

THOUGHTS ON DOING AGILE DATA SCIENCE

DATA SCIENCE NL MEETUP

Edwin Thoen
Data Scientist @ funda
edwinthoen@gmail.com
@edwin_thoen

A RECIPE FOR DISASTER

Ingredients:

- *1 statistics MsC (ideally with no software engineering skills)*
- *1 consulting firm*
- *1 customer site*
- *A bunch of project stakeholders*
- *1 ill-defined project*
- *A handful of undocumented databases*

A RECIPE FOR DISASTER

Preparations:

1. Use the consulting firm to sell the statistics MsC as an experienced data scientist.
2. Place the “data scientist” at the customer site.
3. Have the stakeholders convincingly explain the project and set a deadline in the far future.
4. Be sure the “data scientist” starts trying a great number of models and algorithms.
5. Let gradually simmer for a couple of months.
6. As soon as the project implodes you know it is done.

WHAT WENT WRONG?

Lack of Reproducibility

WHAT WENT WRONG?

Random Walk coding

WHAT WENT WRONG?

Miles away from the business
(actually saved me from myself)

WHOAMI (NOW)

Data Scientist @ funda

Love for interactive analysis (R's Tidyverse),
machine learning, workflow

AGILE DATA SCIENCE

The screenshot shows the Funda Waardecheck interface. At the top, there's a navigation bar with the Funda logo, a menu icon, account information, and a language switch to NL. Below the header, a breadcrumb navigation shows 'Home > Waardecheck'. The main content area has a large question 'Wat is de geschatte waarde van jouw woning?'. Below it, the title 'Waardecheck' is displayed. To the right, there's a form for entering address details. The form is divided into two sections: 'Adres' (marked with a circled '1') and 'Woninggegevens' (marked with a circled '2'). The 'Adres' section contains fields for 'Postcode' (with an empty input field), 'Huisnummer' (with an empty input field), and 'Toevoeging' (with a dropdown menu labeled 'Kies'). Below the address fields is a checkbox labeled 'Opslaan in account als mijn huis'. At the bottom of the form is a large grey button labeled 'Verdergaan'.

funda

☰ Menu

Account

NL

Home > Waardecheck

Wat is de geschatte waarde van jouw woning?

