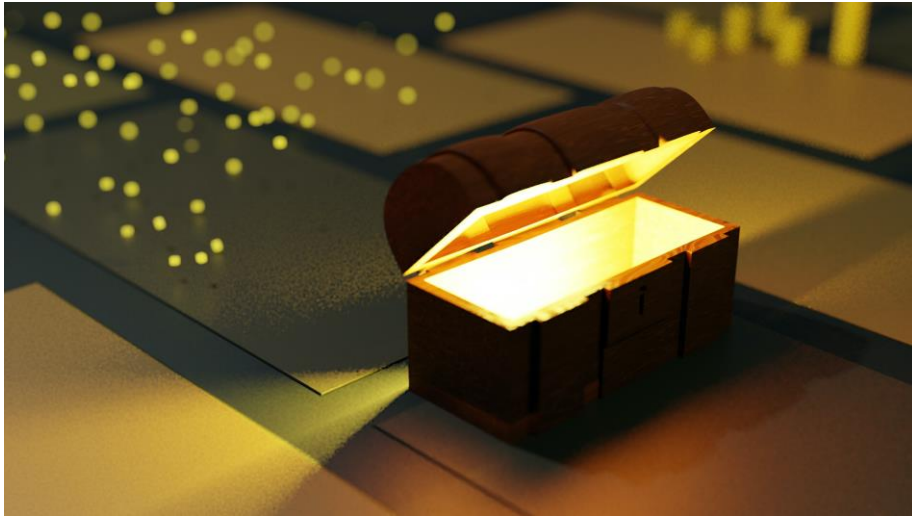


Exploring big data effectively in safety to identify the (human) factors that matter most and under which circumstances

Dr. Julia Burggraaf

20 september, ERA Safety days 2023





Presentation 1 (now):
An introduction to using big data for the 'behavioural' side of safety



Presentation 2 (video):
How to prevent big data safety projects from leading to disappointment

Psychologist



who can write
code





Potential big data treasures for safety

Big data provides the ability to...



Identify **previously unknown factors** that influence behaviour



Obtain **convincing evidence** to support already existing 'hunches'



Prioritisation: Identify which factors have the **largest influence** (effect size)



Identify **under which circumstances** these factors have a large influence and **under which circumstances they don't**



Proactively monitor the effect of innovations (or other changes)

So you can ...




Fix **existing 'dangerous situations'**



Prevent the unintentional introduction of **'dangerous situations' during future (infrastructure) changes**



Go from **'generic' safety measures to 'specific' measures**

'one size fits all' tends to be too big for  some and too small for others



Support innovation in a safe way and with less concerns among all parties



Intervene **before an incident occurs**

What do disappointing big data projects in safety look like?

The results do not lead to any useably insight for safety (example result: 'we cannot say for sure')

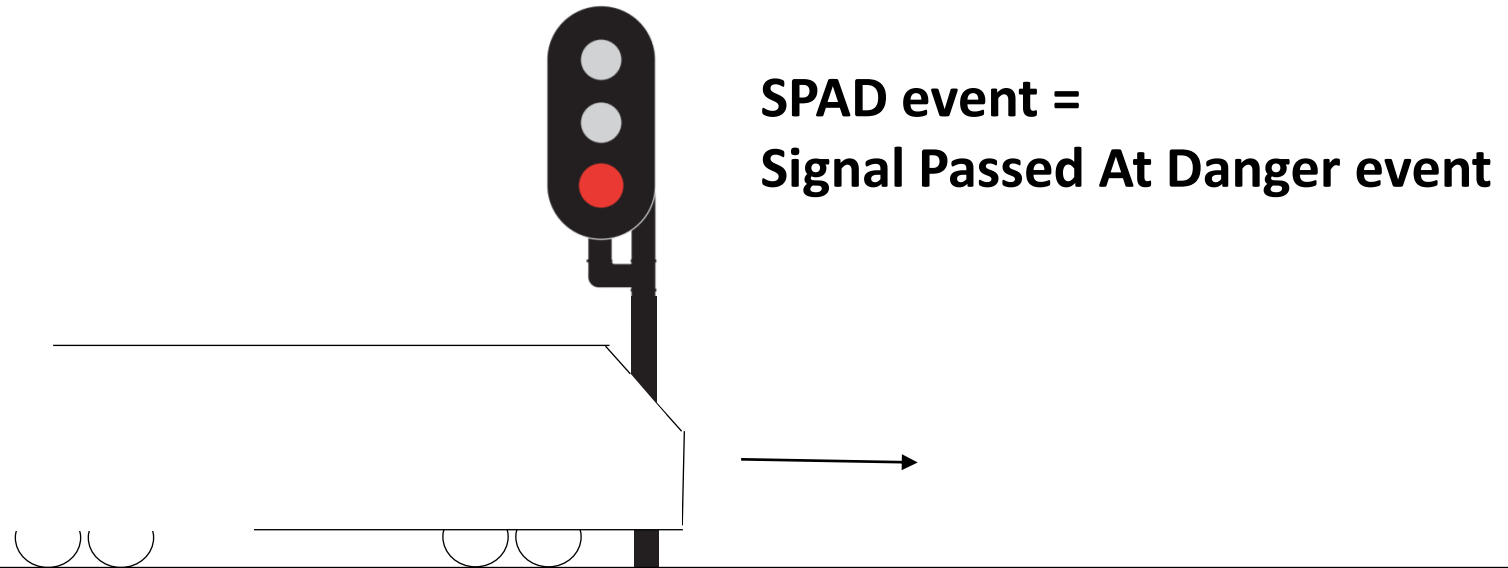
Monitoring project

- Unclear when to intervene whilst monitoring fase has already started.
Discussions among experts whether safety levels are decreasing.
- No firm conclusions at end of monitoring period

