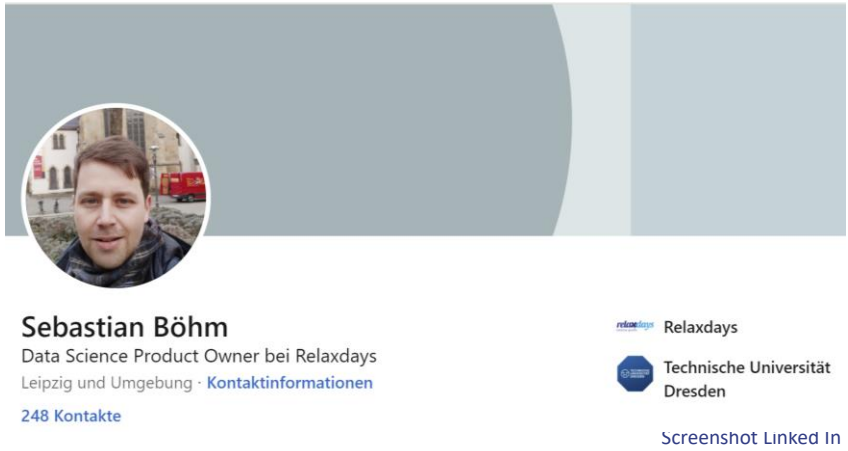
How to create an environment for successful Data Science Projects

Sebastian Böhm, Data Science Product Owner
Relaxdays GmbH

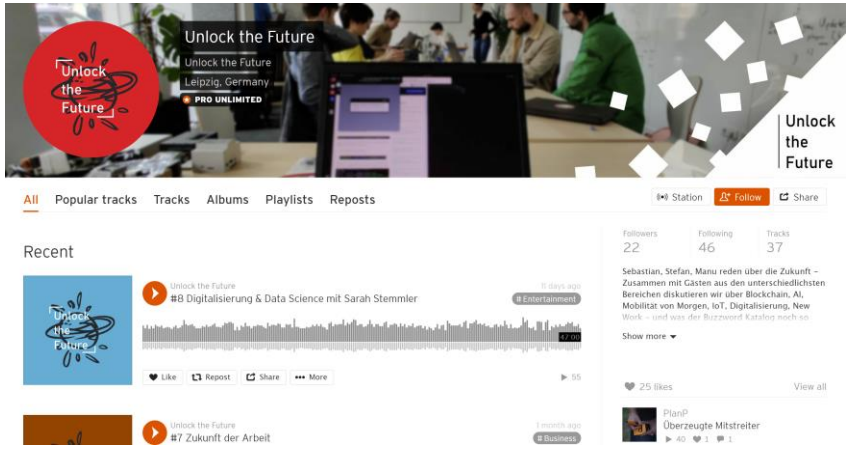
relaxdays
creative goods

About me

- Mathematician
- Podcaster (Unlock the Future)
- 10+ year experience in data analysis (and related projects) with focus on financial reporting, energy trading and energy IoT data
- former PO for an IoT and Data Science Platform for a municipal energy provider
- Speaker at different events like MATLAB Expo, Rebels@Work, Energieforen, GO.Digital



Screenshot LinkedIn



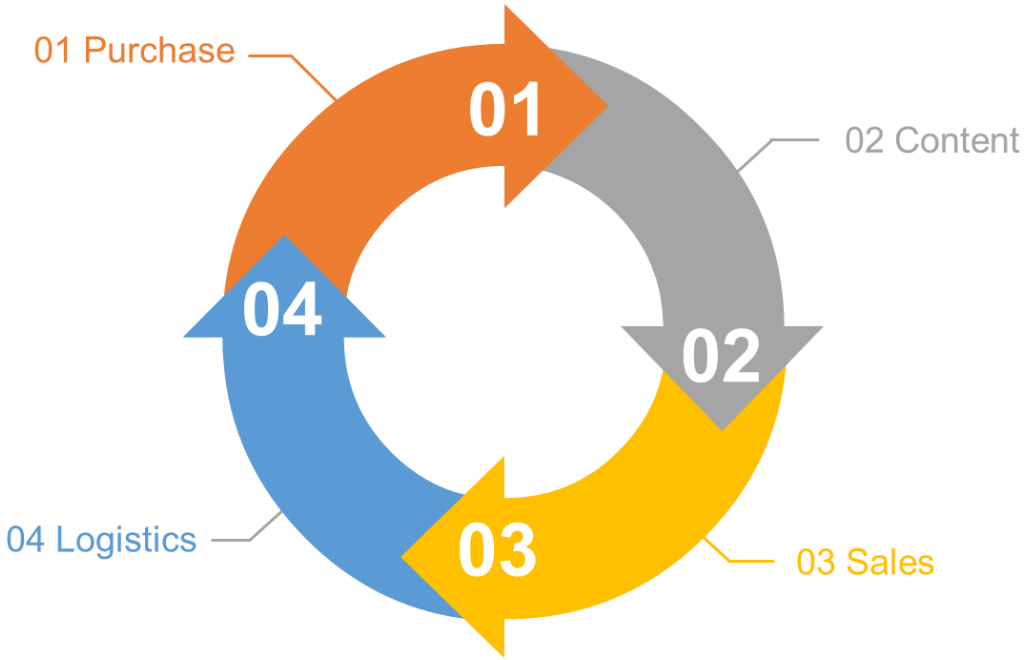
https://soundcloud.com/user-568958765-98202350

Data + Science = Business Value

relaxdays creative goods

What is Relaxdays?

- founded in 2006
- e-Commerce, content creation, logistics, technology
- offices in Halle, Leipzig, Dresden, Prague



