

# SALSA

A Large Scale Driving Data Analysis Tool

[clement.val@ceesar.fr](mailto:clement.val@ceesar.fr)

**WHICH KIND OF DATA?**

### Trips Explorer

Logger ID: 23100  
MeanSpeed: 30  
Driver ID: 0  
Trip Duration: 3600 - 5000  
Trip Length: 80

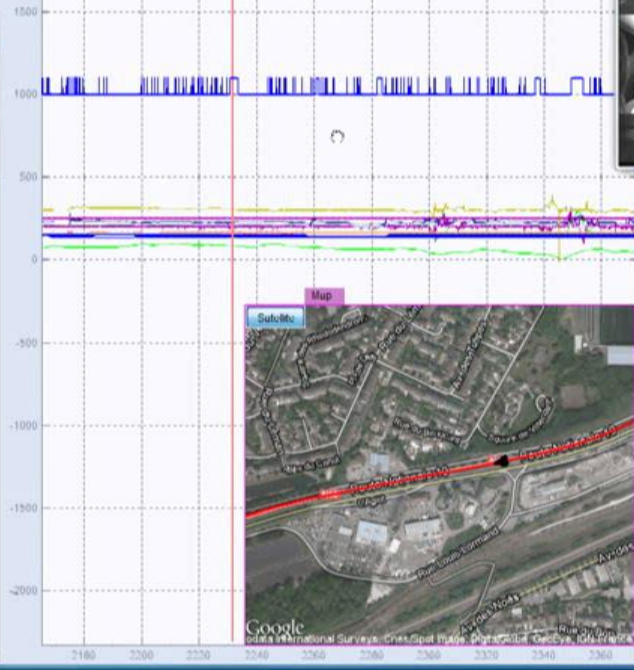
Logger	Config	Date
23100	LapHigh	24-Nov-2012 14:44
23106	LapHigh	16-Dec-2011 08:36
23100	LapHigh	21-Jun-2011 18:08
23100	LapHigh	22-Jun-2011 00:13
23101	LapHigh	25-Jun-2012 11:31
23105	CloHigh	01-Apr-2011 11:50
23106	LapHigh	03-Jan-2012 16:49
23100	LapHigh	18-Jun-2011 17:39
23112	CloHigh	24-Apr-2012 11:26
23106	LapHigh	03-Jan-2012 07:03
23101	LapHigh	09-May-2011 12:41
23112	CloHigh	24-Apr-2012 08:15
23106	LapHigh	27-Jun-2012 20:07
23103	LapHigh	01-Jun-2011 23:09
23106	LapHigh	05-Jan-2012 07:41
23101	LapHigh	09-May-2011 12:06
23105	CloHigh	08-Mar-2012 20:55
23100	LapHigh	01-Apr-2011 14:50
23100	LapHigh	01-Jul-2011 12:59
23105	CloHigh	08-Mar-2012 21:44
23100	LapHigh	16-Jul-2012 07:51
23100	LapHigh	16-Jul-2012 11:11
23100	LapHigh	01-Apr-2011 16:26

### MATLAB R2011b

Checking version of : Db  
Checking version of Prep  
Checking version of : S1  
Checking version of Prod  
Checking version of : So  
Checking version of : P  
-- No version was change

### Trip Viewer - 23100 - 01-Apr-2011 14:50:40, Driver: 0, Phase: 0, Vehicle: 1

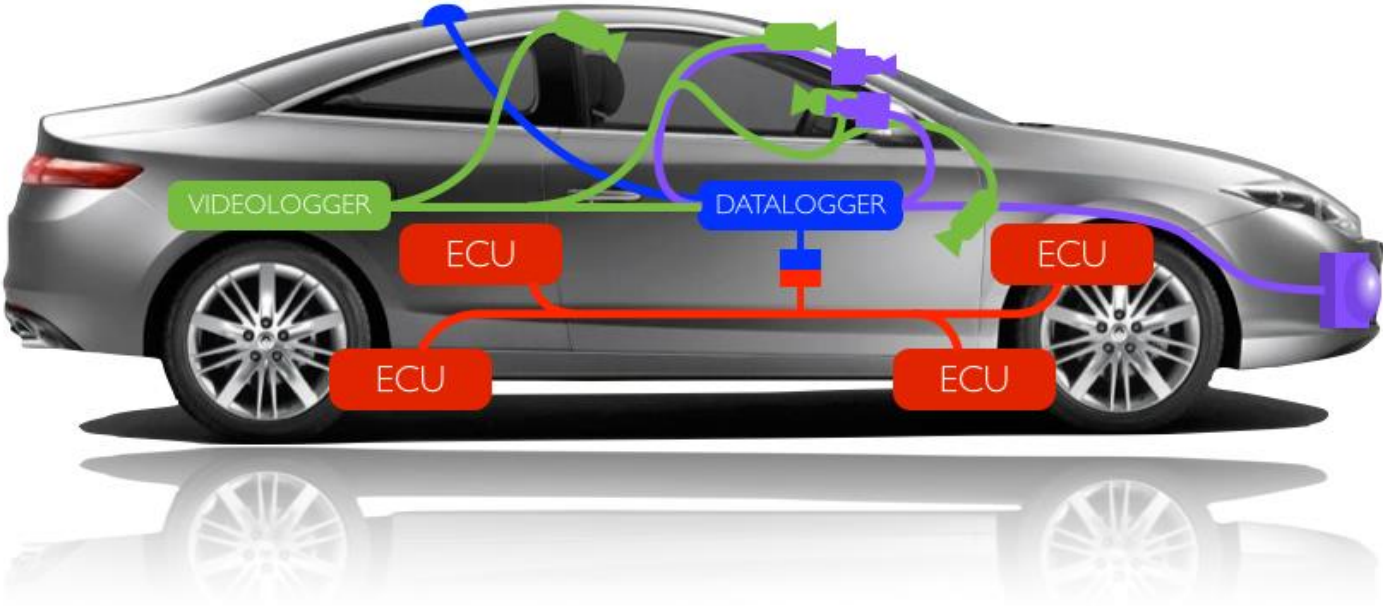
Signal	Name	U...	Gain	Offset	Color	Style	Insp
<input type="checkbox"/>	Veh_CC-SL_State		1	0			<input checked="" type="checkbox"/>
<input type="checkbox"/>	Veh_CC-SL_Speed	kmh	1	0			<input checked="" type="checkbox"/>
<input type="checkbox"/>	Veh_CC-SL_Overapud		1	0			<input type="checkbox"/>
<input checked="" type="checkbox"/>	Veh_Vehicle_Speed	kmh	1	0			<input checked="" type="checkbox"/>
<input type="checkbox"/>	Veh_Vehicle_Long_Accou	m/s <sup>2</sup>	1	0			<input type="checkbox"/>
<input type="checkbox"/>	Veh_Vehicle_Trans_Accel	g	1	0			<input type="checkbox"/>
<input type="checkbox"/>	Veh_Clutch_Switch_Max_Tr		1	0			<input type="checkbox"/>
<input type="checkbox"/>	Veh_Clutch_Switch_Travel		1	0			<input type="checkbox"/>
<input type="checkbox"/>	Veh_Braku_Info_Status		1	0			<input type="checkbox"/>
<input type="checkbox"/>	Veh_Throttle	nf...	1	0			<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	Veh_Steering_Wheel_Angle	dug	1	200			<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	Veh_Steering_Wheel_RSPEED	de...	1	200			<input checked="" type="checkbox"/>
<input type="checkbox"/>	Veh_Engine_RPM	Trf...	1	0			<input type="checkbox"/>
<input type="checkbox"/>	Veh_Fuel_Consumption	mm3	1	0			<input type="checkbox"/>
<input type="checkbox"/>	Veh_Front_Wipwr		200	0			<input type="checkbox"/>
<input type="checkbox"/>	Rad_Yaw_Rate	de...	1	0			<input checked="" type="checkbox"/>
<input type="checkbox"/>	Rad_RelativeVelocity	m/s	1	0			<input type="checkbox"/>
<input type="checkbox"/>	Rad_Range	m	1	0			<input type="checkbox"/>
<input type="checkbox"/>	Rad_Accokruton	m/s <sup>2</sup>	1	0			<input type="checkbox"/>
<input type="checkbox"/>	Rad_CorrectedAzimuth	dug	1	0			<input type="checkbox"/>
<input type="checkbox"/>	Rad_RelativeAzimuth	dug	1	0			<input type="checkbox"/>
<input type="checkbox"/>	Rad_LateralDistanceTrack	m	1	0			<input type="checkbox"/>
<input type="checkbox"/>	Mob_DistanceToLeftLaneMark	m	10	220			<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	Mob_DistanceToRightLaneMark	m	10	220			<input checked="" type="checkbox"/>
<input type="checkbox"/>	Mob_LeftLaneConfidence		10	0			<input type="checkbox"/>
<input type="checkbox"/>	Mob_RightLaneConfidence		10	0			<input type="checkbox"/>
<input checked="" type="checkbox"/>	Mob_roadCurvature	rc...	100	300			<input checked="" type="checkbox"/>
<input type="checkbox"/>	Mob_roadSlope	ra...	100	200			<input type="checkbox"/>
<input type="checkbox"/>	Prc_rl_Driver_noActivity		1	0			<input type="checkbox"/>
<input type="checkbox"/>	Prc_rl_cond_actif_ayv_obs		1	0			<input type="checkbox"/>
<input type="checkbox"/>	Prc_rl_min_nvk_d		1	0			<input type="checkbox"/>
<input type="checkbox"/>	Prc_rl_max_d		1	0			<input type="checkbox"/>
<input type="checkbox"/>	Prc_rl_min_d		1	0			<input type="checkbox"/>
<input type="checkbox"/>	Prc_EyeLdMerged	mm	1	0			<input type="checkbox"/>
<input type="checkbox"/>	Prc_EyeLdMergedQuality		10	0			<input type="checkbox"/>
<input type="checkbox"/>	Prc_Perloc80		100	0			<input type="checkbox"/>



