



# *Let's Compare Them All – TFL Comparisons*



**PHUSE EU Connect**

07.11.2023 / Katja Glaß





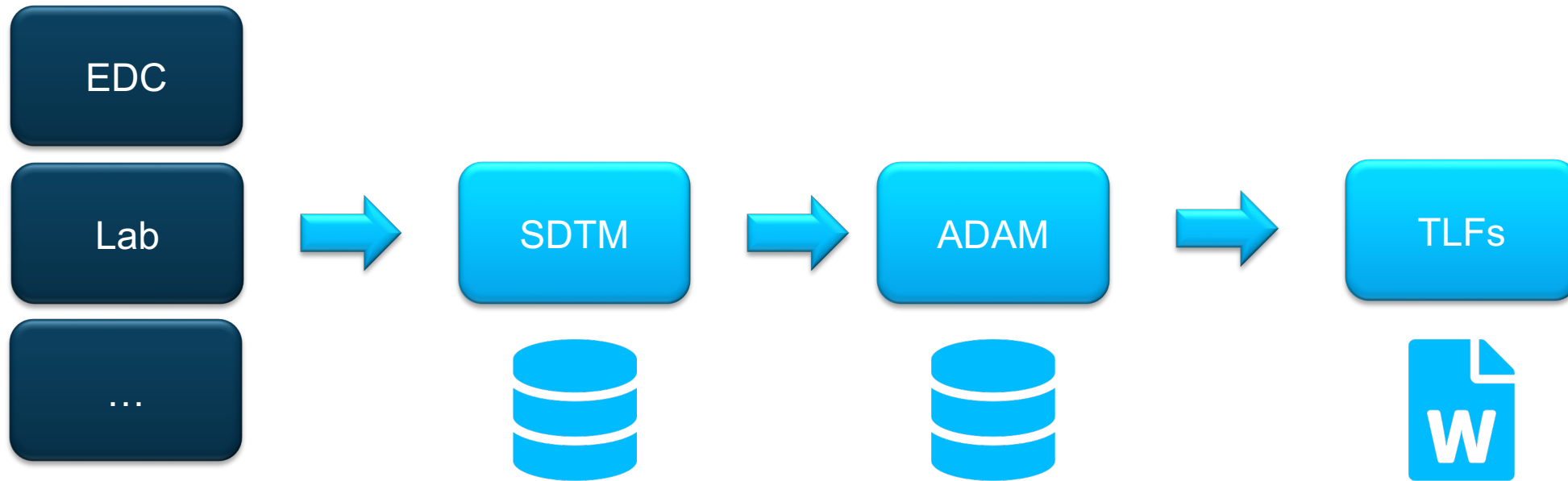
# Agenda

- // Introduction
- // Text Compare
- // Track Changes Compare
- // Detailed Table Compare
- // Images Compare
- // Solution Options
- // Summary

A close-up photograph of a person's hand, wearing a dark blue suit jacket and a white shirt, carefully placing a wooden block on top of a stack of other wooden blocks. The blocks are light-colored wood, possibly birch or maple, and are stacked in a slightly offset manner. The background is a plain, light-colored wall. The overall image conveys a sense of precision, balance, and the cumulative effect of small actions.

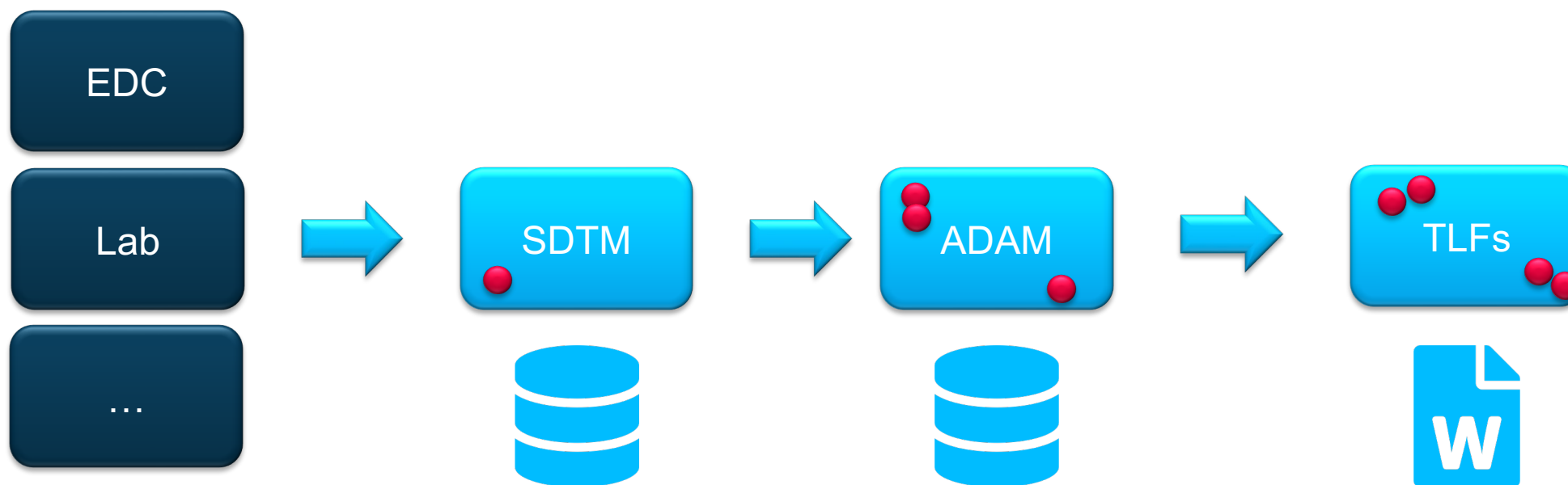
*Minor*  
changes can have  
*Major* impact