480
Mitarbeiter

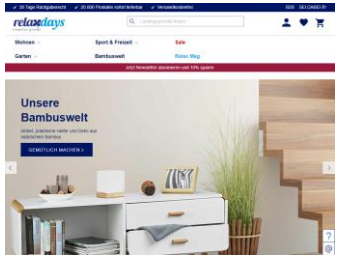
> 40
Teams

> 20.000
Produkte

> 7 Mio.
Kunden

europaweit
Versand

https://relaxdays-unternehmen.de



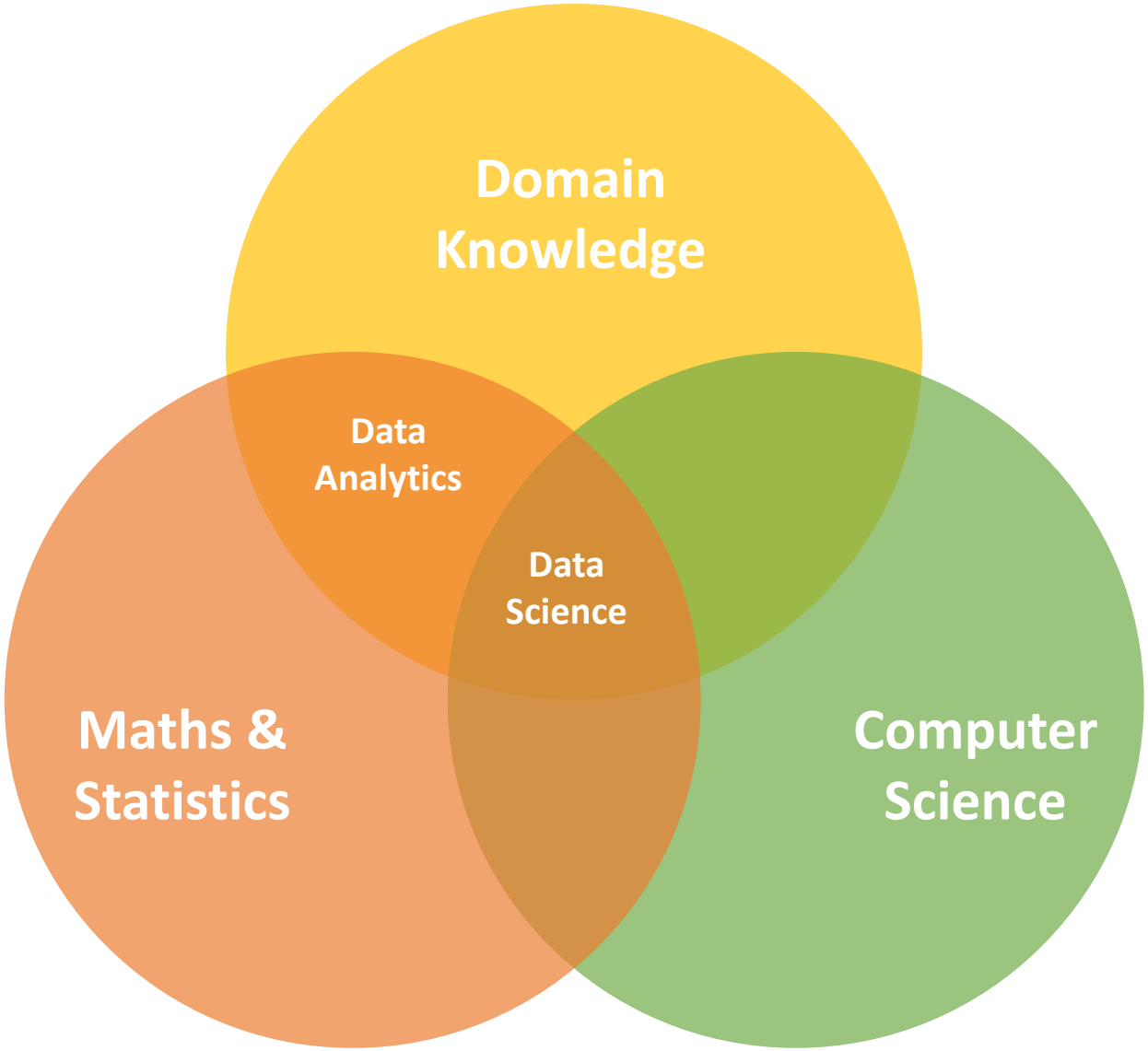
https://relaxdays.de

https://www.youtube.com/c/RelaxdaysDeShop/videos

relaxdays
creative goods

What do I mean with Data Science?

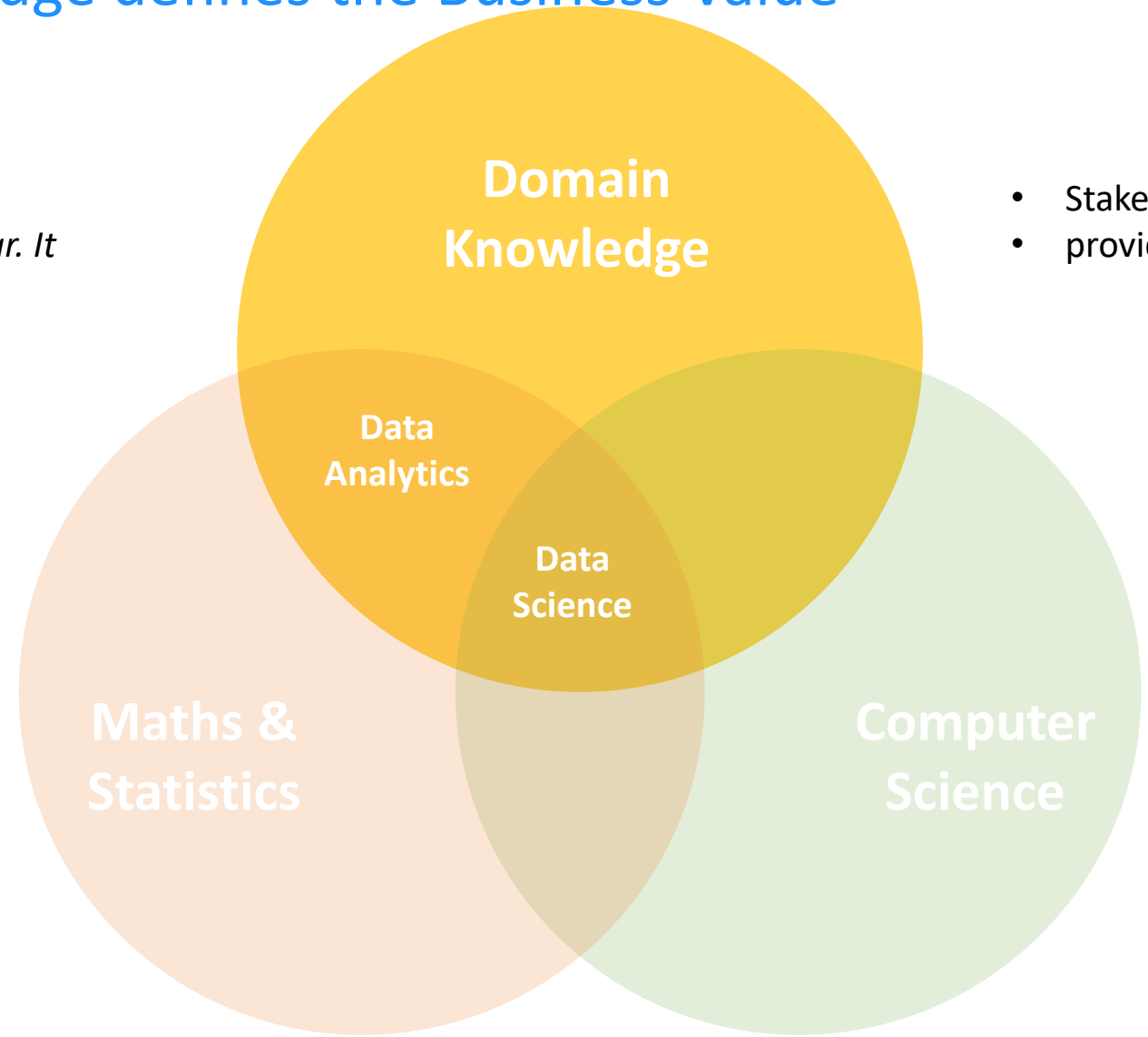
Data Science = usage of valuable analytical functions scalable and 24/7!