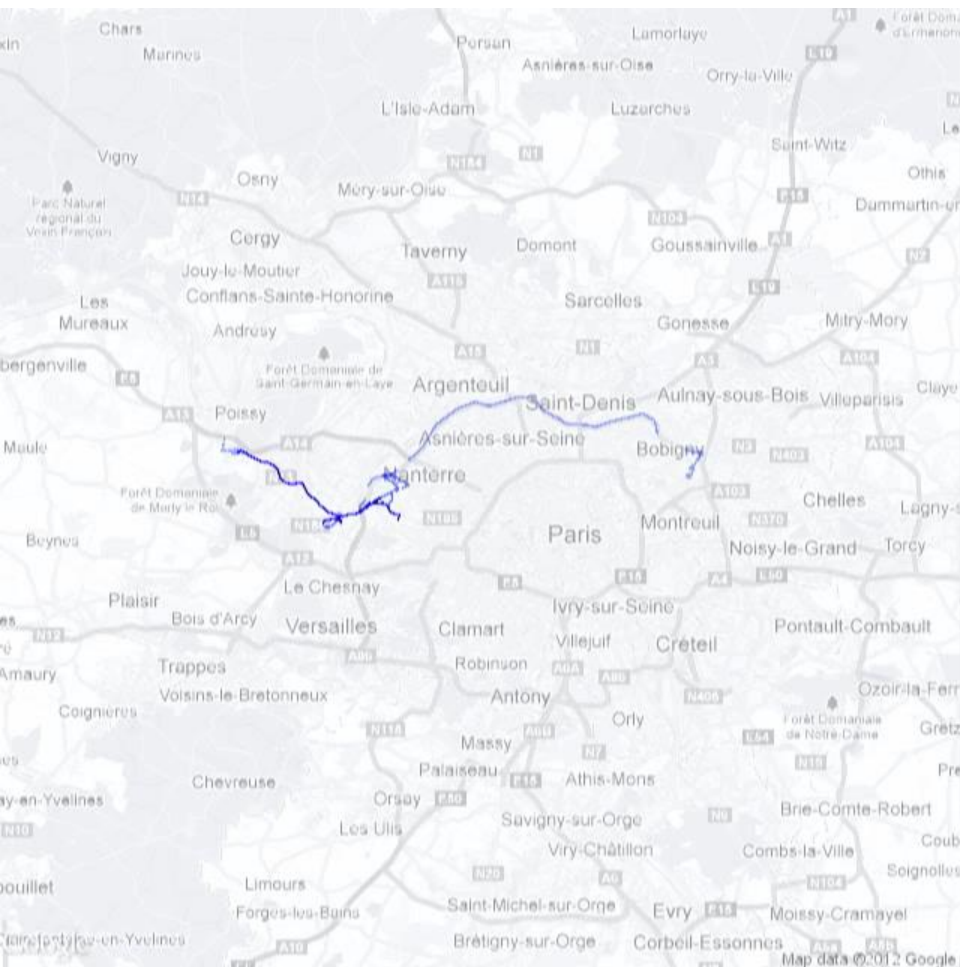
### Inspector

Time: 2231.9  
01-Apr-2011 15:27:52

Veh_CC-SL_State	CC-SL_off
Veh_CC-SL_Speed	254 kmh
Veh_Vehicle_Speed	84.6 kmh
Veh_Throttle	0.22 n/a (0-1)
Veh_Steering_Wheel_RSPEED	6.4 deg/s
Mob_DistanceToLeftLaneMark	-0.8 m
Mob_DistanceToRightLaneMark	0.75 m
Mob_roadCurvature	0.00041 1e6(2radius)
Map_SpeedLimit	90 kmh
St_RoadType	Urban
Prc_rl_Stid_act	1
Prc_rl_DAA_enVhPosWlyMth	-0.15
Prc_rl_DAA_endVhPosWlyMth	0.26
Prc_rl_coupe_virage	0
Prc_rl_chst_vole_rl_en_cours	0
Prc_rl_chst_vole_d_en_cours	0
Prc_GazePitchMerged	Nali rad
Prc_GazeYawMerged	Nali rad
Prc_GazeRollMerged	Nali rad
Prc_GazeQualityMerged	Nali
St_Daylight	Day Time
St_Weather	No rain
Eye_BlinkDuration	3000 ms



**IN WHICH QUANTITY?**



Map data ©2012 Google



Map data ©2012 GeoBasis-DE/BKG (©2009), Google, Tele Atlas

**BUT WHY ?**

# Two Reasons...

- **Road safety research**

- Less, more diverse accidents, complex factors
- Obvious safety measures already taken
- New challenges (distraction, vulnerable road users)

→ Transition from *accidentology* to *incidentology*.



# Two Reasons...

- **Industrial development** of Advanced Driving Assistance Systems and Autonomous Vehicles
  - Continuous action of system, depending on real-time interpretation of a complex environment from sensor measures
  - Extremely rare events may have disastrous impact (black swan theory)
  
- Necessity to characterize driving habits

# ...To invent a new methodology

Experimental approach

Epidemiology



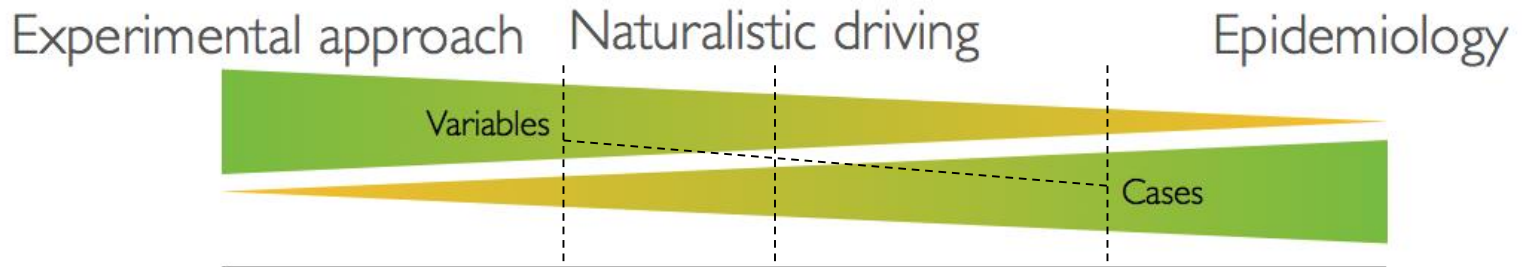
- Can be used early in development
- Can answer **why ? how ? what if... ?**

- Limited situations :
  - ▶ Limited statistical representativeness
  - ▶ Can be blind to unexpected effects

- Assess **real-world** efficiency

- Mostly just descriptive
- **After** innovation has become very common.

