



# To check or not to check

## Challenges in the scoping of a data quality project

---

**Actuariaatcongres 2024**

Inge Lodder AAG, MSc – Founder of ELAVV

# Index



Introduction – Profile & Company



Data Quality – Our experience



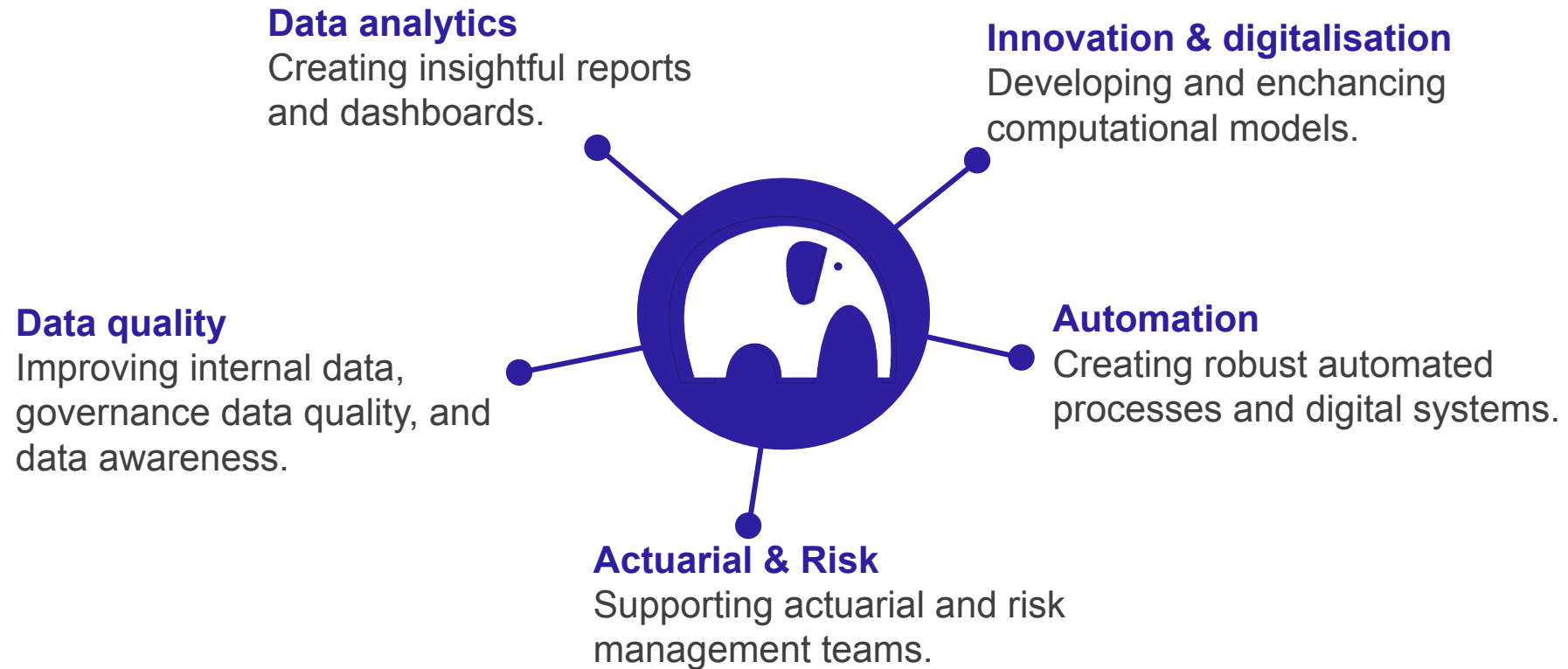
What would you do?

