The use of Big Data in food safety

大数据在食品安全中的应用

Hans Marvin, Yamine Bouzembrak, Martijn Staats and Esmée Janssen (RIKILT Wageningen UR)

RIKILT WAGENINGEN UR

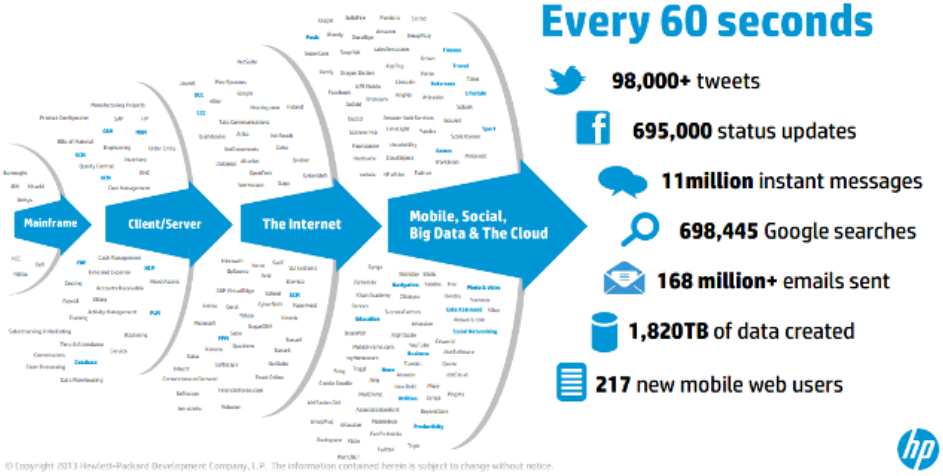
Workshop on Global Trends in Food Safety, Guangzhou, China, 4-5 November 2015

Outline

- Introduction to Big Data
- Use of media in food safety alerts
- Bayesian Networks (BNs) modelling in Big Data research
- Application of BNs to predict food fraud
- Conclusions

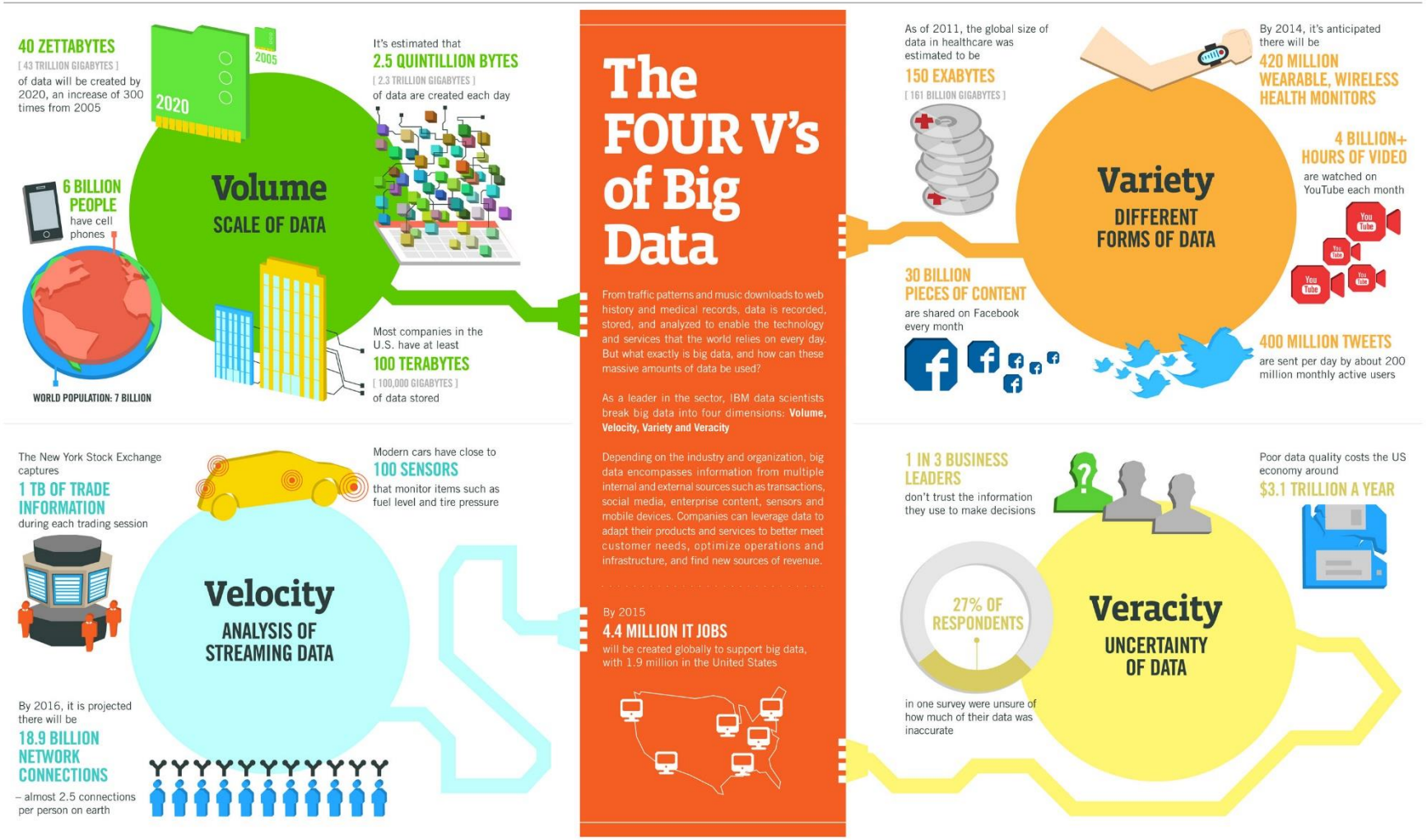
Introduction

- Lots of data is being collected and stored, e.g.:
  - Web data, e-commerce
  - Bank/Credit Card transactions
  - E-mail
  - Data recorded by sensors, scanners, other mobile devices.
  - Chemical monitoring data
  - Omics data



RIKILT
WAGENINGEN UR

Introduction



RIKILT
WAGENINGEN UR

Introduction

Definition of European Commission:

The term "big data" refers to large amounts of different types of data produced with high velocity from a high number of various types of sources. Handling today's highly variable and real-time datasets requires new tools and methods, such as powerful processors, software and algorithms.[7]

[7] Going beyond traditional "data mining" tools designed to handle mainly low-variety, small scale and static datasets, often manually.