# ...To invent a new methodology



- Hundreds of cars, millions of kilometers
- Participants carry out their usual trips
- Open road, minimal instructions and experimental control
- Extensive but inconspicuous and autonomous instrumentation
- Continuous data acquisition

# ...To invent a new methodology

Experimentation

Naturalistic driving

Epidemiology

- Benefits
  - ▶ Observation of ordinary behaviour
  - ▶ Statistical power and representativeness
  - ▶ Rare and unplanned events are recorded

- Challenges :
  - ▶ Vehicle instrumentation
  - ▶ Organisation and logistics
  - ▶ Legal & Ethical constraints
  - ▶ **Data processing and management**

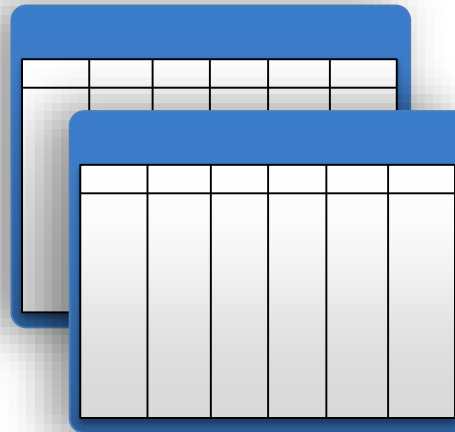
**SO... WHAT DO WE DO WITH THAT  
DATA?**

# THE Process...

« Raw data »



« Aggregated data »



Attributes

- Not too many lines
- Comparable statistical individuals
- In most cases, one line is one sub-part of a trip