relaxdays creative goods

Domain Knowledge defines the Business Value

„I need a forecast of the expected sales of next year. It can‘t be that hard…“

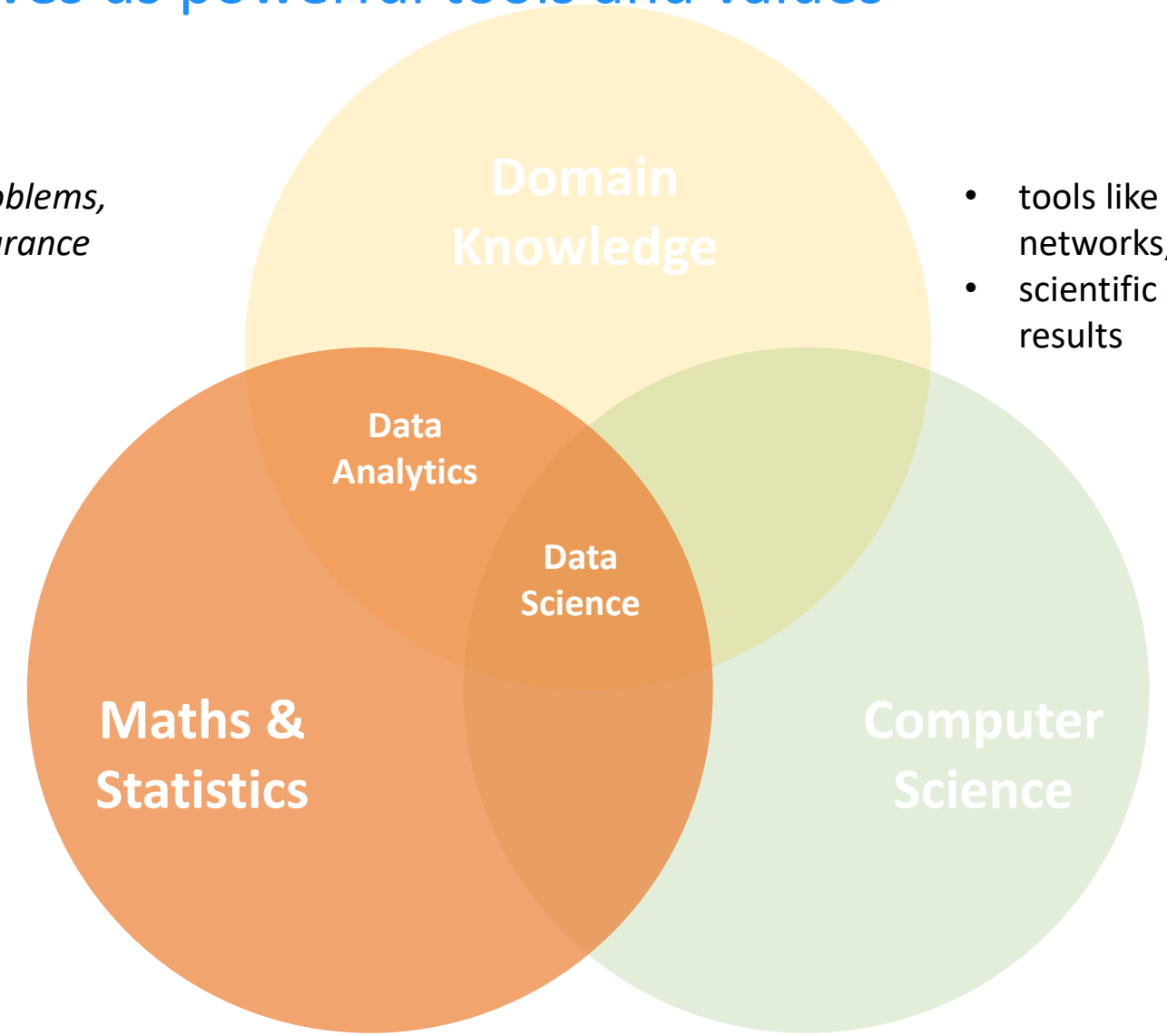


- Stakeholder, Users, Experts
- provide challenges and use cases

relaxdays creative goods

Mathematics gives us powerful tools and values

mindset to solve hard problems,
e.g. persistence and endurance

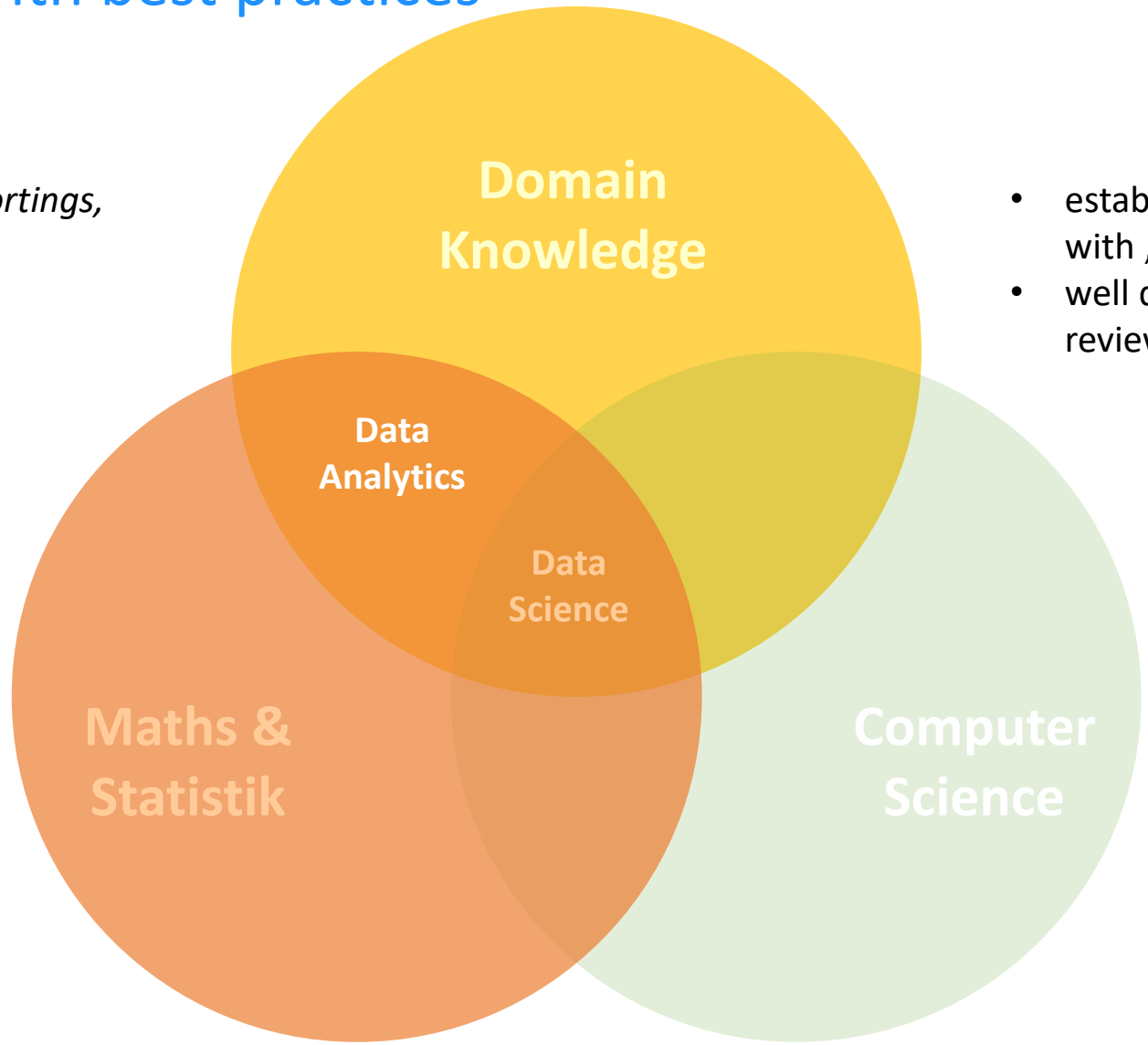


