Trusting Data for Action

Taking Care of Uncertainty - The Power of Trust

Massimo Galli, Vice-President
Pimsoft S.p.A.
Section 1

Introduction
Tusting data for action

Any decision, business wise or operational, should be based upon trustable information

• What defines data as trustable?
• What is certainty?
• Is it something tangible?
• Is it measureable?
• How to be able to trust our data?
Can we trust our analytics?
Section 2

Certainty vs. Uncertainty
Certainty or uncertainty

• “Certainty is perfect knowledge that has total security from error, or the mental state of being without doubt” – Wikipedia

• “Doubt is not a pleasant condition, but certainty is absurd” — Voltaire

Certainty is not of this world: we have to deal with uncertainty

• “In this world nothing can be said to be certain, except death and taxes” — Benjamin Franklin
Measure of uncertainty
Intuition on Uncertainty

• Assume we have an information source, e.g., a flip of a coin. How much information do we receive when the outcome is heads?

  • If it’s a fair coin, i.e., \( P(\text{heads}) = P(\text{tails}) = 0.5 \), we say that the amount of information is **1 bit**.

  • If we already know that it will be (or was) heads or if the coin is not fair, i.e., \( P(\text{heads}) = 1 \), the amount of information is **zero**!
Measure of uncertainty
Shannon’s Information Theory


Shannon’s measure of information is the number of bits to represent the amount of uncertainty (randomness) in a data source, and is defined as entropy of information:

\[ H = - \sum_{i=1}^{n} p_i \log(p_i) \]

Information is always a measure of the decrease of uncertainty at a receiver. It applies to any source of information!!
Sources of uncertainty

- Inherent to Measuring Syst.
- Inherent to use
- Good Data, Bad Results
- Good Data, Bad Usability

- Random
- Context
- Methods
- Condition

Data Disorders
Sources of Uncertainty

- Wrong measurement system
- Poor calibration
- Undetected defective meters
- Manual data entry error
- Transcription error
- Wrong correction factors
- Engineering unit mismatch
- Stalled readings from sensors

- Missing or bad data
- Corrupted records
- Data communication loss
- Wrong date/time reference
- Data asymmetry
- Gross Error, Induced Biases
- Bad presentation
- Wrong aggregation
# Humans cannot fix it

<table>
<thead>
<tr>
<th>Study</th>
<th>Activity</th>
<th>Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baddeley &amp; Longman [1973]</td>
<td>Entering mail codes. Errors after correction. Per mail code.</td>
<td>0.5 %</td>
</tr>
<tr>
<td>Chedru &amp; Geschwind [1972]</td>
<td>Grammatical errors per word</td>
<td>1.1 %</td>
</tr>
<tr>
<td>Dhillon [1986]</td>
<td>Reading a gauge incorrectly. Per read.</td>
<td>0.5 %</td>
</tr>
<tr>
<td>Edmondson [1996]</td>
<td>Errors per medication in hospital, based on data presented in the paper. Per dose.</td>
<td>1.6 %</td>
</tr>
<tr>
<td>Grudin [1983]</td>
<td>Error rate per keystroke for six expert typists. Told not to correct errors, although some did. Per keystroke.</td>
<td>1.0 %</td>
</tr>
<tr>
<td>Rabbit [1990]</td>
<td>Flash one of two letters on display screen. Subject hits one of two keys in response. After correction. Per choice.</td>
<td>0.60%</td>
</tr>
<tr>
<td>Hotopf [1980]</td>
<td>W sample (written exam). Per word</td>
<td>0.9 %</td>
</tr>
<tr>
<td>Hotopf [1980]</td>
<td>10 undergraduates write for 30 minutes, grammatical and spelling errors per word</td>
<td>1.6 %</td>
</tr>
<tr>
<td>Swain &amp; Guttman [1983]</td>
<td>Error reading a graph. Per read.</td>
<td>1 %</td>
</tr>
<tr>
<td>Potter [1995]</td>
<td>Errors in making entries in an aircraft flight management system. Per keystroke. Higher if heavy workload.</td>
<td>10.0 %</td>
</tr>
</tbody>
</table>
Section 4

Sigmafine supporting Data Quality Management
Data Quality Dimensions

Factors affecting an information asset’s utility and ultimately its economic value