Research project

- Not possible to firmly conclude which factors influence the behaviour or risk
- Only very basic factors identified which are already common knowledge and widely accepted; no new knowledge on how much the factors actually contribute due to noise and/or high unaccuracy and/or other factors insufficiently taken into account
- A predictive model with low accuracy for a safety issue that requires high precision and certainty

Examples of successful projects on SPAD probability



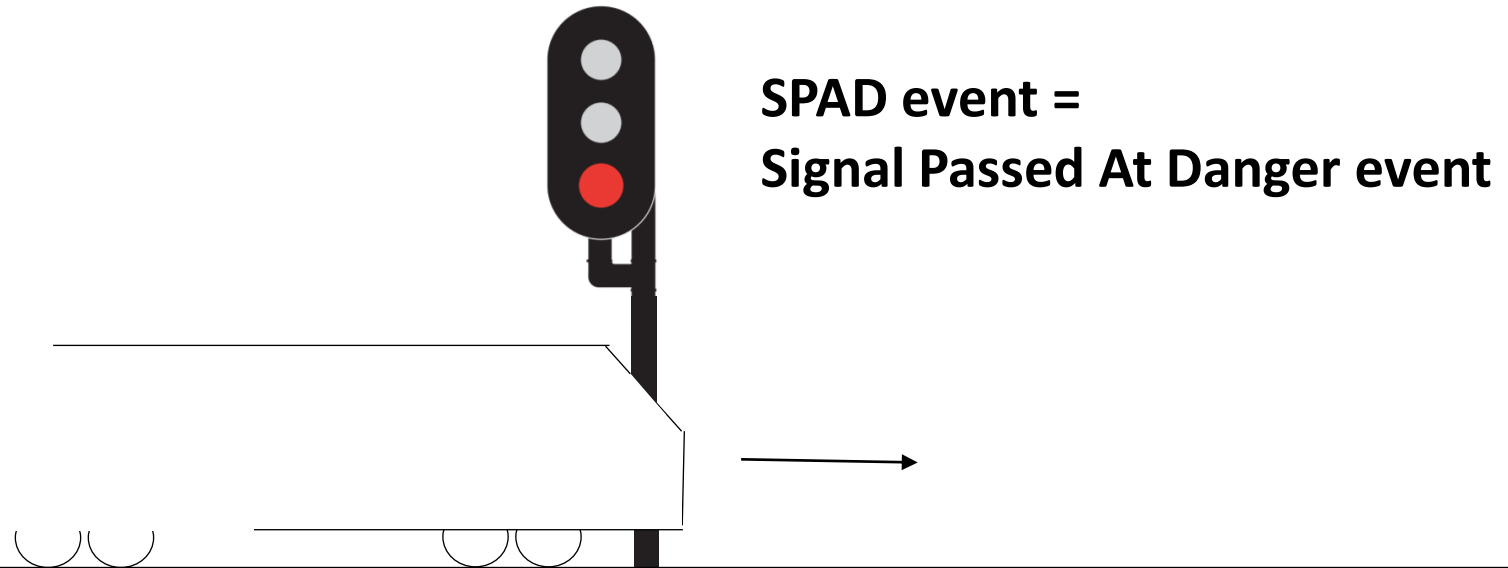
1. Monitoring project

- Does 'infrastructure change x' increase the probability of a SPAD occurring?

2. Research project

- Which human factors increase the probability of a SPAD occurring?

Examples of successful projects on SPAD probability



1. Monitoring project

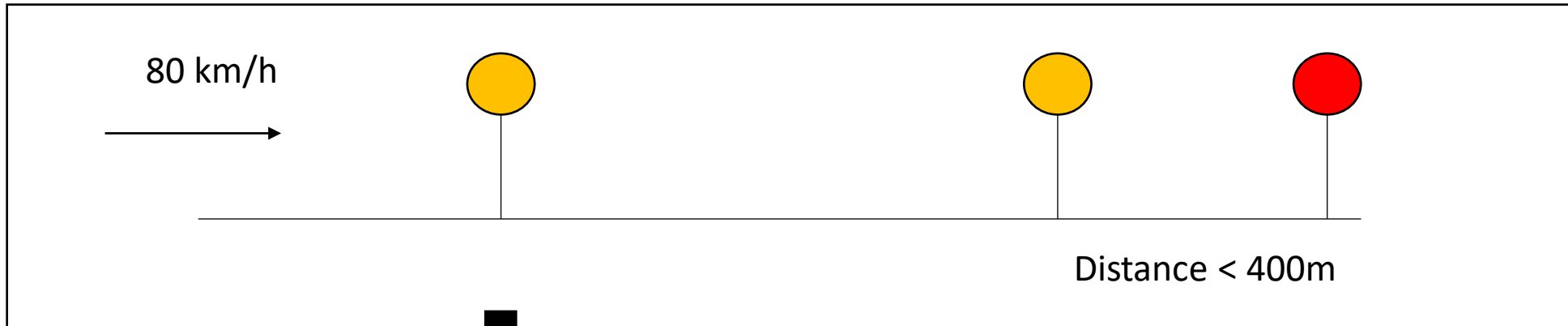
- Does 'infrastructure change x' increase the probability of a SPAD occurring?

