 1.3 for the IT managers' case which shows 4 times higher rate in the business mangers' case. Consultants, although relatively fewer in numbers (12%), had a success rate of 5.

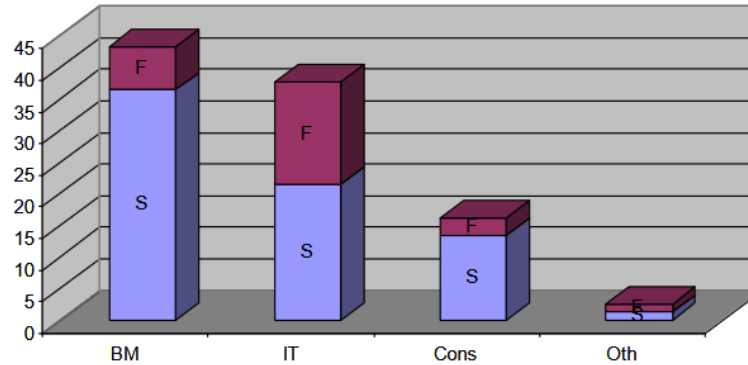


Figure 10: Implementation leadership (%)

BM: Business managers
IT: IT manager
Cons: Consultant.

Going back to the Case: A large B2B company

Assume you have a company

- Data quality project led by IT department
- The IT department wants to follow the “12 Actions to Improve Data Quality” strategy offered by Gartner
- The first step of the strategy is to build data-quality dashboards... and ...

Why I think the project is in a high risk:

1. **“Data quality is fixed at the moment of creation!”**
2. **The data quality KPI is useless and often not actionable.**
 - Quality for data quality cannot be improved even with KPI.
 - A huge amount of data variables are available, and it is impossible to access them all.
 - Which variables should be chosen if they should be assessed via KPI? ---- Data users. Likewise, how important is a data variable, and how good/bad is it? This cannot be assessed by looking at the data values.
 - Not considering business processes.
 - Although the CDO and data steward are involved in the project at a later stage, data users' perspectives are still missing.

The Success Ratio of a Data Quality Project

A High Chance



**The level of data
users' involvement**

A Low Chance