Creating value out of Big Data

European Commission (2014) COM(2014) 442 final

Introduction

- The EC has developed a strategy on Big Data and supports a data-driven economy.
  - Big Data as the centre of the future knowledge economy.
  - Supporting open access of data, global connections and e-infrastructure services (www.openaire.eu)

- The Dutch Government stimulates public-private projects to explore Big Data.
  - Big Data Value Center (www.bdvc.nl)
  - Netherlands eScience center www.esciencecenter.nl

Introduction

Typical Big Data workflow

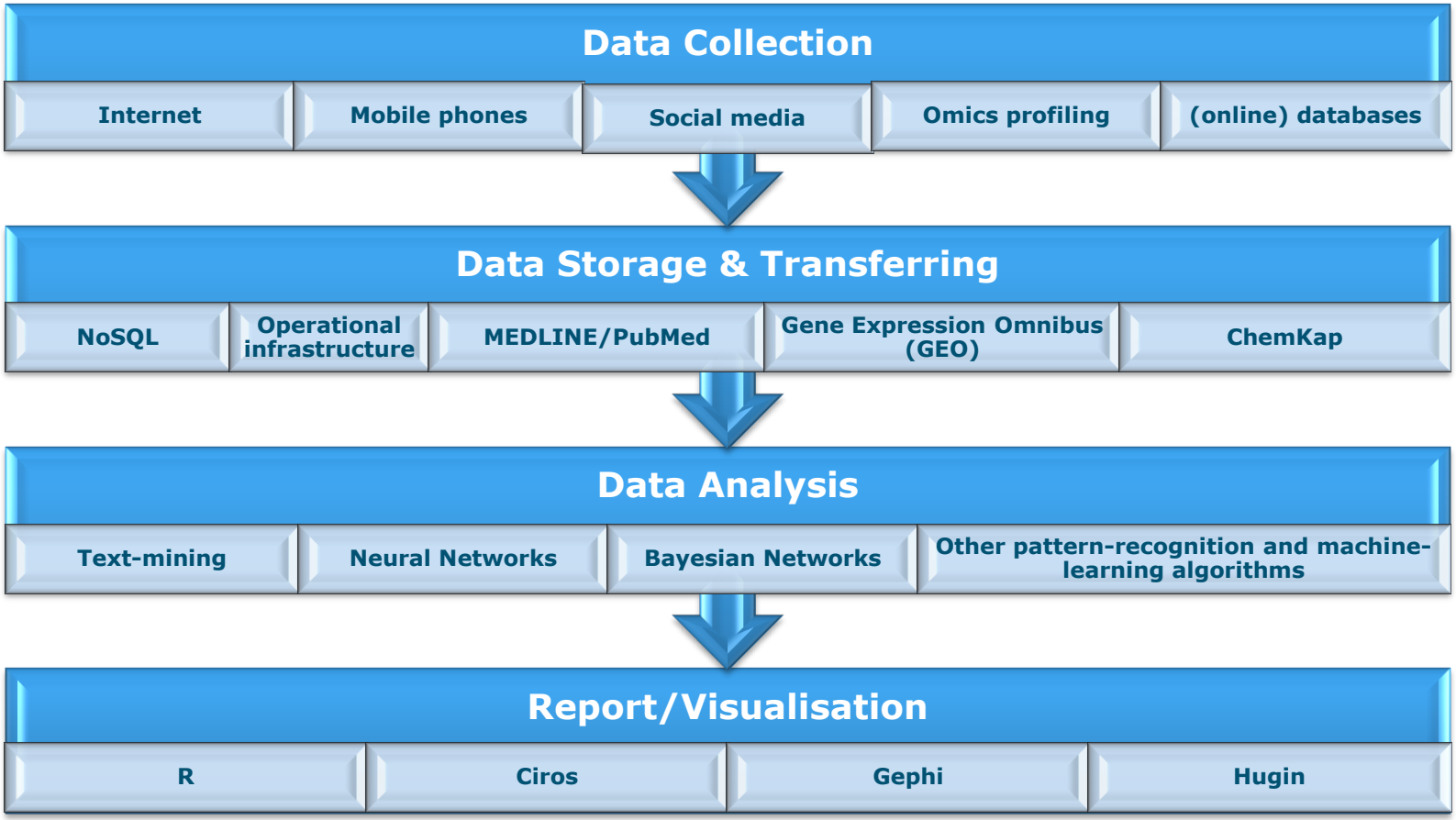
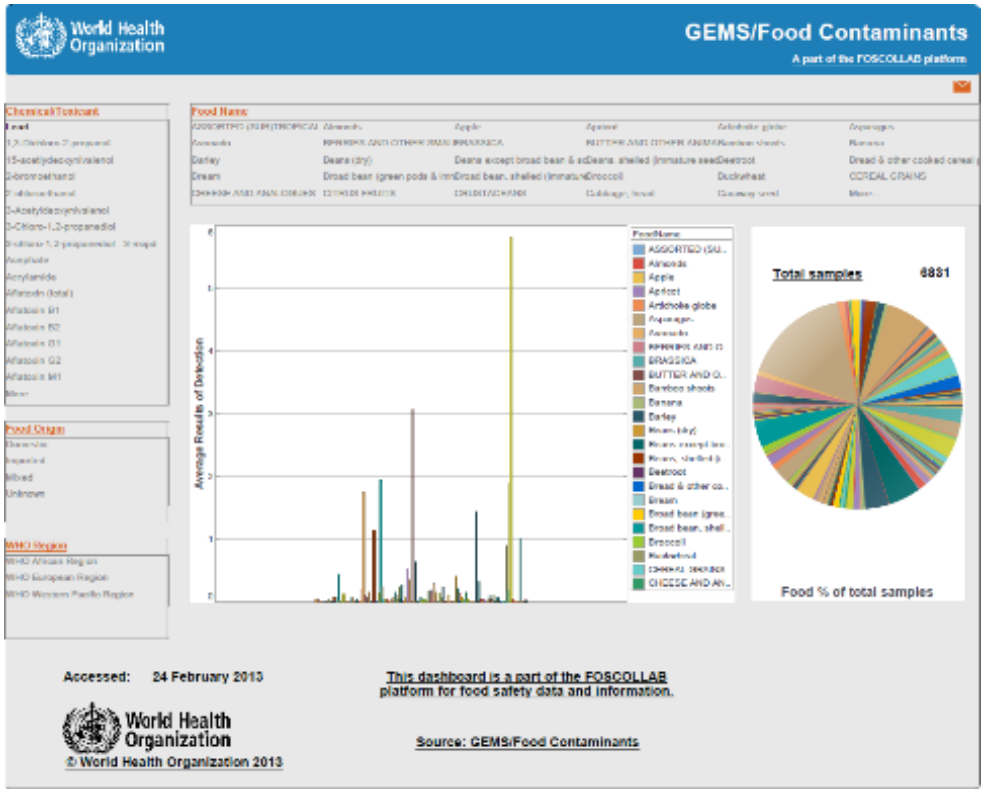