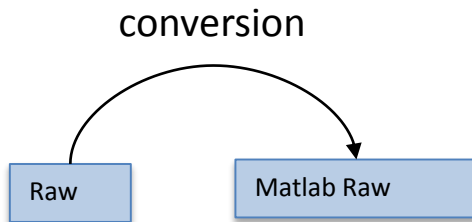
# ...In more details

Per-trip processing

Raw

# ...In more details

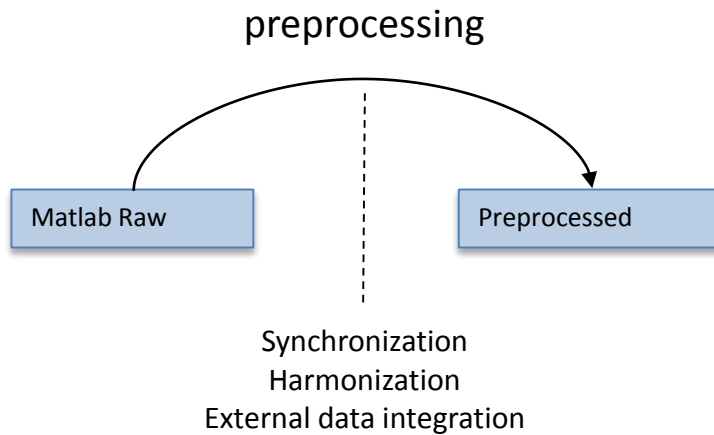
Per-trip processing





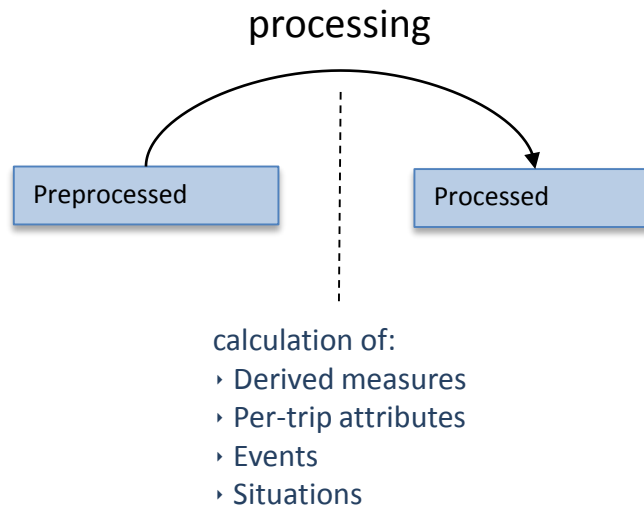
# ...In more details

## Per-trip processing



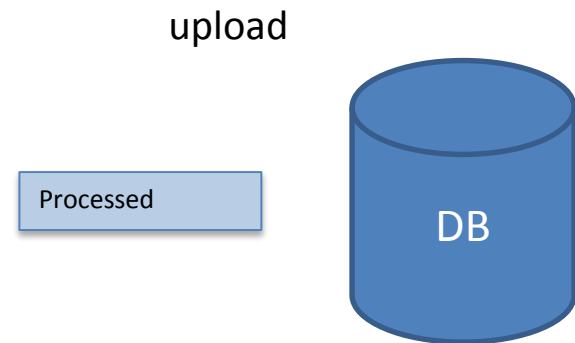
# ...In more details

## Per-trip processing



# ...In more details

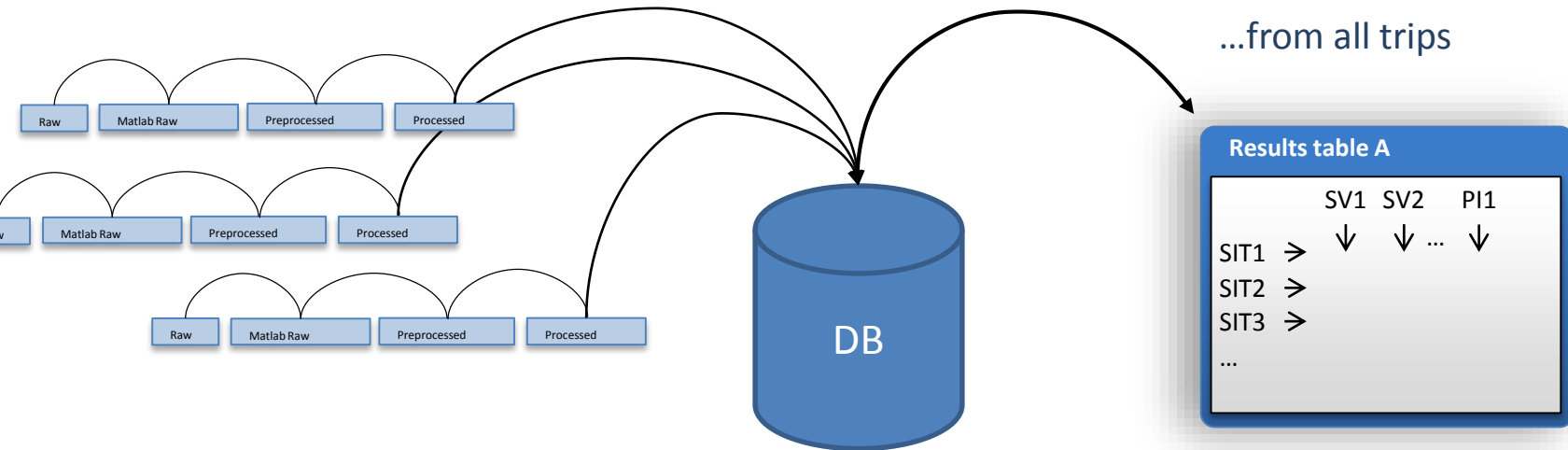
Per-trip processing



# ...In more details

Per-trip processing

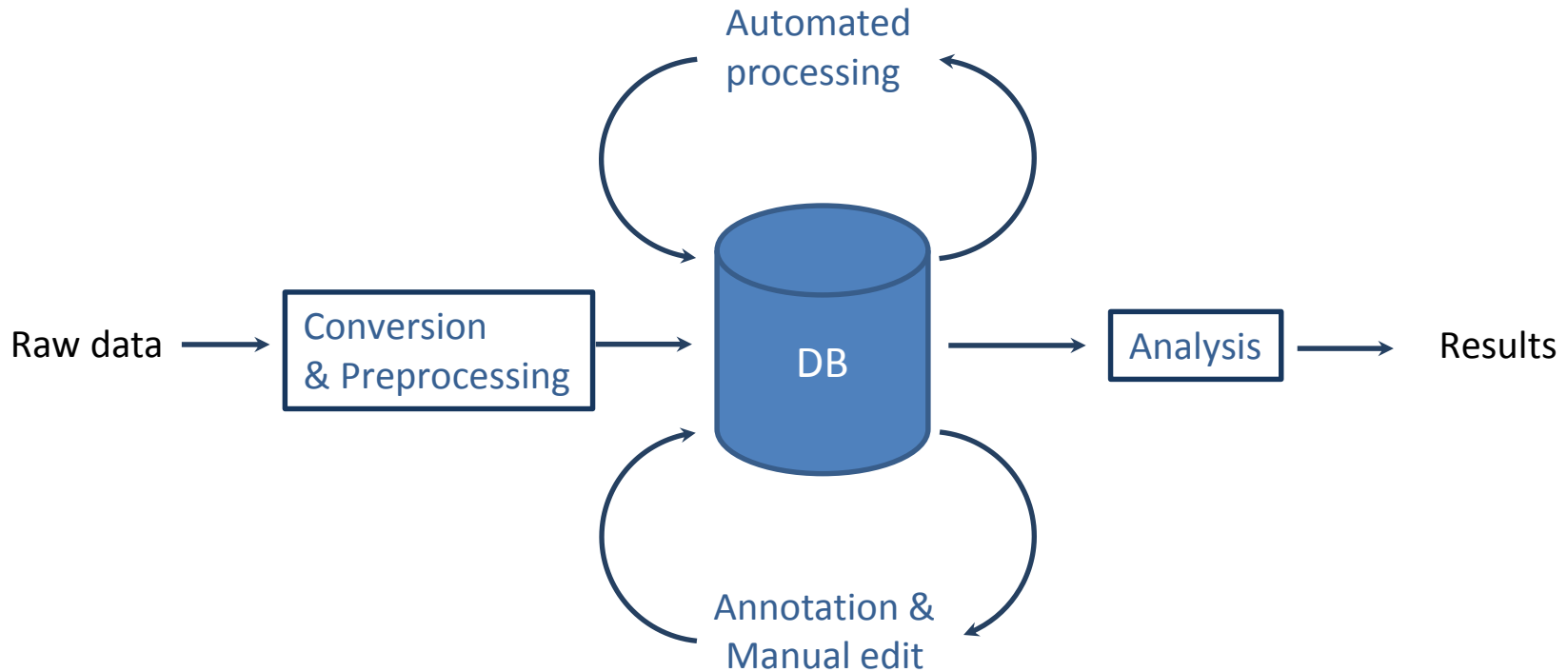
Transversal analysis



**LOOKS A LITTLE BIT TOO EASY...**

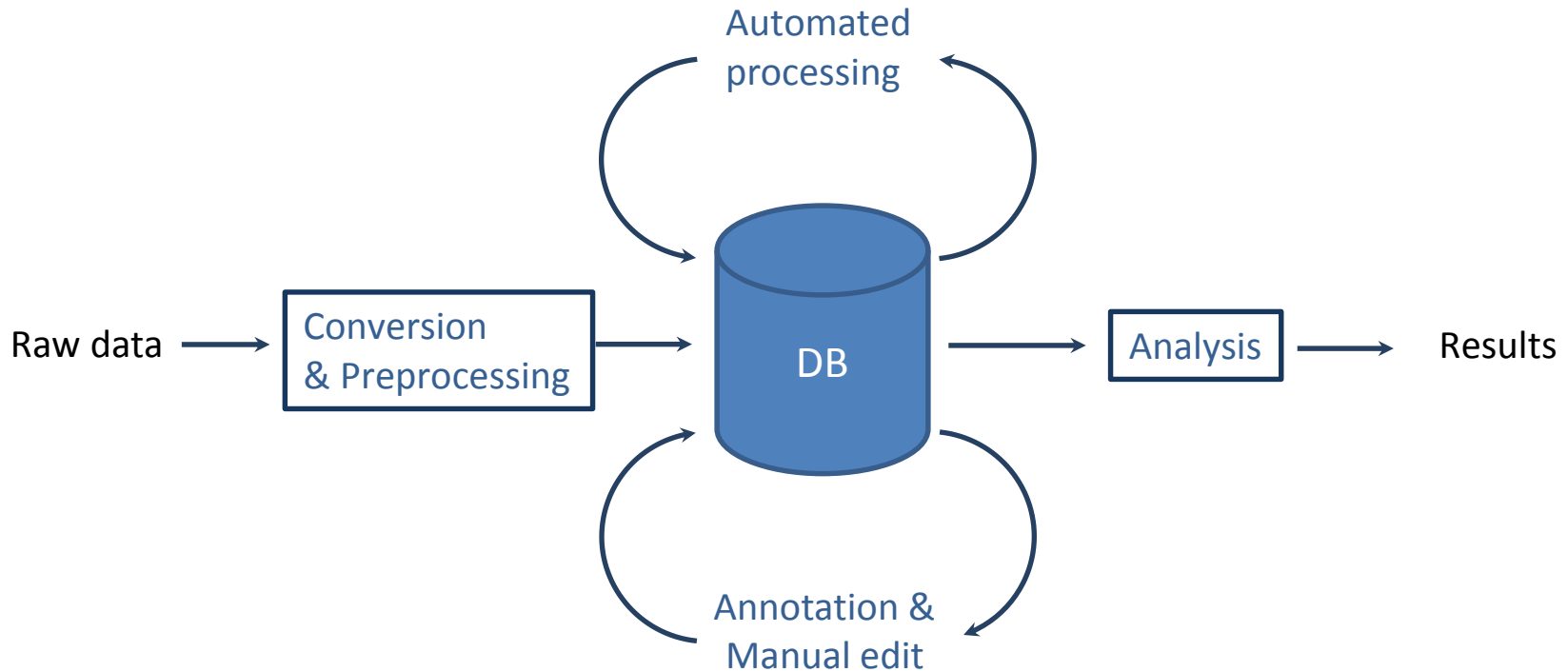
# Challenges

No such a linear process...