- tools like optimization models, neural networks, machine learning, simulations
- scientific approach, measurability of the results

relaxdays creative goods

Data Analytics with best practices

risk management, BI, reportings, marketing KPIs

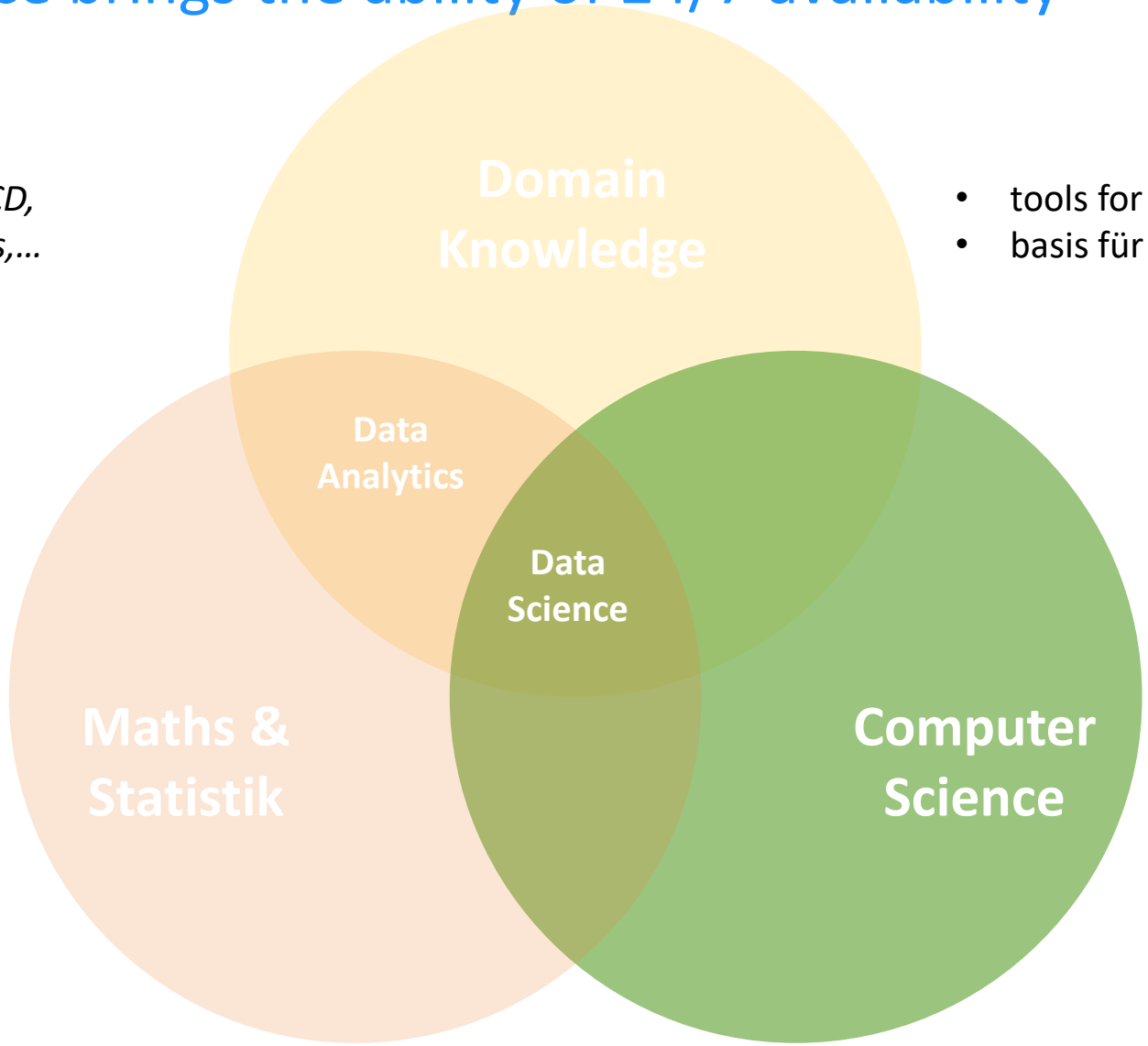


- established orzanizational structures with „Quant-Teams“
- well defined methods with an annual review of these methodes