2. Research project

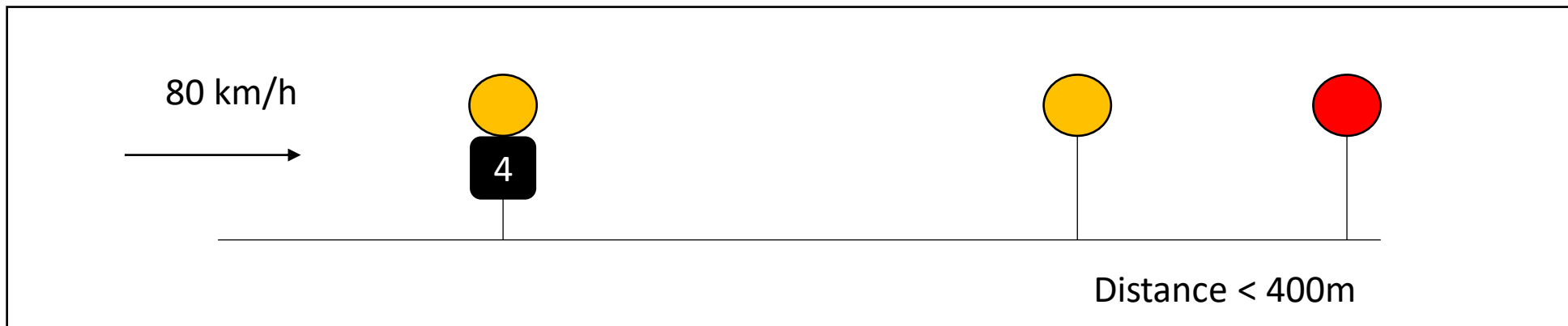
- Which human factors increase the probability of a SPAD occurring?

The to-be-monitored infrastructure change

Old situation:



New situation:



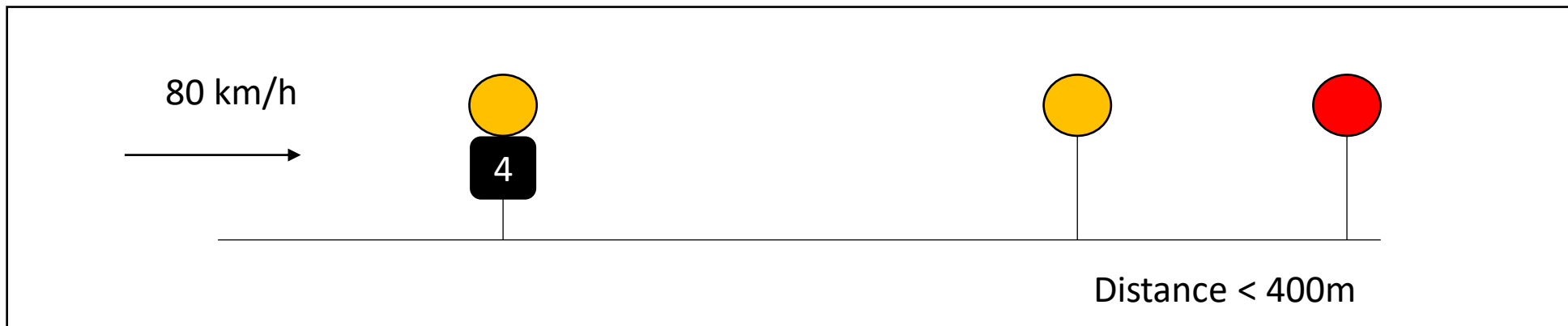
**Higher
probability
of SPADs?**

The to-be-monitored infrastructure change

Old situation:



New situation:



Higher probability of SPADs?

Monitoring measures

Number of incidents

- Number of SPADs (per red aspect approach)

Proactive data

- Percentage of near misses via measure mDtSPAD*
- Changes in train driver behaviour via measure mDtSPAD*

Subjective data

- Questionnaires to train drivers

*mDtSPAD = maximum Deceleration to SPAD, or the maximum required deceleration to to stop in front of the aspect (per approach)

Monitoring measures

Number of incidents

- Number of SPADs (per red aspect approach)