Waardecheck

1 Adres 2 Woninggegevens

Postcode

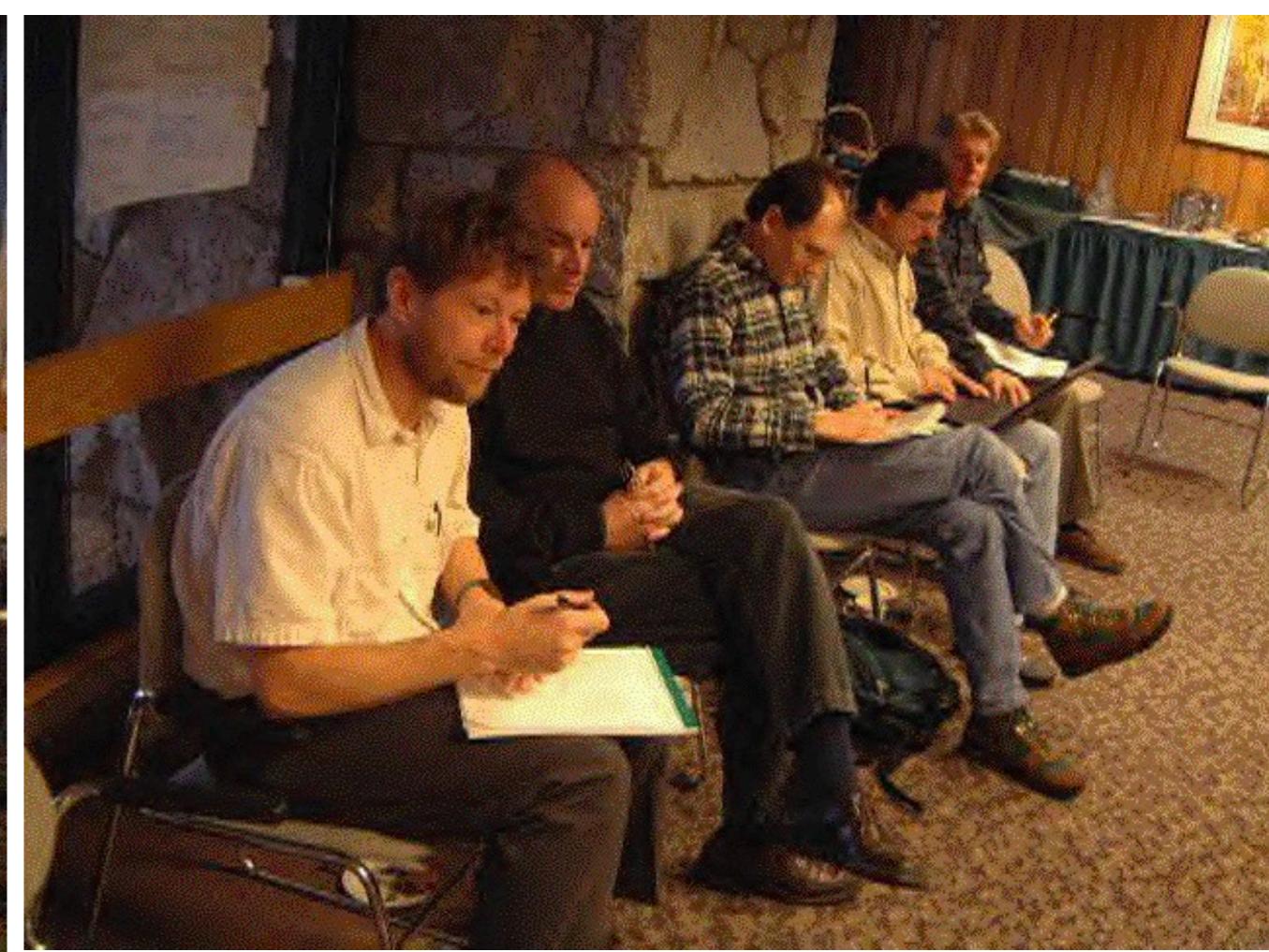
