

Data Thinking & Canvas

Prof. Dr. Yves Staudt

Agenda

1. Why Data Projects Fail
2. What is the Problem?
3. Process Models
4. What is Data Thinking?
5. What is the Data?
6. From Data Thinking to Data Strategy
7. Think Like a Data Head
8. Data Leadership & Canvas
9. Summary

Outline

1. Why Data Projects Fail
2. What is the Problem?
3. Process Models
4. What is Data Thinking?
5. What is the Data?
6. From Data Thinking to Data Strategy
7. Think Like a Data Head
8. Data Leadership & Canvas
9. Summary

- Decision-makers need reliable information “at the push of a button” — but analysts cannot provide it instantly.
- Need for **always-on**, automated and scalable analytics.
- Many possible use cases — but only a few are economically viable.
- **Acceptance of results** increases the success of implementation.

(Szugat, 2022)

Two Failures

Failure 1 — no common direction

60–80% of data-science projects fail. Leading causes: **lack of a common strategy**, unclear goal definitions, unclear users.

Failure 2 — no standardized method

82% of data-science teams do not use a standardized method (*Saltz & Shamshurin, 2019*).

Why Data Projects Fail

- Projects fail for many reasons: lack of funding, tight timelines, wrong expertise, unreasonable expectations.
- Adding data and analysis **increases** the risk: more points of failure, failures hidden behind complex analysis.
- Worst case: a team applies a method it cannot explain, on data it does not understand, to solve a problem that does not matter — and still believes it succeeded.

(Gutman & Goldmeier, 2021)

Barriers Faced by Data Scientists *(Kaggle Survey)*

- 30.4% — lack of a clear question to answer
- 24.3% — results not used by decision-makers
- 19.6% — lack of domain-expert input
- 15.8% — unrealistic expectations of project impact
- 13.6% — difficulty integrating findings into decisions

Outline

1. Why Data Projects Fail
2. What is the Problem?
3. Process Models
4. What is Data Thinking?
5. What is the Data?
6. From Data Thinking to Data Strategy
7. Think Like a Data Head
8. Data Leadership & Canvas
9. Summary

What is the Problem?

Quote from Charles Kettering

“A problem well stated is a problem half solved.”

Many data projects start because companies *like the idea* of implementation, without understanding why the project matters. **First step: clearly define the business problem.**

Five guiding questions:

- ① Why is the problem important?
- ② Who does this problem affect?
- ③ What if we don't have the right data?
- ④ When is the project over?
- ⑤ What if we don't like the results?

(Gutman & Goldmeier, 2021)

Working the Five Questions

- **Important?** Describe the problem to solve, not the technology; make a **stakeholder analysis** (who is affected and how their work will change).
- **Right data?** Every dataset has limited information — no method goes beyond it.
- **When over?** Without clear expectations, teams generate reports no one reads. Find the final deliverable.
- **Don't like results?** The data may answer confidently, but not as stakeholders wished — have a plan for that.

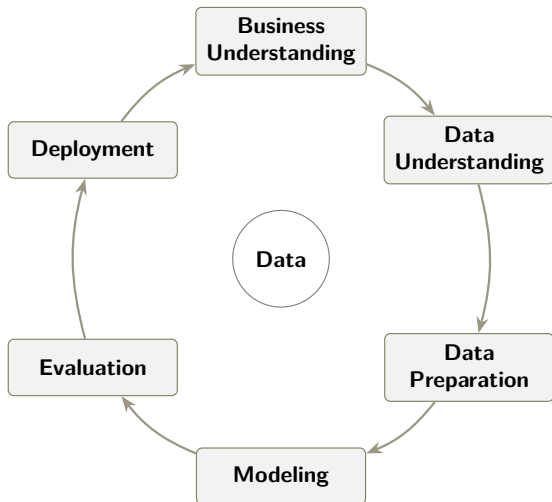
(Gutman & Goldmeier, 2021)

Outline

1. Why Data Projects Fail
2. What is the Problem?
3. **Process Models**
4. What is Data Thinking?
5. What is the Data?
6. From Data Thinking to Data Strategy
7. Think Like a Data Head
8. Data Leadership & Canvas
9. Summary