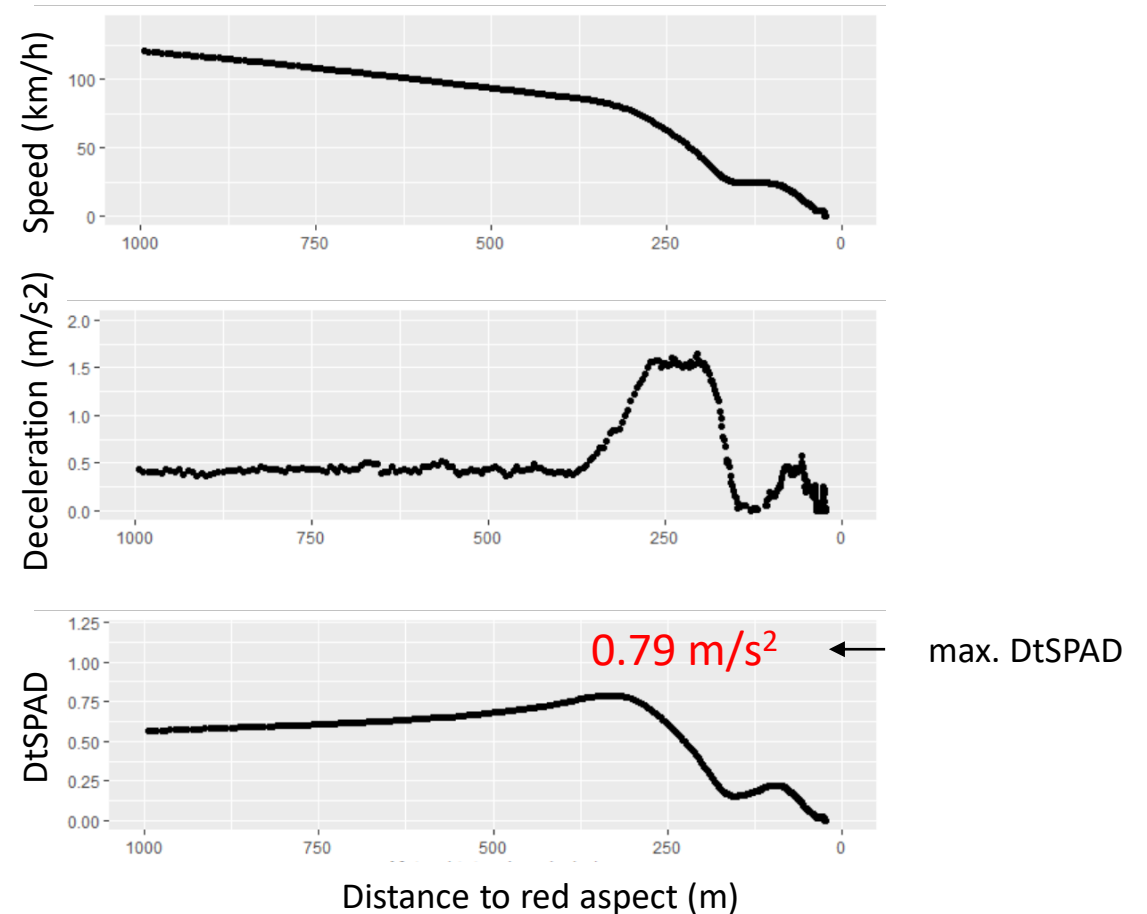
Proactive data

- Percentage of near misses via measure mDtSPAD
- Changes in train driver behaviour via measure mDtSPAD

Subjective data

- Questionnaires to train drivers

One mDtSPAD value per red aspect approach



*mDtSPAD = maximum Deceleration to SPAD, or the maximum required deceleration to to stop in front of the aspect (per approach)

Monitoring measures

Number of incidents

- Number of SPADs (per red aspect approach)

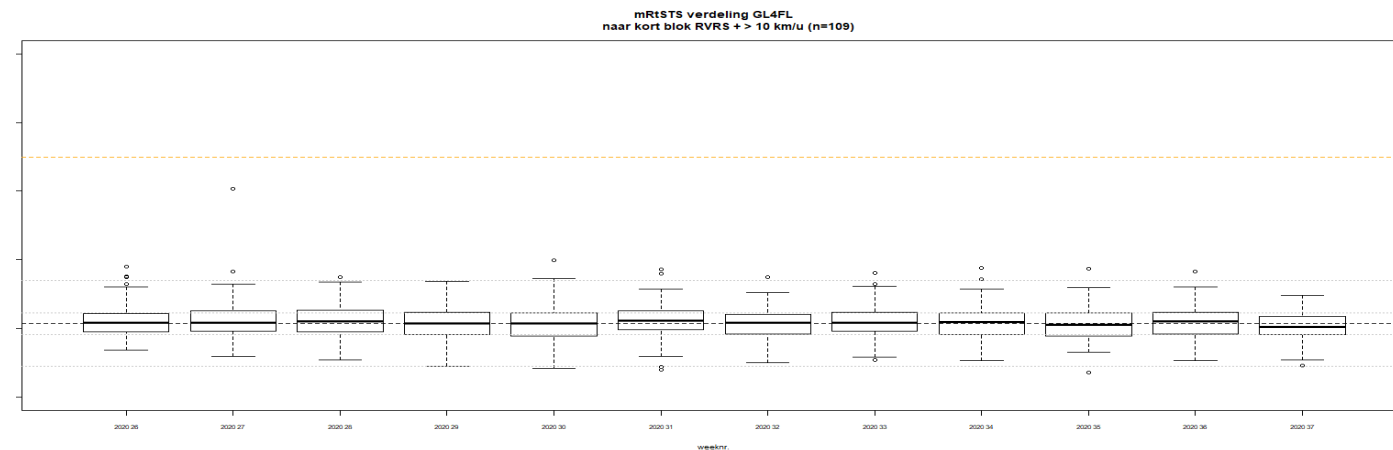
Proactive data

- Percentage of near misses via measure mDtSPAD
- Changes in train driver behaviour via measure mDtSPAD

Subjective data

- Questionnaires to train drivers

During monitoring: Visualisations with data per week + 'near miss' threshold



*mDtSPAD = maximum Deceleration to SPAD, or the maximum required deceleration to to stop in front of the aspect (per approach)

Monitoring measures

Number of incidents

- Number of SPADs (per red aspect approach)

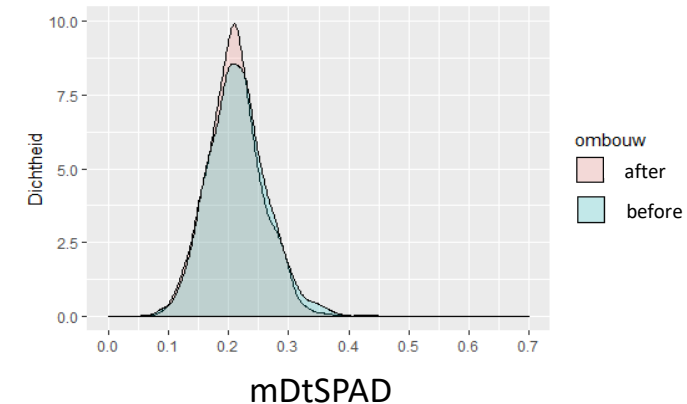
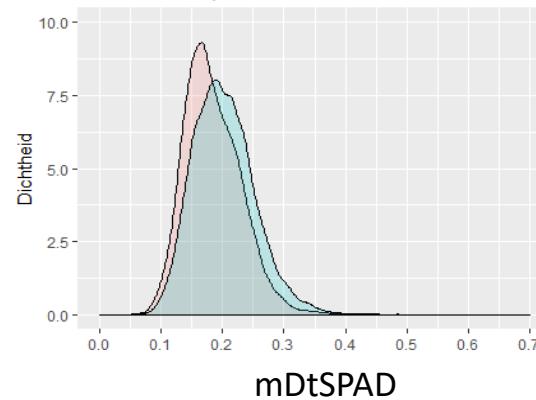
Proactive data