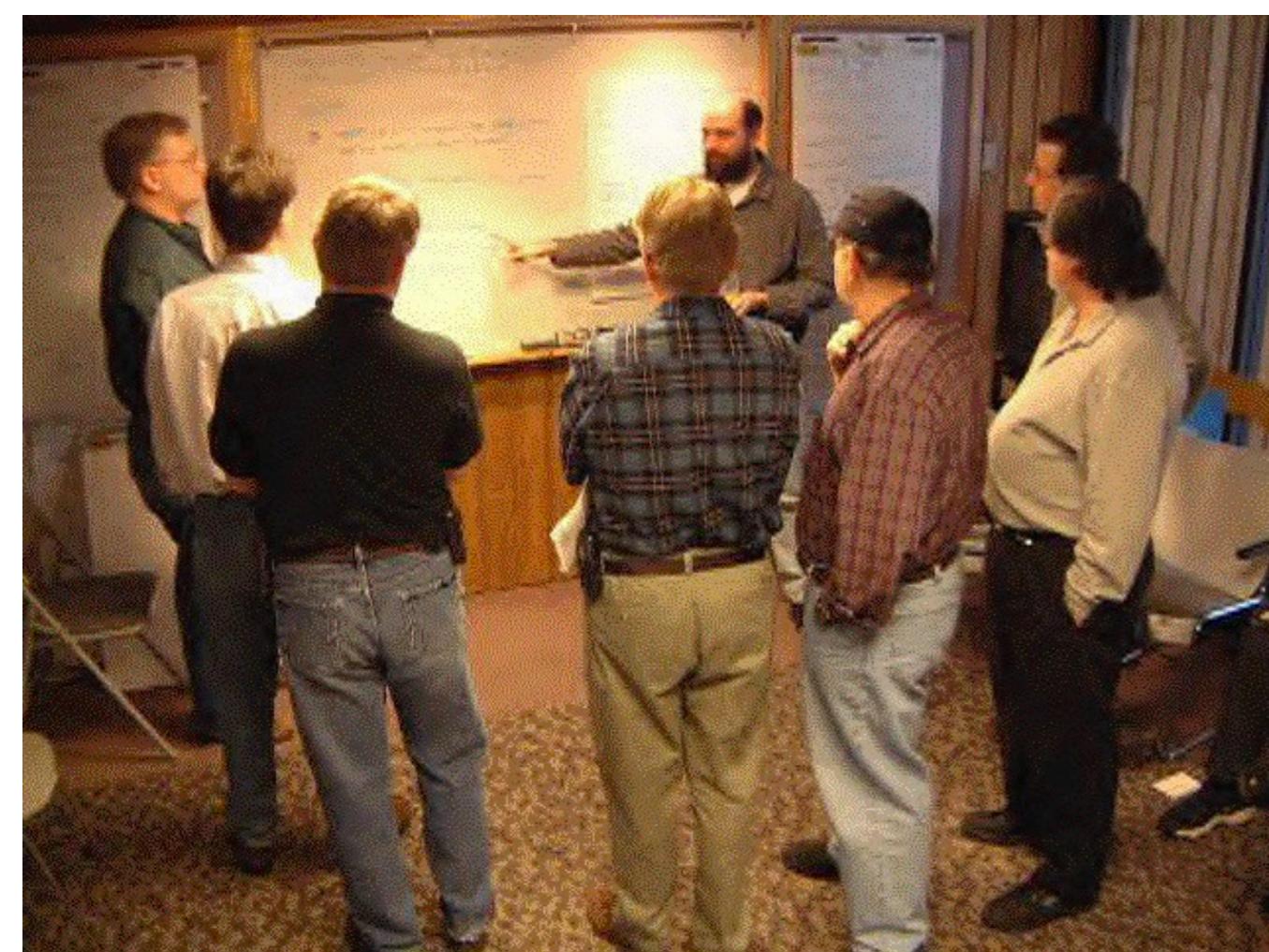
Huisnummer