# Introduction





# Introduction





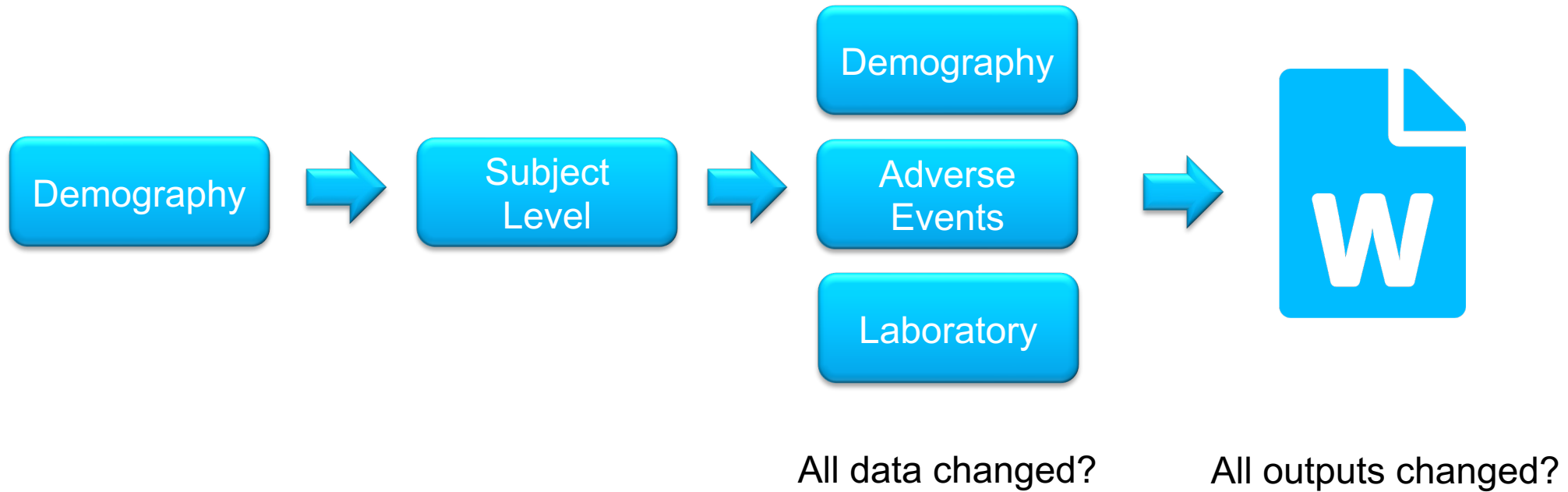
# *Traceability*

could be the answer



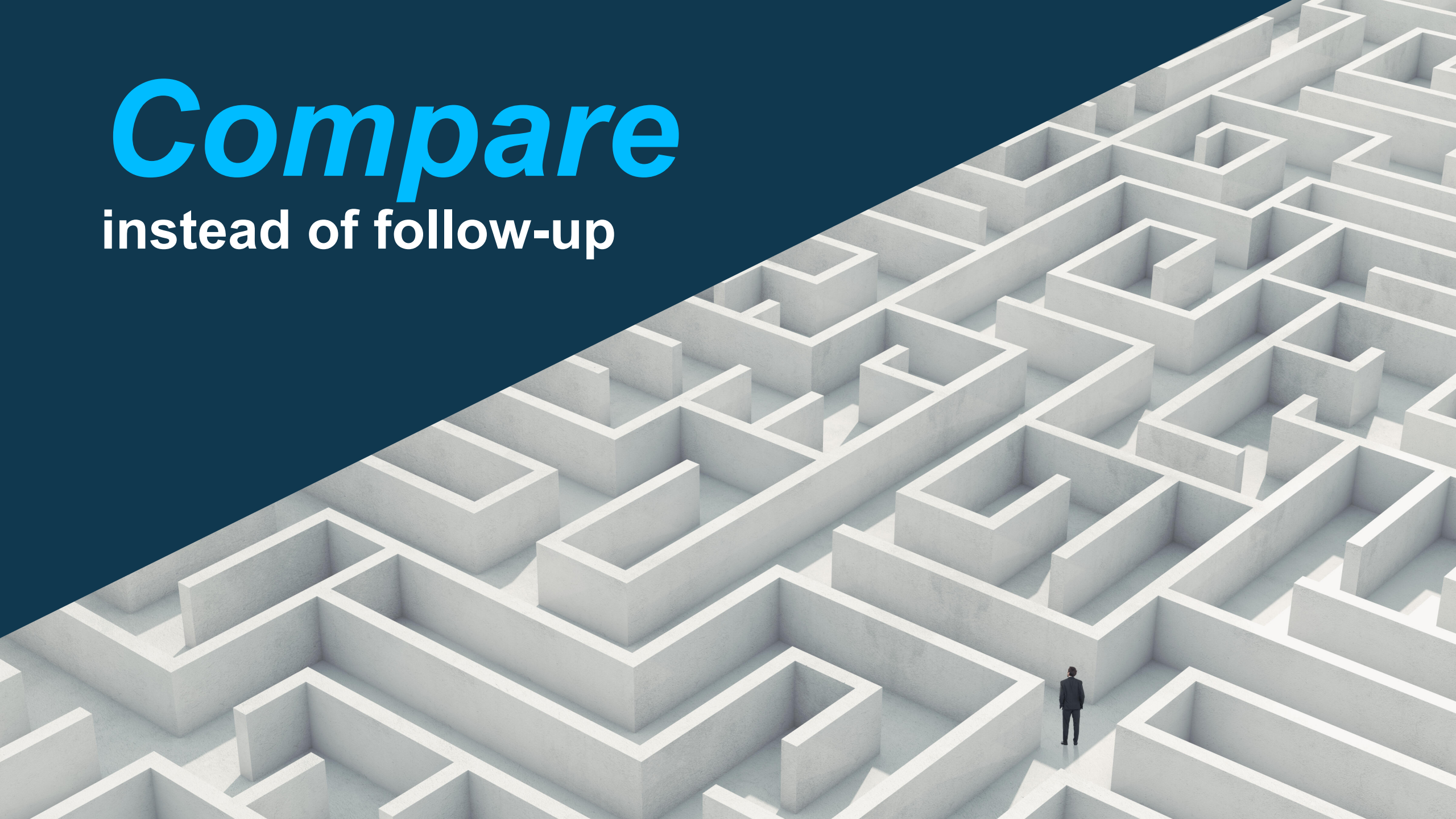
# Introduction

## Traceability



# *Compare*

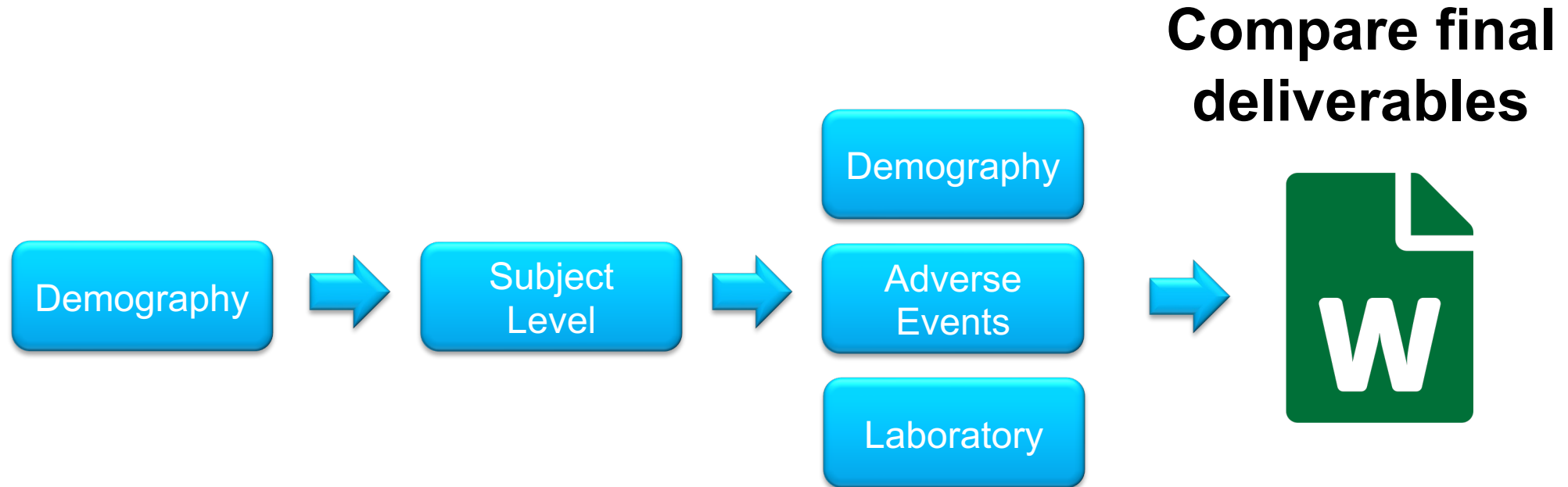
instead of follow-up





# Introduction

## Comparison



# Introduction

## Comparison

### Types Documents:

- Tables (Section 8)
- Listings (Section 10)
- Single RTFs
- SAPs
- Narratives
- Specifications
- ...



### Areas:

- Tables
- Texts
- Figures



### Exclusion:

- Document header/footer
- Standard table footer (idfoot)



**Did document texts change?**

*Text Compare*



# Text Compare

➤ Typically we expect no / minimal changes

➤ Quick full compare

## Table of contents

14. Tables and figures referred to but not

14.1 Demographics.....

14.1.1 Disposition.....

Table 14.1.1 / 1: Study sample  
switched to

Table 14.1.1 / 2: Distribution

Table 14.1.1 / 3: Distribution

Table 14.1.1 / 4: Subject validity and primary reasons for exclusion from analysis (all  
randomized subjects)..... 121

Table 14.1.1 / 5: Subject disposition (all enrolled subjects)..... 122

Table 14.1.1 / 6: Disposition: End of run-in (all enrolled subjects)..... 123

Table 14.1.1 / 7: Disposition: End of screening (all subjects who completed run-in)..... 124

Table 14.1.1 / 8: Disposition: End of treatment (all randomized subjects)..... 125