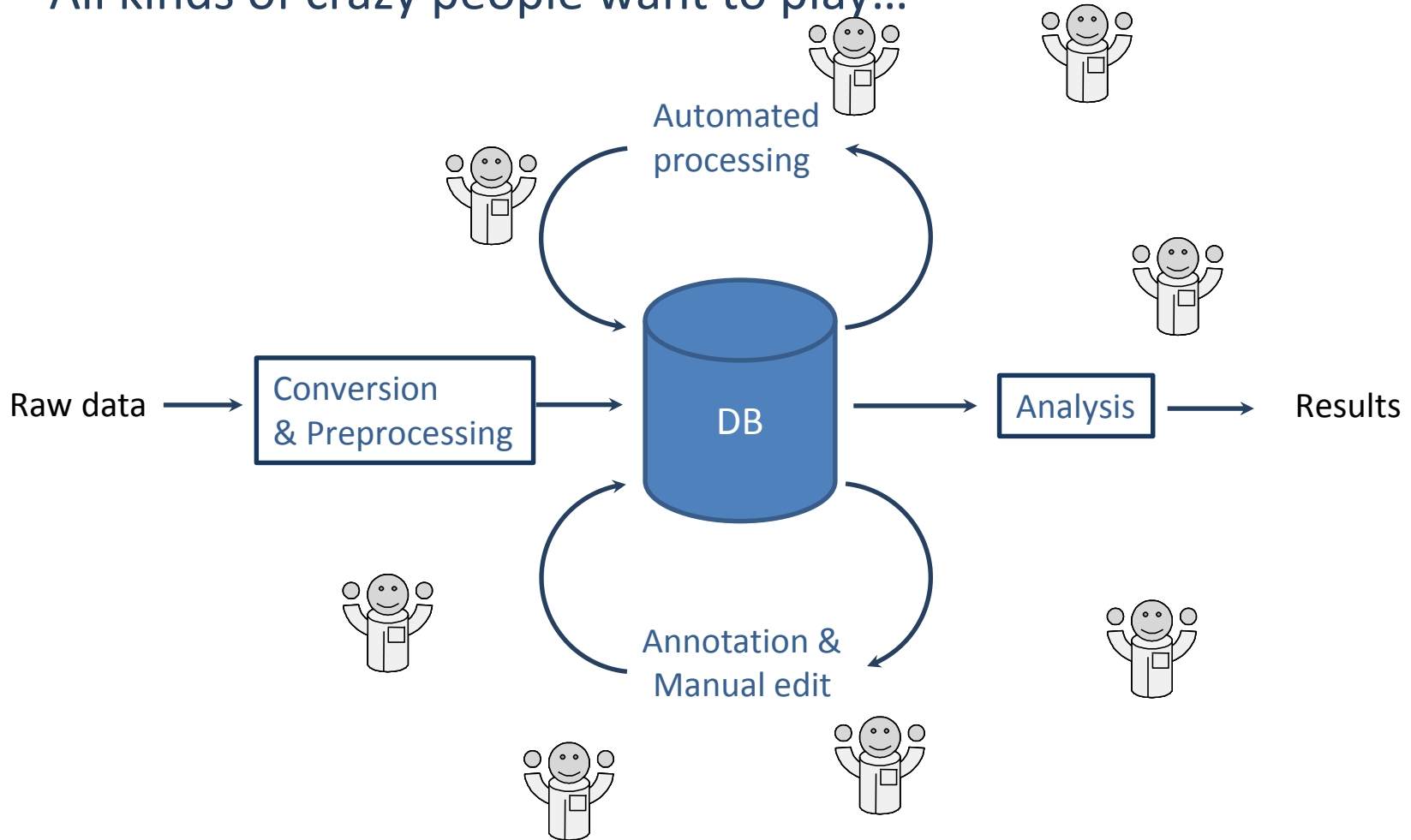
# Challenges

Dataset is huge... and unpredictable !



# Challenges

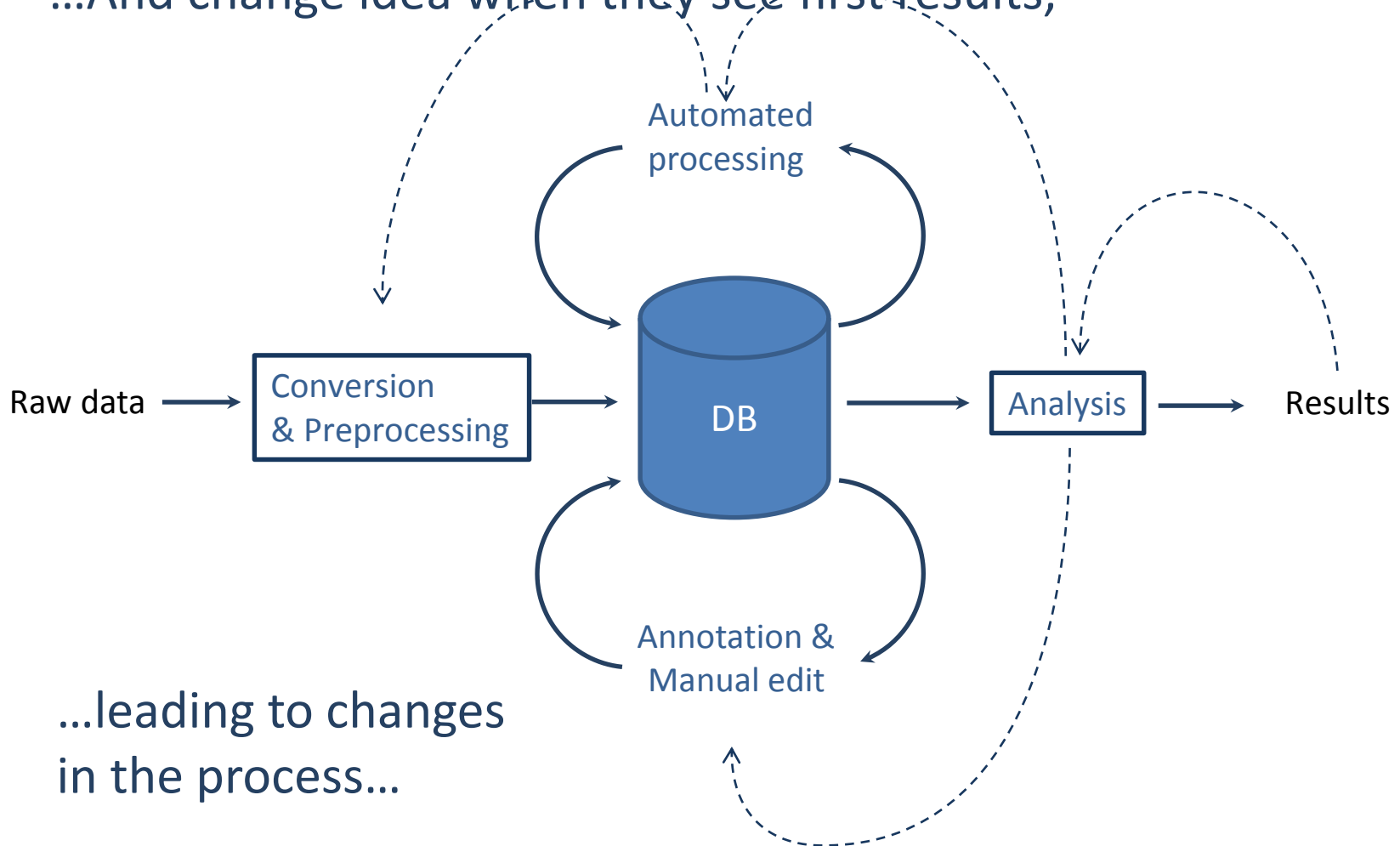
All kinds of crazy people want to play...





# Challenges

...And change idea when they see first results,



# Challenges

While new data continues to be acquired !



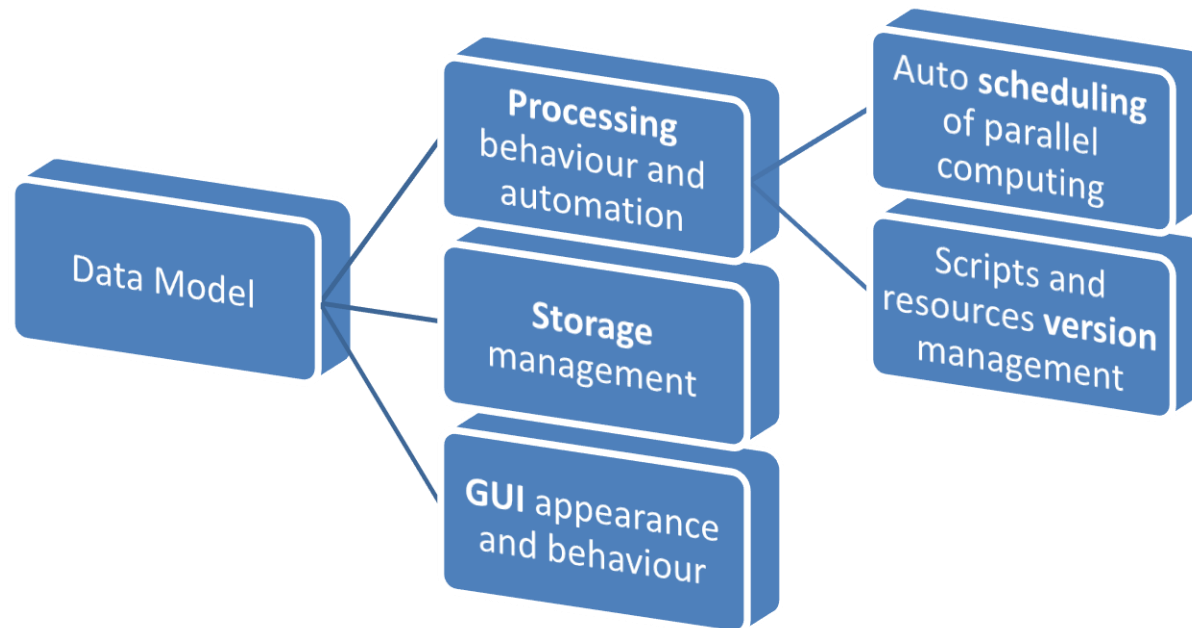
**HOW TO ADDRESS THEM?**

# DRIVEWARE

- An integrated framework & tool taking care of:
  - Data management
  - Indexing and research
  - Harmonization and synchronization
  - External data integration and algorithmic enrichment
  - Reduction
  - Visualization
  - Annotation
- Dataflow management system for large scale, driving data batch processing
- In MATLAB