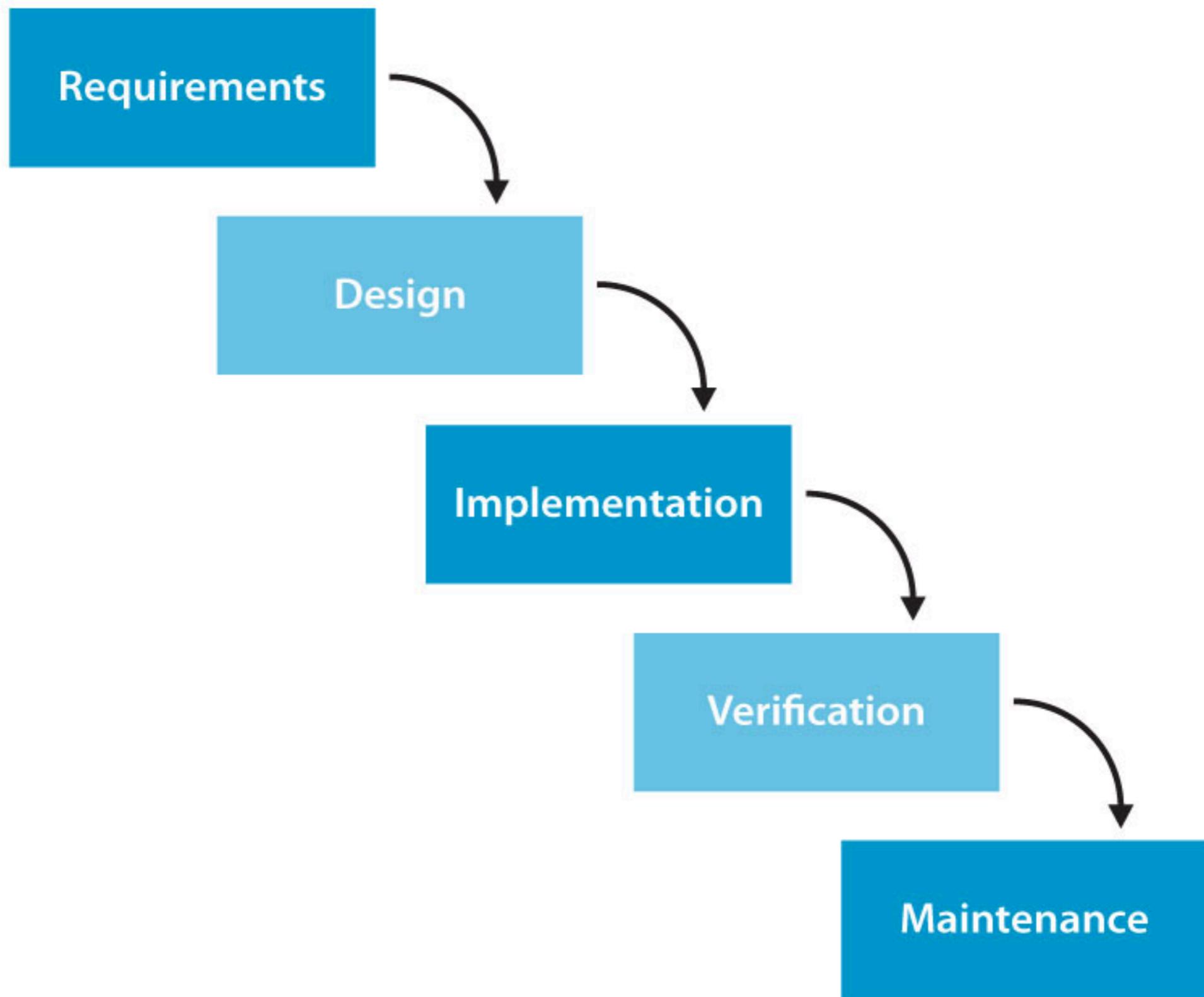
Toevoeging

Kies ▾

Opslaan in account als mijn huis

Verdergaan





We are uncovering better ways of developing software by doing it and helping others do it.

Through this work we have come to value:

Individuals and interactions over processes and tools

Working software over comprehensive documentation

Customer collaboration over contract negotiation

Responding to change over following a plan

That is, while there is value in the items on the right, we value the items on the left more.

We follow these principles:

Our highest priority is to satisfy the customer through **early and continuous delivery** of valuable software.

Welcome changing requirements, even late in development. Agile processes **harness change** for the customer's competitive advantage.

Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.

Business people and developers must **work together daily** throughout the project.

Build projects **around motivated individuals**. Give them the environment and support they need, and trust them to get the job done.

The most efficient and effective method of conveying information to and within a development team is **face-to-face conversation**.

Working software is the primary measure of progress.

Agile processes promote **sustainable development**. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.

Continuous attention to **technical excellence** and good design enhances agility.

Simplicity--the art of maximizing the amount of work not done--is essential.

The best architectures, requirements, and designs emerge from **self-organizing teams**.

At regular intervals, the team **reflects** on how to become more effective, then tunes and adjusts its behavior accordingly.



AI

SPONSORED

Why do 87% of data science projects never make it into production?

VB STAFF JULY 19, 2019 4:10 AM



Microsoft Azure

Bouw en ontwikkel apps met Azure. Gratis tot je iets anders wilt.

[Probeer Azure gratis uit >](#)



MOST READ



We follow these principles:

Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.

Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.

Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.

Business people and developers must work together daily throughout the project.

Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.

The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.

Working software is the primary measure of progress.

Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.

