# Data Collection and Management

## Software for Data Processing, Visualization and Analysis at the French VMC

**Clément VAL**
CEESAR

Final Event
26-27 June 2012
Autoworld, Brussels

# Aim



**raw data**



**aggregated data**

- **Trips Logs:** CAN, GPS, Video files
  - ▷ unstructured, in any kind of context, depends on vehicle/instrumentation level
- **Context:** Experimental context database, Geographical database, Questionnaires

- per trip attributes
- events and per events attributes
- situations and per situations attributes
  - ▷ reduced, structured, contextualized
  - ▷ accessible across all trips using relations
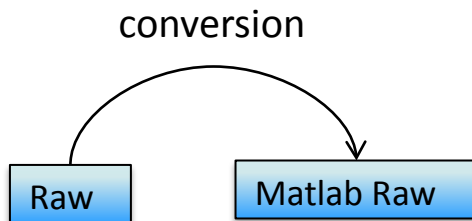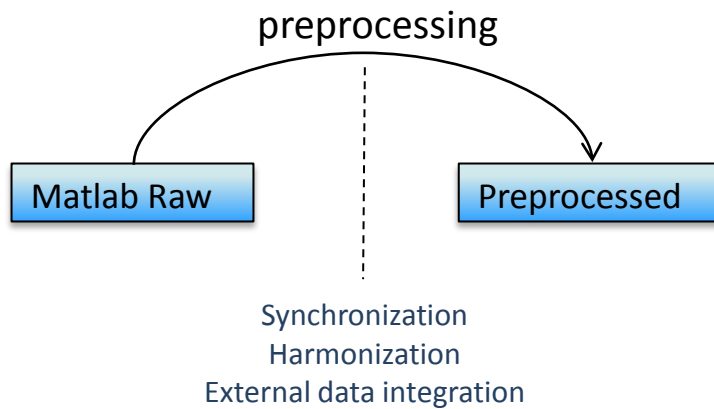
# Process