# DRIVEWARE

- It's entire behaviour is based on the description of the data (a.k.a as *data model*) and the corresponding user algorithms





User



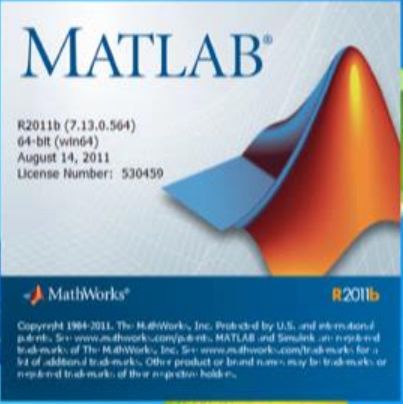
Ordinateur



Bordelites

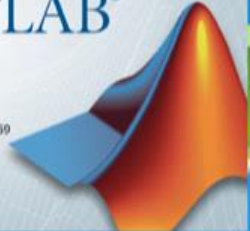




DriveWare



**MATLAB®**

R2011b (7.13.0.564)  
64-bit (win64)  
August 14, 2011  
License Number: 530459



 MathWorks®  R2011b

Copyright 1984-2011. The MathWorks, Inc. Protected by U.S. and international patents. See [www.mathworks.com/patents](http://www.mathworks.com/patents). MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See [www.mathworks.com/trademarks](http://www.mathworks.com/trademarks) for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.



Corbeille





```
MATLAB R2011b  
File Edit Debug Desktop Window Help  
ForceNoColorFeedback = true  
Execution mode found is: Parallel  
Experiment options:  
HasOEMSupport = 0  
ZipPreviousLog = 0  
KeepPreviousLog = 1  
Loading local database...
```

Trips Explorer (Process debugging) ---

Available: 46742, Filtered: 6156, Selected: 0

Logger	Config	Date	Driver	Phase	Length	Errors	Select
23100	LugHigh	03-Feb-2011 19:23:53	2	1	00:30, 13 km		
23100	LugHigh	04-Feb-2011 05:27:00	2	1	00:02, 0 km		
23100	LugHigh	04-Feb-2011 05:32:39	2	1	00:34, 27 km		
23100	LugHigh	04-Feb-2011 13:16:48	2	1	00:02, 0 km		
23100	LugHigh	04-Feb-2011 13:41:55	2	1	00:01, 0 km		
23100	LugHigh	04-Feb-2011 18:26:49	2	1	00:32, 22 km		
23100	LugHigh	04-Feb-2011 20:21:17	2	1	00:19, 13 km		
23100	LugHigh	05-Feb-2011 11:17:07	2	1	00:20, 15 km		
23100	LugHigh	05-Feb-2011 11:56:17	2	1	00:18, 15 km		
23100	LugHigh	05-Feb-2011 17:54:07	2	1	00:10, 3 km		
23100	LugHigh	05-Feb-2011 18:12:27	2	1	00:06, 0 km		
23100	LugHigh	05-Feb-2011 18:23:07	2	1	00:16, 10 km		
23100	LugHigh	05-Feb-2011 19:14:14	2	1	00:04, 0 km		
23100	LugHigh	05-Feb-2011 19:29:03	2	1	00:04, 0 km		
23100	LugHigh	05-Feb-2011 13:14:14	2	1	00:29, 23 km		
23100	LugHigh	06-Feb-2011 17:50:14	2	1	00:29, 22 km		
23100	LugHigh	06-Feb-2011 18:20:49	2	1	00:05, 2 km		
23100	LugHigh	07-Feb-2011 05:05:38	2	1	00:36, 27 km		
23100	LugHigh	07-Feb-2011 12:49:49	2	1	00:02, 0 km		
23100	LugHigh	07-Feb-2011 13:13:25	2	1	00:02, 0 km		
23100	LugHigh	07-Feb-2011 18:26:46	2	1	00:37, 25 km		
23100	LugHigh	08-Feb-2011 05:12:37	2	1	00:37, 27 km		
23100	LugHigh	08-Feb-2011 18:09:50	2	1	00:36, 26 km		

Welcome to DriveWare beta

Trips Explorer  
List trips, reproduce them,  
Filter them according to search criteria

Experiment Builder  
Define new data sources, link algorithms,  
Manage the database

Trip Data Viewer  
Display aggregate & aggregated data from  
one trip

Experiments  
EuroFOT  
Hidden Projects  
MairieParis

Window on top  Out MATLAB with application





Trips Explorer (Process debugging) ---

Available: 46742, Filtered: 119, Selected: 0

Loggur	Config	Date	Driver	Phusu	Length	Errors	Subject
23100	LugHigh	22-May-2012 18:38:48	33	2	00:49, 28 km		
23100	LugHigh	27-May-2012 10:40:35	34	2	00:12, 58 km		
23100	LugHigh	27-May-2012 14:27:37	34	2	00:25, 51 km		
23100	LugHigh	27-May-2012 15:13:13	34	2	00:36, 206 km		
23100	LugHigh	28-May-2012 17:58:24	34	2	00:32, 23 km		
23100	LugHigh	30-May-2012 16:11:51	34	2	00:32, 24 km		
23100	LugHigh	30-May-2012 22:09:40	34	2	00:46, 35 km		
23100	LugHigh	01-Jun-2012 11:51:07	34	2	00:30, 24 km		
23100	LugHigh	01-Jun-2012 14:58:26	34	2	00:44, 302 km		
23100	LugHigh	01-Jun-2012 16:50:04	34	2	00:16, 189 km		
23100	LugHigh	02-Jun-2012 13:41:27	34	2	01:05, 91 km		
23100	LugHigh	02-Jun-2012 15:02:44	34	2	00:24, 23 km		
23101	LugHigh	17-Apr-2012 07:31:04	0	0	01:08, 27 km		
23101	LugHigh	22-Apr-2012 10:38:46	30	2	00:33, 31 km		
23101	LugHigh	22-Apr-2012 16:45:33	30	2	00:36, 29 km		
23101	LugHigh	27-Apr-2012 21:04:55	30	2	00:36, 193 km		
23101	LugHigh	01-May-2012 16:04:36	30	2	00:36, 280 km		
23101	LugHigh	02-May-2012 06:29:08	30	2	00:30, 20 km		
23101	LugHigh	05-May-2012 13:43:43	27	2	00:49, 90 km		
23101	LugHigh	05-May-2012 14:48:21	27	2	01:06, 125 km		
23101	LugHigh	05-May-2012 15:58:37	27	2	01:13, 122 km		
23101	LugHigh	09-May-2012 14:26:56	27	2	01:15, 120 km		
23101	LugHigh	09-May-2012 15:11:50	27	2	00:51, 108 km		

```
ForceNoColorFeedback = true
Execution mode found is: Parallel
Experiment options:
HaaOEMSupport = 0
ZipPreviousLog = 0
KeepPreviousLog = 1
Loading local database...
```

Welcome to DriveWare beta

Experiments

- EuroFOT
- Hidden Projects
- MairieParis

Trips Explorer  
List trips, reproduce them,  
Filter them according to search criteria