Figure adapted from Huang et al. (2015)

Introduction; example Big Data in food safety

- FOSCOLLAB: Global platform for food safety data and information.

- Support risk managers and decision-makers
- Integrate data and information from animal/agricultural, food, and human health
- Promote better data generation



RIKILT WAGENINGEN UR

http://www.who.int/foodsafety/foscollab_dashboards/en/

World Health Organization

Outline

- Introduction to Big Data
- Use of media in food safety alerts
- Bayesian Networks (BNs) modelling in Big Data research
- Application of BNs to predict food fraud
- Conclusions

Use of media in food safety alerts

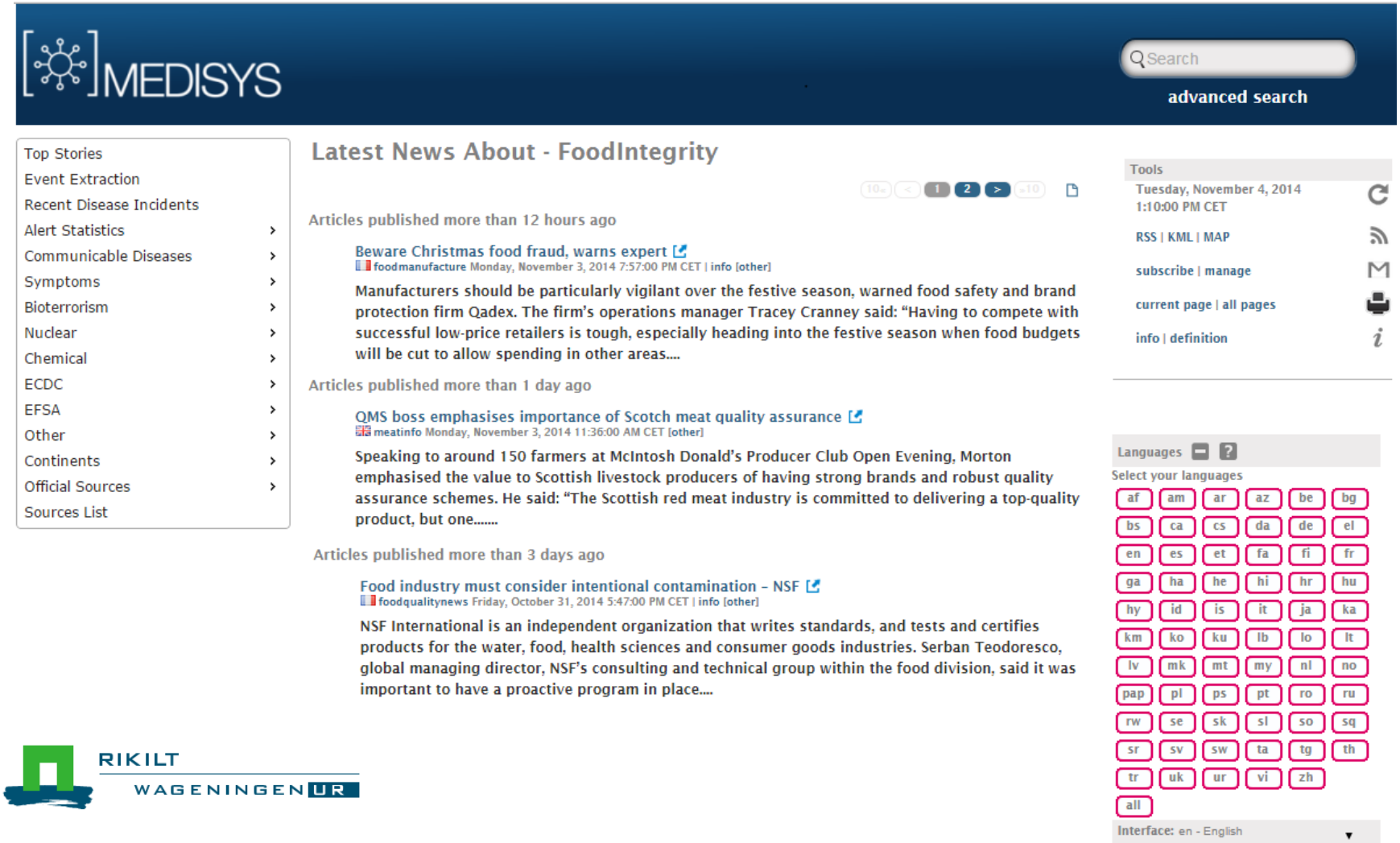
The Europe Media Monitor (EMM) provides advanced analysis systems for monitoring of both traditional and social media.

