Can auto-text recognition software for coding injuries replace manual coding?

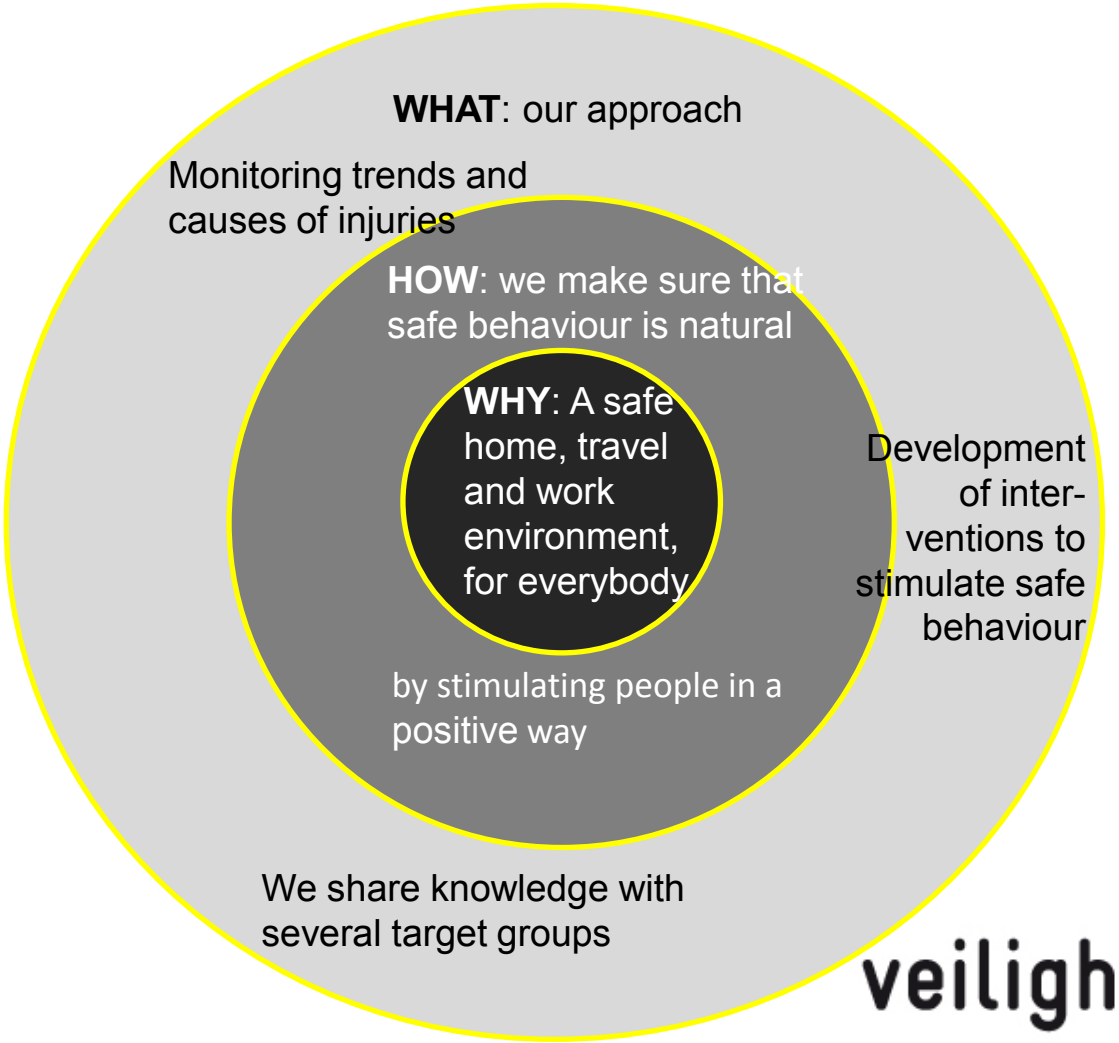
Findings from IDB/DISS in the Netherlands

Susanne Nijman, Consumer Safety Institute, Amsterdam, the Netherlands
(s.nijman@veiligheid.nl)

Birgitte Blatter, Consumer Safety Institute, Amsterdam, the Netherlands

What is VeiligheidNL?