- Percentage of near misses via measure mDtSPAD
- Changes in train driver behaviour via measure mDtSPAD

Subjective data

- Questionnaires to train drivers

Final analysis after monitoring period



*mDtSPAD = maximum Deceleration to SPAD, or the maximum required deceleration to to stop in front of the aspect (per approach)

Successful aspects of the project

Monitoring phase



Clear rules for what 'dangerous' approaches are and when intervention was necessary

Versus



"We'll know what to do when see it"

Final analysis



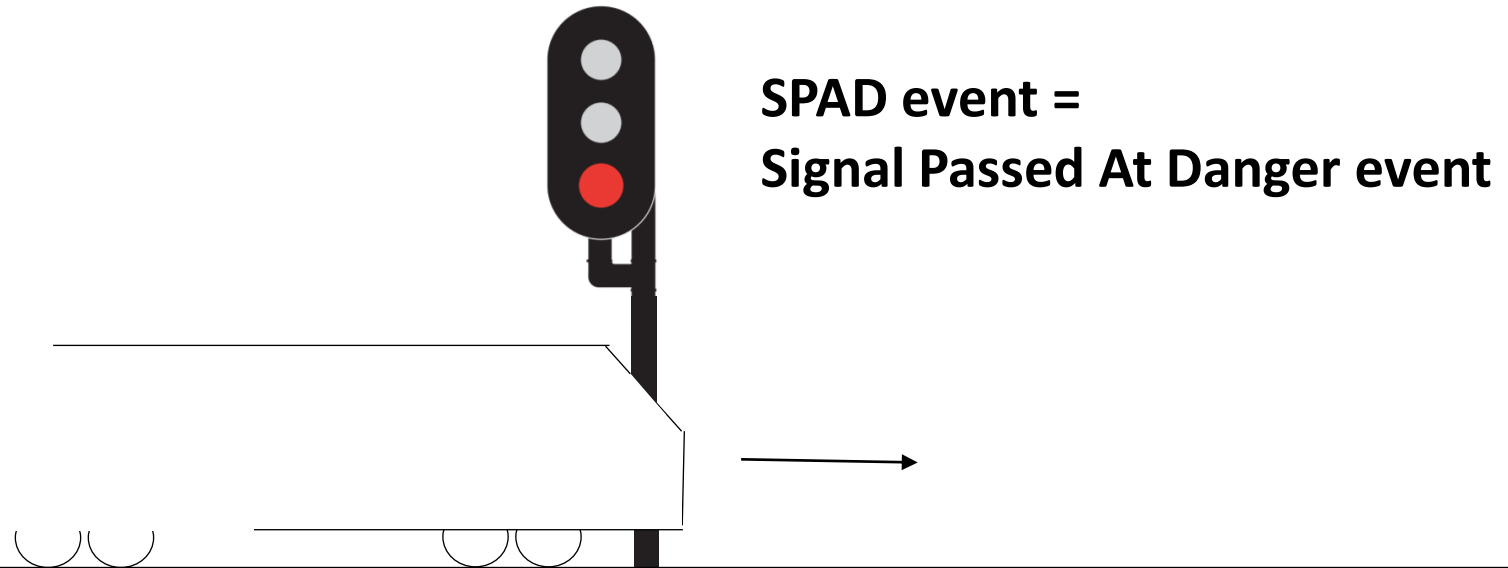
Firm conclusion was possible for passenger trains: "the safety level does not decrease"

Versus



"There were no incidents (SPADs) but unclear whether this is simply because there are always very few SPADs"

Examples of successful projects on SPAD probability



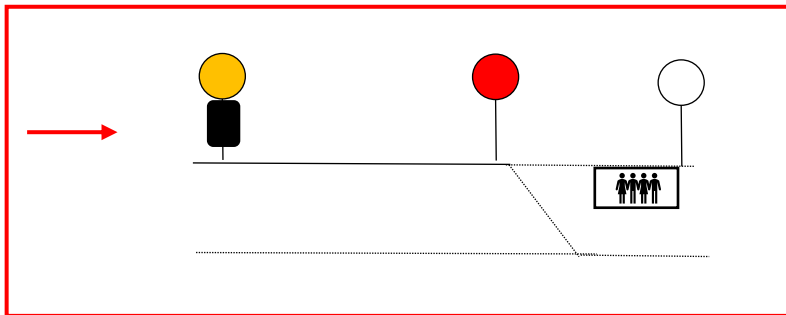
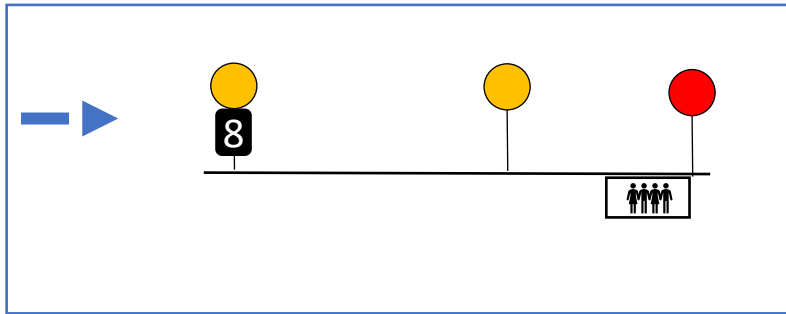
1. Monitoring project

- Does 'infrastructure change x' increase the probability of a SPAD occurring?

2. Research project

- Which human factors increase the probability of a SPAD occurring?

Does previous exposure to different signal aspect at same location, increase SPAD probability?