Experiment Builder  
Define new data umbra, link algorithms,  
Manage the database

Trip Data Viewer  
Display agrupta & aggregated data from  
one trip

Window on top  Out MATLAB with application



```
MATLAB R2011b  
File Edit Debug Desktop Window Help  
Checking version of Conversion parameters  
Checking version of : DbcFiles  
Checking version of Preprocessing parameters  
Checking version of : Signals configuration  
Checking version of Processing parameters  
Checking version of : Scripts  
Checking version of : Process graph model
```

Trips Explorer

Logbook

Logbook	Config	Date	Driver	Phasu	Lun
23100	LanHigh	16-May-2011 10:07:34	0	0	00:00,0
23100	LanHigh	30-May-2011 10:15:17	0	0	00:03,1
23100	LanHigh	31-May-2011 08:39:24	0	0	00:01,0
23100	LanHigh	31-May-2011 11:38:45	0	0	00:03,1
23100	LanHigh	31-May-2011 11:43:17	0	0	00:10,4
23100	LanHigh	31-May-2011 16:21:49	0	0	00:07,1
23100	LanHigh	31-May-2011 16:30:51	0	0	00:11,4
23100	LanHigh	31-May-2011 17:20:31	0	0	00:02,0
23100	LanHigh	31-May-2011 17:27:31	0	0	00:06,1
23100	LanHigh	17-Jun-2011 15:29:36	0	0	00:02,0
23100	LanHigh	17-Jun-2011 17:50:36	0	0	00:01,0
23100	LanHigh	17-Jun-2011 18:39:44	0	0	00:08,2
23100	LanHigh	18-Jun-2011 11:52:21	0	0	01:11,2
23100	LanHigh	18-Jun-2011 17:39:13	0	0	00:40,4
23100	LanHigh	18-Jun-2011 23:30:44	0	0	01:11,2
23100	LanHigh	20-Jun-2011 10:03:23	0	0	00:10,2
23100	LanHigh	20-Jun-2011 10:16:08	0	0	00:05,1
23100	LanHigh	21-Jun-2011 08:53:53	0	0	00:04,1
23100	LanHigh	21-Jun-2011 09:21:45	0	0	00:10,2
23100	LanHigh	21-Jun-2011 10:54:20	0	0	00:01,0
23100	LanHigh	21-Jun-2011 11:09:01	0	0	00:06,2
23100	LanHigh	21-Jun-2011 11:17:20	0	0	00:09,2
23100	LanHigh	21-Jun-2011 17:26:13	0	0	00:16,2

Experiment Builder

Data Objects

- Signal Tables
- Signals
- Events
  - EV\_stg\_attention
    - Generator Inputs
    - Generator Parameters
    - Refresher Inputs
    - Refresher Parameters
    - Attributes
    - Sub Events
  - EV\_std\_niveau\_attention
  - EV\_std\_niveau\_attention\_win
  - EV\_BaisseAttention**
    - Generator Inputs
    - Generator Parameters
    - Refresher Inputs
    - Refresher Parameters
    - Attributes
    - Sub Events
- Situations
- Situations Calculations
- Trip Attributes
- Processes
- Drivers Attributes
- Vehicles Attributes

Events: EV\_BaisseAttention  
TripOutputModuleEvent(4)

Name	EV_BaisseAttention
Description	
Trips Per Partition in DB table	1000
PursuitFunc	<input checked="" type="checkbox"/>
Generator Script File	ALY_EV_GeneratorBaisseAttention.m
Refresher Script File	ALY_EV_RefresherBaisseAttention.m
Event Instance Input Variable	Instancu

Edit Generator Script File  
Browse Generator Versions  
Edit Refresher Script File  
Browse Refresher Versions

Delete this Item



# DRIVEWARE

- Pros
  - Takes care of computer science challenges
  - Secures data (no direct access to original data)
  - Enforces structured approach
  - Readily implements usual tasks (visualization, filtering...)
  - Facilitates data access
  - Facilitates reuse of data
  - Facilitates reuse of algorithms
  - Avoids unnecessary recalculations
  - Very generic and adaptable
- Limits
  - Does not contain « scientific » algorithms. These have to be implemented by users.
  - Handles each record independently (allows parallelization)
  - Not « everything » can be done within this environment. Only continuous data enrich/reduce.

