Master Data Health Check

Rationale for Service

consulting4bit.com

Why Master Data Management matters

A study by Forbes + SAP clearly showed why master data management matters and reasons why companies need to invest into this.

© Fotolia, Thinkstockphotos

Costs	•Data-related problems cost the majority of companies more than $5 million annually. One-fifth estimate losses in excess of $20 million per year.
Key to Success	•95% of organizations agree that strong information management is critical for business success.
Problem Sources	•Line-of-business executives and IT frequently disagree about the source of data-related problems and the potential solutions.
Ownership	•Fragmented data ownership is the single biggest roadblock to an enterprise information management program.
Communication	•CIOs and their IT managers need to communicate more frequently and effectively about information management projects and their benefits to lines of business.

* Source: http://images.forbes.com/forbesinsights/StudyPDFs/SAP_InformationManagement_04_2010.pdf

Selected Study Results (1)

Fig. 1: The consequence of poor data quality

Fig. 2: Where does primary responsibility for quality of data reside at your organization?

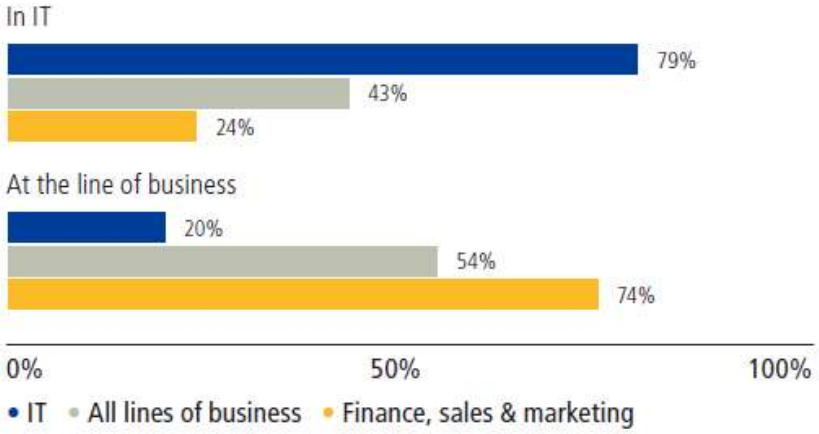
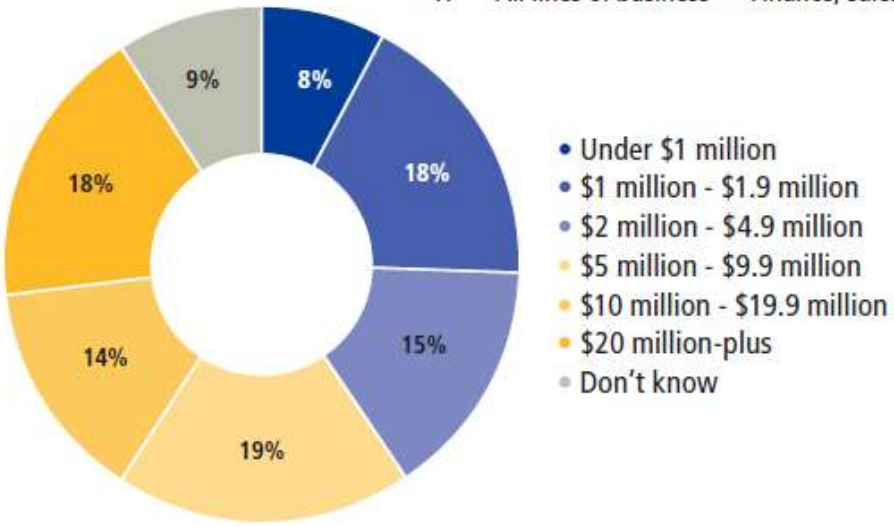