# Profile & company



# Who are we?

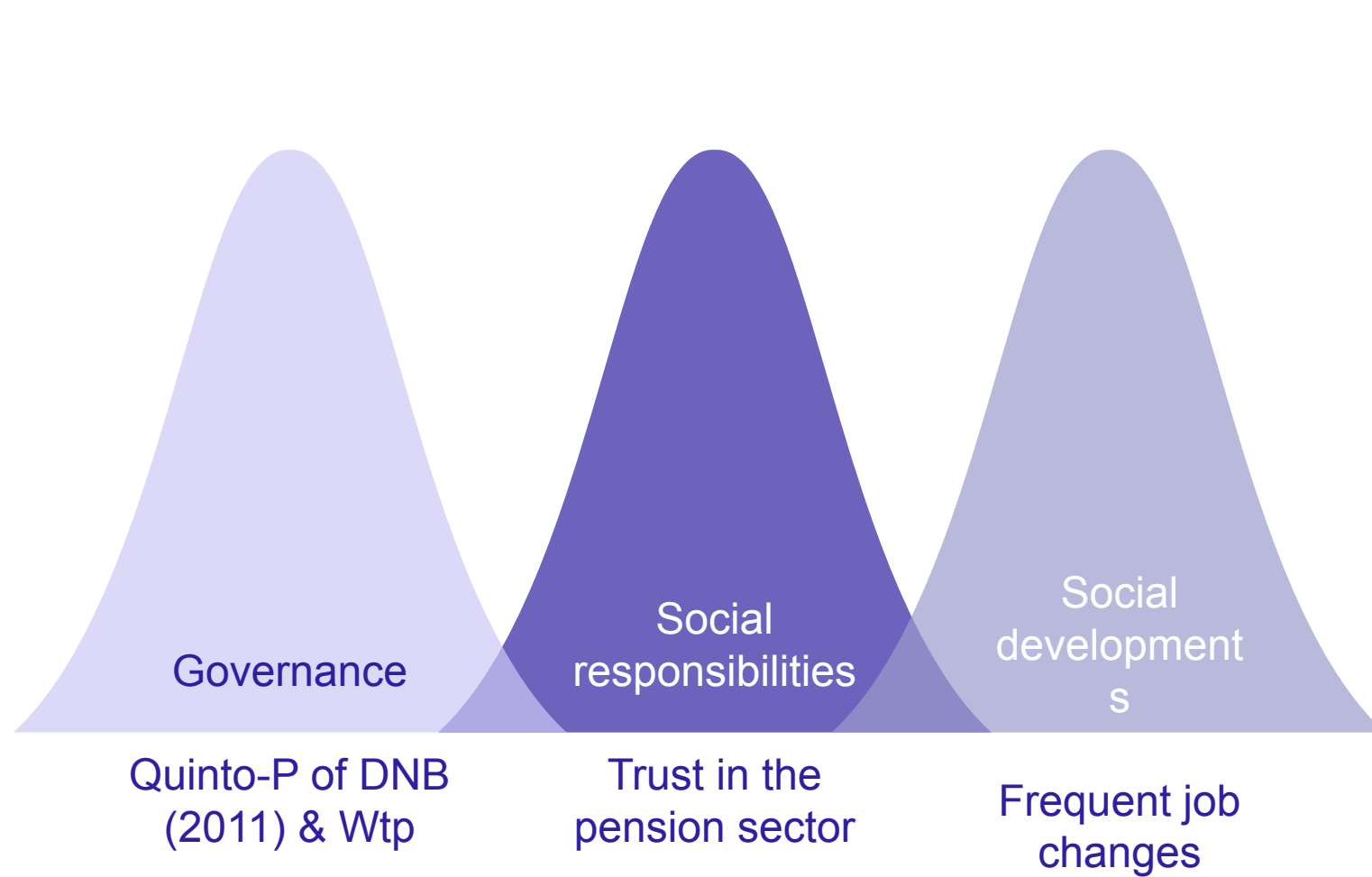
**We stand for innovation, digitalisation & advising for complex financial systems**



# **Data Quality - Our experience**

---

# Why do we need data quality checks?



# Lessons learned in data quality investigations



- **Data profiling:** completeness, correctness, accuracy, data distribution
  - **Generic & Specific analyses**
-

# Lessons learned in data quality investigations



- **Data profiling:** completeness, correctness, accuracy, data distribution
- **Generic & Specific analyses**

- 
- **Verifying all benefits** based on the mutation data

+ Additional checks



# Lessons learned in data quality investigations



- **Data profiling:** completeness, correctness, accuracy, data distribution
- **Generic & Specific analyses**

- 
- **Verifying all benefits** based on the mutation data

+ Additional checks

---

2 pension funds  
*medium-small sized,  
Self-administrating*

- **Predefined KDE** (critical data elements)
- Database analyses
- Time horizons benefits verification +/- 4 years

# Top 6 Challenges



How to obtain a reasonable time & effort estimation of a project?



Do you immediately “get clean”?



When do you perform a deep dive?



How do you determine a margin of error?



What are specific analyses and what are generic analyses?



Is an explained finding a solved finding?

# Top 6 Challenges



How to obtain a reasonable time & effort estimation of a project?



Do you immediately “get clean”?



When do you perform a deep dive?



How do you determine a margin of error?



What are specific analyses and what are generic analyses?



Is an explained finding a solved finding?

SCOPING

# **Data Quality - What would you do?**

---

# Case study: One pension fund, limited resources

*Which task has your priority?*



## Data formatting

Is all the data formatted properly? E.g. have dates the correct notation?



## Data distribution

Is the distribution of the data in line with your expectations?



## Data correctness KDE

Is the data correct based on the defined KDE's? These are simple, crucial checks.



## Benefits correctness

Do the benefits align with your expectations going back for 4 years?



## Data completeness

Do you find all the data you expect to find?

# Case study: One pension fund, limited resources

*My ranking*



1. Data completeness



2. Data formatting



3. Benefits correctness



4. Data correctness KDE



5. Data distribution

# Case study: One pension fund, limited resources

*Which observation has your priority?*



## Missing participant data

You have found missing addresses & partner names. You will look for more missing data and try to recover the data.