Expertise center for safe behaviour in a safe environment



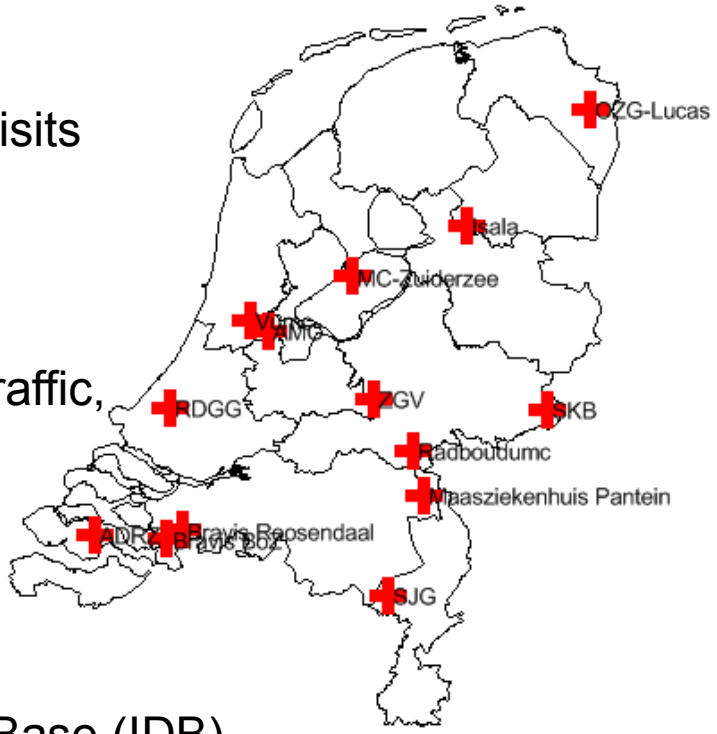
veiligheid nl

Dutch Injury Surveillance System (DISS)

- Since 1997
- Registration of injuries at EDs
- Representative, 14 of 87 EDs, 11% of visits
- Extrapolation to national figures
- Injuries/intoxications:

  cause (home and leisure, work, sport, traffic,

  violence, self-harm) + reasons



- Annual upload to European Injury DataBase (IDB)

veiligheid nl

Background

- Until a few years ago all variables in IDB/DISS (such as injury mechanism, product involved, type of injury and body part involved) were coded manually by the staff of the ED

- To reduce the administrative burden on EDs we developed a system for automatic text recognition software

veiligheid nl

What do we ask from EDs?

Information that is already registered in their own Hospital Information System:

- Personal characteristics: age, gender and postal code
- Diagnosis
- Hospitalization yes/no

Additional information in open text fields (integrated in their hospital information system) for all injuries and intoxications:

- What happened?
- Where/when dit it happen?
- What products were involved?

veiligheid nl

Aim

Can automatic text recognition software for coding injuries replace manual coding?

Examples:

1. Car driver, collision against tree, high **speed** accident
    - Desired output:
        - Injury mechanism: contact with object
        - Products involved: car, tree