Process Models: CRISP-DM and End-to-End ML

CRISP-DM (IBM, 2016)

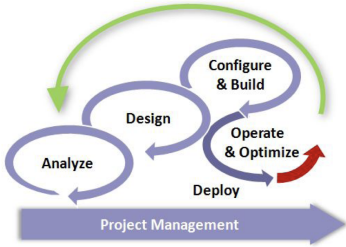


End-to-end ML (Géron, 2017)

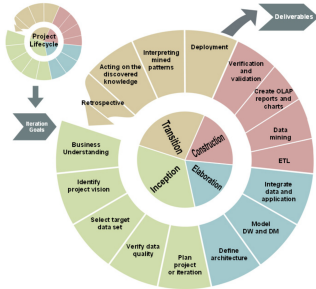
- Look at the big picture
- Get the data
- Discover & visualize to gain insights
- Prepare the data for ML algorithms
- Select a model and train it
- Fine-tune the model
- Present the solution
- Launch, monitor & maintain

- Agile methods break down the rigid CRISP-DM phases and involve stakeholders more closely (*Szugat, 2022*).
- Work proceeds in short, **iterative increments** instead of a strict linear sequence — returning to earlier phases when needed.
- **But:** the customer benefit is still not deeply integrated — a gap that *Data Thinking* sets out to close.

Process Models — Illustrations



ASUM (IBM)



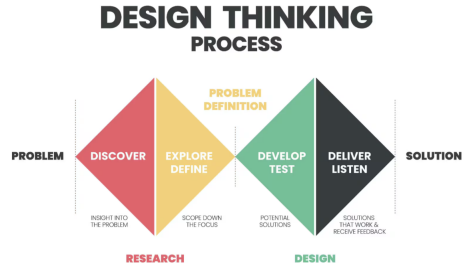
Agile KDD (*Oliveira, 2012*)



Design Thinking + CRISP-DM
(Rodrigues, 2024)

Design Thinking as an Innovation Approach

- **Human-centered:** people's needs, technological feasibility, business viability (*Brown, 2008; Lewrick, 2019*).
- **Double Diamond:** problem space and solution space.
- **Prototyping & rapid iteration** drive product innovation.
- Over 150 methods (e.g. the “5 Whys”).



Design Thinking process (*Stanton, 2024*)

Outline

1. Why Data Projects Fail
2. What is the Problem?
3. Process Models
- 4. What is Data Thinking?**
5. What is the Data?
6. From Data Thinking to Data Strategy
7. Think Like a Data Head
8. Data Leadership & Canvas
9. Summary

- Origin (2013): introduced by Faria & Panigassi — a mental framework to identify, organize, and describe system components (parallels to Design Thinking).
- A **mindset**: use data, algorithms, and technology to drive business decisions and innovation (*Bollhöfer, 2017*).
- Bridges **business and IT**; builds data literacy.
- **Goal**: transform traditional companies into **data-driven** organizations, with data at the core of decision-making.

Data Thinking = Data Science + Design Thinking

Iterative product development tailored to user needs (*Szugat, 2022*). Combines three ways of thinking (*Mike et al., 2022*):

- **Computational thinking** — decomposition, abstraction, systematic approaches.
- **Statistical thinking** — why/how methods apply; variance; samples.
- **Domain thinking** — knowledge of the target domain.

Practical application

Multidisciplinary teams (innovation coach, business, data scientists/engineers). **Proofs of Concept** detect implementation challenges early and reduce failure rates (*Schnakenburg & Kuhn, 2020*).

Many Definitions — One Common Structure

- Faria & Panigassi (2013); University of Lucerne (2017); Biasibetti & Fomenko (2020); StackFuel (2021); Handorf (2023) — each defines its own steps.
- They share an **iterative** arc: from problem identification and definition, to evaluation of findings, to implementation and continuous refinement.
- Aligns with Design Thinking: **Discover** → **Define** → **Develop** → **Deliver**.

Outline

1. Why Data Projects Fail
2. What is the Problem?
3. Process Models
4. What is Data Thinking?
- 5. What is the Data?**
6. From Data Thinking to Data Strategy
7. Think Like a Data Head
8. Data Leadership & Canvas
9. Summary

Quote from Jmi Barksdale

“If we have data, let’s look at data. If all we have are opinions, let’s go with mine.”