- EMM presents the latest news and classifies it according to subject.
- It is updated every 10 minutes, 24 hours per day.
- It gathers reports from news portals world-wide in 60 languages,
- EMM classifies the articles, analyses the news texts, aggregates the information and issues alerts.
- EMM applies text mining techniques to screen different types of media on the world wide web: websites, databases, blogs, ..etc.
- EMM contain 3 portals: NewsBrief, NewsExplorer and MedISys

(http://emm.newsbrief.eu/overview.html)

Use of media in food safety alerts

MedISys collects food safety news and RIKILT adapted it also to detect food fraud



RIKILT WAGENINGEN UR

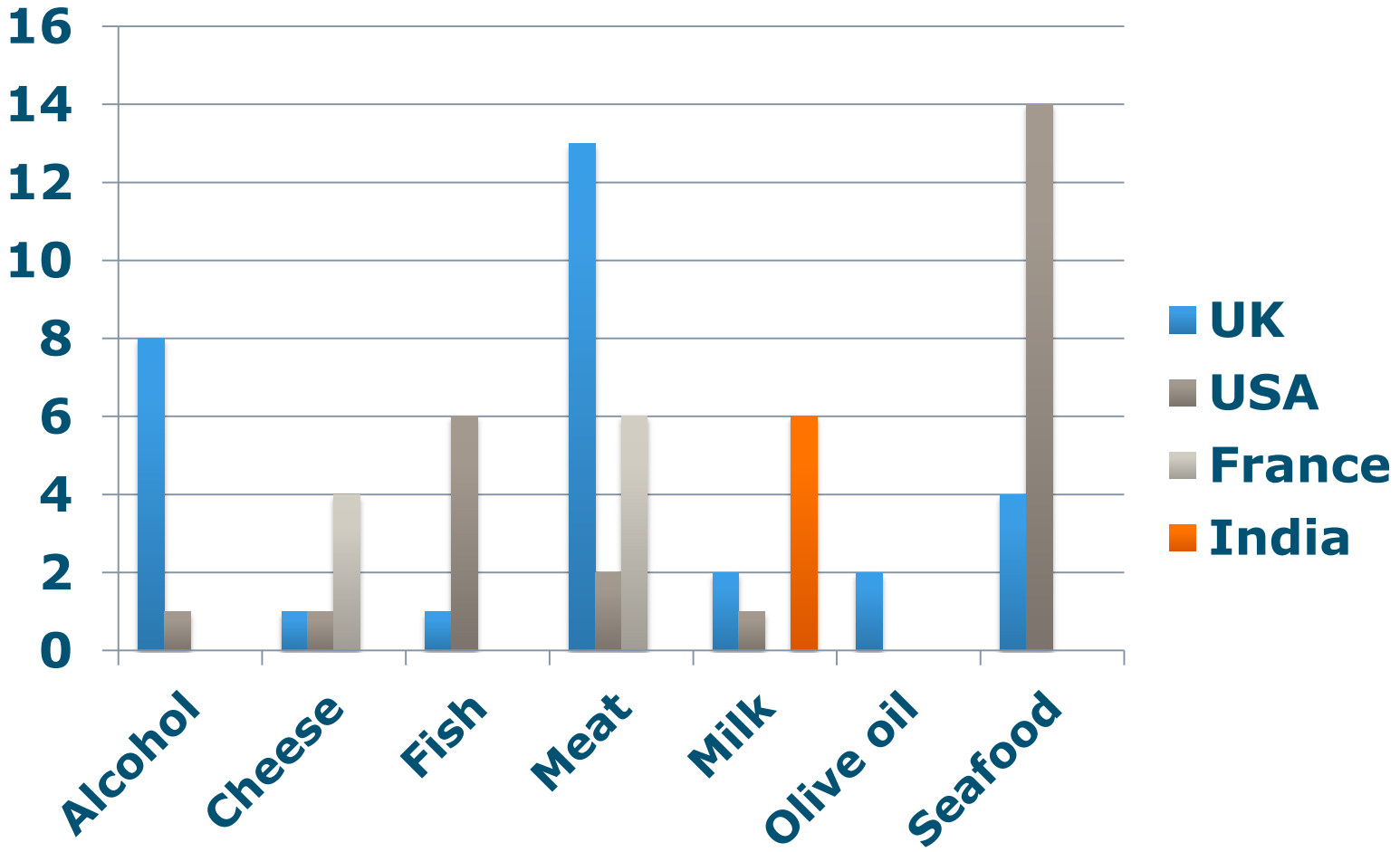
Use of media in food safety alerts

Food fraud reports in MedISys (period September 2014 to June 2015; N > 600)

RIKILT
WAGENINGEN UR

Use of media in food safety alerts

Food fraud reports in MedISys (period September 2014 to February 2015)