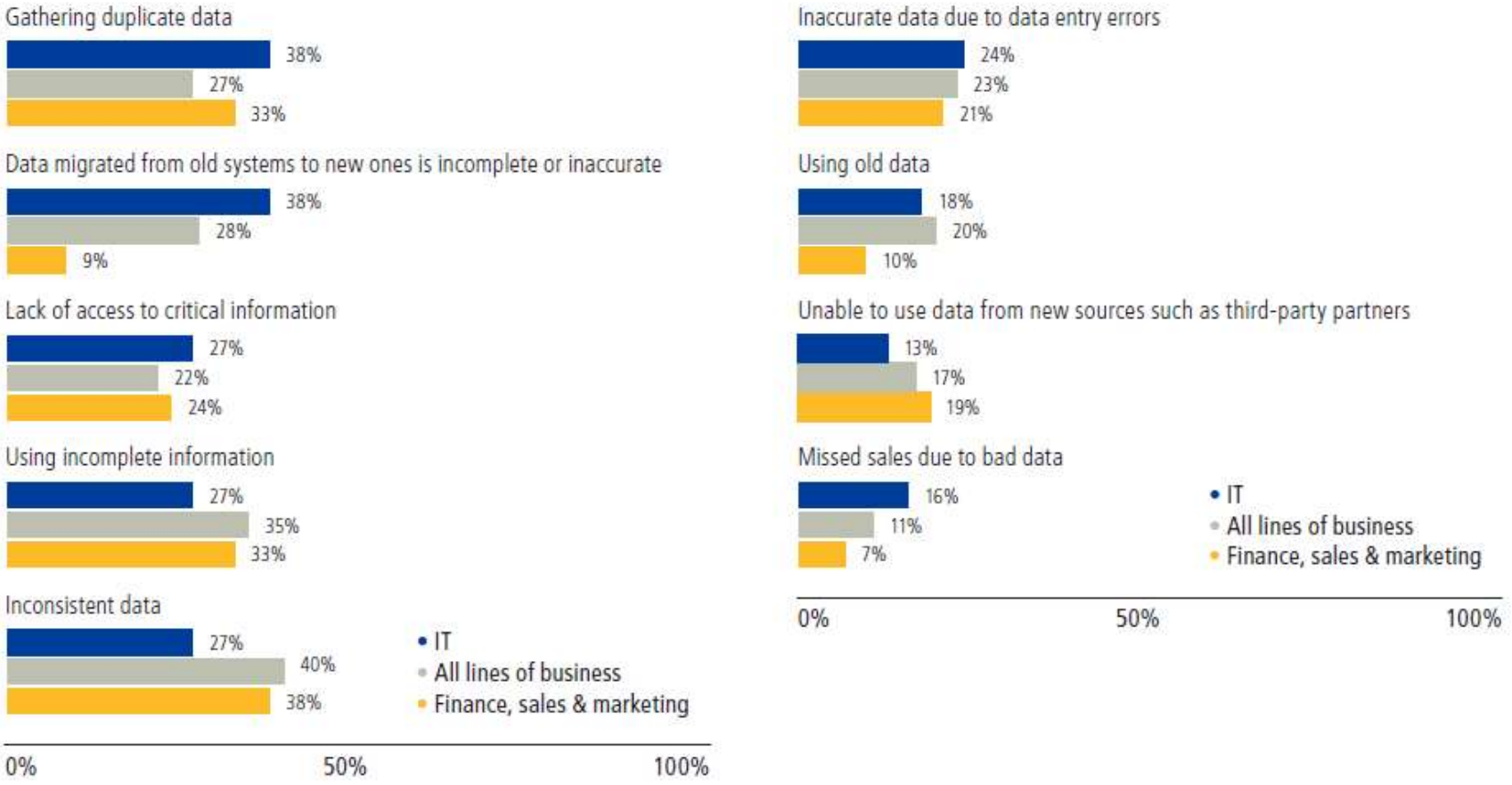
Fig. 3: What do you estimate that data-related issues cost your company annually?



* Source: http://images.forbes.com/forbesinsights/StudyPDFs/SAP_InformationManagement_04_2010.pdf

Selected Study Results (2)

What are the primary data quality issues experienced by business managers at your company?



* Source: http://images.forbes.com/forbesinsights/StudyPDFs/SAP_InformationManagement_04_2010.pdf

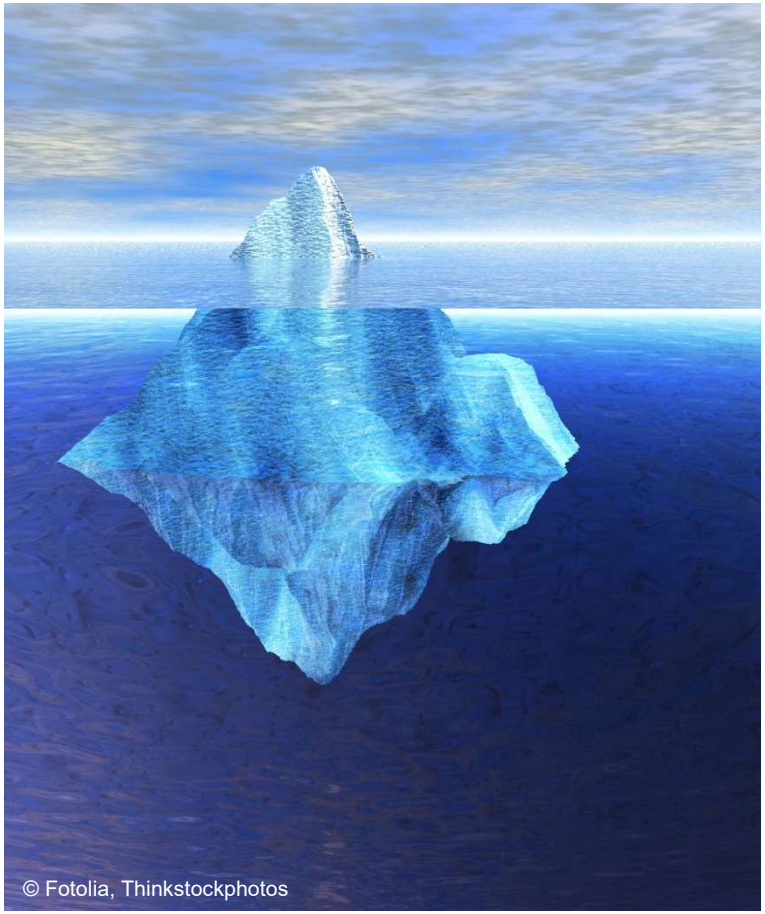
If Master Data Quality were visible . . .

If not measured and made visible master data quality is elusive. It becomes visible when you see the impacts of bad quality. But then it is too late.

. . . which room would then resemble the quality of your master data ?

The Tip of the Iceberg

Only a fraction of master data quality deficiencies is visible to you. Most of the deficiencies you are probably not aware of.