Per-trip processing
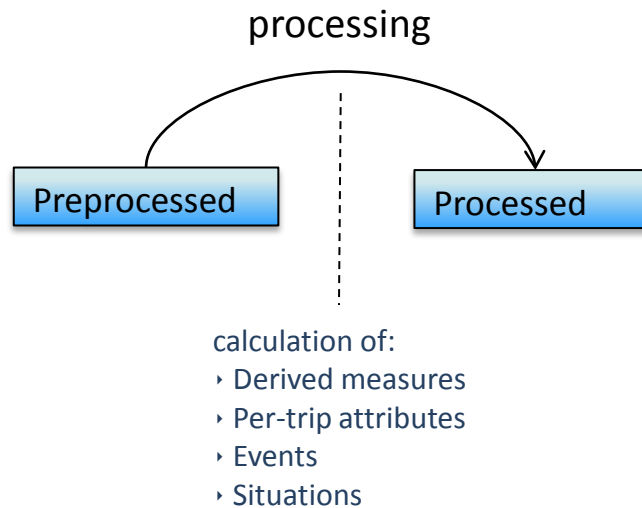
Raw

# Process

## Per-trip processing

conversion

```
Raw  →  Matlab Raw
```

# Process

## Per-trip processing

preprocessing

Matlab Raw → Preprocessed

Synchronization
Harmonization
External data integration

# Process

## Per-trip processing

processing

Preprocessed → Processed

calculation of:
‣ Derived measures
‣ Per-trip attributes