relaxdays creative goods

Computer Science brings the ability of 24/7 availability

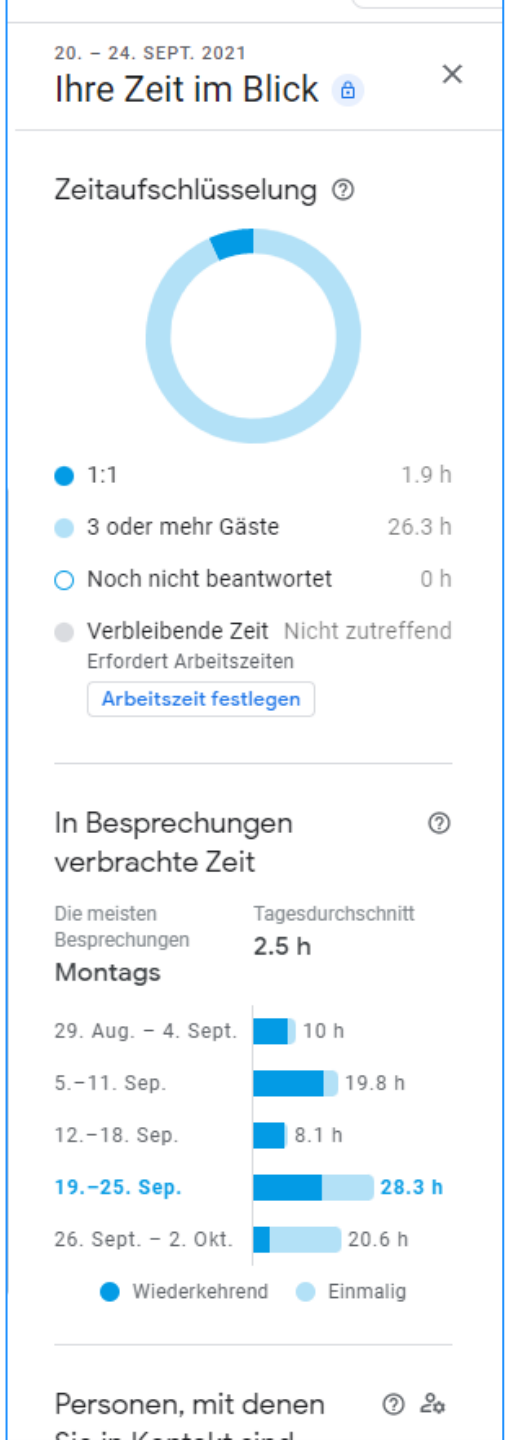
Scrum, Kanban, Agile, CI/CD, DevOps, other frameworks,…



- tools for monitoring and operations
- basis für scalable, high-quality solutions

relaxdays creative goods

Example



Google Calendar Screenshot

Key drivers to success

Important for any Data Science Use Case

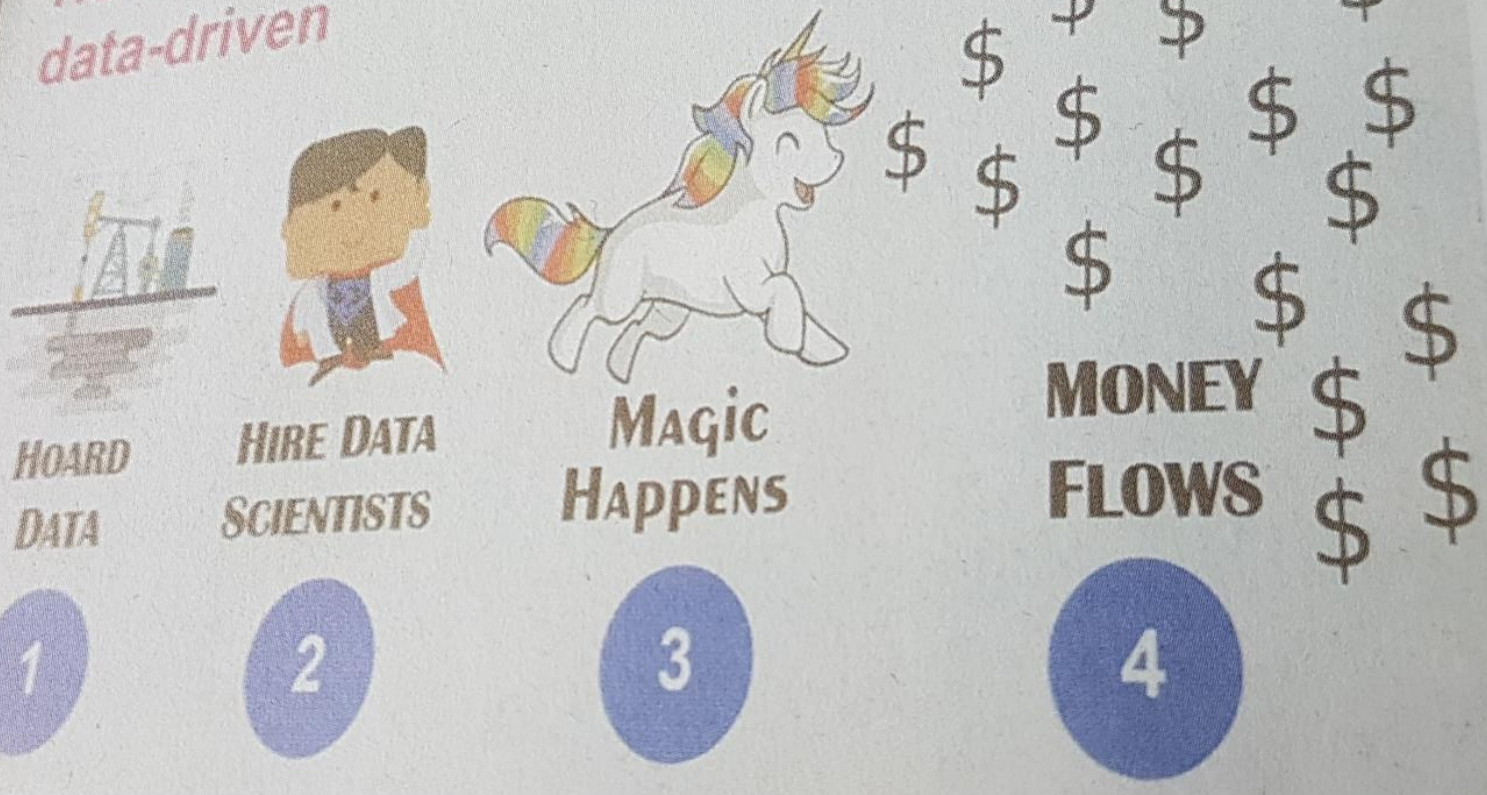
- You have a CDO or at least C-level support
- Your Data Science use cases support the strategic goals of the company
- You have a fantastic WHY which brings you through hard times
- You don‘t think in silos any more
- Your company develops a data-driven-culture
- You have a specific idea and project
- You are able to integrate Data Science approaches into you existing business processes
- You have a good technical basis
- You have all the data needed
- You have the right people with the right skills
- etc.