Data:

- First analysed with train behaviour data
- then analysed with 6 years of incident data and red aspect approaches

Successful aspects of the project



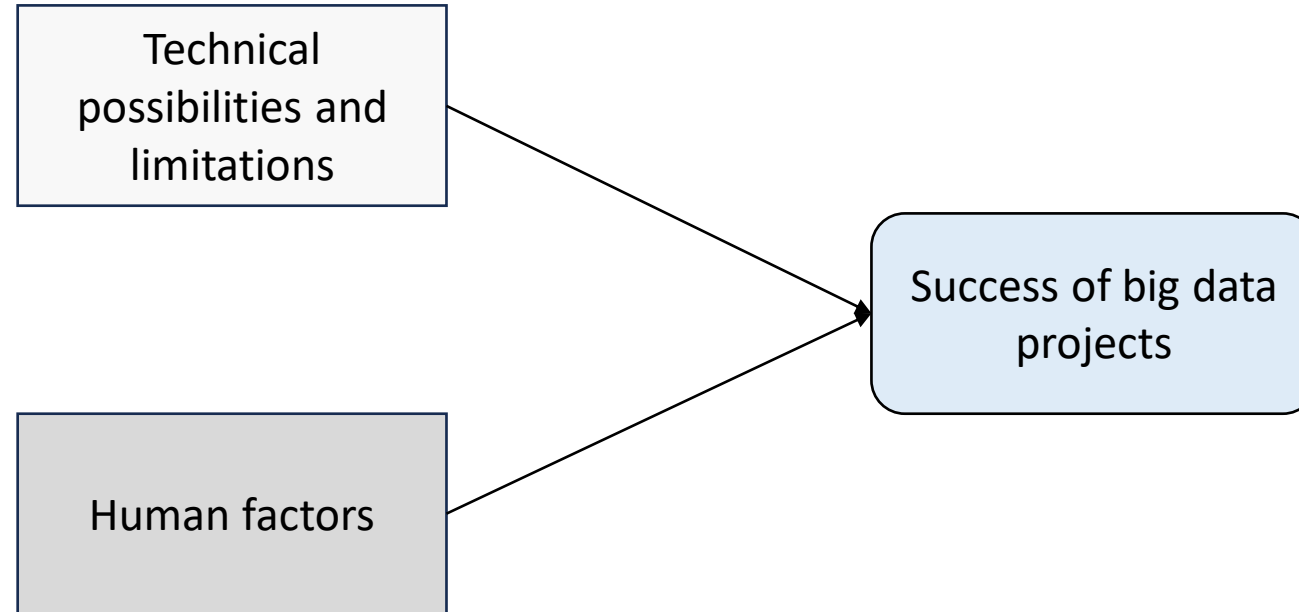
- Firm evidence that effect is present
- Indication of effect size (large)
- Identified under which circumstances the effect occurs (very specific combination of factors)
- and under which circumstances the effect does not occur (e.g. at lower track speeds)

Versus

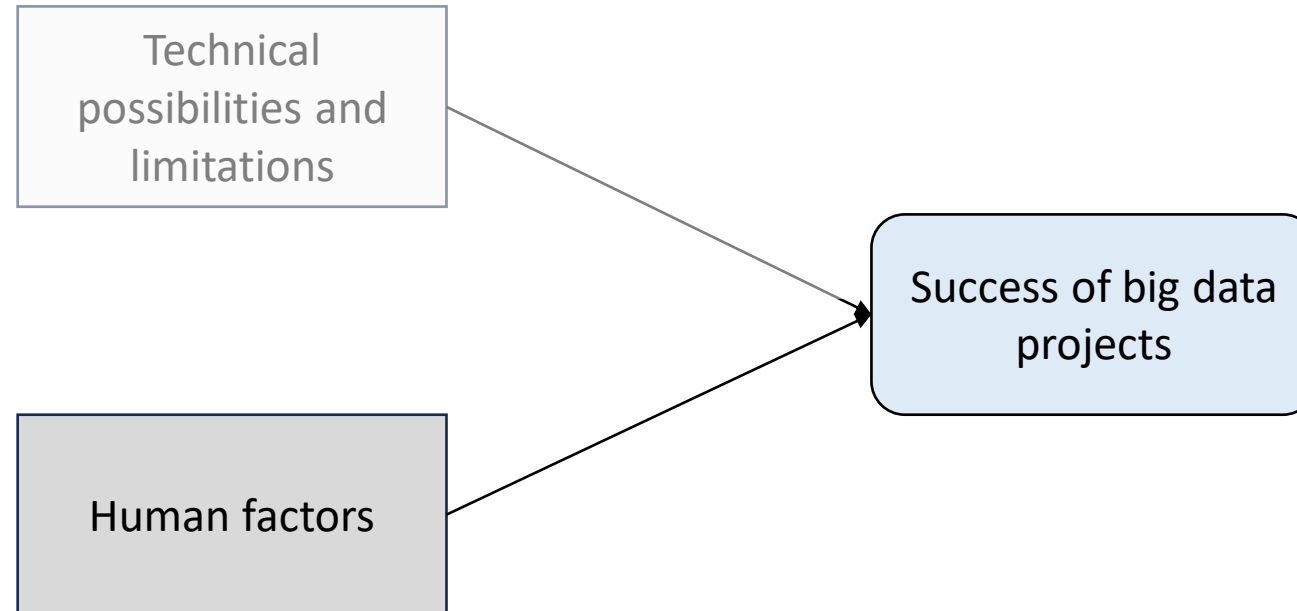


- “There seems to be an effect but we are not sure”
- “There is an effect but unclear if it is a large effect compared to other factors”
- “O nooo, there is an effect. We can never use these kind of signal aspects again!!”