Continuous attention to technical excellence and good design enhances agility.

Simplicity--the art of maximizing the amount of work not done--is essential.

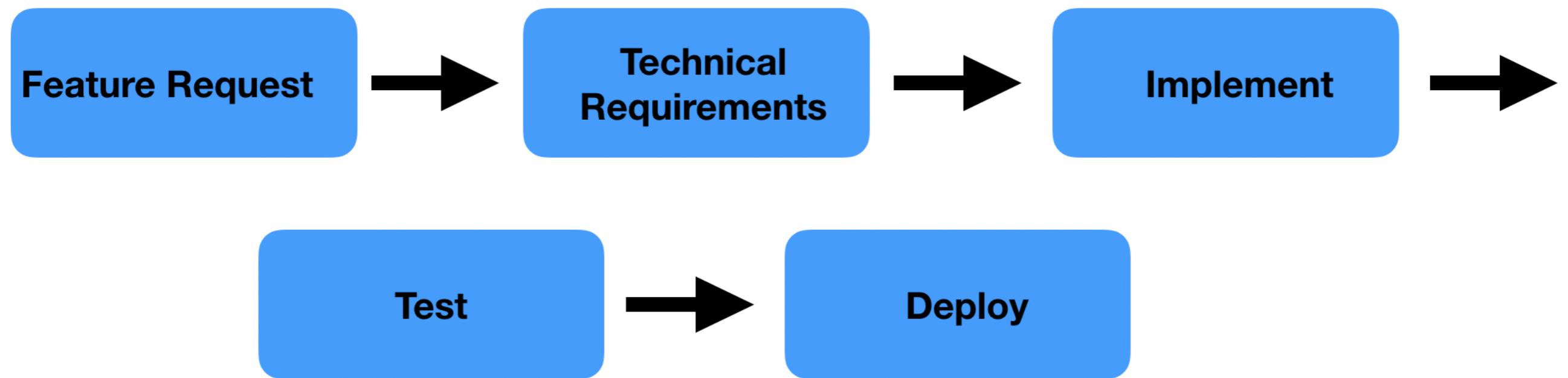
The best architectures, requirements, and designs emerge from self-organizing teams.

At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.