## Table of contents

16.2.4 Demographics and baseline data[Part1] ..... 1

Listing 16.2.4 / 1: Inclusion/exclusion criteria..... 2

Listing 16.2.4 / 2: Inclusion and exclusion criteria not met..... 5

Listing 16.2.4 / 3: Informed consent ..... 100

Listing 16.2.4 / 4: Enrollment and randomization (full analysis set)..... 460

Listing 16.2.4 / 5: Demographics ..... 1246

Listing 16.2.4 / 6: Current cigarette smoking status ..... 1774

Listing 16.2.4 / 7: Alcohol use ..... 2158

# Text Compare

## ➤ Read in Word Documents into SAS datasets\*

|    | VAR1    | VAR2      | VAR3      | VAR4      | VAR5      | VAR6       | VAR7   |
|----|---------|-----------|-----------|-----------|-----------|------------|--------|
| 1  | Age (y) | n         | 86        | 84        | 84        | 254        | 0.5934 |
| 2  |         | Mean      | 75.2      | 75.7      | 74.4      | 75.1       | .      |
| 3  |         | SD        | 8.59      | 8.29      | 7.89      | 8.25       | .      |
| 4  |         | Median    | 76.0      | 77.5      | 76.0      | 77.0       | .      |
| 5  |         | Min       | 52.0      | 51.0      | 56.0      | 51.0       | .      |
| 6  |         | Max       | 89.0      | 88.0      | 88.0      | 89.0       | .      |
| 7  |         |           |           |           |           |            | .      |
| 8  |         | <65 yrs   | 14 ( 16%) | 8 ( 10%)  | 11 ( 13%) | 33 ( 13%)  | 0.1439 |
| 9  |         | 65-80 yrs | 42 ( 49%) | 47 ( 56%) | 55 ( 65%) | 144 ( 57%) | .      |
| 10 |         | >80 yrs   | 30 ( 35%) | 29 ( 35%) | 18 ( 21%) | 77 ( 30%)  | .      |
| 11 |         |           |           |           |           |            | .      |
| 12 | Sex     | n         | 86        | 84        | 84        | 254        | 0.1409 |
| 13 |         | Male      | 33 ( 38%) | 34 ( 40%) | 44 ( 52%) | 111 ( 44%) | .      |
| 14 |         | Female    | 53 ( 62%) | 50 ( 60%) | 40 ( 48%) | 143 ( 56%) | .      |

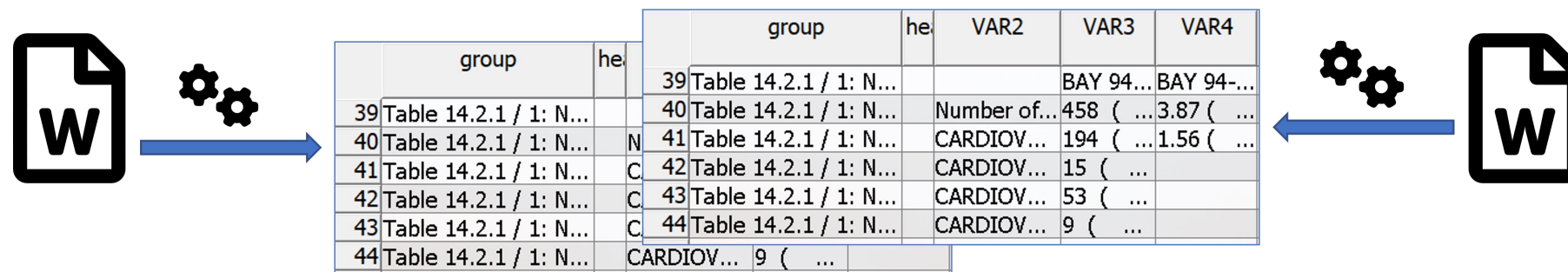
  

|                             |           |           |           |           |  |  |  |
|-----------------------------|-----------|-----------|-----------|-----------|--|--|--|
| Protocol: CDISCPIL01        |           |           |           |           |  |  |  |
| Population: Intent-to-Treat |           |           |           |           |  |  |  |
| Summary                     |           |           |           |           |  |  |  |
|                             |           |           |           |           |  |  |  |
|                             |           |           |           |           |  |  |  |
| Age (y)                     | n         |           |           |           |  |  |  |
|                             | Mean      |           |           |           |  |  |  |
|                             | SD        |           |           |           |  |  |  |
|                             | Median    |           |           |           |  |  |  |
|                             | Min       |           |           |           |  |  |  |
|                             | Max       |           |           |           |  |  |  |
|                             |           |           |           |           |  |  |  |
|                             | <65 yrs   | 14 ( 16%) | 8 ( 10%)  | 11 ( 13%) |  |  |  |
|                             | 65-80 yrs | 42 ( 49%) | 47 ( 56%) | 55 ( 65%) |  |  |  |
|                             | >80 yrs   | 30 ( 35%) | 29 ( 35%) | 18 ( 21%) |  |  |  |
|                             |           |           |           |           |  |  |  |
| Sex                         | n         | 86        | 84        | 84        |  |  |  |
|                             | Male      | 33 ( 38%) | 34 ( 40%) | 44 ( 52%) |  |  |  |
|                             | Female    | 53 ( 62%) | 50 ( 60%) | 40 ( 48%) |  |  |  |

\*[Let's get Groovy - Word and PDF processing in SAS](#),  
PHUSE EU Connect 2022

# Text Compare

## Process



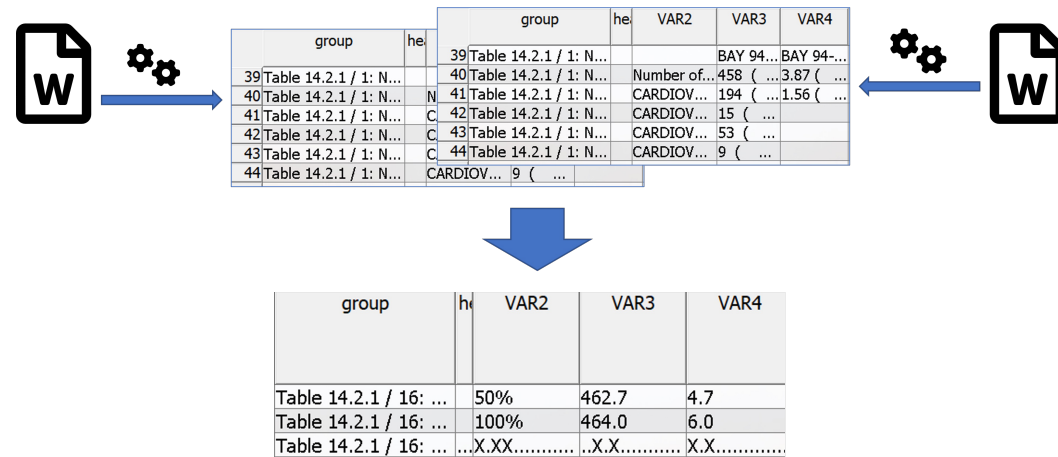
- Read in documents to SAS
- Replace timestamps
- Use simple PROC COMPARE
- Print information to log

| group                  | he | VAR2         | VAR3       | VAR4     |
|------------------------|----|--------------|------------|----------|
| Table 14.2.1 / 16: ... |    | 50%          | 462.7      | 4.7      |
| Table 14.2.1 / 16: ... |    | 100%         | 464.0      | 6.0      |
| Table 14.2.1 / 16: ... |    | ...X.XX..... | ..X.X..... | X.X..... |

# Text Compare

## Summary

- Sometimes difficult to see where exactly the change was
- When too many changes are available, result could be too big
- Super fast -> 41 seconds for 80MB, ~8.500 pages





# *Track Changes Compare*

**What has changed?**



# Track Changes Compare

➤ Changes available, want to see them

Reference Number: RD-OI-0129  
Supplement Version: 3

Program specification for Macro MOSTO\_PDF\_CREATE\_OVERLAY  
Version: 6.4  
Document Version: 1.0  
Date Completed: 01MAR2017



| Type of check | Object     | Descriptions                         | Message / Reaction                                          |
|---------------|------------|--------------------------------------|-------------------------------------------------------------|
| FORMAL        | <parmbuff> | Only allowed parameter must be used. | A warning is printed and additional parameters are ignored. |
| FORMAL        | FILE       | Must not be missing                  | Abort with an error message                                 |
| FORMAL        | LOGO       | Must not be missing                  | Abort with an error message                                 |
| FORMAL        | LOGO       | Must be existing or be "DEFAULT"     | Abort with an error message                                 |



Reference Number: RD-SOP-1129

Program specification for Macro MOSTO\_PDF\_CREATE\_OVERLAY  
MoSTO Version: 7  
Document Version: 1.1  
Date Completed: 13SEP2021