Big data projects do not automatically succeed

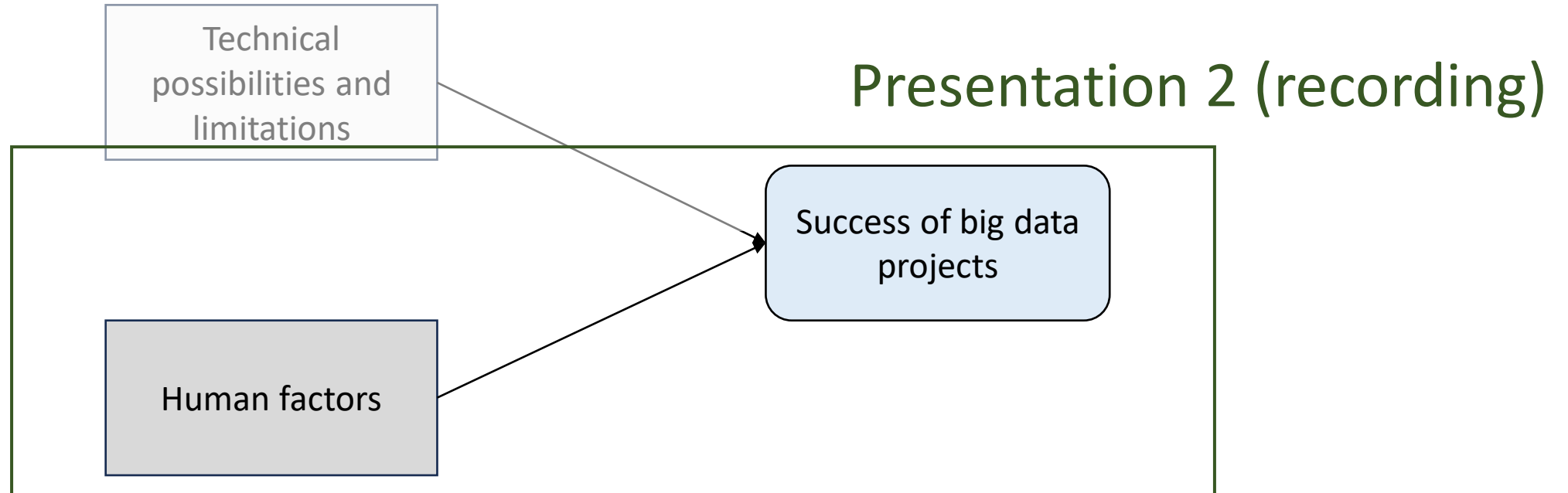


Big data projects do not automatically succeed



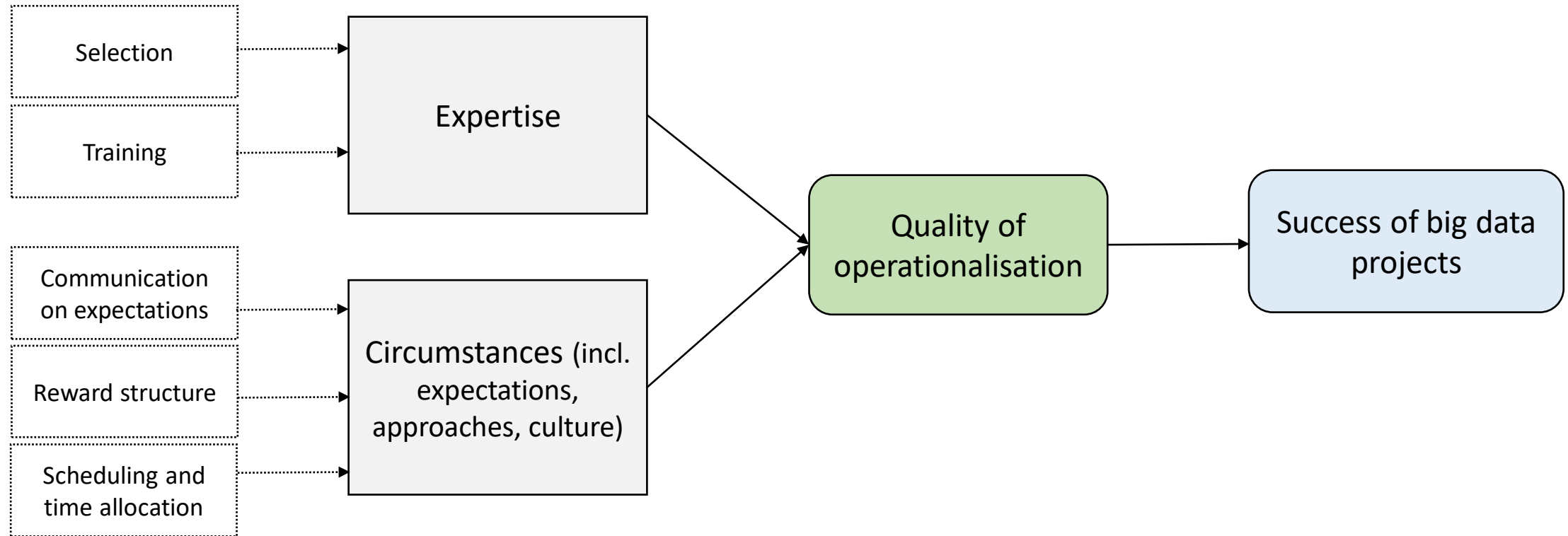
The persons involved in big data projects are humans that are influenced by 'human factors'. The field of Human Factors is thus also relevant to reduce the probability of errors and increase the probability of optimal performance in big data projects for safety.