‣ Events
‣ Situations

# Process

## Per-trip processing

upload

Processed

DB
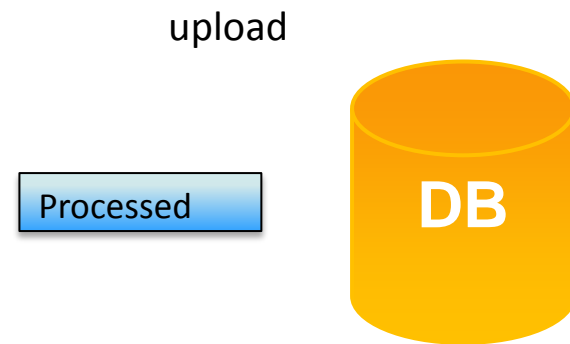
# Process

## Per-trip processing

## Transversal analysis

…from all trips

| Raw | Matlab Raw | Preprocessed | Processed |

| Matlab Raw | Preprocessed | Processed |

| Raw | Matlab Raw | Preprocessed | Processed |

**DB**

**Results table A**

|  | SV1 | SV2 | PI1 |
| --- | --- | --- | --- |
| SIT1 | ↓ | ↓ … | ↓ |
| SIT2 | | | |
| SIT3 | | | |
| … | | | |

www.eurofot-ip.eu
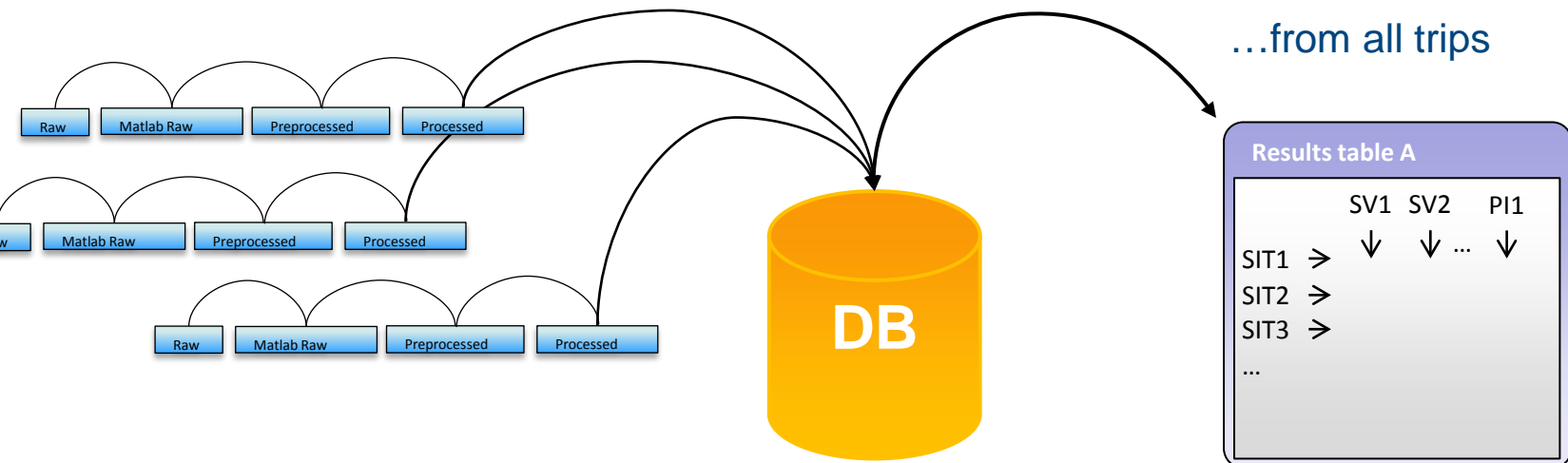
# Challenges

No such a linear process…



27 June 2012 euroFOT Final Event – Brussels

www.eurofot-ip.eu
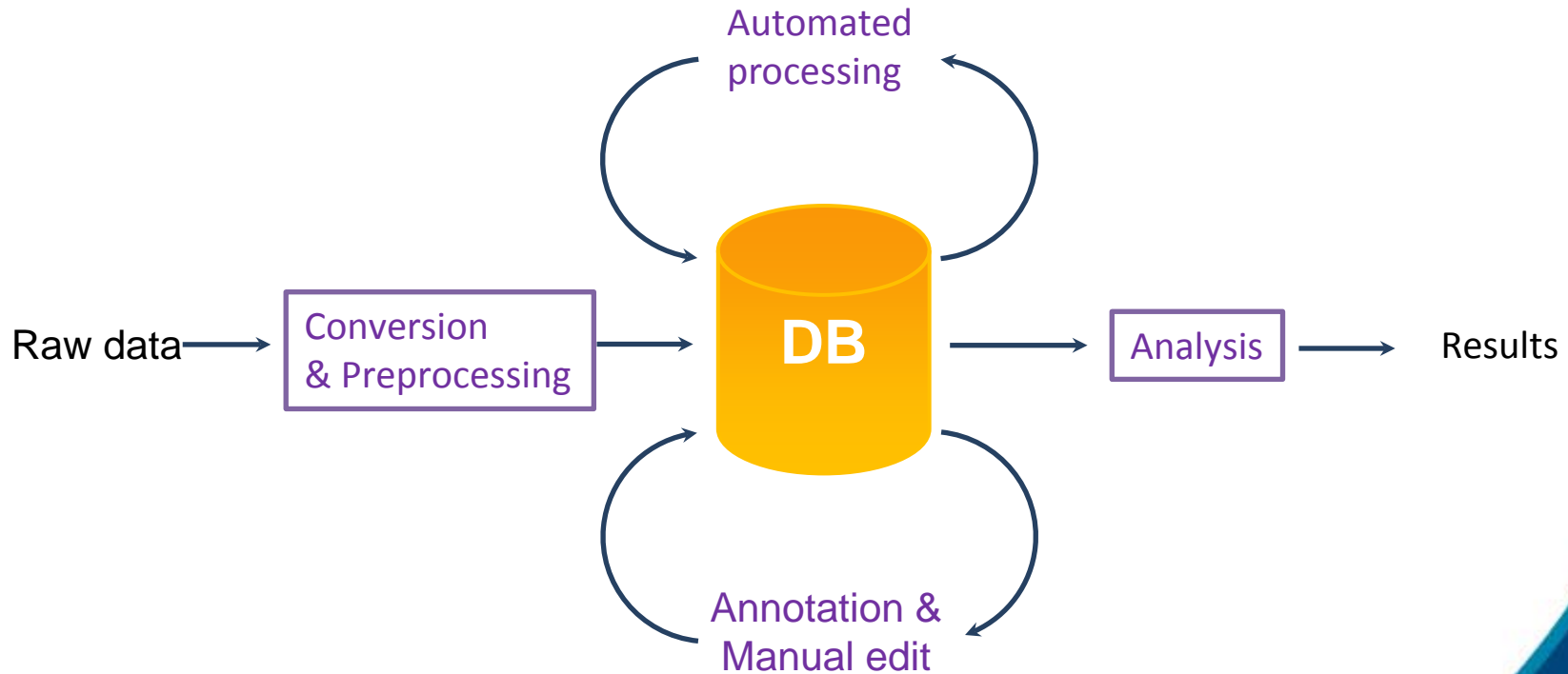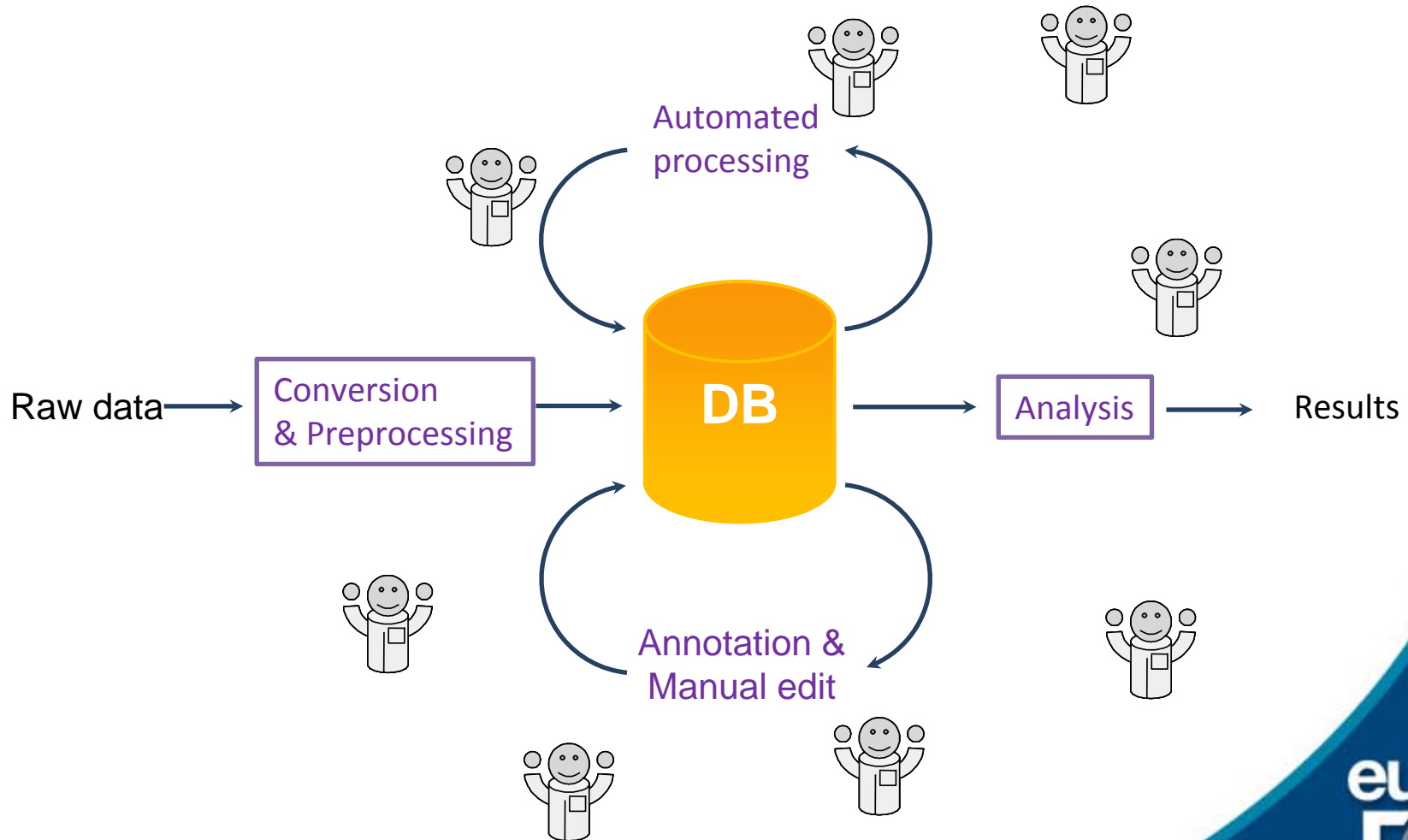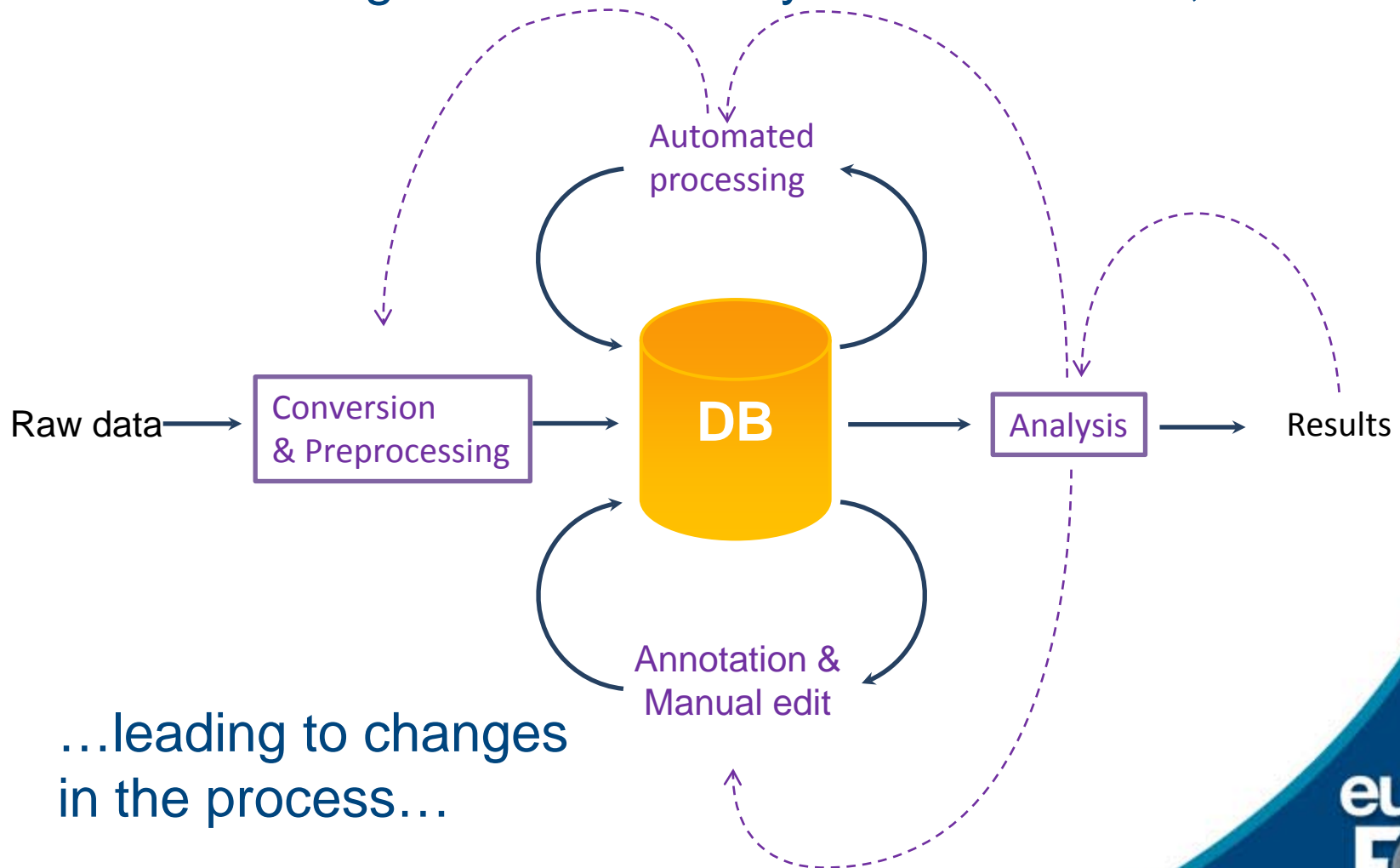
# Challenges