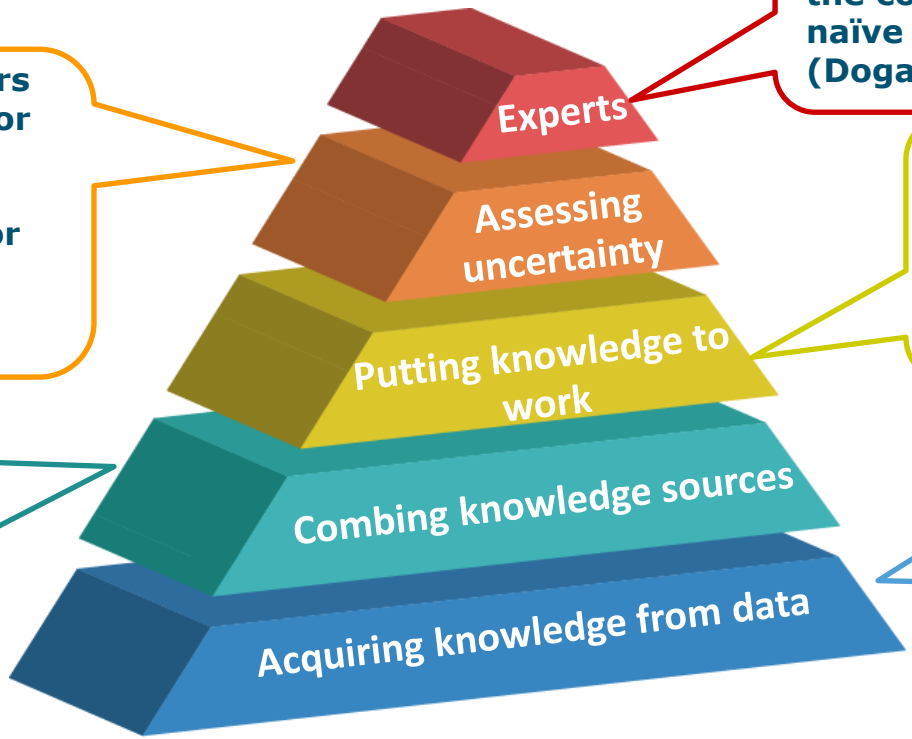


RIKILT
WAGENINGEN UR

Outline

- Introduction to Big Data
- Use of media in food safety alerts
- Bayesian Networks (BNs) modelling in Big Data research
- Application of BNs to predict food fraud
- Conclusions

Bayesian Networks (BNs) modelling in Big Data research



BN requires less parameters than the conventional naïve method (Dogan, 2012).

BN approach offers a high potential for representing ambiguous knowledge and for performing reasoning under uncertainty

BN can combine expert and domain knowledge that allows flexible inference even with partial and limited information (Lauritzen,1995).

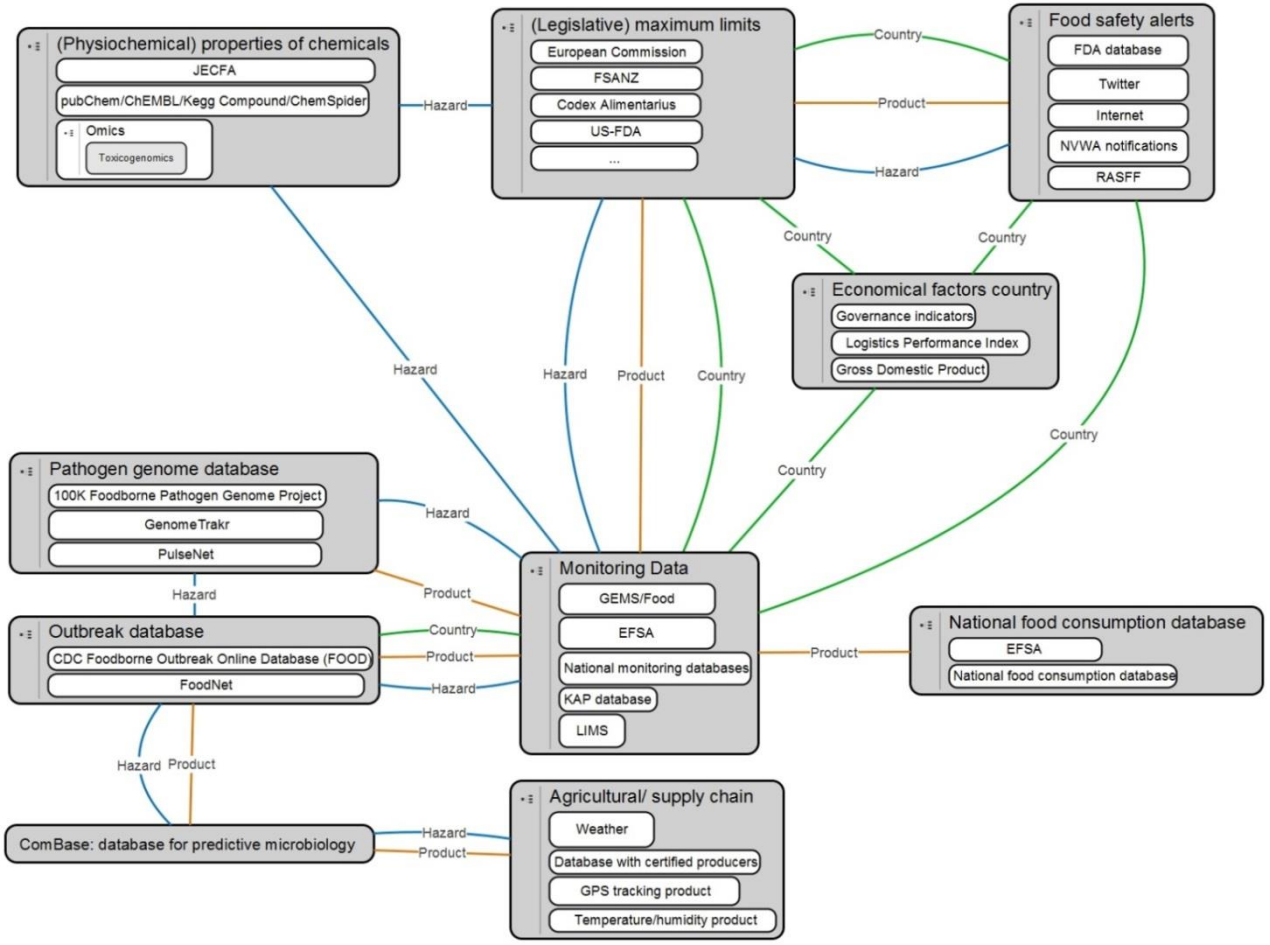
BN takes into account a number of quantitative and qualitative criteria under uncertainty (Dogan and Aydin, 2011)

BN are very powerful for making inferences and drawing conclusions based on available information (Jensen, 1996).