Let‘s assume we have all of this

relaxdays creative goods

…PLUS The usual construction plan

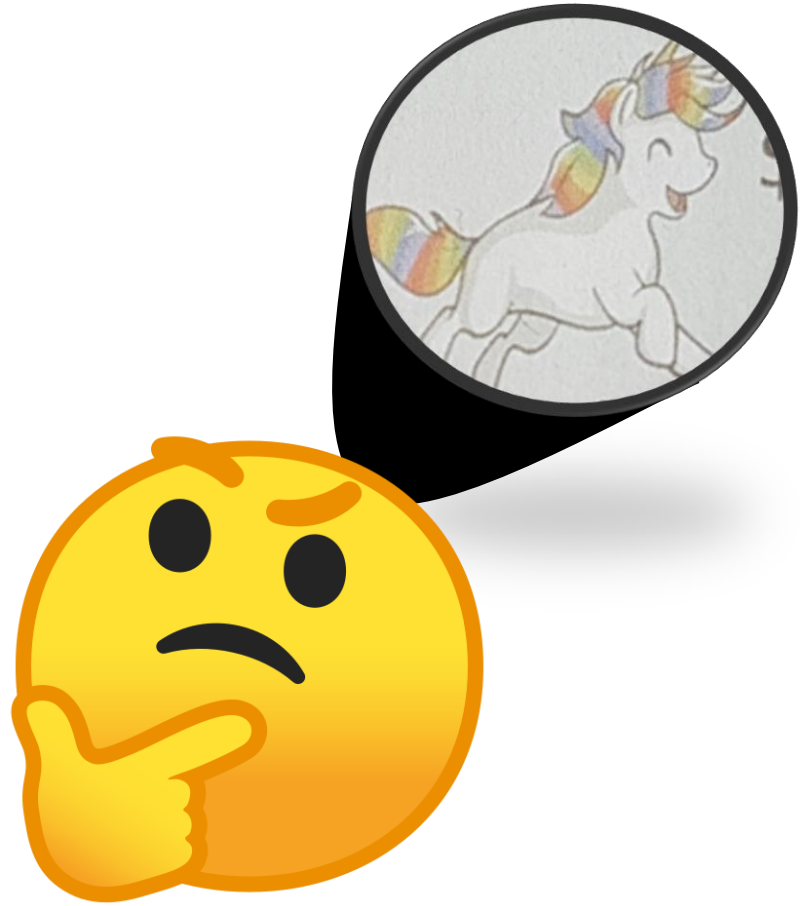
How businesses think they become data-driven



Source lost… a picture taken from a magazine somewhen in 2018

Starting lineup

- a team consisting of several data scientists: physicists, mathematicians, biologists, computer scientists
- which uses the scrum framework
- and Docker, K8s, Gitlab, Python, etc.
- and gots exiting challenges from the business unit
- and a lot of data



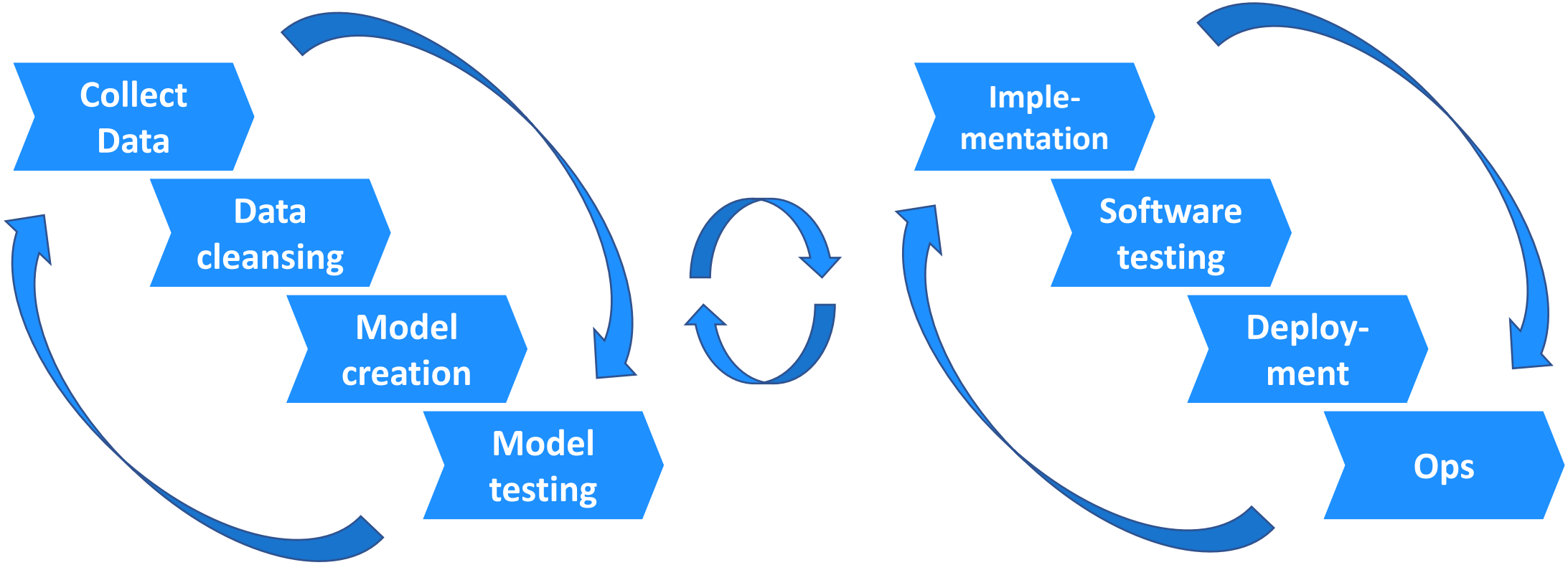
What happens?

„Duck-tales“ or typical software development pitfalls

- „it works, but we can‘t use it in production actually“
- „small“ changes take „a lot“ of time
- team talks about computing power, memory usage etc. (and NOT business value)
- the customer/stakeholder has to listen to long technical explanations instead of getting business value
- there is library or framework for everything, but you don‘t know which one fits perfect to your problem
- scientist forget about scientific approaches (hypothesis -> proof)
- the model worked fine during training, but times are a-changing and we now get different data