### 3.6 Checks and Error Handling

The following checks are performed:

| Type of check | Object     | Descriptions                         | Message / Reaction                                          |
|---------------|------------|--------------------------------------|-------------------------------------------------------------|
| FORMAL        | <parmbuff> | Only allowed parameter must be used. | A warning is printed and additional parameters are ignored. |
| FORMAL        | FILE       | Must not be missing                  | Abort with an error message                                 |
| FORMAL        | HEADLINE   | Must not be missing                  | Abort with an error message                                 |
| FORMAL        | LOGO       | Must be existing or be "DEFAULT"     | Abort with an error message                                 |

Reference Number: RD-SOP-1129OI-0129  
Supplement Version: 3

Program specification for Macro MOSTO\_PDF\_CREATE\_OVERLAYMOSTO\_PDF\_CREATE\_OVERLAY  
MoSTO Version: 76.4  
Document Version: 1.40  
Date Completed: 13SEP202401MAR2017



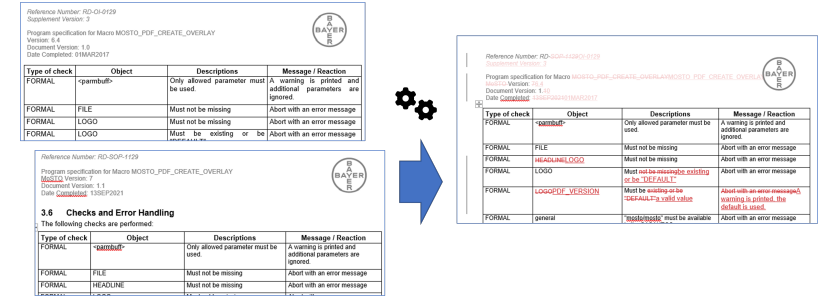
| Type of check | Object          | Descriptions                                   | Message / Reaction                                                    |
|---------------|-----------------|------------------------------------------------|-----------------------------------------------------------------------|
| FORMAL        | <parmbuff>      | Only allowed parameter must be used.           | A warning is printed and additional parameters are ignored.           |
| FORMAL        | FILE            | Must not be missing                            | Abort with an error message                                           |
| FORMAL        | HEADLINELOGO    | Must not be missing                            | Abort with an error message                                           |
| FORMAL        | LOGO            | Must not be missingbe existing or be "DEFAULT" | Abort with an error message                                           |
| FORMAL        | LOGOPDF_VERSION | Must be existing-or-be "DEFAULT"a valid value  | Abort with an error messageA warning is printed, the default is used. |
| FORMAL        | general         | "mosto/mosto" must be available                | Abort with an error message                                           |



# Track Changes Compare

## Process

- Utilize VBA
- Use Java Options
  - We use commercial Java library “Aspose for Word”
- Read in documents
- Create “track-changes” outputs (if differences available)
- Print summary in log where changes are available
- Optionally: remove timestamps and similar before compare





# Track Changes Compare

## Summary

- Nice change view
- Review of complete document
- Too many/difficult changes could result in “all changed”
- Issues with “big” files

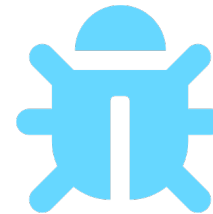
- Fast
  - 80MB, ~8.500 pages -> out of memory
  - 30MB, 990 pages -> 2:14 minutes

- More for “tables”, not for “listings”

| Type of check | Object    | Descriptions                        | Message / Reaction                                         |
|---------------|-----------|-------------------------------------|------------------------------------------------------------|
| FORMAL        | *paramDef | Only allowed parameter must be used | A warning is printed and additional parameters are ignored |
| FORMAL        | FILE      | Must not be missing                 | About with an error message                                |
| FORMAL        | LOGO      | Must not be missing                 | About with an error message                                |
| FORMAL        | LOGO      | Must be existing or be              | About with an error message                                |

| Type of check | Object         | Descriptions                               | Message / Reaction                                         |
|---------------|----------------|--------------------------------------------|------------------------------------------------------------|
| FORMAL        | *paramDef      | Only allowed parameter must be used        | A warning is printed and additional parameters are ignored |
| FORMAL        | FILE           | Must not be missing                        | About with an error message                                |
| FORMAL        | HEADLINE/LOGO  | Must not be missing                        | About with an error message                                |
| FORMAL        | LOGO           | Must not be missing, existing or be USEFUL | About with an error message                                |
| FORMAL        | LOGOED_VERSION | Must be existing as a DEFAULT value        | About with an error message                                |
| FORMAL        | general        | 'headlines' must be available              | About with an error message                                |



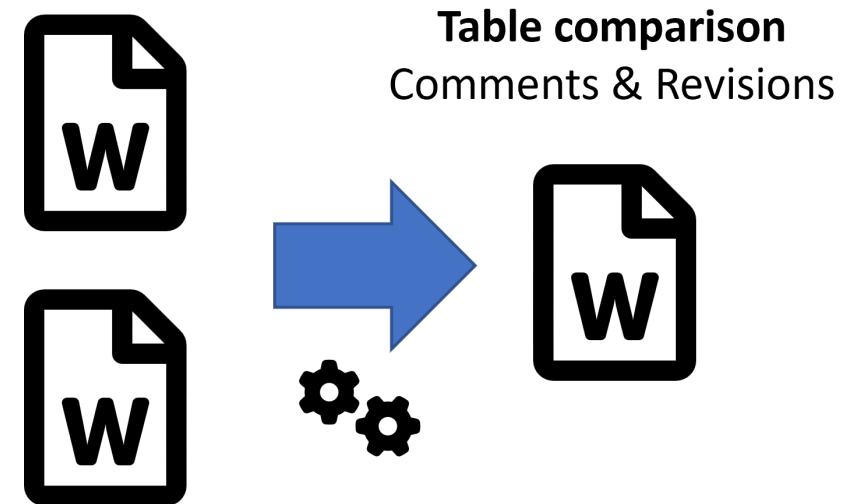


# *Detailed Table Compare*

**Did tables change?**

# Detailed Table Compare

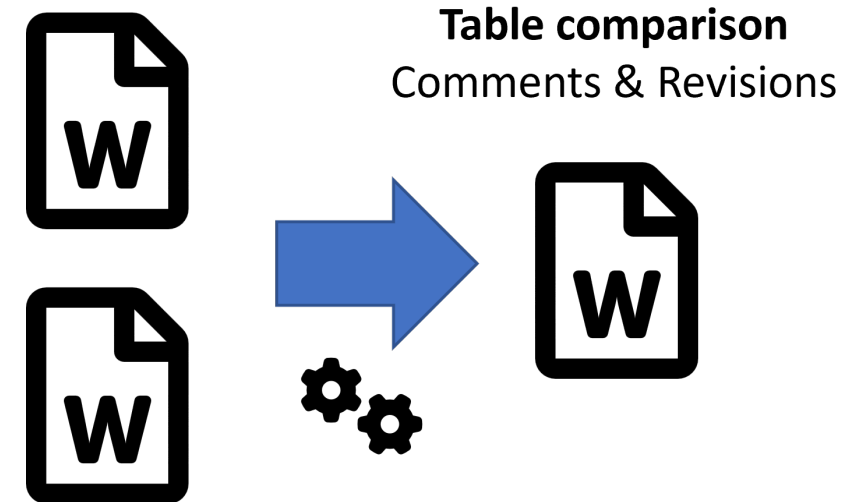
- Table title changed
  - New or deleted tables
  - New or deleted rows in a table
  - New or deleted columns in a table
  - Changes in single table cells
  - Changes in the footer (excluding identifier)
- 
- Track changes output with comments



# Detailed Table Compare

## Processing

- Read in all tables from old / new
- Match them up, if they do not match
  - Similarity investigation      -> title changed
  - Otherwise                      -> added / deleted
- Check for row counts
  - Added / deleted row
- Check for column headers
  - Added / deleted column
- Check cell content
- ...