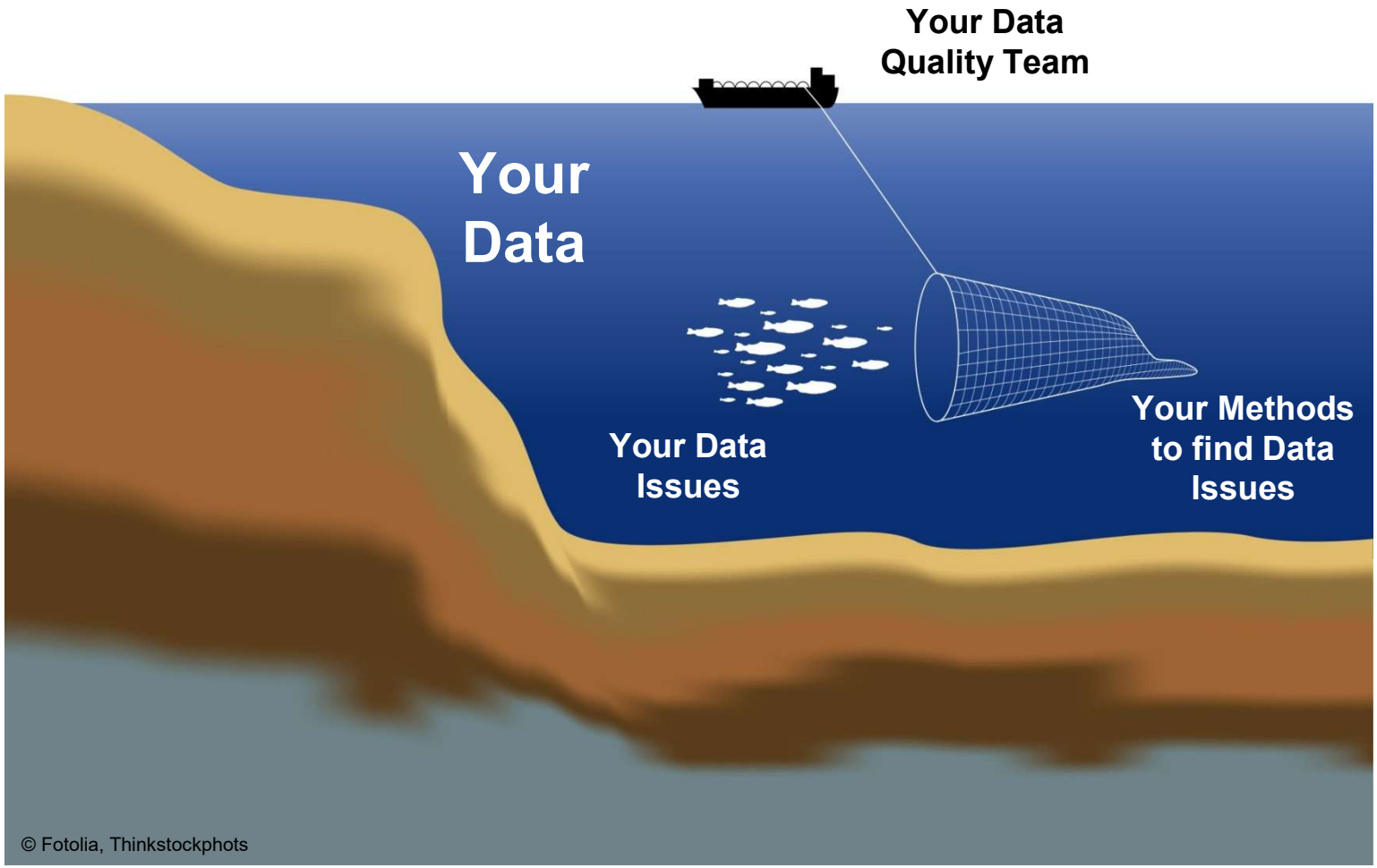


Master data deficiencies you already know

Deficiencies that can be disclosed by a Master Data Health Check

Fishing for Errors

Key for an effective master data health check are business rules. Business rules filter the issues from your data like a net that is used to catch fish.



Measuring Master Data Quality

Quality must be measured or described. Appropriate criteria for a master data health check are those that can be measured with reasonable efforts.

Content

- Reputation
- Free of Error
- Objectivity
- Believability

Utilization

- Timeliness
- Value-Added
- Completeness
- Appropriate Amount of Data
- Relevancy