RIKILT
WAGENINGEN UR

Bayesian Networks (BNs) modelling in Big Data research

BN is able to combine data from different nature and location



RIKILT
WAGENINGEN UR

Outline

- Introduction to Big Data
- Use of media in food safety alerts
- Bayesian Networks (BNs) modelling in Big Data research
- Application of BNs to predict food fraud
- Conclusions

Application of BNs to predict food fraud

- (Bouzembrak and *Marvin,* 2015) used BNs prediction model to aid the border inspector to detect the type of food fraud.

ARTICLE IN PRESS JFCO4660_proof ■ 29 September 2015 ■ 1/8

Food Control xxx (2015) 1–8

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

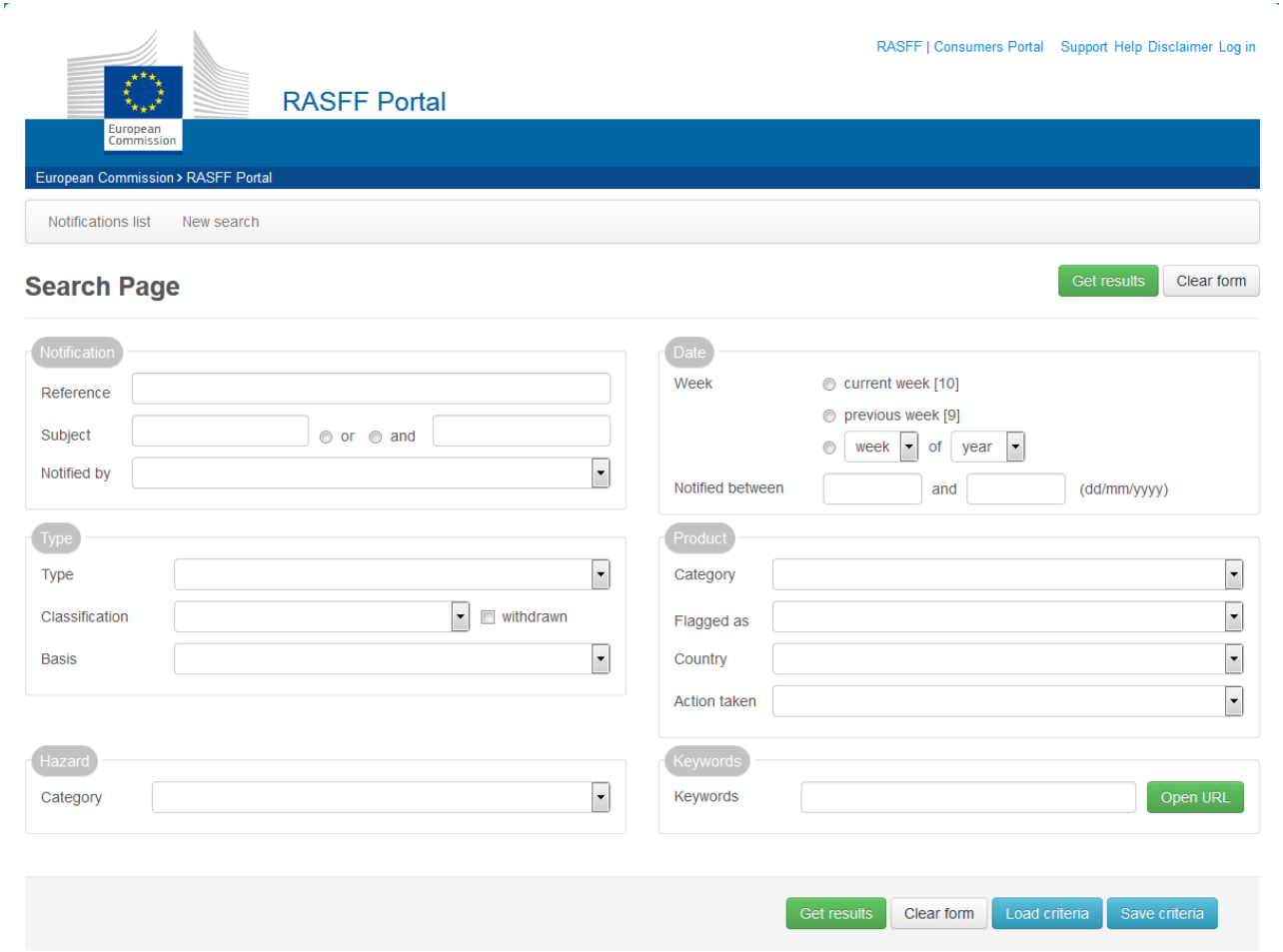
Prediction of food fraud type using data from Rapid Alert System for Food and Feed (RASFF) and Bayesian network modelling