# Detailed Table Compare

## Examples

### Removed Rows

|        |         |       |           |           |      |           |
|--------|---------|-------|-----------|-----------|------|-----------|
| Europe | Total   |       | 28SEP2010 | 14APR2019 | 4875 | Total     |
|        |         |       |           |           |      | Treatment |
|        |         |       |           |           |      | Placebo   |
|        | AUSTRIA | Total | 27OCT2010 | 19MAR2019 | 126  | Total     |
|        |         |       |           |           |      | Treatment |
|        |         |       |           |           |      | Placebo   |
|        |         |       | 27OCT2010 | 19MAR2019 | 126  | Total     |
|        |         |       |           |           |      | Treatment |
|        |         |       |           |           |      | Placebo   |
|        | BELGIUM | Total | 24NOV2010 | 02APR2019 | 91   | Total     |
|        |         |       |           |           |      | Treatment |
|        |         |       |           |           |      | Placebo   |

### Deleted Table, comment at TOC

|   |         |                                              |
|---|---------|----------------------------------------------|
| 2 | ComBear | Deleted: Disposition: Number of subjects who |
|---|---------|----------------------------------------------|

### Deleted Column

|         |                         |
|---------|-------------------------|
| ComBear | Column deleted: Placebo |
|         | N=2841 (100%)           |

Table 3: Distribution of subjects among countries (full analysis)

|         | Treatment<br>N=2833 (100%) | Total<br>N=5674 (100%) |
|---------|----------------------------|------------------------|
|         | 1182 ( 41.7%)              | 2358 ( 41.6%)          |
| Europe  | 619 ( 21.8%)               | 1251 ( 22.0%)          |
| AUSTRIA | 29 ( 1.0%)                 | 62 ( 1.1%)             |
| BELGIUM | 31 ( 1.1%)                 | 54 ( 1.0%)             |
| DENMARK | 59 ( 2.1%)                 | 111 ( 2.0%)            |

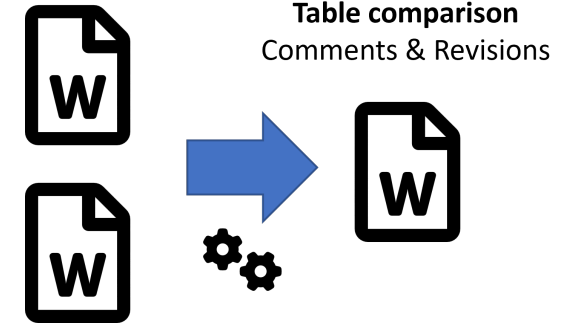
COMBEAR - ----- Summary -----  
COMBEAR - 6 out of 13 compared tables changed.  
COMBEAR - 1 tables are added.  
COMBEAR - 1 tables are removed.  
COMBEAR - Table changes detected in the document.



# Track Changes Compare

## Summary

- Detailed overview of changes
- Review of complete document
- Track complex changes
- Fast
  - 30MB, 990 pages -> 2:09 minutes
- Only compares tables, not other content, for example section headings



**Did images change?**

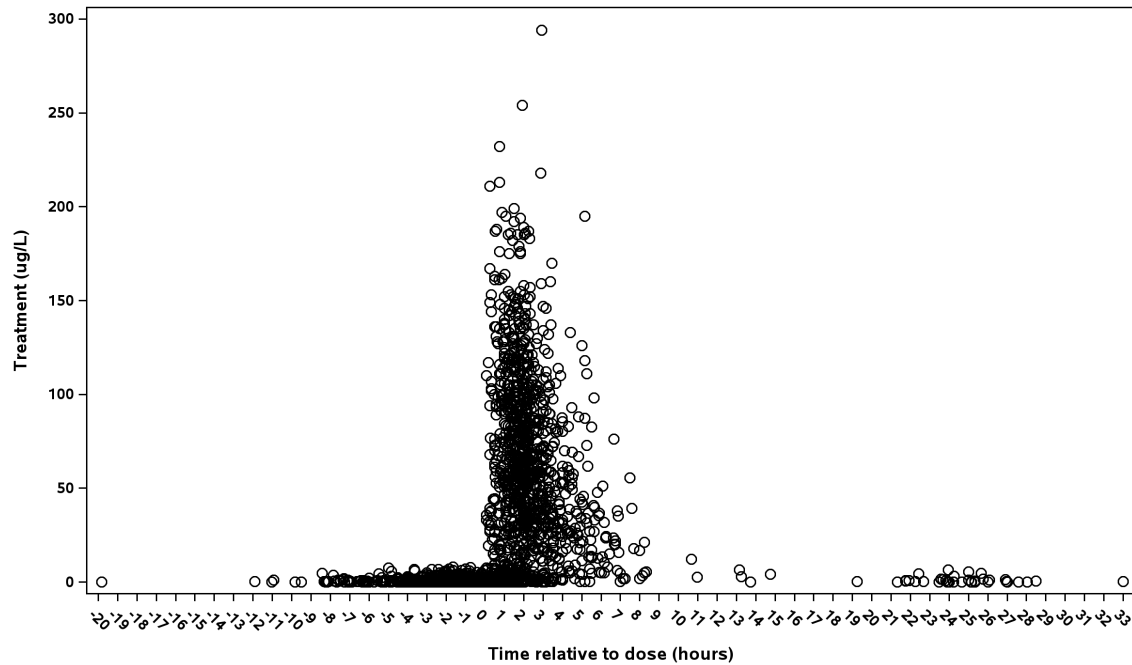
*Image Compare*



# Image Compare

Individual concentrations of treatment in plasma by dose (pharmacokinetic analysis set)

Dose = 10 mg  
LIN-LIN plot

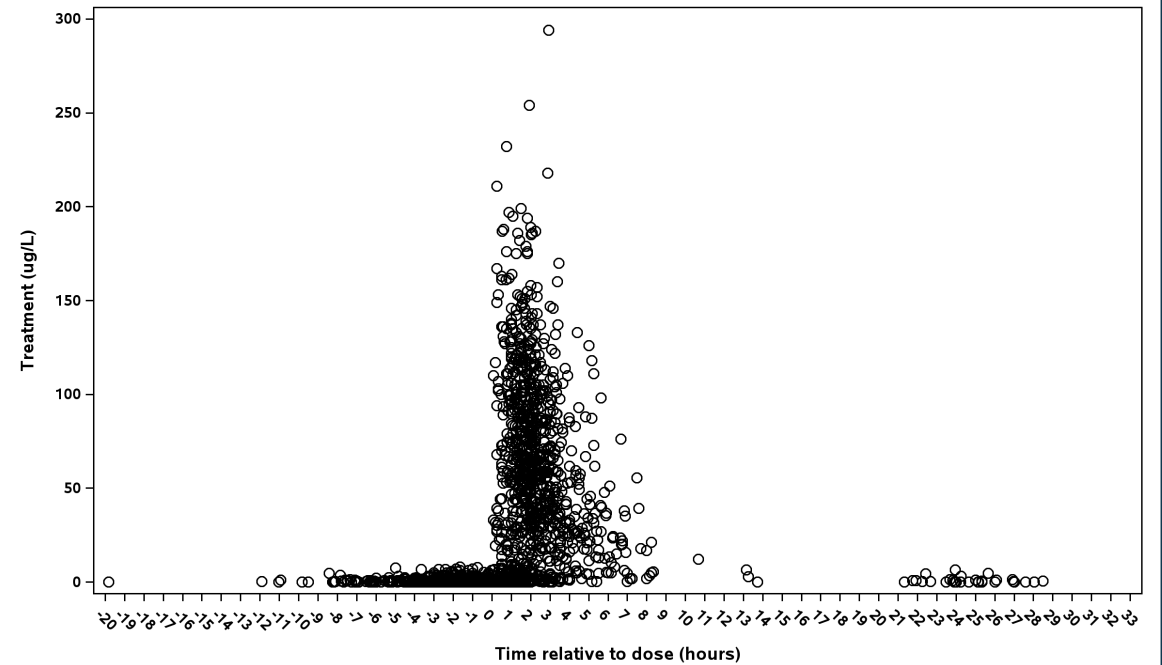


Measurements that were taken after study drug administration for Visit 3.

Bayer: /var/swan/root/bhc/general/playground/phuse/dev/pgms/2023\_compare.sas 05OCT2023 9:55

Individual concentrations of treatment in plasma by dose (pharmacokinetic analysis set)

Dose = 10 mg  
LIN-LIN plot



Measurements that were taken after study drug administration for Visit 3.