relaxdays
creative goods

Why do we end up there? Look at the data science process…

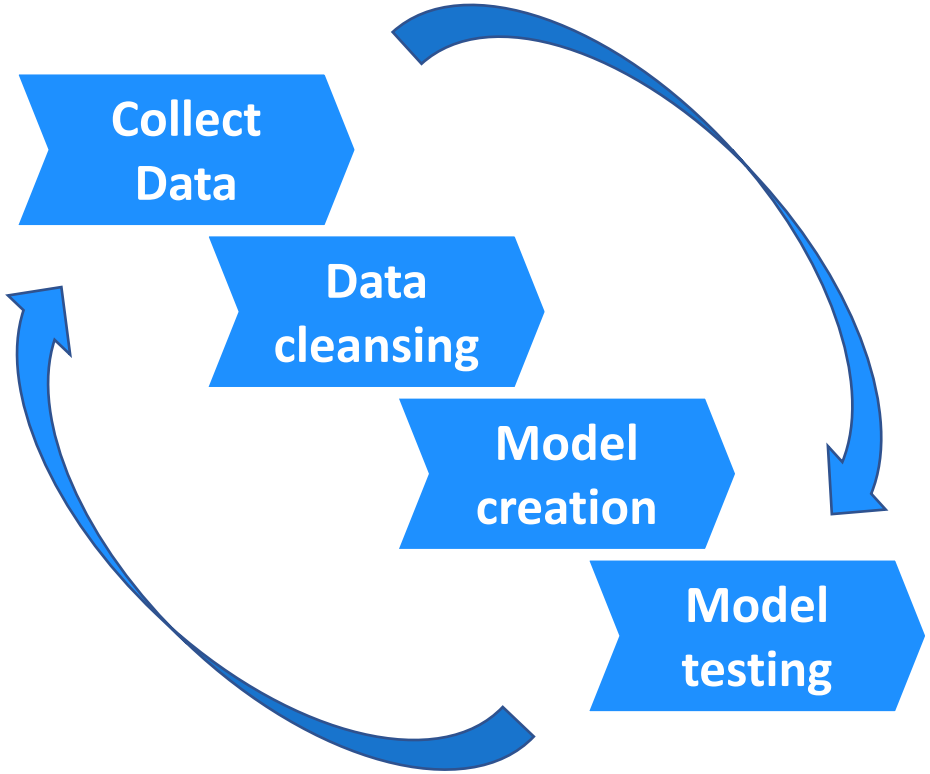


Data Science Cycle

Software Development Cycle

relaxdays creative goods

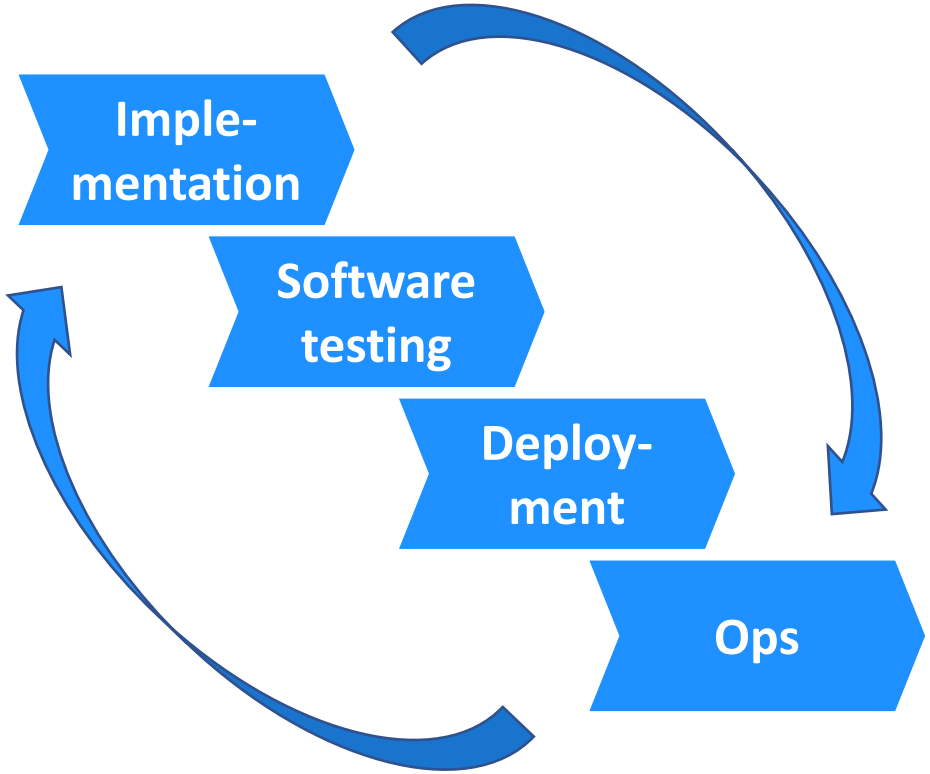
Data Science Cycle creates solutions



Data Science Cycle

- Goal: find solutions for business problems
- fail fast and go on to the next idea
- for this cycle we need data scientists
- In an ideal world:
  - concentrate on the important steps like model creation and model testing
  - automate data aquisition and cleansing