### Dataset is huge… and unpredictable !
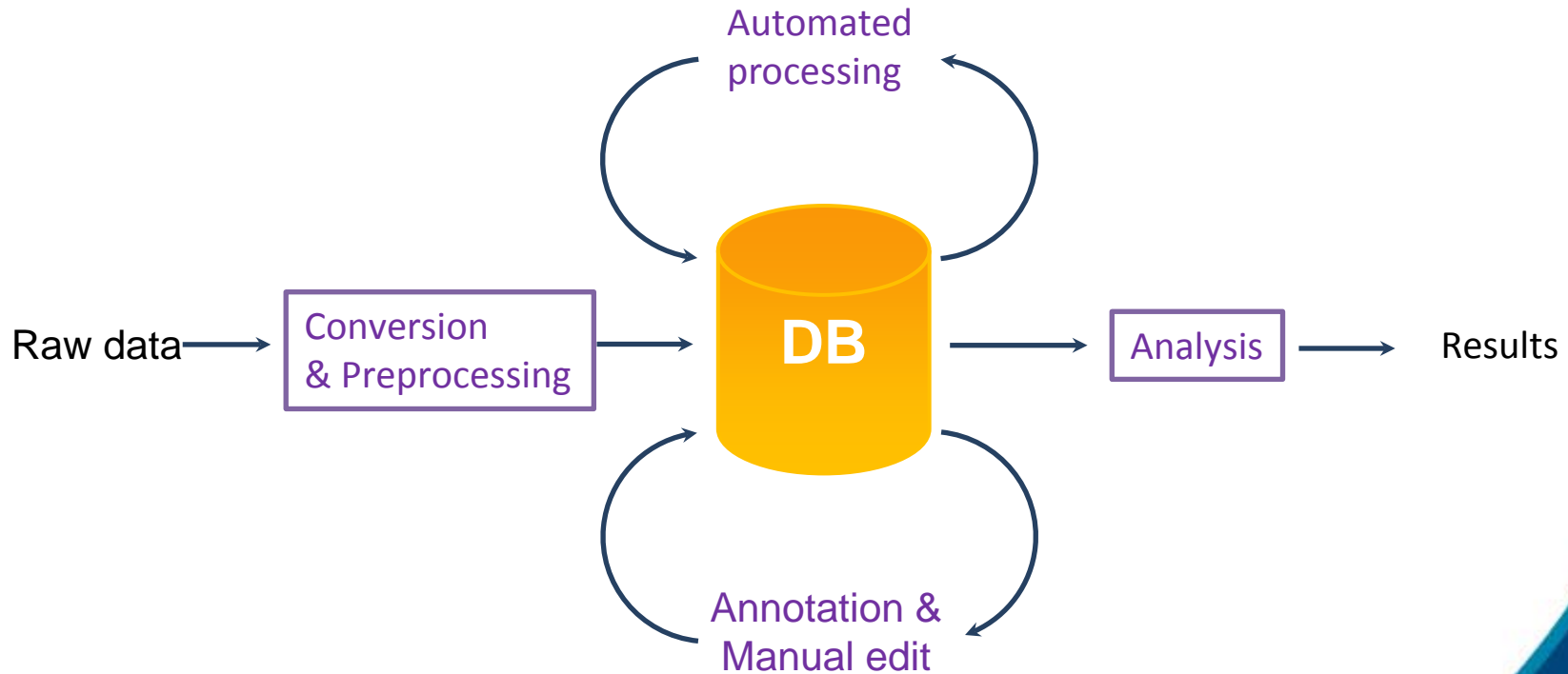
# Challenges

All kinds of crazy people want to play with data…