<table>
<thead>
<tr>
<th>Objective factors</th>
<th>Subjective factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td><strong>Relevance</strong></td>
</tr>
<tr>
<td>The data accurately represents reality or a verifiable source.</td>
<td>The data is applicable to one or more business processes or decisions.</td>
</tr>
<tr>
<td><strong>Integrity</strong></td>
<td><strong>Usability</strong></td>
</tr>
<tr>
<td>Appropriate links and relationships exist among data.</td>
<td>Business process(es) and/or individuals understand and are able to leverage this data.</td>
</tr>
<tr>
<td><strong>Consistency</strong></td>
<td><strong>Believability</strong></td>
</tr>
<tr>
<td>Each type of data has a single representation.</td>
<td>Data is deemed credible by those using it.</td>
</tr>
<tr>
<td><strong>Completeness</strong></td>
<td><strong>Clarity</strong></td>
</tr>
<tr>
<td>Records are not missing fields. Data sets are not missing instances.</td>
<td>Data has a unique credible and can be easily comprehended.</td>
</tr>
<tr>
<td><strong>Accessibility</strong></td>
<td><strong>Objectivity</strong></td>
</tr>
<tr>
<td>Data is easily retrieved and/or integrated into business processes.</td>
<td>Data is unbiased and impartial and not depend on the judgment, interpretation, or evaluation of individuals.</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td><strong>Scarcity</strong></td>
</tr>
<tr>
<td>Data is recorded with the precision required by the business.</td>
<td>Is the data proprietary, secret, and difficult to come by?</td>
</tr>
</tbody>
</table>
Our Mission: making data trustable

“Creating value for our customers by reducing the Uncertainty and increasing Reliability and Usability of process & manufacturing data in order to unlock the latent economic value of information”
**Sigmafine data quality indicators**

**Measurement based – Test indicators and more**
- Test1 index of correction
- Test3 index of imbalance
- Solvability

**Model based – DX indicators**
- DX0 index of redundancy
- DX1 index of imbalance
- DX2 index of correction
- DX3 index of accuracy
Section 4

Sigmapfine & Data Quality: supporting asset life cycle
Sigafine supporting asset lifecycle

Data Quality
• Intrinsic

Accuracy
• Intrinsic

Usability
• Extrinsic

© Pimsoft Inc. – 2015 – Do not reproduce without authorization - Sigafine® is a registered trademark of Pimsoft Inc.
Plan: optimal placement for a new meter

• CAPEX investment on a new meter

The optimal placement is the one minimizing the overall amount of uncertainty of the measurements system.
Operate: Should I trust my operational KPI?

Reliability KPI: How much my Performance Indicator is trustable

Performance KPI: How much I am performing
Audit: supporting compliance to regulations

COMPLIANCE - SOX
KEY CONTROL – DAILY MASS BALANCING (CONT.)

- We track the DX1, DX2 and DX3 daily.
- If we have a case which does not meet the criteria set above, we provide explanation for that day.
- Exception reporting - occasionally happens when unit startup/shutdowns occur or other unit upsets
- Finance Manager reviews the monthly monitoring spreadsheet and signs as evidence of the control.
Replace/Repair: meters bias and drift
Replace/repair: health status of instrumentation system

Overall DX1 Trend (Q4 2008 - Q1 2010)
Section 5

Pimsoft prescription: the Power of Trust
Pimsoft prescription

• Define
  • Analyze different areas of error you may be subjected to including but not limited to measurement networks, IT systems integration, human error, time synchronization
  • Leverage Sigmafine indicators to establish your specific DQ indicators

• Track
  • Make it pervasive: bind each pieces of valuable indicators to data quality metrics

• Analyze
  • Historize DQ metrics, too
  • On a hourly, daily, monthly basis
  • Understand hidden trends (degradation, drifts, upset…)

• Act
  • Take actions on your real assets
  • Take actions on your information asset
  • Adjust your DQ metrics
We believe there is a value given by the Information Gain companies can capture by applying Sigmanine together with Data Quality practices
Today’s presentation

- Prof. Bagajewitz / Roberto Linares: the value of information
- Phillips 66: expanding the usage of Sigmafine
- ESM @ Suncor: leveraging data quality indicators
- Iplom: trustable data in compliance to energy efficiency regulations
Contact Information

+39 335 5270680
massimo.galli@pimsoftinc.com
https://www.linkedin.com/company/pimsoft
https://twitter.com/pimsoft
THANK YOU!