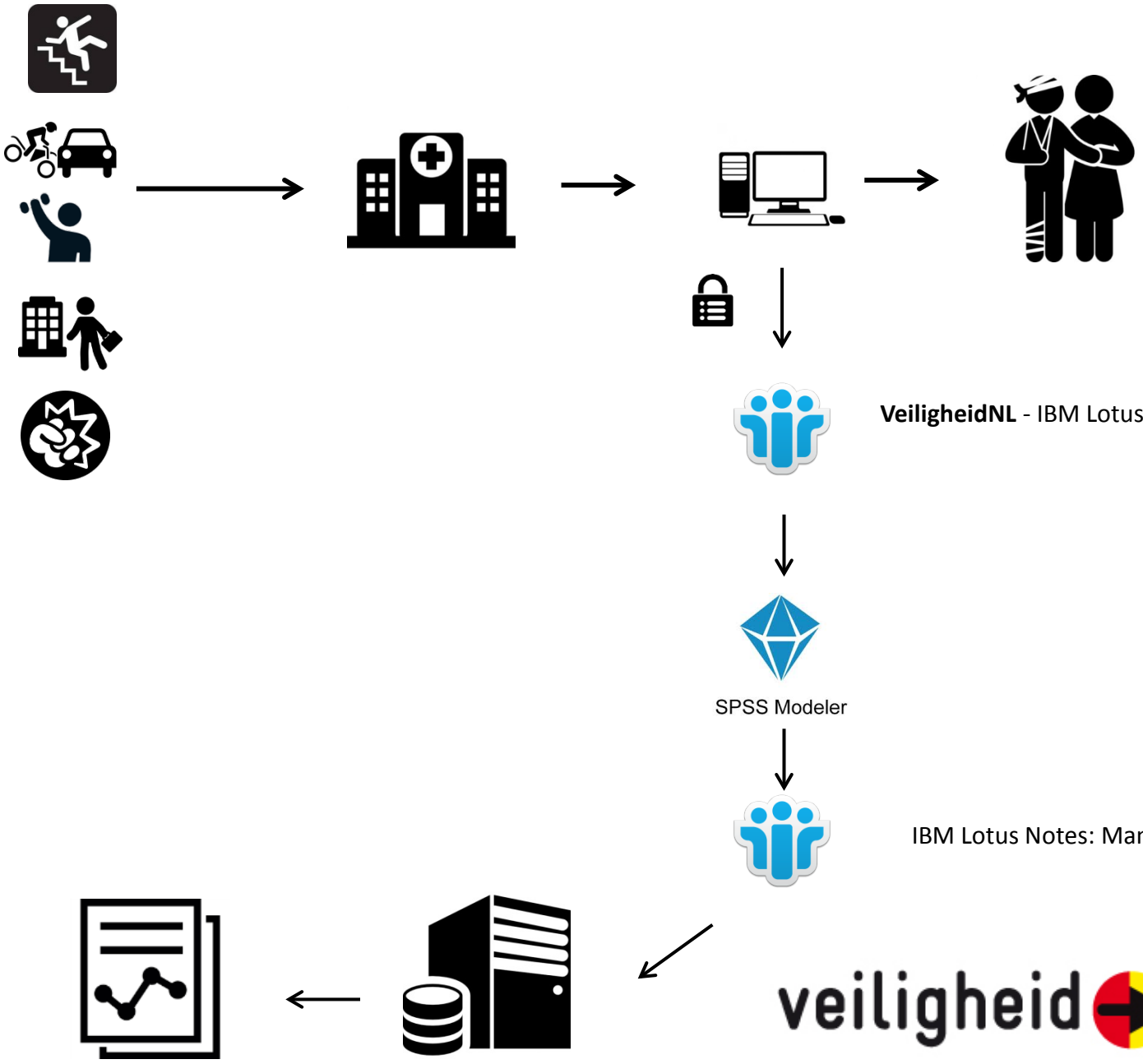
2. Patient found intoxicated, used alcohol and **speed**
    - Desired output:
        - Injury mechanism: chemical mechanism
        - Products involved: alcohol, speed