IT System

- Accessability
- Ease of Manipulation

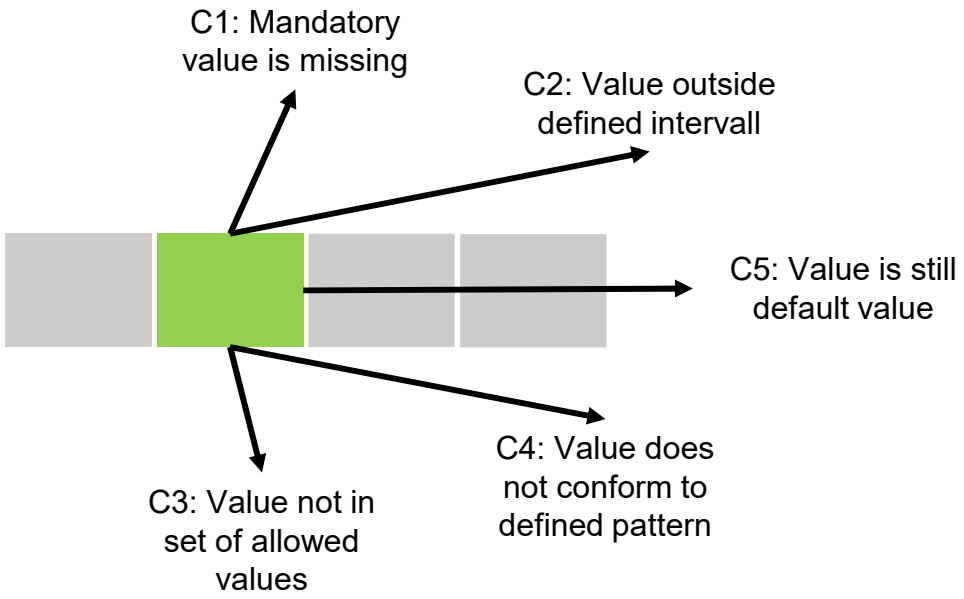
Presentation

- Understandability
- Concise Representation
- Consistent Representation
- Interpretability

* Source: Hildebrand, Gebauer, Hinrichs, Mielke (2008), Daten- und Informationsqualität, Vieweg und Teubner, see also www.dgiq.de

Business Rules – Single Field Checks

Checks for single fields are independent of other fields and are usually easy to implement.



Examples

C1: ZIP Code in postal address is missing

C2: Price cannot be negative

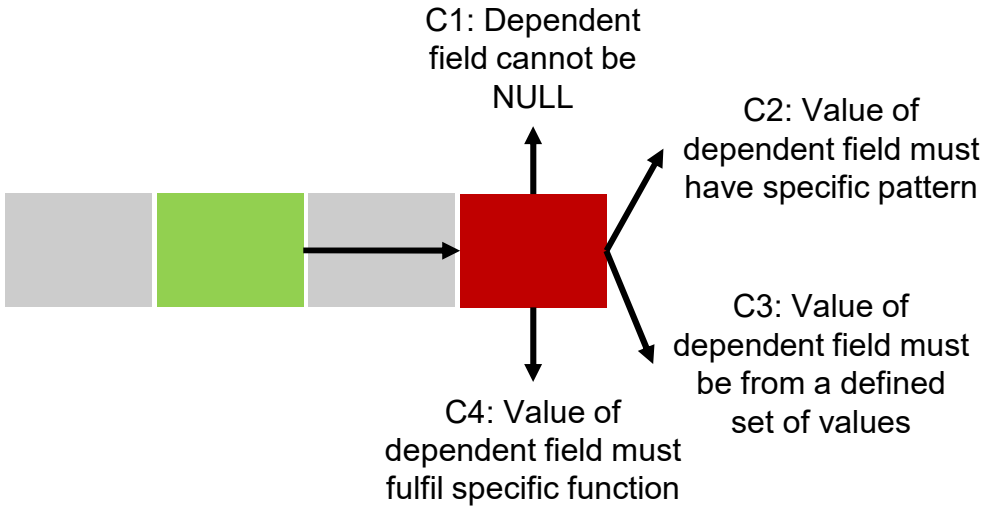
C3: Unit of quantity is not an ISO code

C4: email address has invalid format

C5: Attribute in new record has not yet been maintained

Business Rules – Data Dependency Checks

Although it is comparatively easy to implement data dependency checks finding valid rules requires in-depth knowledge of data semantics and data analysis.



Examples

C1: If net weight is maintained a unit kg must be specified

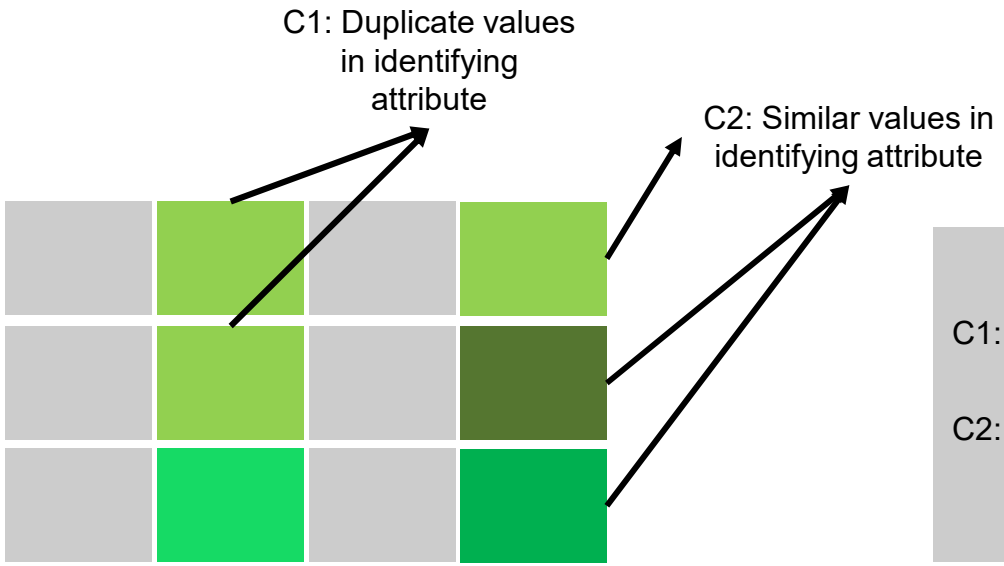
C2: If country is DE then phone pattern starts with +49

C3: Price list applies only for specific customer groups

C4: Net weight must be <= gross weight

Business Rules – Duplicate Checks

While finding duplicates for unique values is simple finding duplicates that are not equal but similar is much more complex.