# Challenges

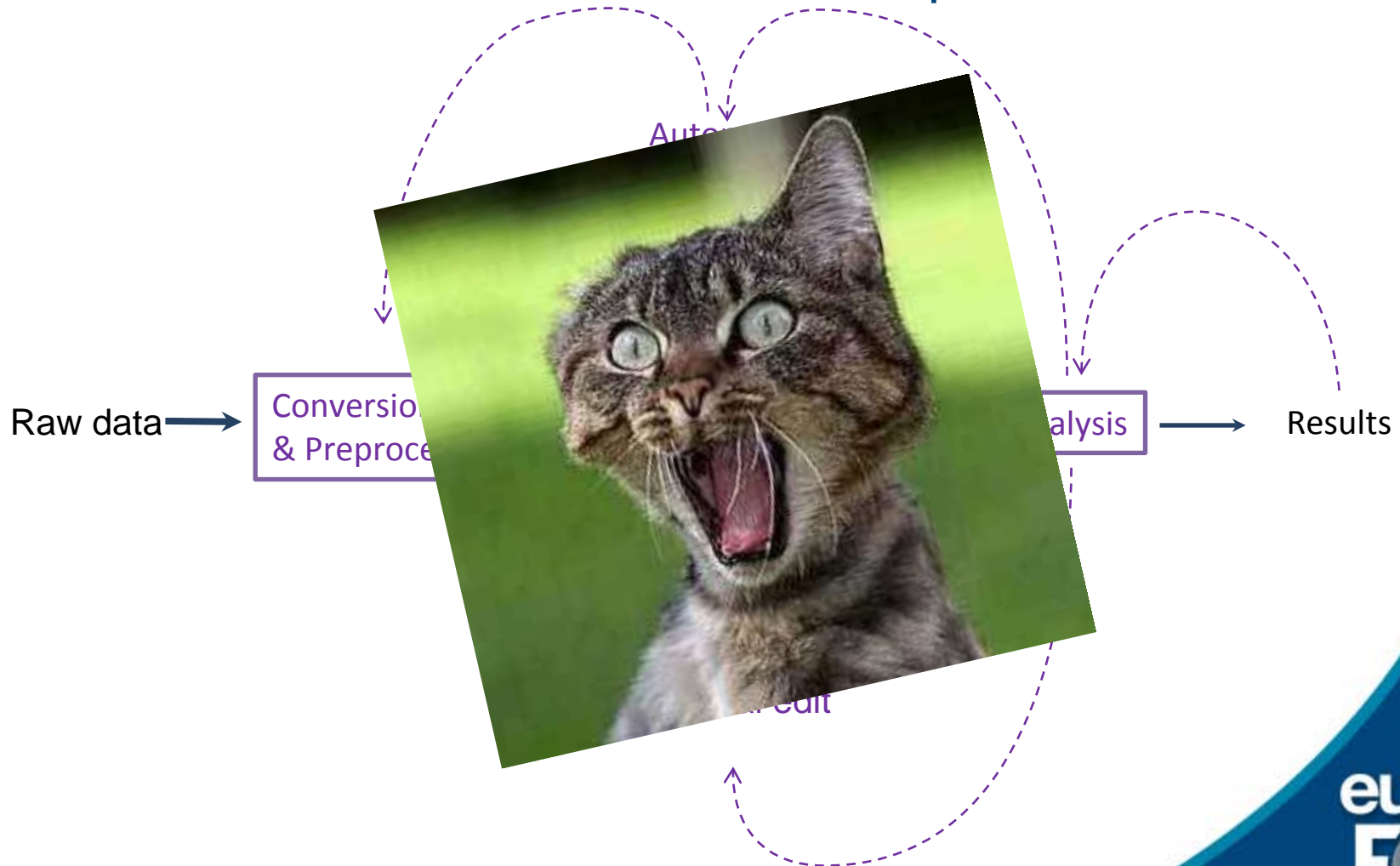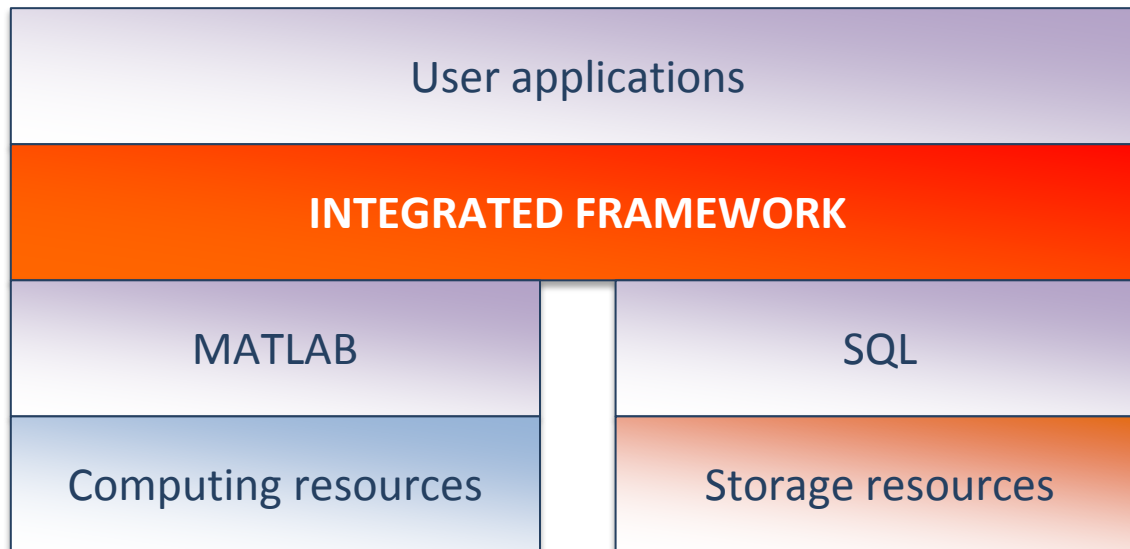…And change idea when they see first results,