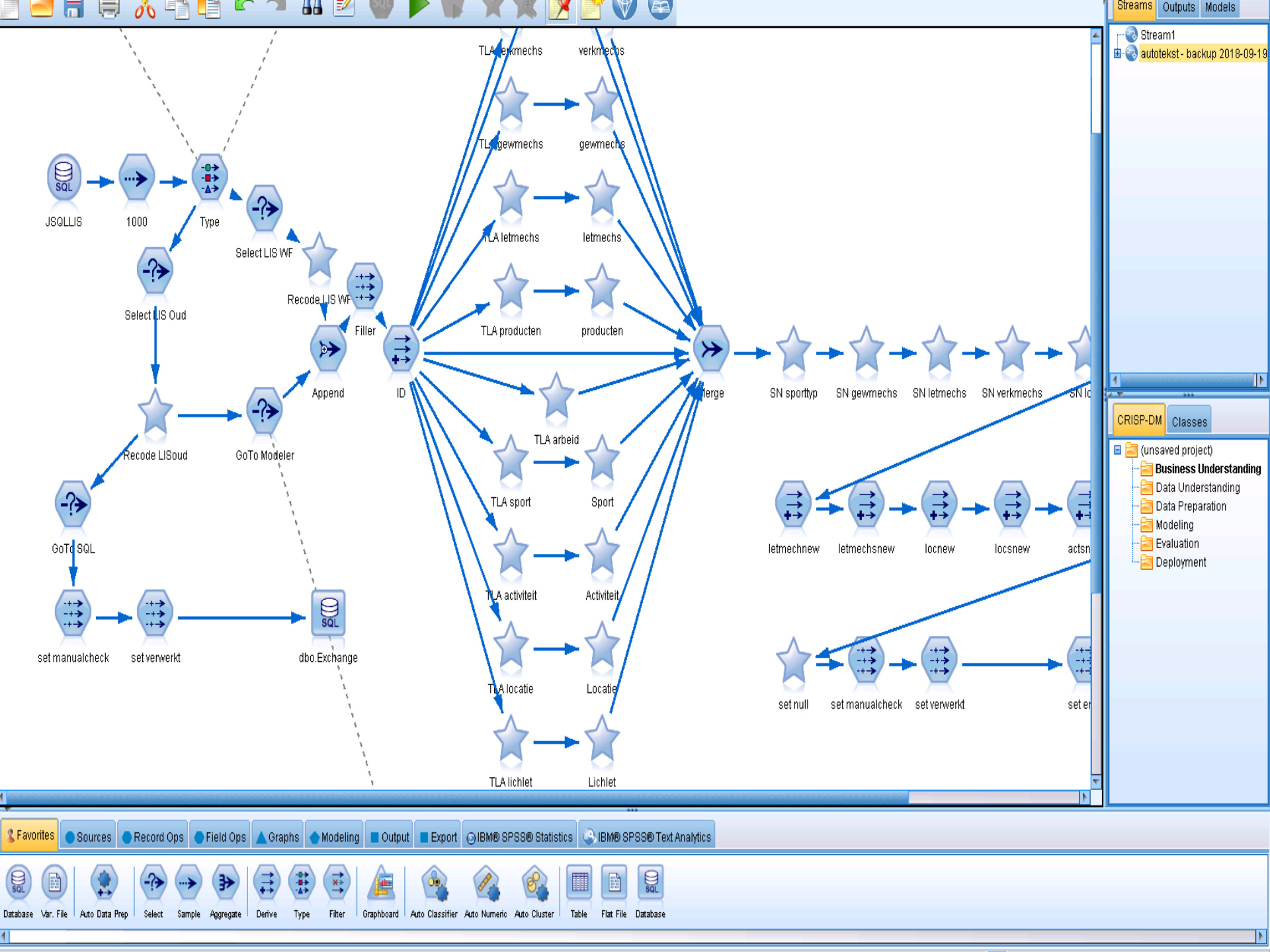
veiligheid nl

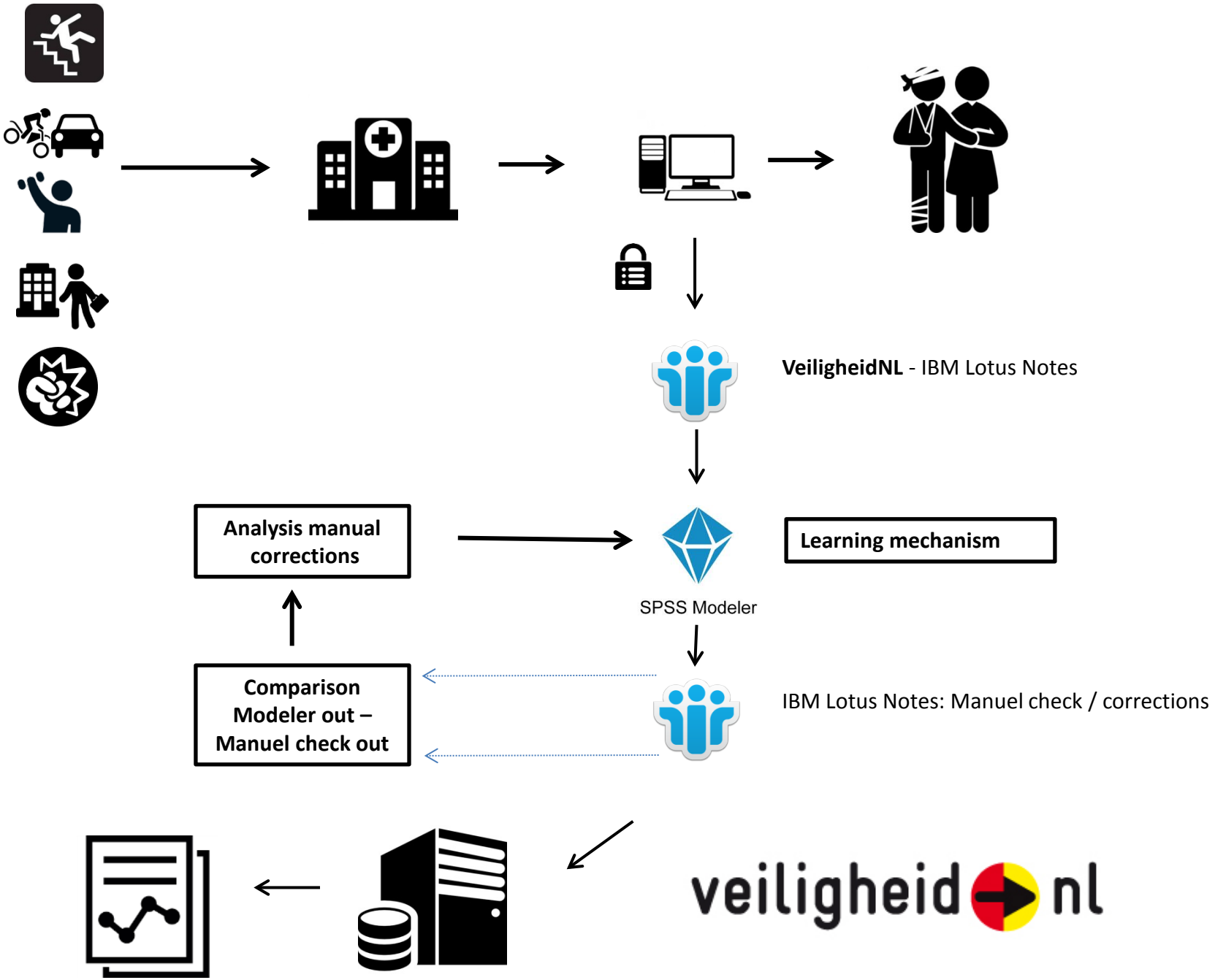
Methods

- After assessment of several tools, we chose IBM SPSSModeler
- We taught the system from scratch how to code information on accidents and injuries
- All possible words were classified into libraries and the system was taught how to interpret sentences
- Comparison: IBM Modeler out – manual check out

veiligheid nl


VeiligheidNL - IBM Lotus Notes
SPSS Modeler
IBM Lotus Notes: Manuel check / corrections
veiligheid nl