Examples

C1: Different materials have same vendor material number

C2: Customers have same zip code and address with last name and street is almost identical

Business Rules – Consistency Checks

Consistency checks can refer to attributes within a record or to the consistency of different records.

C1: Values of associated data are inconsistent

C2: Values of different records are inconsistent

Examples

C1: ZIP Code does not match with city name

C2: Same material is assigned to different material group in different records, same zip codes but names of cities are different

Master Data Health Check – Illustration

First Name	Last Name	Country Code	Area Code	Phone	City	Street	ZIP Code	Salary	Marital Status
Werner	Schmidt	49	40		Hamburg	Elbchaussee	22765	60k €	Single
Otto	Maier	49	069	234656	Hamburg	Hafenstr.	60307	96k €	Married
Joachim	Schwarz	0041	44	789123	Zurich	Bahnhofstr.	8001	90k €	Married
Maria E.	Weiss	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Single
Maria E.	Schwarz	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Married
Maria E.	Schwarz	49	69	635875	Frankfurt	Untermainkai	60311	60k €	Married

Which issues can you find in the following data ?

Incomplete Data

First Name	Last Name	Country Code	Area Code	Phone	City	Street	ZIP Code	Salary	Marital Status
Werner	Schmidt	49	40		Hamburg	Elbchaussee	22765	60k €	Single
Otto	Maier	49	069	234656	Hamburg	Hafenstr.	60307	96k €	Married
Joachim	Schwarz	0041	44	789123	Zurich	Bahnhofstr.	8001	90k €	Married
Maria E.	Weiss	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Single
Maria E.	Schwarz	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Married
Maria E.	Schwarz	49	69	635875	Frankfurt	Untermainkai	60311	60k €	Married

Data are missing:

Landline phone no. might not be existing but mobile no. could be existing

Data are not standardized

First Name	Last Name	Country Code	Area Code	Phone	City	Street	ZIP Code	Salary	Marital Status
Werner	Schmidt	49	40		Hamburg	Elbchaussee	22765	60k €	Single
Otto	Maier	49	069	234656	Hamburg	Hafenstr.	60307	96k €	Married
Joachim	Schwarz	0041	44	789123	Zurich	Bahnhofstr.	8001	90k €	Married
Maria E.	Weiss	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Single
Maria E.	Schwarz	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Married
Maria E.	Schwarz	49	69	635875	Frankfurt	Untermainkai	60311	60k €	Married

Data are not standardized:

Formats for Country Code and Area Code must be standardized

Multiple First Names should be separated or not abbreviated

Inconsistent Data

First Name	Last Name	Country Code	Area Code	Phone	City	Street	ZIP Code	Salary	Marital Status
Werner	Schmidt	49	40		Hamburg	Elbchaussee	22765	60k €	Single
Otto	Maier	49	069	234656	Hamburg	Hafenstr.	60307	96k €	Married
Joachim	Schwarz	0041	44	789123	Zurich	Bahnhofstr.	8001	90k €	Married
Maria E.	Weiss	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Single
Maria E.	Schwarz	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Married
Maria E.	Schwarz	49	69	635875	Frankfurt	Untermainkai	60311	60k €	Married

Data are inconsistent:

Area Code + ZIP Code refer to Hafenstrasse in Frankfurt, City seems to be wrong

Obvious Data Duplicates

First Name	Last Name	Country Code	Area Code	Phone	City	Street	ZIP Code	Salary	Marital Status
Werner	Schmidt	49	40		Hamburg	Elbchaussee	22765	60k €	Single
Otto	Maier	49	069	234656	Hamburg	Hafenstr.	60307	96k €	Married
Joachim	Schwarz	0041	44	789123	Zurich	Bahnhofstr.	8001	90k €	Married
Maria E.	Weiss	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Single
Maria E.	Schwarz	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Married
Maria E.	Schwarz	49	69	635875	Frankfurt	Untermainkai	60311	60k €	Married

Obvious Data Duplicates:

Two records have the same address + phone no.

Potential Data Duplicates + Data Timeliness

First Name	Last Name	Country Code	Area Code	Phone	City	Street	ZIP Code	Salary	Marital Status
Werner	Schmidt	49	40		Hamburg	Elbchaussee	22765	60k €	Single
Otto	Maier	49	069	234656	Hamburg	Hafenstr.	60307	96k €	Married
Joachim	Schwarz	0041	44	789123	Zurich	Bahnhofstr.	8001	90k €	Married
Maria E.	Weiss	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Single
Maria E.	Schwarz	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Married
Maria E.	Schwarz	49	69	635875	Frankfurt	Untermainkai	60311	60k €	Married

Potential Data Duplicates + Data not up to date ?:

Two records have the same name. If these are duplicates, one record is outdated and no longer valid. Invalid record could be record with salary 50k as salary could have been increased to 60k when moving to Germany