Software development cycle brings solutions to production

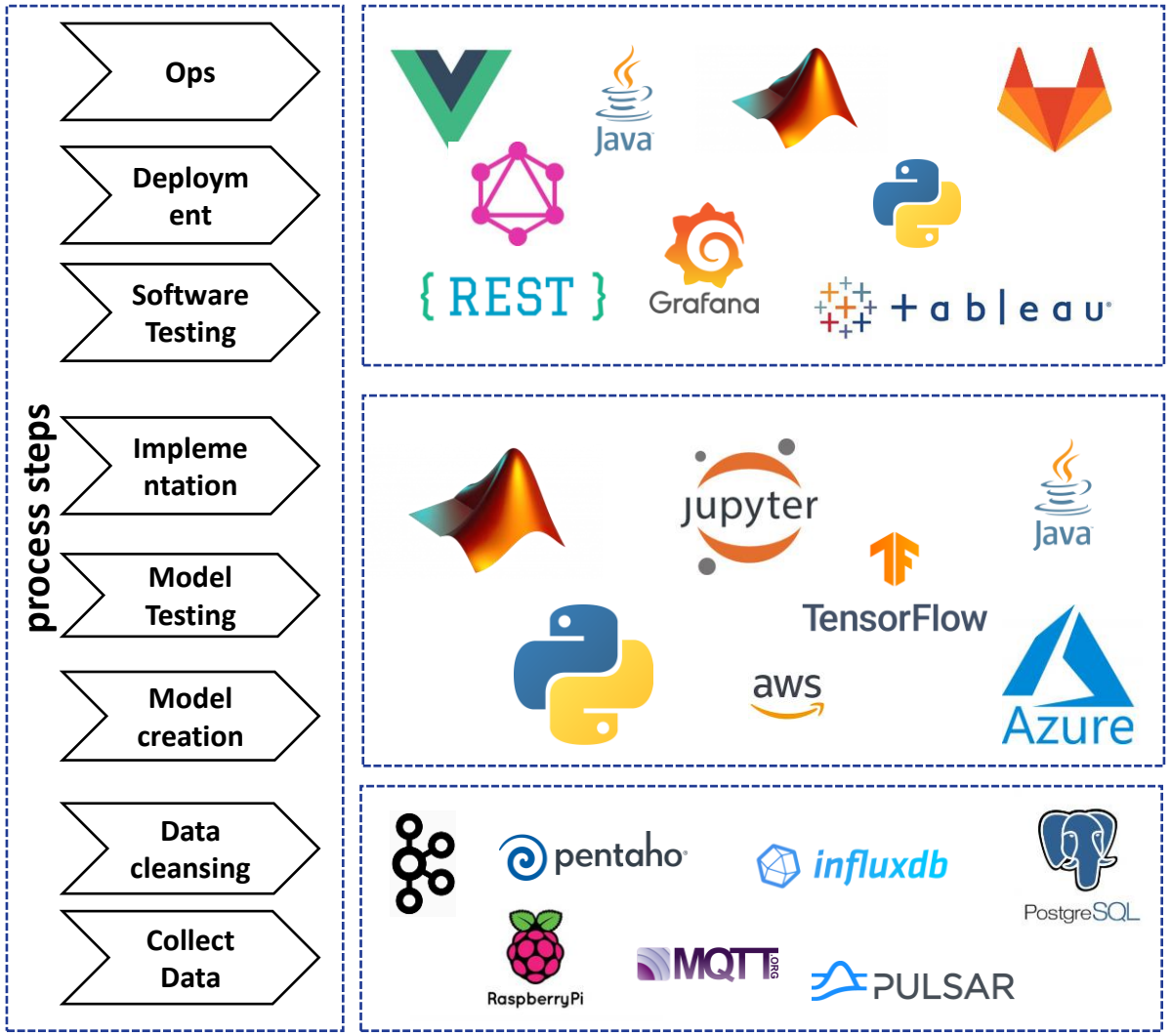


Software Development Cycle

- Goal: bring the solution in production so one can use it
- for this cycle we don‘t really need data scientists
- focus on stability and scalabilty
- In an ideal world:
  - ➢ automate it!

relaxdays
creative goods

Find a proper solution for every step of the procss

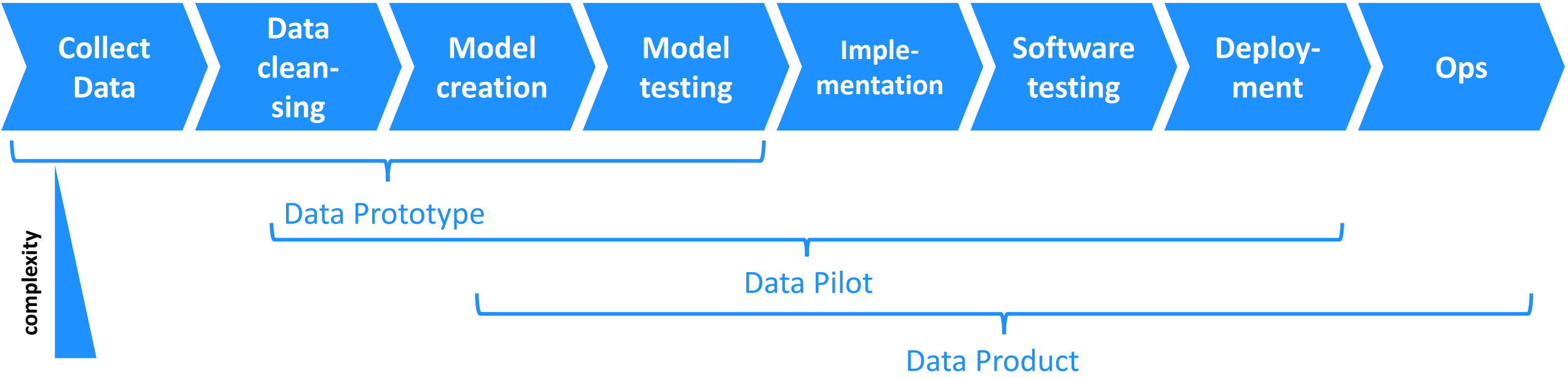


- define principles for you work, like „we prefer microservices instead of monolyths“
- Be open for new technolgies and third party solutions which can help you to automate a process step
- Build it step by step and have always a concrete use case in mind which helps you to create business value

Extensive automation of software development steps enables concentration on professional data science oriented challenges, not on technology

relaxdays creative goods

And one more thing…

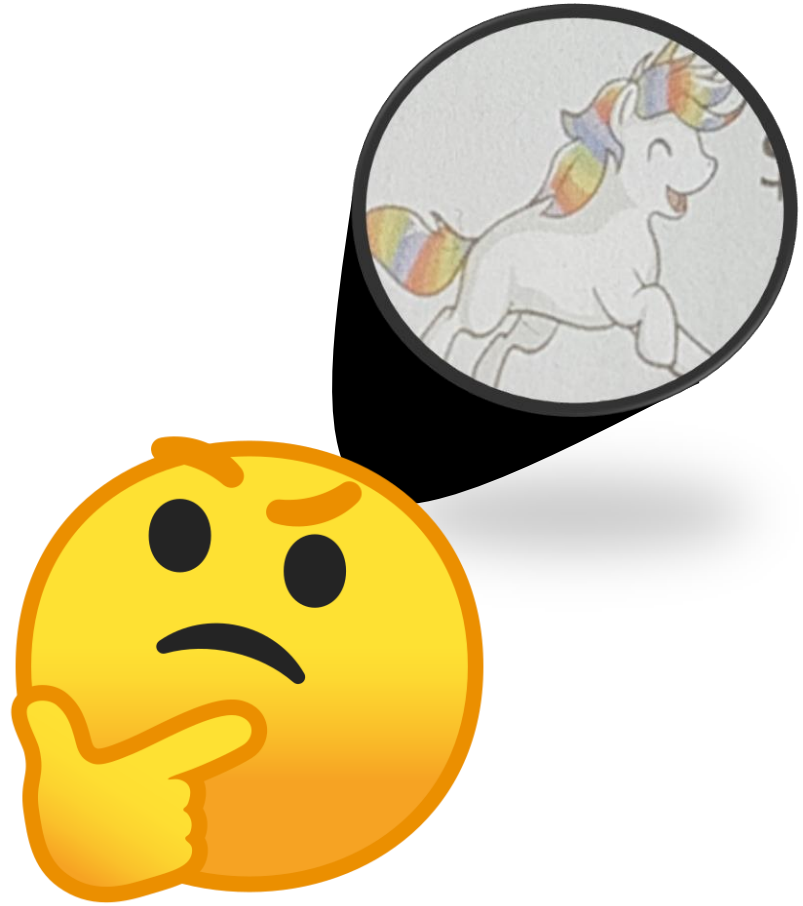


- Add complexity piecewise – we don‘t need scalability in the data prototype phase. We need correct results.
- Use this to speed-up your work and learn fast to deliver business value

relaxdays creative goods

What do we have achieved?

- We created an environment where data scientists can concentrate on the data science part of their work
- Software development related tasks are mostly automated
- We understood that our work consists of two important cycles (data science and software development cycle)
- We can use this insight to create data products with respect to the product stages (prototype – pilote – product)



It‘s not a unicorn, actually – but it looks like a horse

- We empored analytical guys to do analysis and automated the rest
- Now we can concetrate to develop the team with useful roles, improved processes, etc.

- The team now can start it‘s evolution and seek for the promised data science land, where
  - data magic happens and 1 + 1 = 3
  - you are able to build rapid protoypes and bring it to production in no time
  - you have time to learn new stuff and use it in your everyday work
  - all the boring stuff is automated, etc.

How to get there? This can be part of another talk…

Many thanks for your attention!

Sebastian Böhm,
Data Science Product Owner
Relaxdays GmbH
sebastian.boehm@relaxdays.de