Y. Bouzembrak, H.J.P. Marvin*

RIKILT Wageningen UR, Akkermaalsbos 2, 6708 WB Wageningen, The Netherlands

Application of BNs to predict food fraud

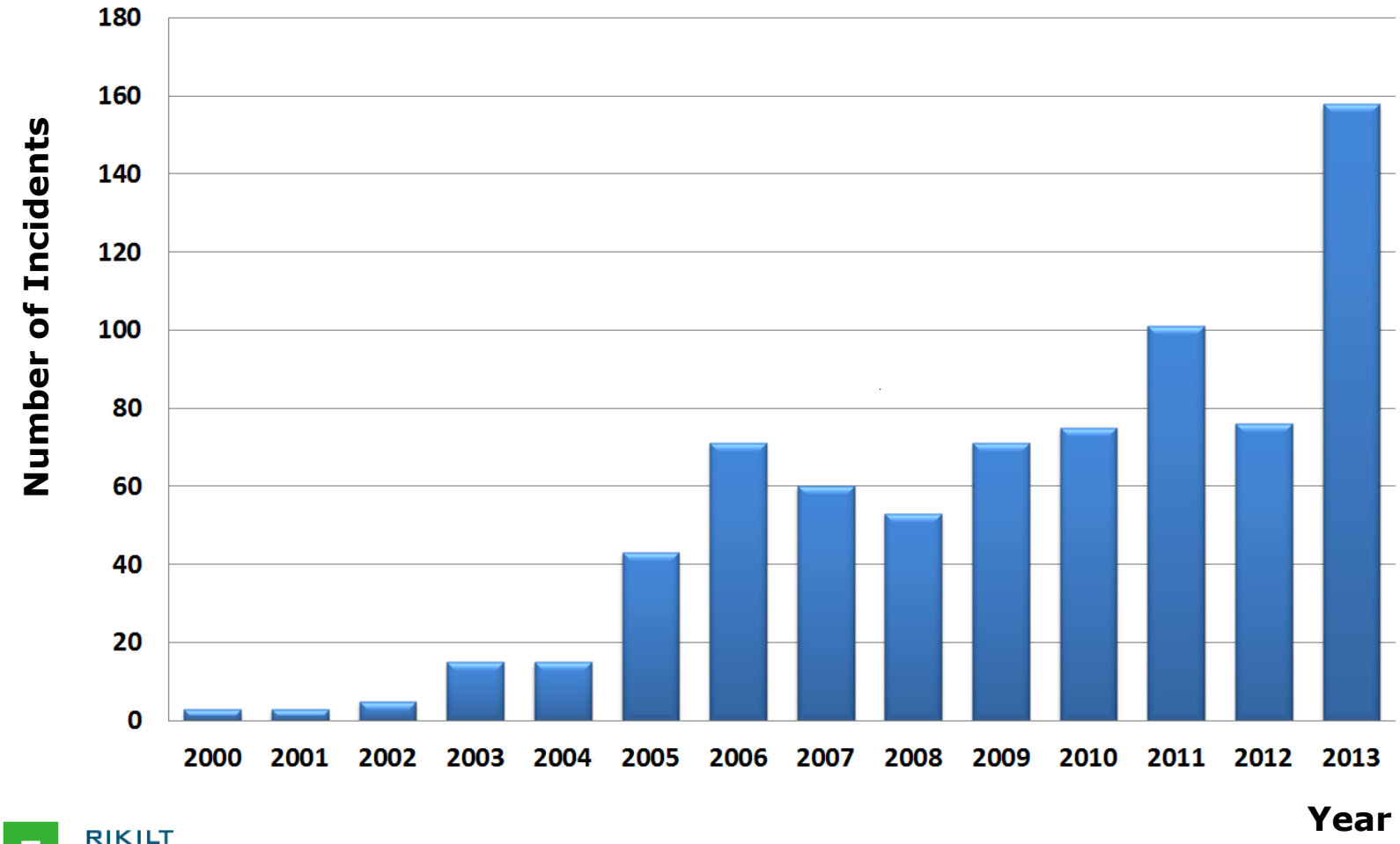
Prediction of food fraud type in RASFF



RIKILT WAGENINGEN UR

Application of BNs to predict food fraud

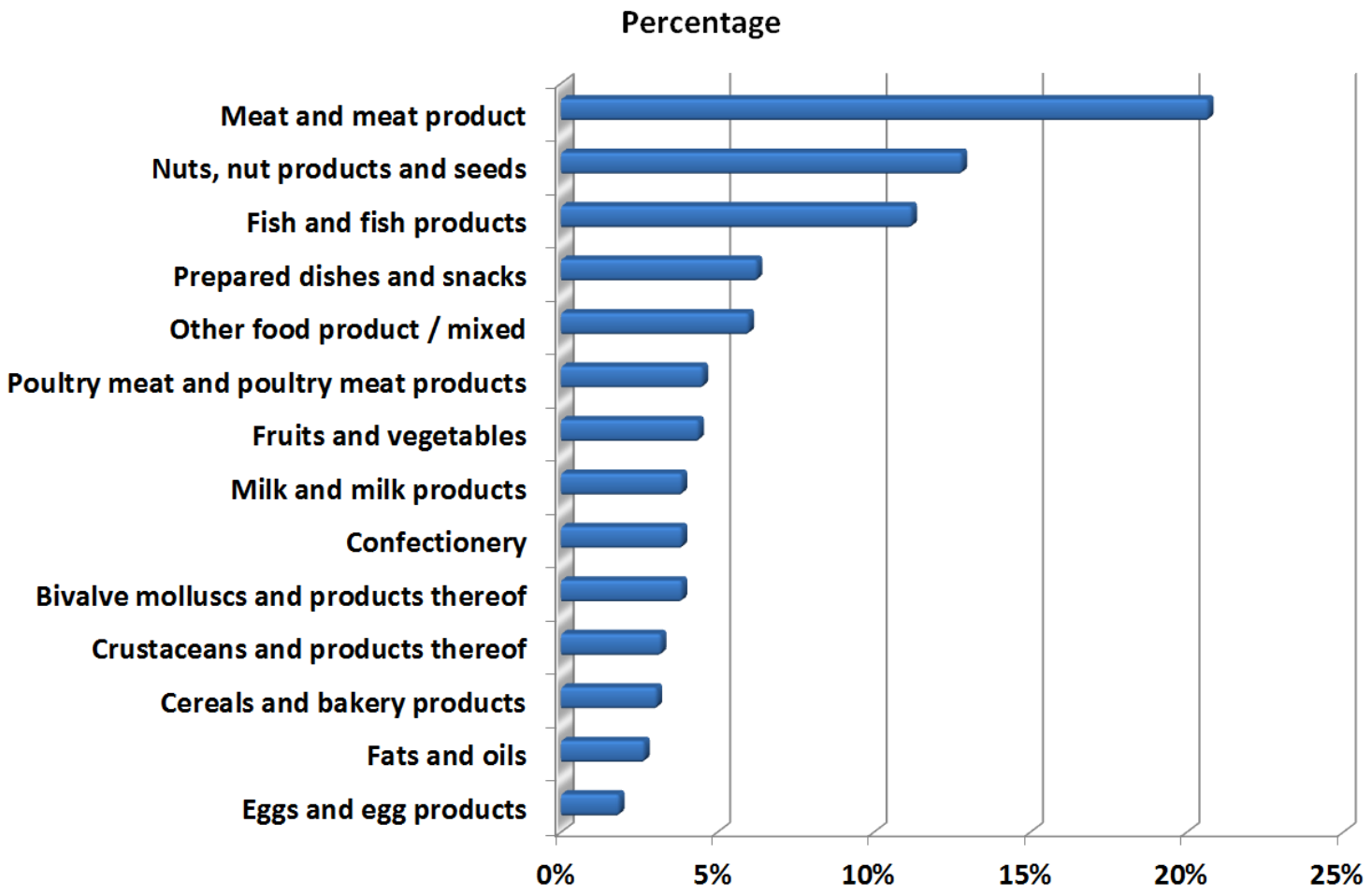
Food fraud notifications in RASFF (2000 to 2013).