Relations between Data + Data Timeliness

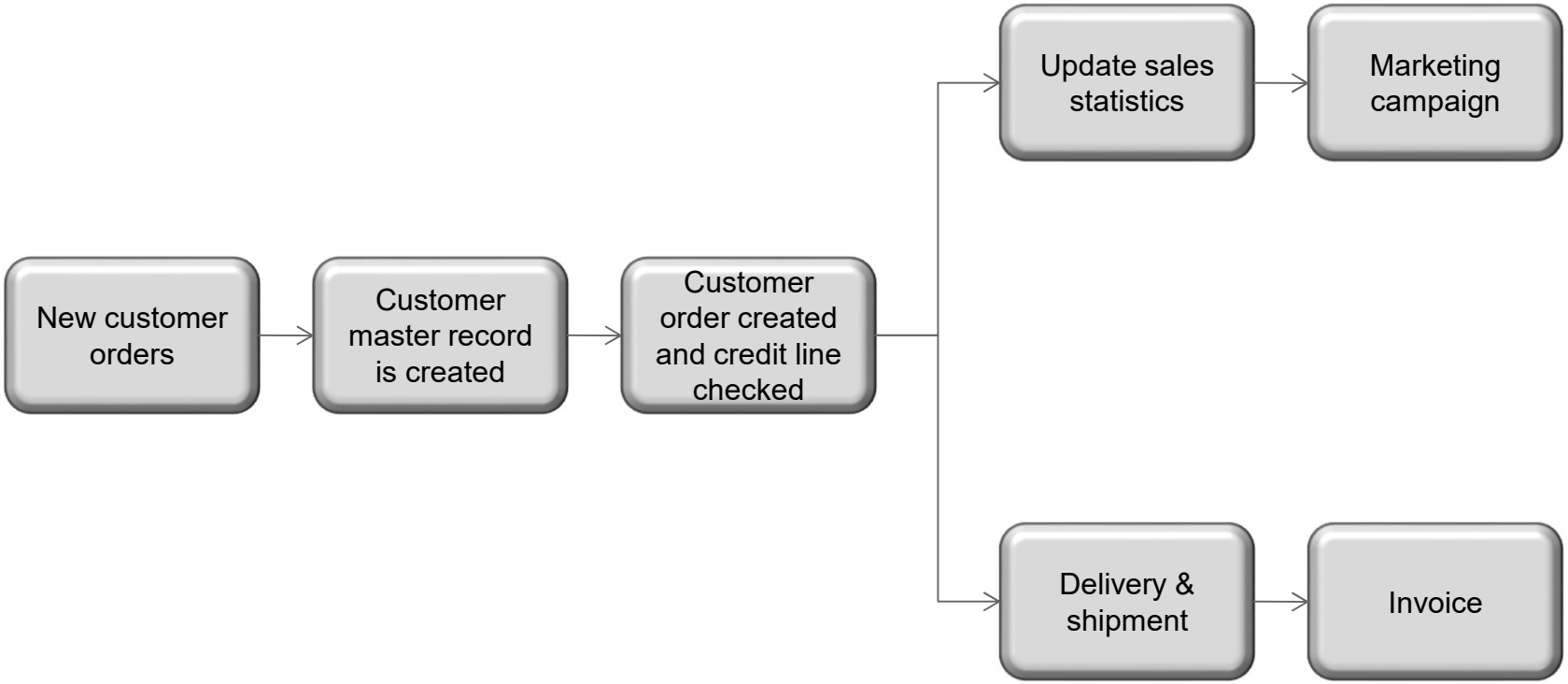
First Name	Last Name	Country Code	Area Code	Phone	City	Street	ZIP Code	Salary	Marital Status
Werner	Schmidt	49	40		Hamburg	Elbchaussee	22765	60k €	Single
Otto	Maier	49	069	234656	Hamburg	Hafenstr.	60307	96k €	Married
Joachim	Schwarz	0041	44	789123	Zurich	Bahnhofstr.	8001	90k €	Married
Maria E.	Weiss	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Single
Maria E.	Schwarz	41	44	789123	Zurich	Bahnhofstr	8001	50k €	Married
Maria E.	Schwarz	49	69	635875	Frankfurt	Untermainkai	60311	60k €	Married

Relationships between Data + Data not up to date ?:

Two records have the same adress and marital status changed. Is Maria E. married to Joachim ? Is Joachim then still living in Zurich provided that Maria E. in Frankfurt is identical with Maria E. in Zurich ?