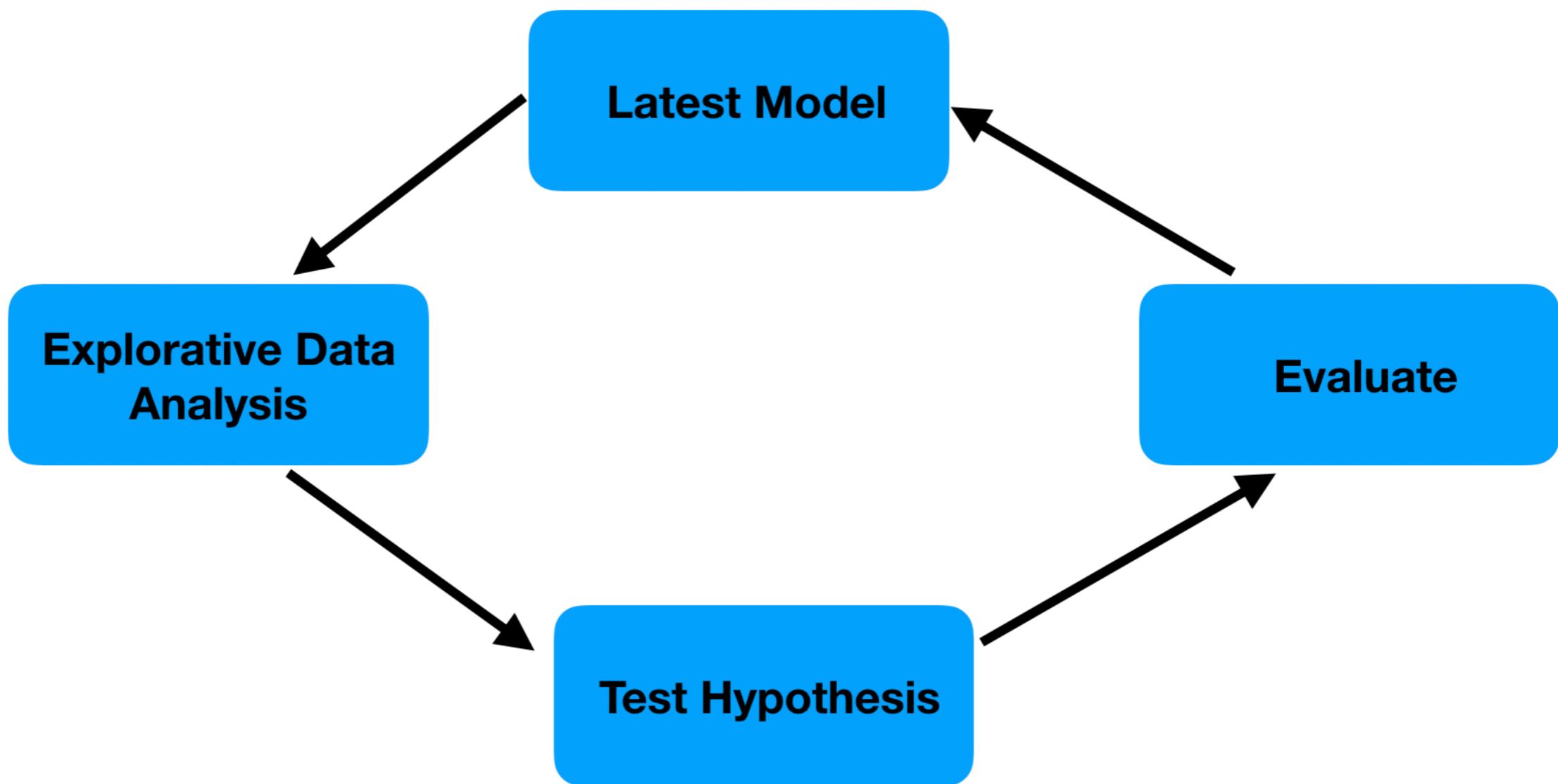
AGILE WORKFLOW

Agile workflows are always means to an end

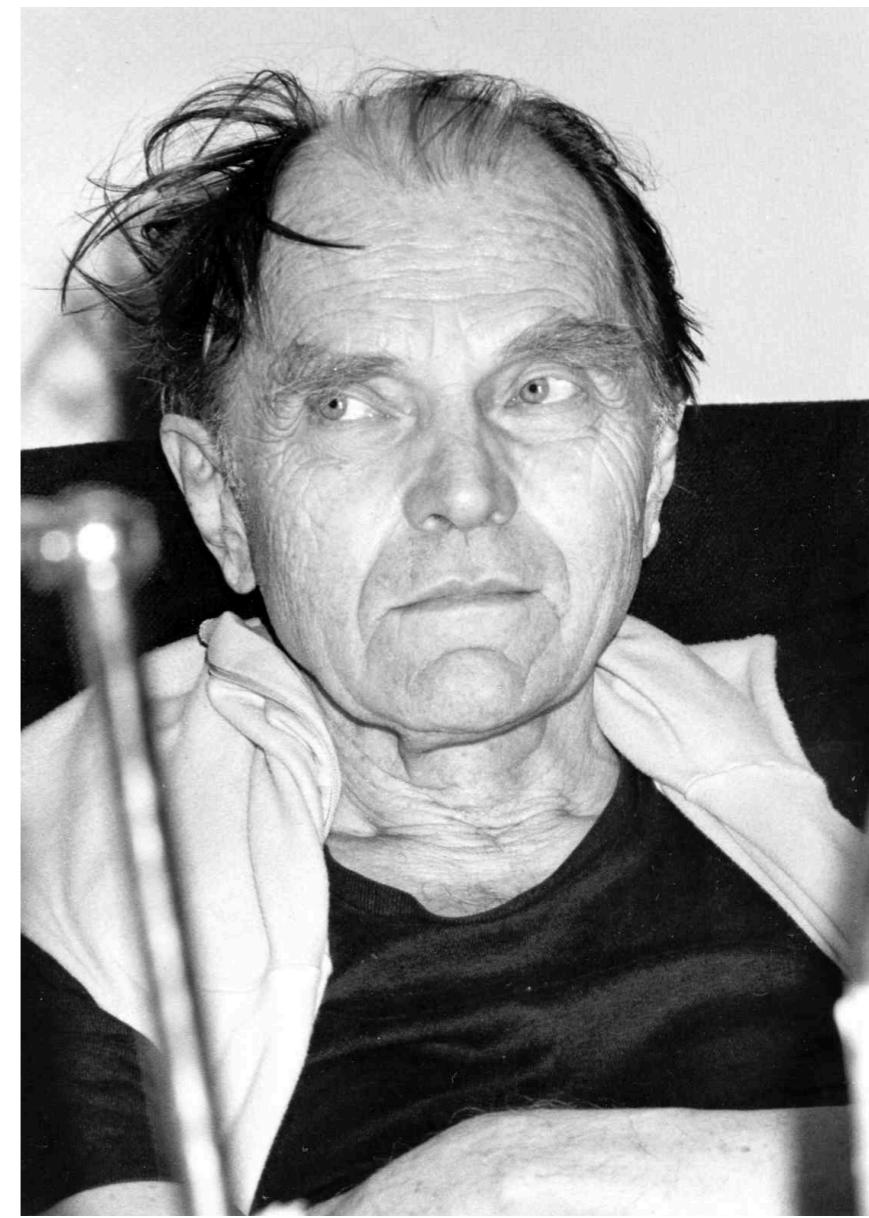
LINEAR TASK



CIRCULAR TASKS



A TWO-WORLD WORKFLOW

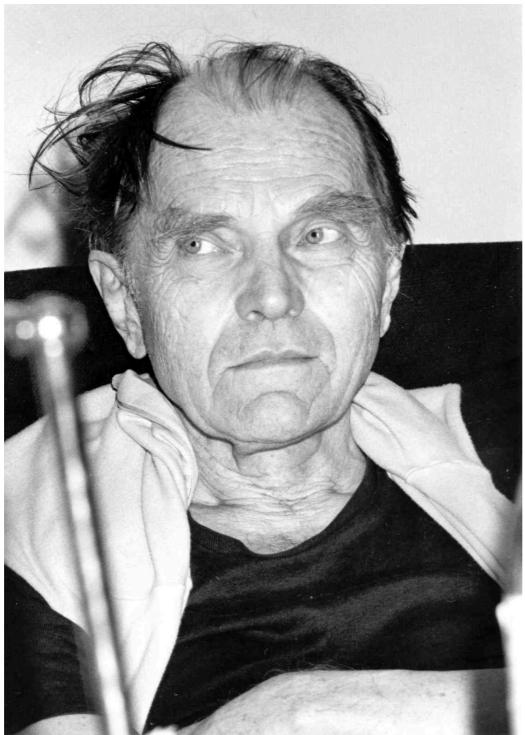


A TWO-WORLD WORKFLOW



Unit testing
Versioning
Build a software package
Completely automated pipeline
Reproducible results
Always production ready

A TWO-WORLD WORKFLOW



Quickly test hypotheses
Don't touch product code
No direct reproducibility

When product can be improved
move back to Marie Kondo world.

WHY THIS IS AGILE

Your product is always ready

Quick to update

Each change added to production right away

KANBAN

IMHO Kanban is more suitable than Scrum for data science

Tasks are in one of the two worlds

THANK YOU

edwinthoen@gmail.com

@edwin_thoen

github.com/EdwinT

For the text *Agile Data Science with R*:

edwinth.github.io/ADSwR