Bayer: /var/swan/root/bhc/general/playground/phuse/dev/pgms/2023\_compare.sas 05OCT2023 9:55



# Image Compare

Pre-requisite:

- Same size, color, file format

Algorithm:

- Compare each pixel
  - All equal
  - Differences
    - Show differences in picture
    - Show number of pixel differences in log
    - Show warning in log



# Image Compare

Implementation in Java:

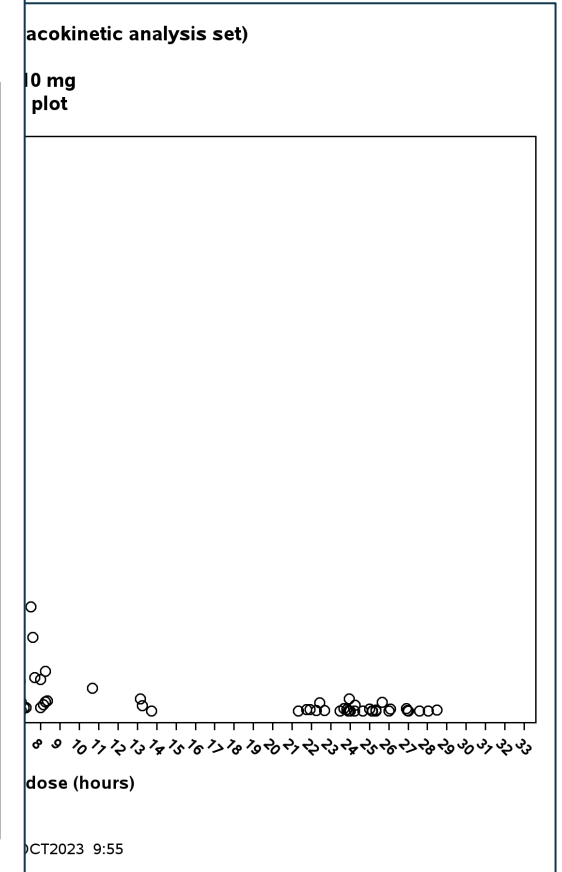
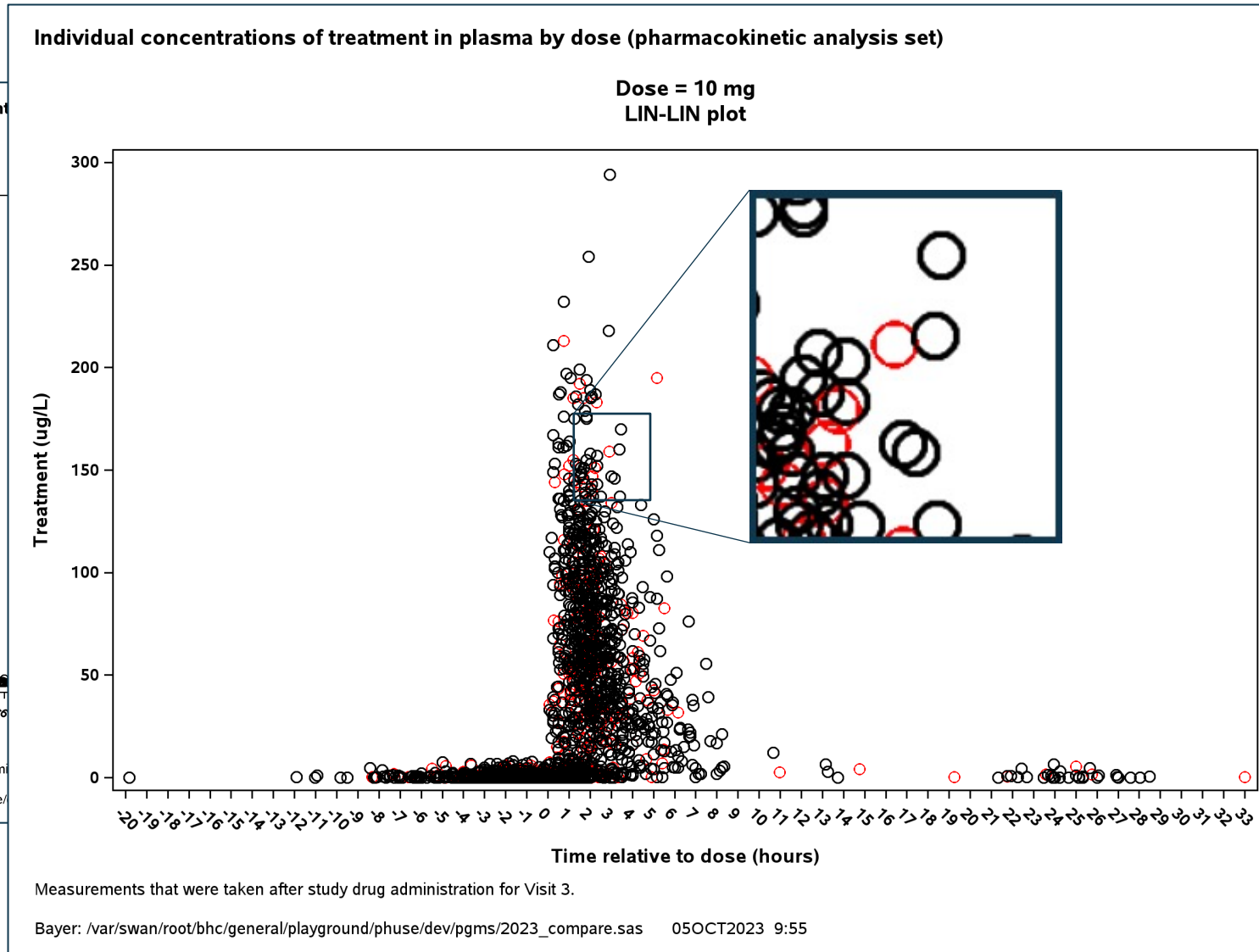
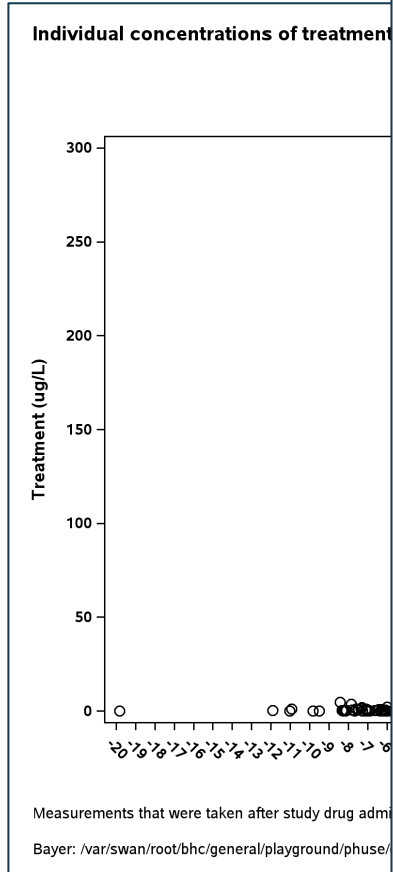
```
BufferedImage oldImage = ImageIO.read(oldFile);
BufferedImage newImage = ImageIO.read(newFile);

int pixelDiff = 0;
if (oldImage.getWidth() == newImage.getWidth() &&
    oldImage.getHeight() == newImage.getHeight()) {
    for (int x = 0; x < oldImage.getWidth(); x++) {
        for (int y = 0; y < oldImage.getHeight(); y++) {
            if (oldImage.getRGB(x, y) != newImage.getRGB(x, y)) {
                pixelDiff++;
                newImage.setRGB(x, y, RED_COLOR);
            }
        }
    }
}
```