- Information is derived knowledge. To communicate or store it, we **encode** it — and encoded information becomes **data**.
- Collection: **observational** (passive — most company data) vs. **experimental** (scientific method, random assignment — the gold standard).

(Gutman & Goldmeier, 2021)

Data Types and Structure

Numeric: continuous vs. discrete (count).

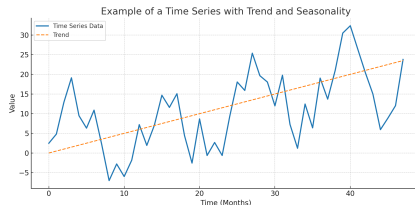
Categorical: ordinal (ordered) vs. nominal (unordered).

(Details in M04 “Data Formats & Modalities”).

Structured — rows and columns; easy to store, search, analyse.

Unstructured — text, images, video, audio; needs special techniques.

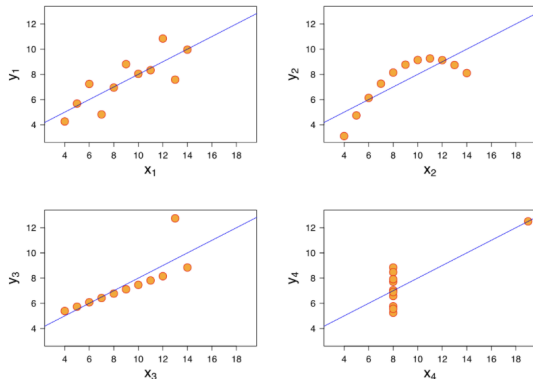
Time series — trend, seasonality, cyclic patterns, noise.



Data Beyond Spreadsheets

- Data is often summarized through **summary statistics** (mean, median, mode; variance, range, standard deviation).
- But statistics alone can mislead — **Anscombe's quartet**: same statistics, very different data.

(Gutman & Goldmeier, 2021)



Anscombe data set

Outline

1. Why Data Projects Fail
2. What is the Problem?
3. Process Models
4. What is Data Thinking?
5. What is the Data?
- 6. From Data Thinking to Data Strategy**
7. Think Like a Data Head
8. Data Leadership & Canvas
9. Summary

- A **long-term plan** for how a company collects, organizes, analyzes, and uses data to achieve its objectives (*Leybold et al., 2021; Loth, 2017*).
- **Why needed:** prevents fragmented data capabilities, data duplication, and conflicting “versions of the truth”; aligns business and IT; sets company-wide metrics and standards.

Data Strategy as Part of Data Thinking

Key elements (*DalleMule & Davenport, 2017; Eckerson, 2011*):

- **Clear objectives, policies**, and data-handling practices.
- **Data architecture** — how data is collected, stored, processed, distributed.
- **Data governance** — policies and procedures so data is managed, protected, and used correctly.
- **Data management** — tools for easy access across sources.
- **Change management**.

Outline

1. Why Data Projects Fail
2. What is the Problem?
3. Process Models
4. What is Data Thinking?
5. What is the Data?
6. From Data Thinking to Data Strategy
- 7. Think Like a Data Head**
8. Data Leadership & Canvas
9. Summary

Quote from Frank Harrell

“Statistical thinking is a different way of thinking that is part detective, skeptical, and involves alternate takes on a problem.”

- **Always ask questions** — it's easy to spot flaws in others' arguments, harder to question your own.
- In boardrooms numbers are treated as “truth in black ink” — and skepticism turns off reflexively.

(Gutman & Goldmeier, 2021)

Two kinds of variation:

- measurement variation (how we collect/measure),
- random variation (inherent randomness).

Is variation just noise, or a true process we can address?

(Gutman & Goldmeier, 2021)

Probability vs. statistics:

- Probability: know the bag → guess your hand (drills *down*).
- Statistics: know your hand → infer the bag (drills *up*).

Casinos use probability; political polls use statistics.

Descriptive vs. Inferential — and Skepticism

- **Descriptive** statistics summarize data;
- **inferential** statistics make principled guesses about the world: *“from the world to the data, and back to the world.”*

“75% of Americans believe UFOs exist!”

Source: 20 tourists at a UFO museum.