## Differences in pension base (PG)

You will look more closely into the underlying explanation of these differences.



## No data deferred orphan pension

You will look more closely into the underlying explanation of this missing data.



## No detailed analyses on disabilities

You will look more closely at the disabled participants and their data.



## Outliers in accrued old age pension

You will verify if these are valid outliers or not.

# Case study: One pension fund, limited resources

*My ranking*



1. No detailed analyses of disabled



2. Differences in pension base (PG)



3. Outliers in accrued old age pension



4. No data latent orphan pension



5. Missing participant data



# One question: What do you think?

## How extensive should data quality checks be?

Is it Ethical to make cuts in data quality checks?

# One question: What do you think?

## How extensive should data quality checks be?

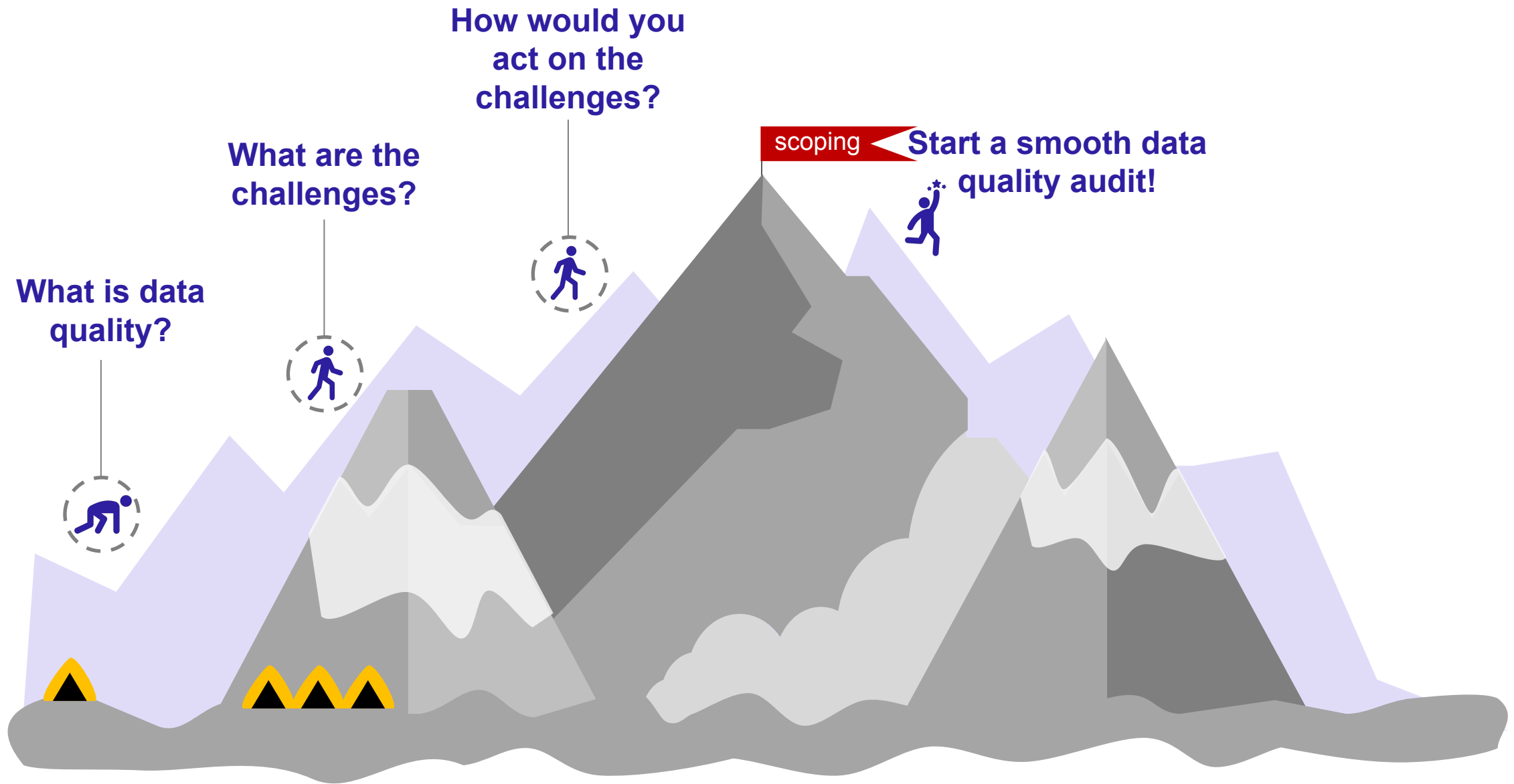
Is it Ethical to make cuts in data quality checks?

Do we need more than the basics?

Are the basics more than sufficient?

# Wrap up

---





Contact info!



I hope you leave the room with an inspiring impression  
&  
With the knowledge of the challenges you may face in Data Quality  
projects