SAS macro available in Paper



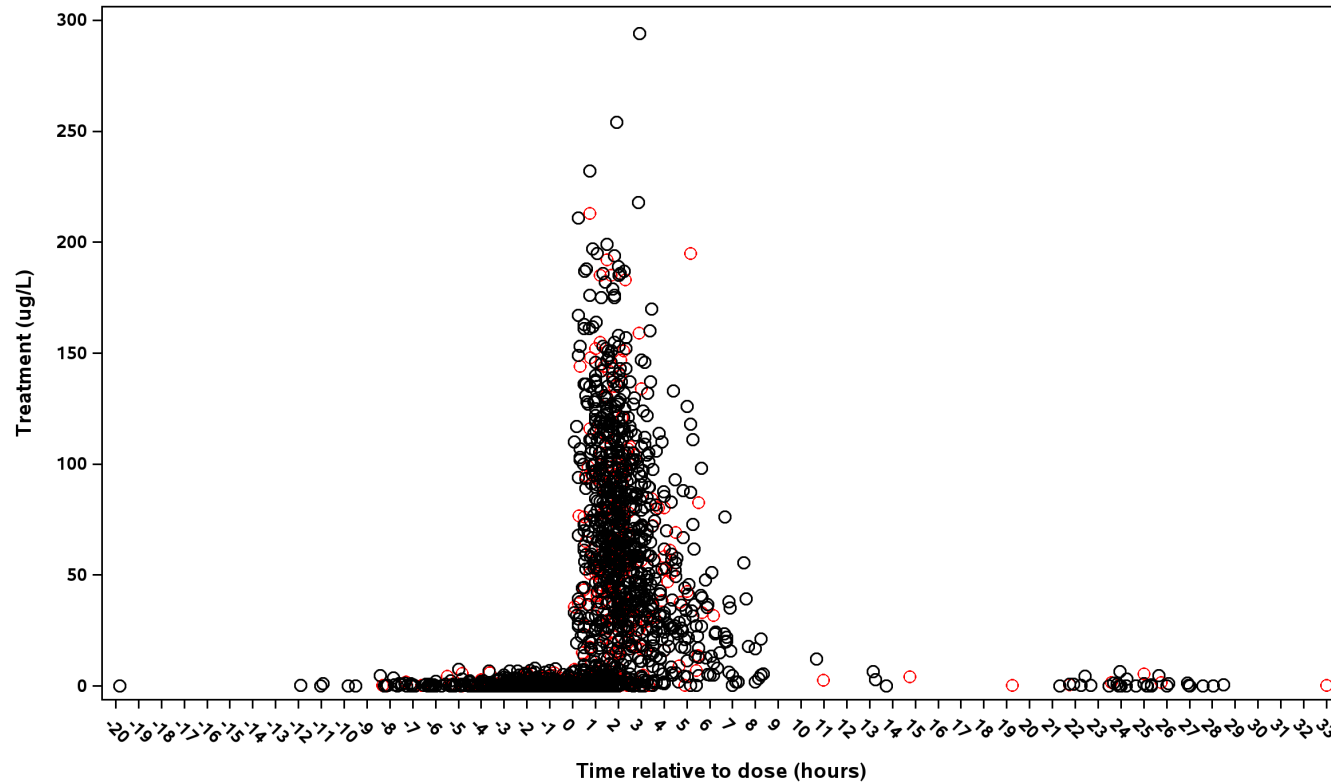
# Image Compare



# Image Compare

Individual concentrations of treatment in plasma by dose (pharmacokinetic analysis set)

Dose = 10 mg  
LIN-LIN plot



Measurements that were taken after study drug administration for Visit 3.

Bayer: /var/swan/root/bhc/general/playground/phuse/dev/pgms/2023\_compare.sas 05OCT2023 9:55

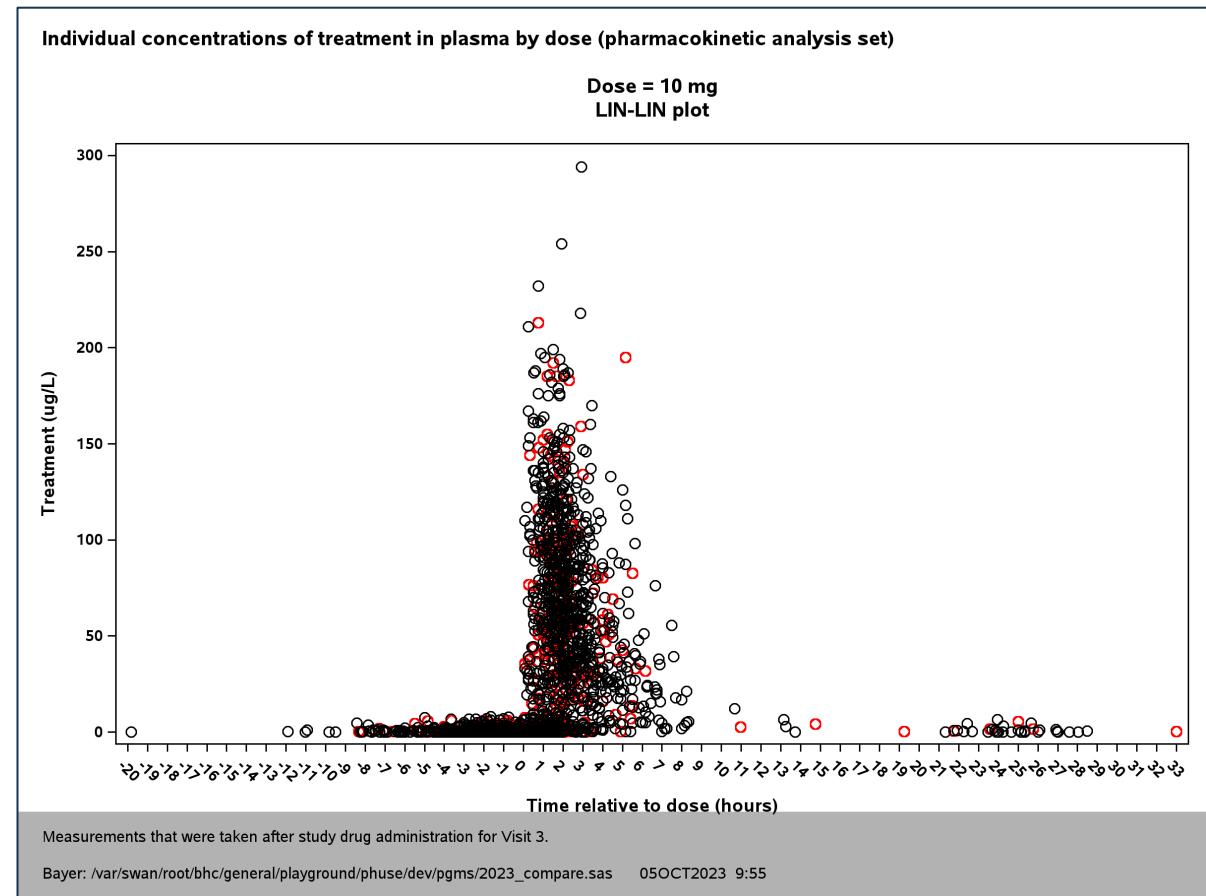
**WARNING: 8225  
changes found**



# Image Compare

## Extras

Ignore the „bottom“ part, for example for date/time stamps





# *Considerations*

# Considerations

**One size fits all?**



# Considerations

**No!**





# Considerations

- What is the scope?
- What do you want to know?
- Know your use-cases
- Provide solutions for this

## **Types Documents:**

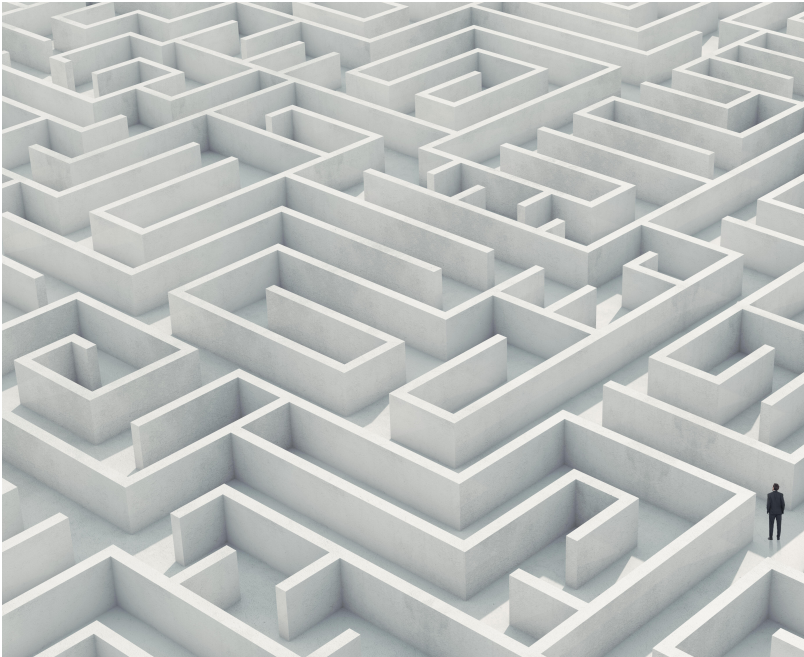
- Tables (Section 8)
- Listings (Section 10)
- Single RTFs
- SAPs
- Narratives
- Specifications
- ...