Biased sample, tiny sample, false assumption — a Data Head’s skepticism fires immediately.

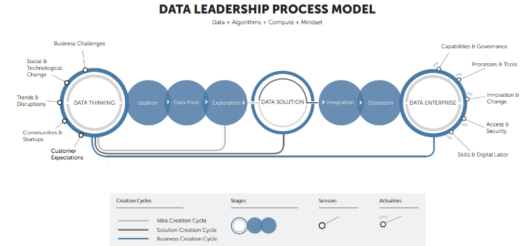
Don’t accept conclusions at face value; don’t blindly trust data or intuition — **think statistically, ask questions.** (*Gutman & Goldmeier, 2021*)

Outline

1. Why Data Projects Fail
2. What is the Problem?
3. Process Models
4. What is Data Thinking?
5. What is the Data?
6. From Data Thinking to Data Strategy
7. Think Like a Data Head
- 8. Data Leadership & Canvas**
9. Summary

Data Leadership Process Model

- 1 **Data Thinking** — the foundation; how to use data and new tech.
- 2 **Data Solutions** — possible and developed projects.
- 3 **Data Enterprise** — transformation into a fully data-driven organization (data in business models, processes, strategy).

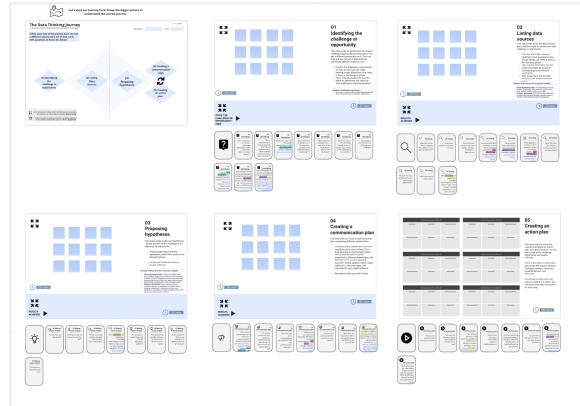


Anmerkung: Quelle Bollhoefer (2017)

The Data Thinking Canvas

A toolbox of canvases supports the workshop (Datentreiber; Thurler):

- Data & AI Design Workshop, Use Case, Business Model,
- Data Management, Data Landscape, Value Chain,
- Stakeholder Analysis, Customer Touchpoints,
- Analytics & AI Maturity, Generative AI, Data Monetization.



Data Thinking Canvas (*Thurler*)

Outline

1. Why Data Projects Fail
2. What is the Problem?
3. Process Models
4. What is Data Thinking?
5. What is the Data?
6. From Data Thinking to Data Strategy
7. Think Like a Data Head
8. Data Leadership & Canvas
- 9. Summary**

- Most data projects fail for non-technical reasons — start by **defining the problem**.
- **Data Thinking** = Data Science + Design Thinking: iterative, user-centered, interdisciplinary.
- Combine computational, statistical, and domain thinking; back it with a **data strategy** and a **Data Head** mindset.

Questions



Figure: Question-mark symbol, from <https://www.qnigge.de/news/detail/modul-v/#images> (accessed 12.07.2021).

Selected References I

- Gutman, A. J. & Goldmeier, J. (2021). *Becoming a Data Head*. Wiley.
- Szugat, S. (2022). *Data Thinking*. Wiley-VCH.
- Schnakenburg, I. & Kuhn, H. (2020). *Data Thinking*. Springer.
- Mike, K. et al. (2022). Computational thinking in the era of data science. *Comm. ACM*, 65(8).
- Géron, A. (2017). *Hands-On Machine Learning*. O'Reilly.
- IBM (2016). *CRISP-DM*; ASUM with Agile Principles.
- DalleMule, L. & Davenport, T. H. (2017). What's Your Data Strategy? *HBR*, 95(3).
- Eckerson, W. (2011). *The Data Governance Imperative*.
- Brown, T. (2008). Design Thinking. *HBR*, 86(6). Lewrick, M. (2019). *The Design Thinking Playbook*. Wiley.
- Faria, M. & Panigassi, R. (2013). Introduction of Data Thinking [unpublished].
- Datentreiber (2024); Thurler — Data Thinking Canvas (Miro board).