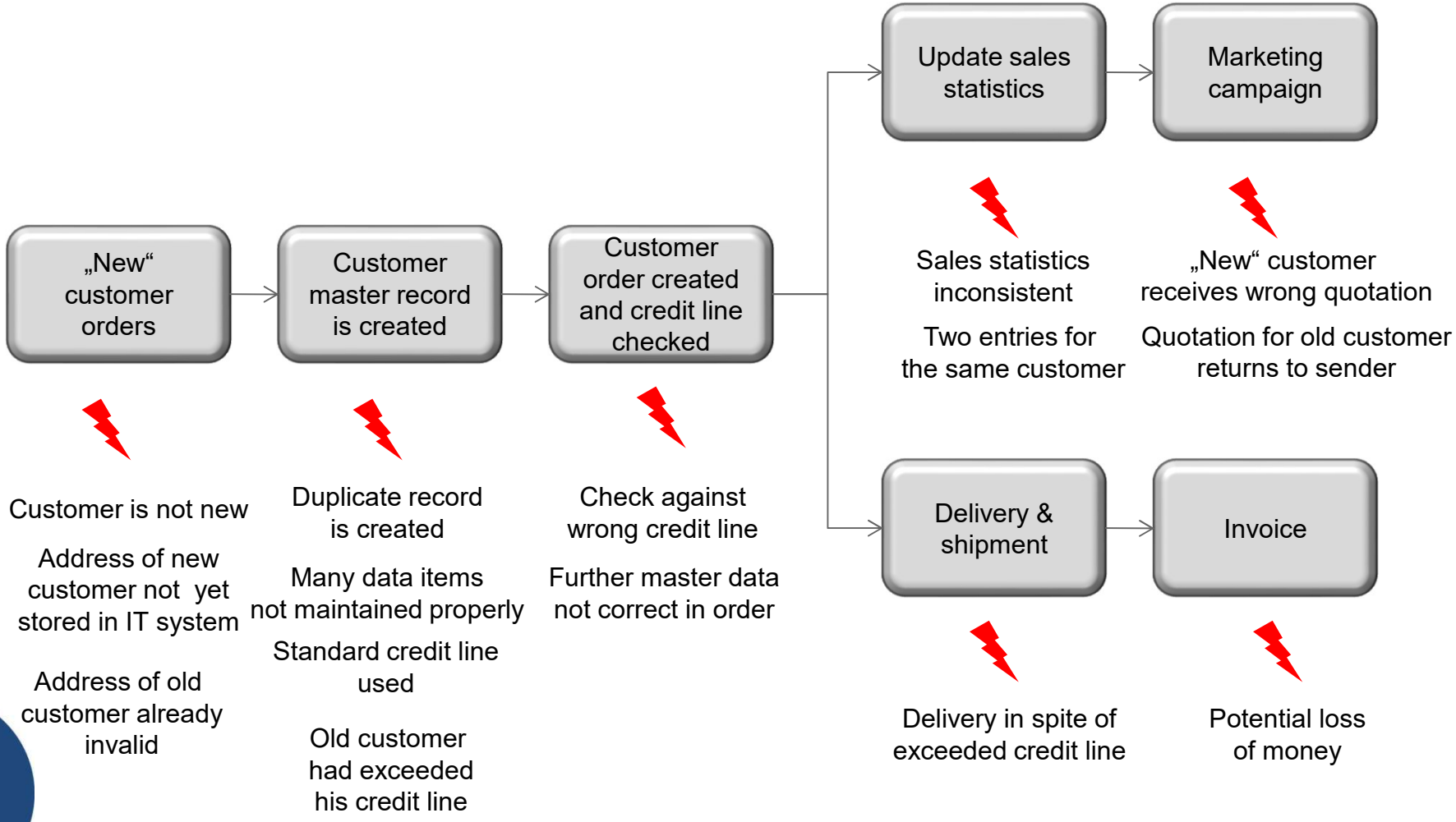
Master Data Duplicates – Exemplary Scenario

A typical scenario when a new customer places an order includes the creation of a customer master data record



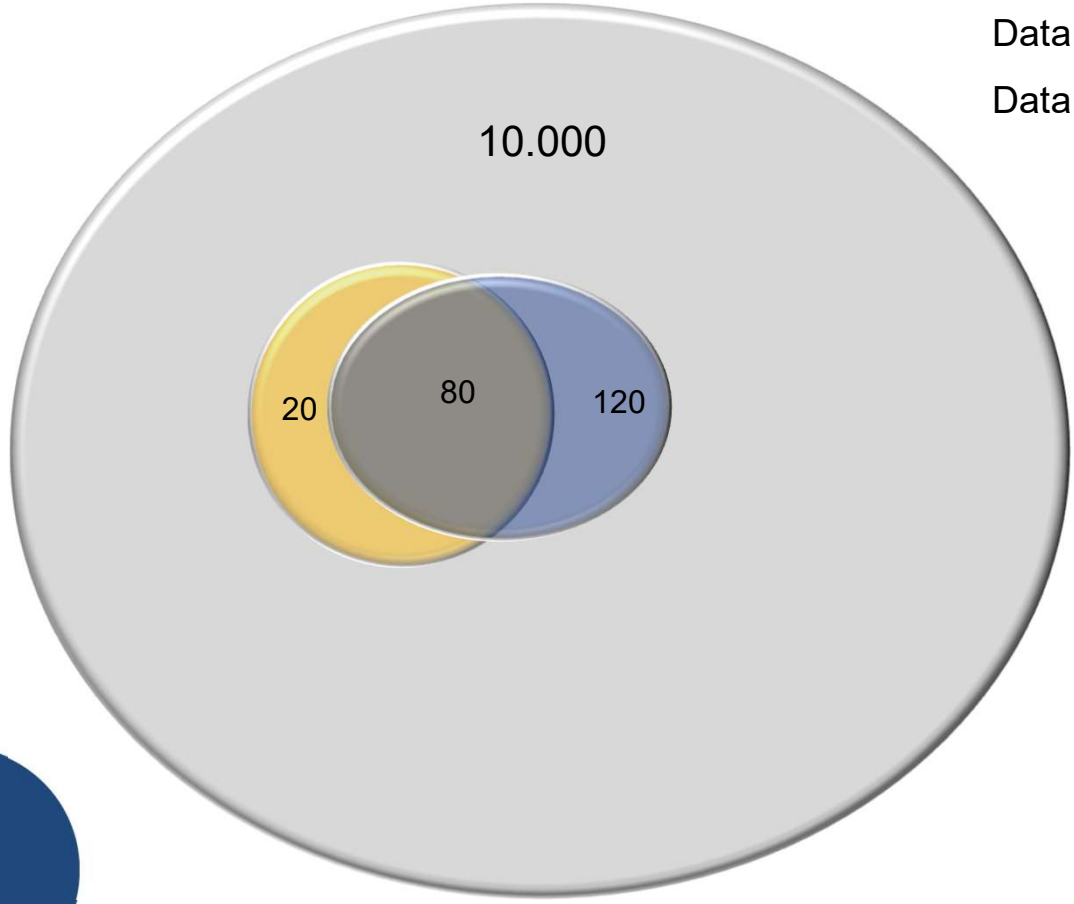
Issues and Risks of Duplicates

A duplicate customer record creates risks and can trigger issues in various steps of the exemplary scenario



Detecting Duplicates automatically

Automated duplicate checks must cope with two types of errors: Actual duplicates that are not found and records marked as duplicates that are actually unique.