# *Summary*

# Summary

## Traceability is good



## Comparison is better



\*4 changes added



# Summary

## Comparison Scope

### ➤ Log-Messages

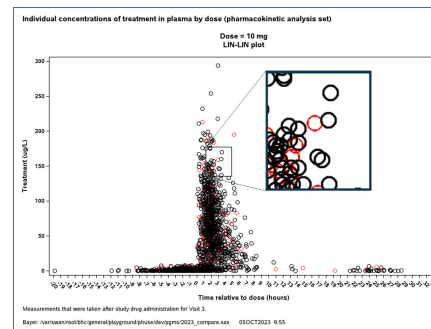
Matched up 266 tables. 6 of them with changes!

**WARNING: COMBEAR - Changes available: /var/swan**

### ➤ Dataset Summary

|   | _type_     | _OBS_  | group                    | h                     | VAR2  |
|---|------------|--------|--------------------------|-----------------------|-------|
| 7 | cd_oldFile | 251074 | Listing 16.2.5 / 2: S... | 16244510706006/64/F/W | 61.03 |
| 8 | cd_newFile | 251074 | Listing 16.2.5 / 2: S... | 16244510706006/64/F/W | 57.02 |
| 9 | DIF        | 3959   | Listing 16.2.5 / 2: S... | .....XX..X.....       |       |

### ➤ Compare visualization

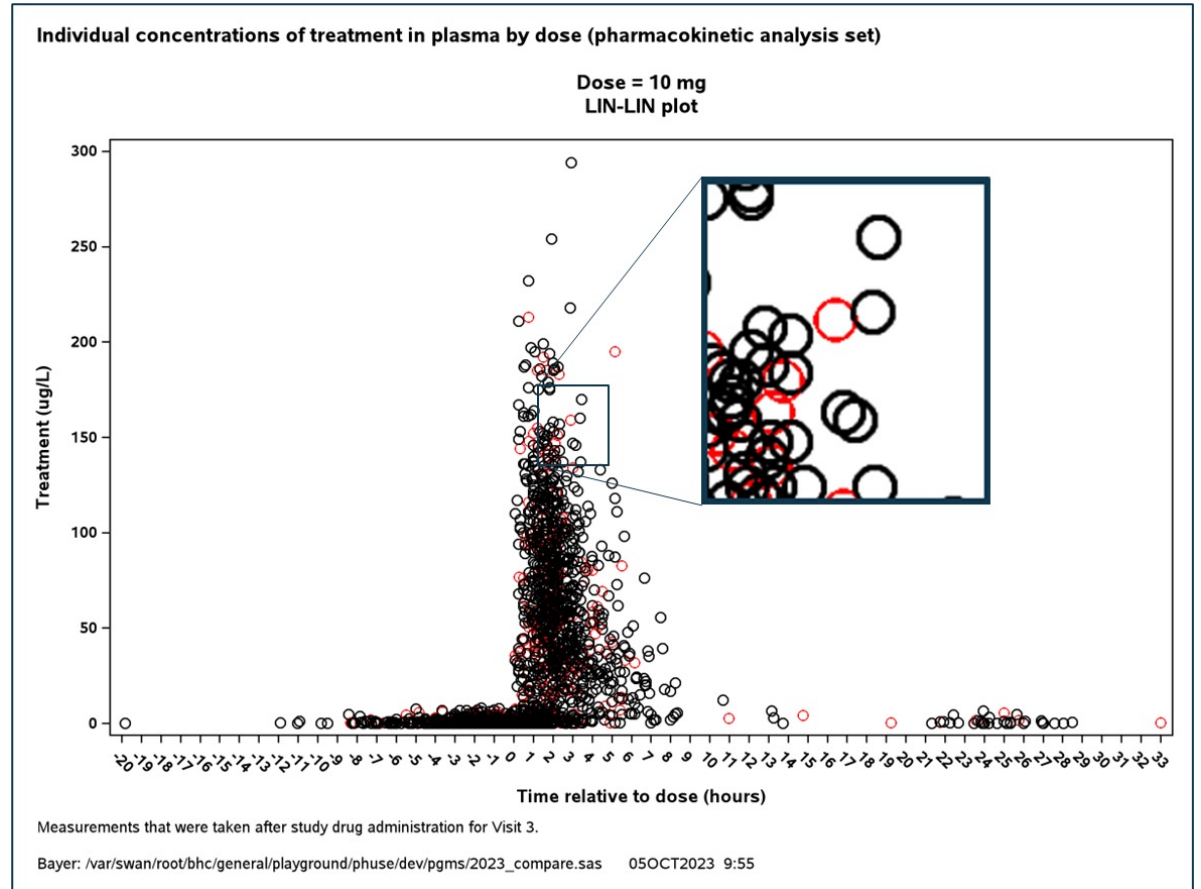


|        |         |       |           |           |      |           |
|--------|---------|-------|-----------|-----------|------|-----------|
| Europe | Total   |       | 28SEP2010 | 14APR2019 | 4875 | Total     |
|        |         |       |           |           |      | Treatment |
|        |         |       |           |           |      | Placebo   |
|        | AUSTRIA | Total | 27OCT2010 | 19MAR2019 | 126  | Total     |
|        |         |       |           |           |      | Treatment |
|        |         |       |           |           |      | Placebo   |
|        |         |       | 27OCT2010 | 19MAR2019 | 126  | Total     |
|        |         |       |           |           |      | Treatment |
|        |         |       |           |           |      | Placebo   |
|        | BELGIUM | Total | 24NOV2010 | 02APR2019 | 91   | Total     |
|        |         |       |           |           |      | Treatment |
|        |         |       |           |           |      | Placebo   |

# Summary

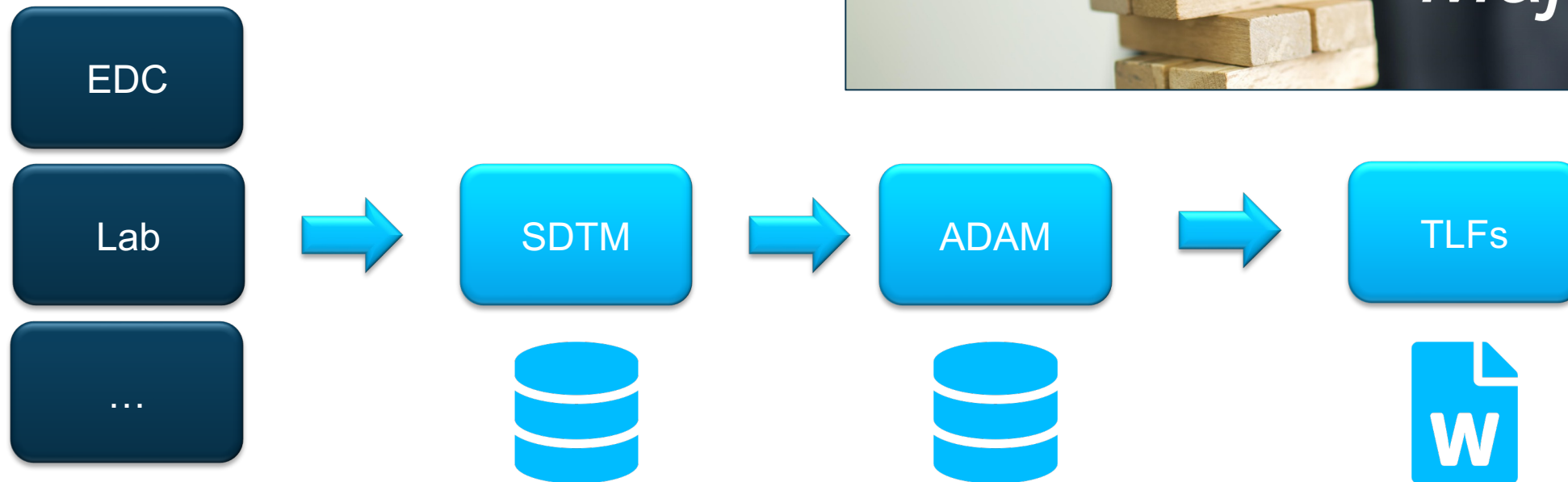
## Implementation Options

- Compare Images
- Simple Text Comparison
- Track-Changes Comparison
- Special Consideration Comparison



# Summary

Detect changes safely!





# *Thank you!*



**Katja Glaß**

**Katja.Glass@bayer.com**