Streams
Outputs
Models
Stream1
autotekst - backup 2018-09-19
CRISP-DM
Classes
(unsaved project)
Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
JSQLLIS
1000
Type
Select LIS WF
Select LIS Oud
Recode LIS WF
Filler
Append
ID
Recode LISoud
GoTo Modeler
GoTo SQL
set manualcheck
set verwerkt
dbo.Exchange
TLA verkmechs
verkmechs
TLA gewmechs
gewmechs
TLA letmechs
letmechs
TLA producten
producten
TLA arbeid
TLA sport
Sport
TLA activiteit
Activiteit
TLA locatie
Locatie
TLA lichlet
Lichlet
Merge
SN sporttyp
SN gewmechs
SN letmechs
SN verkmechs
letmechnew
letmechsnew
locnew
locsnew
set null
set manualcheck
set verwerkt
Favorites
Sources
Record Ops
Field Ops
Graphs
Modeling
Output
Export
IBM® SPSS® Statistics
IBM® SPSS® Text Analytics
Database
Var. File
Auto Data Prep
Select
Sample
Aggregate
Derive
Type
Filter
Graphboard
Auto Classifier
Auto Numeric
Auto Cluster
Table
Flat File
Database


VeiligheidNL - IBM Lotus Notes
Analysis manual corrections
Learning mechanism
SPSS Modeler
Comparison Modeler out – Manuel check out
IBM Lotus Notes: Manuel check / corrections
veiligheid nl

Findings: 81% of injury mechanism coded correctly

	Injury mechanism - after manuel check autotext									
Injury mechanism autotext	Fall	Contact with object	Contact with person or animal	Foreign body	Threat to breathing	Chemical mechanism	Thermal mechanism	Electricity, radiation, explosion	Physical over-exertion	Unspecified
Fall	42.500	1.009	289	12	8	33	16	3	216	147
Contact with object	1.979	13.370	182	104	8	44	27	9	85	140
Contact with person or animal	779	366	3.219	18	0	15	1	1	66	78
Foreign body	41	88	32	790	1	75	6	1	1	18
Threat to breathing	10	8	6	13	68	3	0	0	0	11
Chemical mechanism	156	68	8	10	0	1.469	3	1	14	38
Thermal mechanism	35	30	0	2	1	85	401	4	3	9
Electricity, radiation, explosion	5	10	2	1	0	3	4	69	2	1
Physical over-exertion	182	64	45	2	1	2	3	1	868	67
Unspecified	2.544	2.711	864	167	19	1.221	89	14	726	3.720
Total	48.231	17.724	4.647	1.119	106	2.950	550	103	1.981	4.229
% true	**88%**	**75%**	**69%**	**71%**	**64%**	**50%**	**73%**	**67%**	**44%**	**88%**
% unkown	5%	15%	19%	15%	18%	41%	16%	14%	37%	
% false	**7%**	**9%**	**12%**	**14%**	**18%**	**9%**	**11%**	**19%**	**20%**	**12%**

Total true **81%** of the cases was coded correctly for injury mechanism by autotext recognition software

Analysis

Injury mechanism autotext	Fall	Contact with object	Contact with person or animal
Fall	42.500	1.009	289
Contact with object	1.979	13.370	182
Contact with person or animal	779	366	3.219
Foreign body	41	88	32
Threat to breathing	10	8	6
Chemical mechanism	156	68	8

veiligheid nl

Next steps

- Start with analysis of false and unknown cases (largest numbers and/or percentage false): manual text analysis, search for patterns
- Make adjustments in SPSS Modeler based on analysis of false and unknown cases
- Check if adjustments have the desired effect

veiligheid nl

Conclusions (1)

- It takes a lot of time to prepare proper text analysis
- For only products we have imported 9.000 terms (including synonyms and misspelled words)
- Words with double meaning cause difficulties (takes a lot of time)

veiligheid nl

Conclusions (2)

- First analysis showed: 81% of injury mechanism coded correctly
- We still check every record manually and correct if necessary
- The work that is done by the software makes coding at VeiligheidNL easier
- In the future we will be able to reduce the number of checks
- And most important: we have managed to reduce the administrative burden for ED's!

veiligheid nl

Thanks for your attenion!
Questions?

Susanne Nijman, Consumer Safety Institute, Amsterdam, the Netherlands
s.nijman@veiligheid.nl
Birgitte Blatter, Consumer Safety Institute, Amsterdam, the Netherlands
b.blatter@veiligheid.nl

www.veiligheid.nl/en