# SALSA

A Large Scale Driving Data Analysis Tool

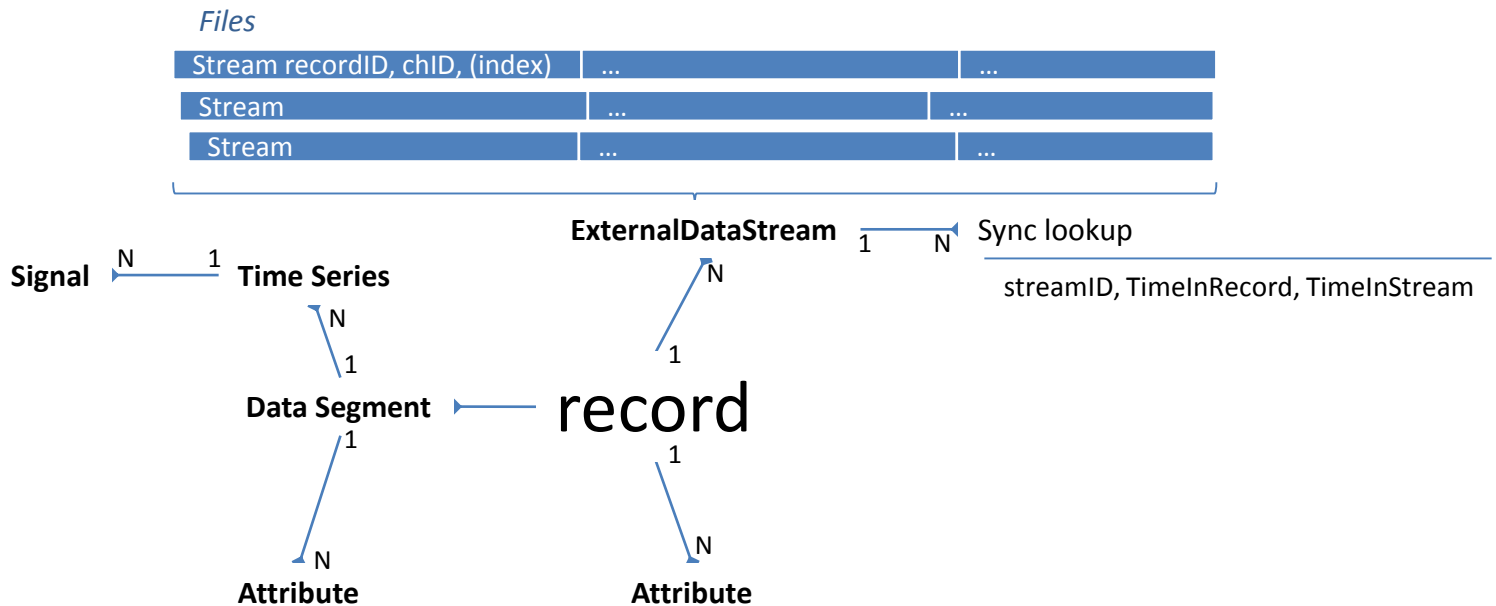
# SALSA

A General Purpose, Large Scale Data Reduction Tool

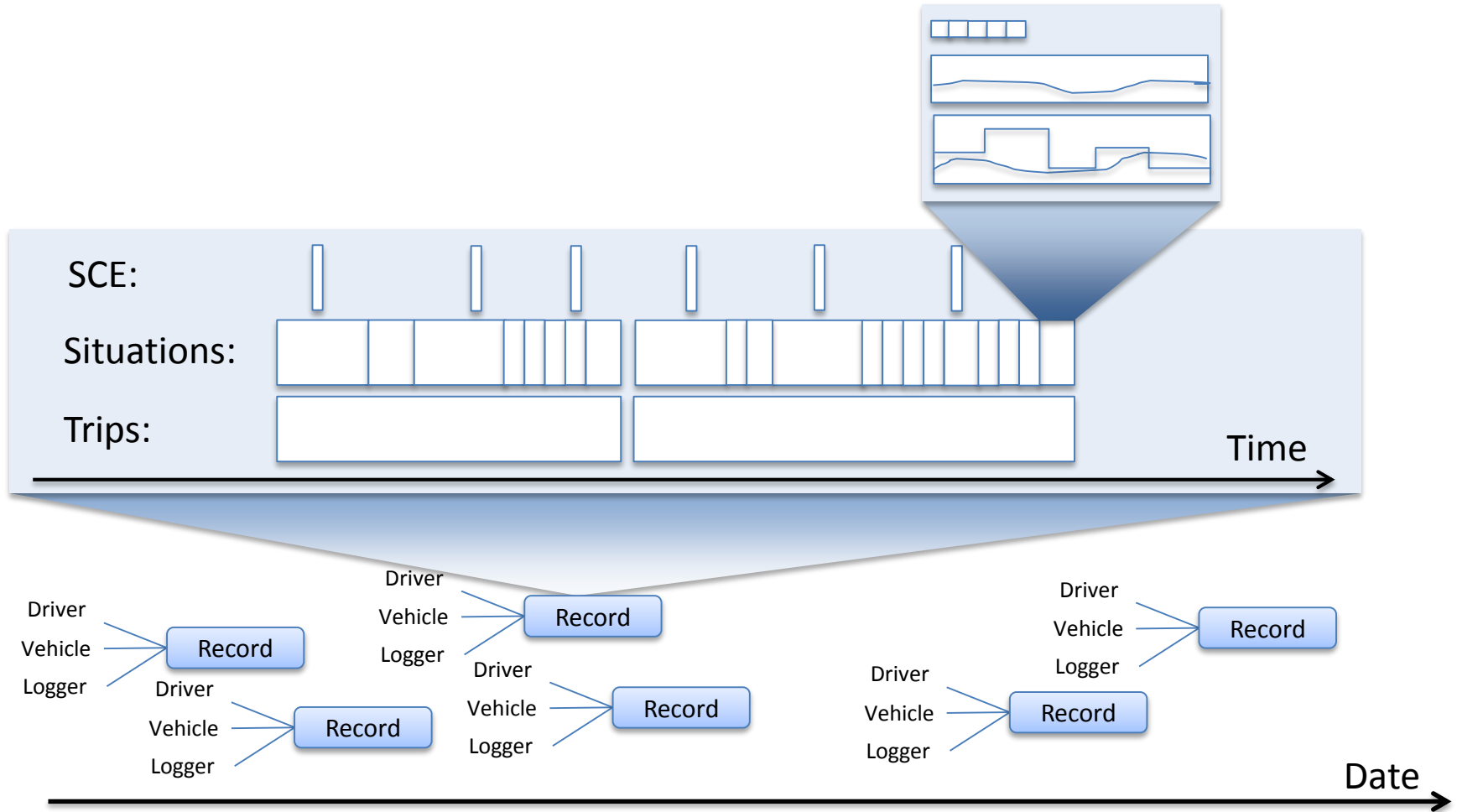
# SALSA

- Same approach as DRIVEWARE
- **Is general purpose**
- Much more open and modular architecture
- Much better collaboration approach, users management and large scale annotation tools
- Integrates 5 years of user experience feedback
- Will be battle tested within UDRIVE project (<http://www.udrive.eu>)
- Still integrates within MATLAB

# SALSA's DATA MODEL



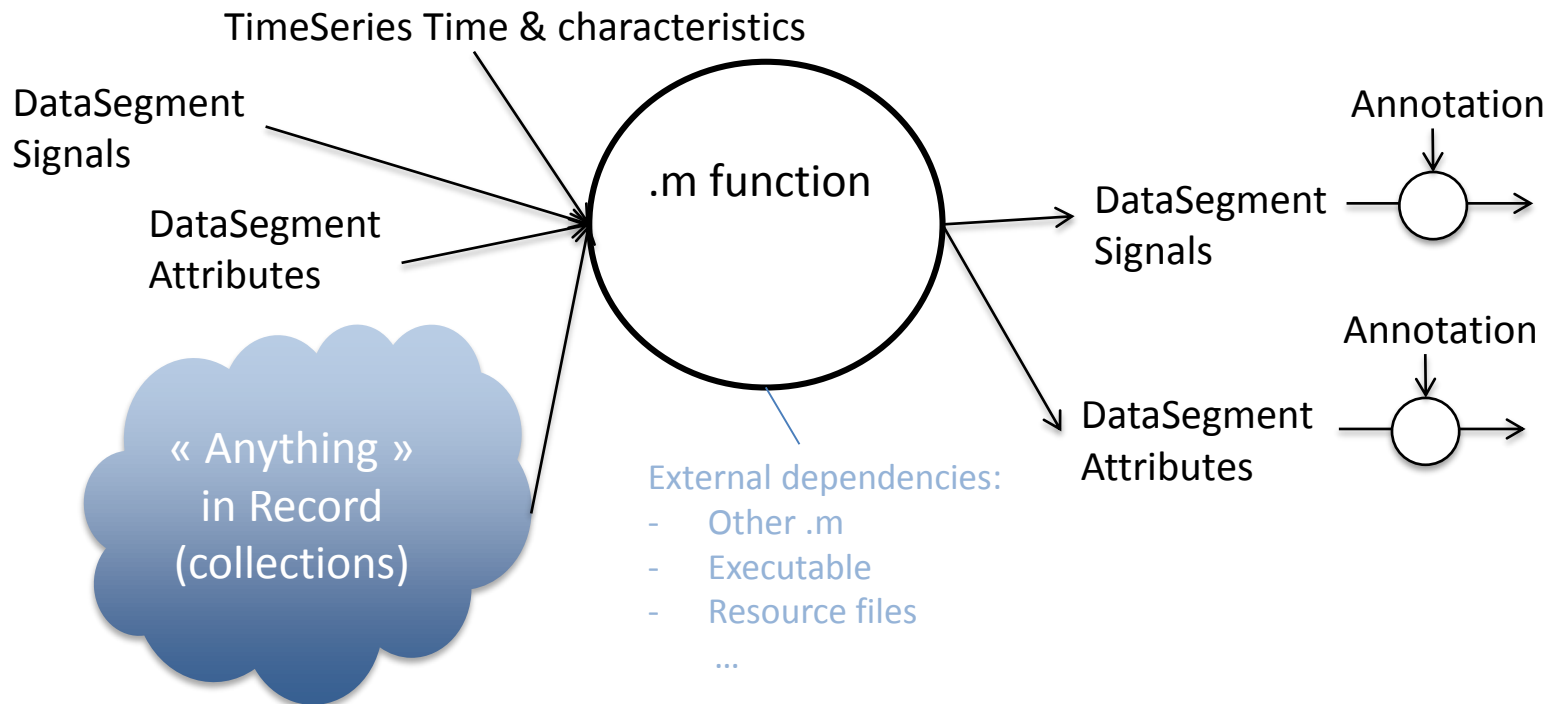
# SALSA's DATA MODEL





# Algorithms

- Contained in nodes
- Executed on each instance of a DataSegment
- Possibility to select a reference time, with on-the-fly resampling



**Demo planned early/spring 2016**

clement.val@ceesar.fr

**ANY MORE QUESTION?**

# Pourquoi Matlab ?

- Utilisé par la majorité de nos utilisateurs pour le traitement de leurs données
- Calcul parallèle très facilement mis en place (PCT)
- Intégration avec l'IDE
- « A peu près » apte au développement d'une telle application (OOP)
- Néanmoins, avec plus de temps, on aurait sans doute développé une appli plus généraliste présentant, entre autre, une interface vers MATLAB.

# Pourquoi MariaDB ?

- Coût d'entrée
- Familiarité avec MySQL
- Convaincus par l'équipe consulting de SkySQL
- Intégration de TokuDB
- Ouverture :
  - Sharding (spider)
  - Connect engine
  - Dynamic columns
  - Cassandra access

# TokuDB

- Compatibilité
- Indexation rapide
- Compression
- Ecriture/update rapide
- Changement du schéma 'en ligne'