Actual Data Duplicates: 100

Data Duplicates proposed by Tool: 200

Data Duplicates remaining after Cleansing: 20

All Data

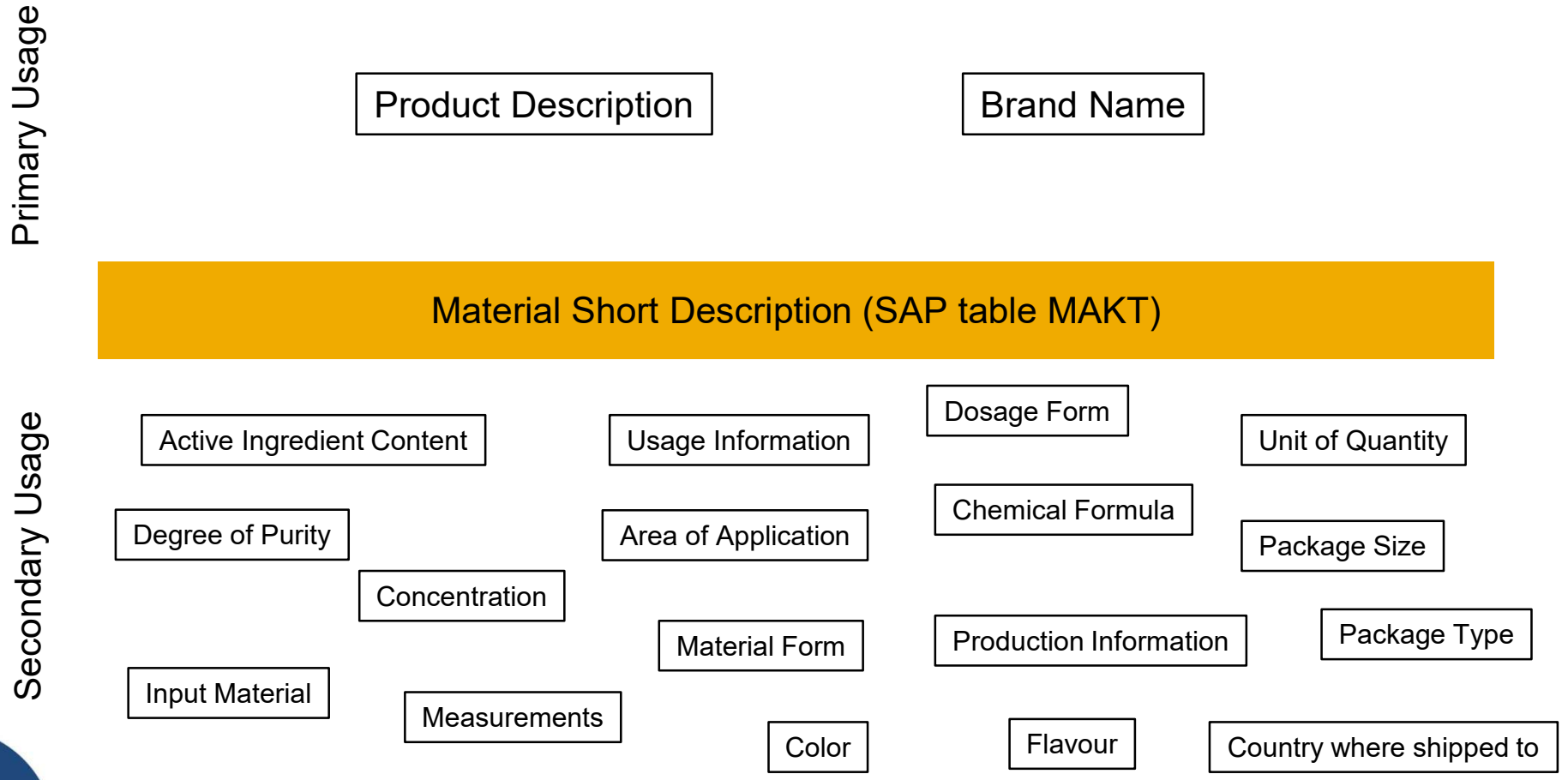
Actual Data Duplicates but not found

Actual Data Duplicates found

Data erroneously marked as Duplicates

Text Analysis – Example Material Short Text

The material short text does often include much more information than just the product name *. Information in texts can be inconsistent with information stored in attributes.



*** Based on inspection of MAKT usage from chemical + pharmaceutical company with more than > 150.000 records**

Master Data Health Check - Summary

A master data health check can be done with a comparatively small effort. Investments will pay off very soon.

- A master data health check delivers a valuable insight into the quality of your master data
- For an effective check the right business rules must be applied. Data owners have the best knowledge about these rules.
- Data quality will be measured by appropriate KPIs. Not everything can be measured. Measurement must be set against the costs of measuring.
- The health check is a starting point. It delivers arguments why further investments into data quality should be made.
- There are always root causes for bad quality. The results of the check can be used to find them.

Contact

Oliver Jutke

Dipl.-Wirtschaftsinformatiker (TU)

email: oj@consulting4bit.com

Web: www.consulting4bit.com

business consulting IT

bringing business and IT together