Raw data → Conversion & Preprocessing → DB → Analysis → Results

Automated processing

Annotation & Manual edit

…leading to changes in the process…

# Challenges

## While new data continues to be acquired !



Raw data → Conversion & Preprocessing ... alysis → Results

# The solution



| User applications |
|---|
| **INTEGRATED FRAMEWORK** |

| MATLAB | SQL |
|---|---|
| Computing resources | Storage resources |

# Implementation

- Complete process is described in a user-defined, evolving, data-model
- Separates data management and scientific tasks
- All interactions are done through GUIs or a public API
- Versions & dependency control for each algorithm or entity definition
- Controlled evaluation of scripts:
  - ▷ execution plan derived from dependencies / versions
  - ▷ sandboxed execution
  - ▷ logs & diagnostic connection
- Database is automagically maintained by SQL interface
- GUIs content and behavior based on data model

# Achieves:

- Reliability « no human or programming error should affect and crash the whole system »
- Efficiency « avoid unnecessary calculations, allow reusability of algorithms »
- Quality, consistency and traceability « Ensure data consistency. Enforce traceability and documentation »
- Interactivity « allow feedback loops and modification of automatically computed entities »
- Scalability « computing should be parallelized »
- Genericity « we can use the same system to process data from different experiments »

# GUIs



27 June 2012    euroFOT Final Event – Brussels

www.eurofot-ip.eu

8 Functionalities, 28 Partners, 1000 Vehicles

1 Field Operational Test, **8 Functionalities**

28 Partners, 1000 Vehicles, 1 Field Operational Test

8 Functionalities, **28 Partners**, 1000 Vehicles

1 Field Operational Test, 8 Functionalities

28 Partners, **1000 Vehicles**, 1 Field Operational Test

8 Functionalities, 28 Partners, 1000 Vehicles

**1 Field Operational Test**