Big data projects do not automatically succeed



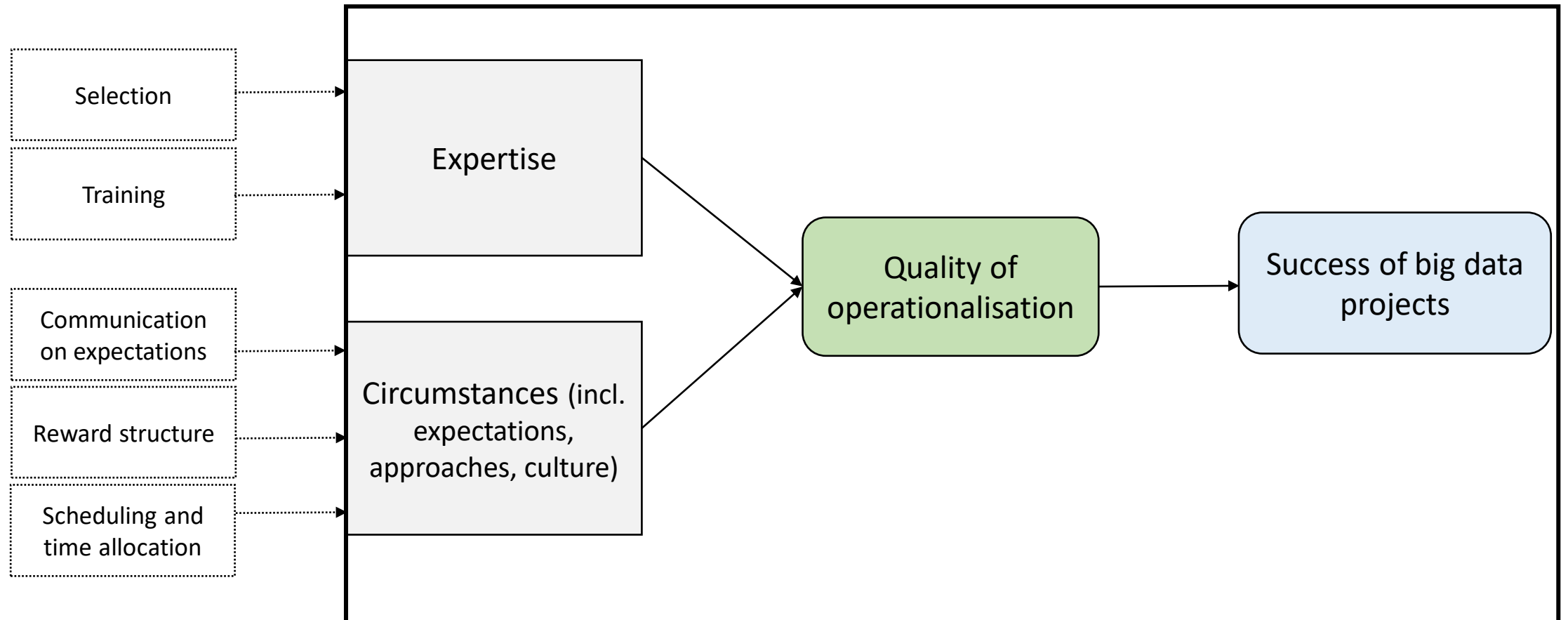
The persons involved in big data projects are humans that are influenced by ‘human factors’. The field of Human Factors is thus also relevant to reduce the probability of errors and increase the probability of optimal performance in big data projects for safety.

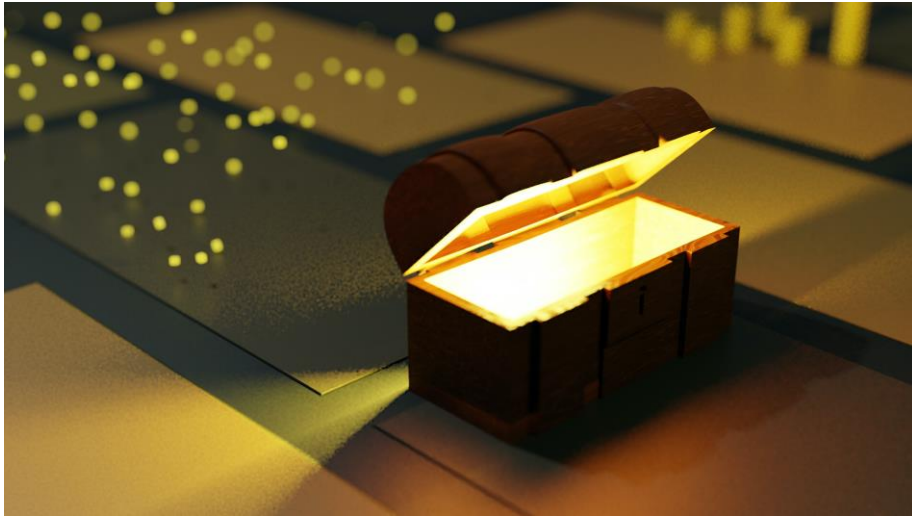
Correct operationalisation is critical for the success of a big data project in safety



Correct operationalisation is critical for the success of a big data project in safety

Will be clarified via case study from ProRail





Step 1.

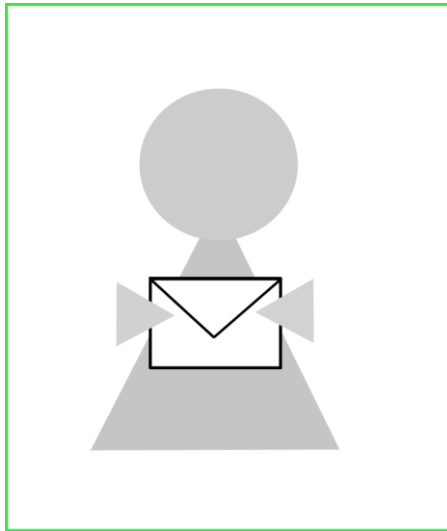
Get the enthusiasm going to use big data for safety and see how it can simultaneously support safety and other KPI's

Step 2.

Creating a map to actually use big data effectively for safety and avoid disappointment



Contact info



Dr. Julia Burggraaf 

julia@juliaburggraaf.com

juliaburggraaf.com/DataForSafety

Presentation 2 on:

juliaburggraaf.com/ERA2023

Papers and thesis:

juliaburggraaf.com/DataForSafety