RIKILT
WAGENINGEN UR

Application of BNs to predict food fraud

Products related to food fraud in RASFF



RIKILT
WAGENINGEN UR

Application of BNs to predict food fraud

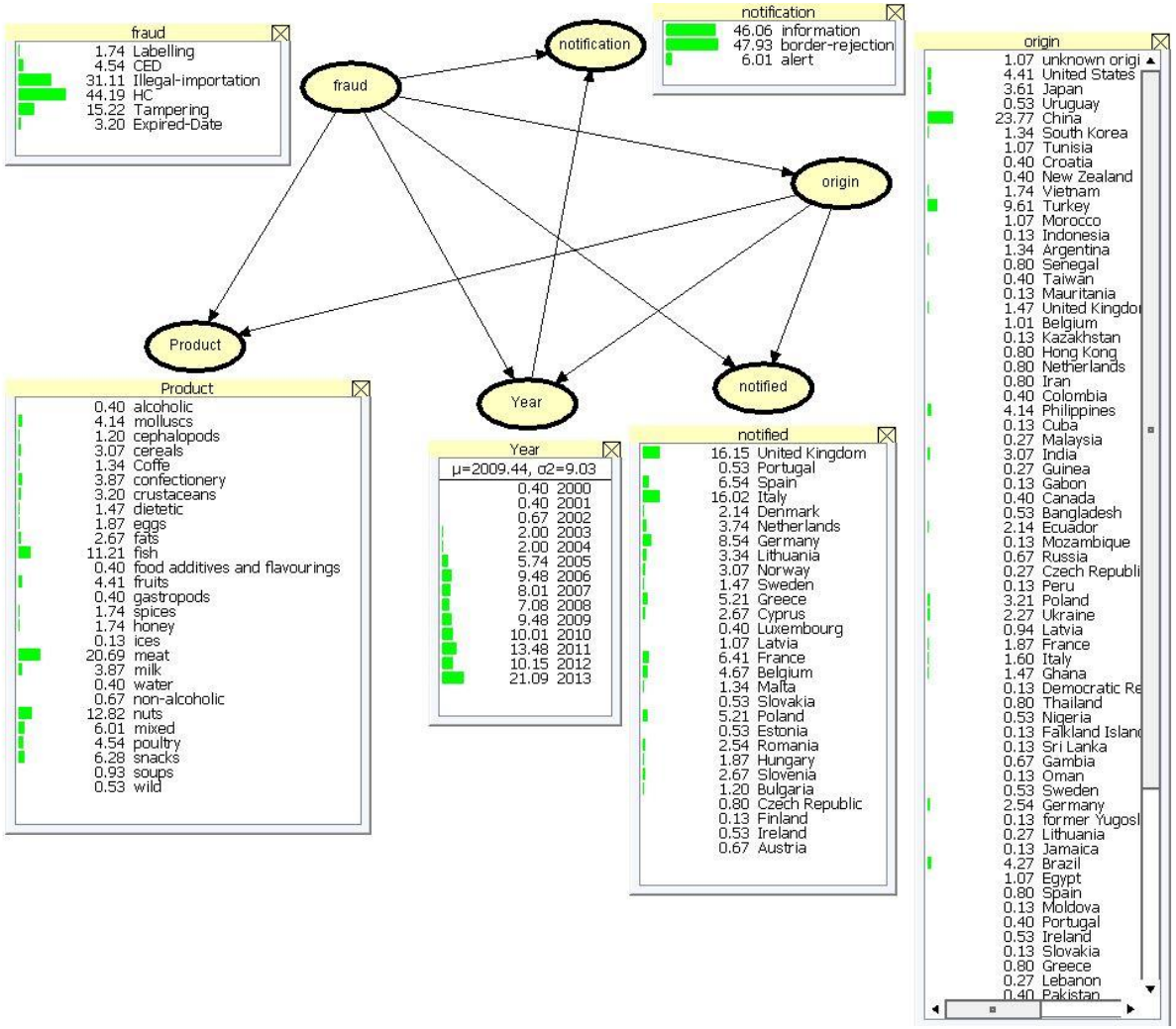
Food fraud types in RASFF

Fraud type	Description	Count of notification type	Percentage
HC	Improper or fraudulent or missing or absent of health certificate	370	44%
Illegal importation	Illegal or unauthorized import or trade or transit	256	31%
Tampering	Adulteration or fraud or tampering	126	15%
CED	Improper or expired or fraudulent or missing of common entry document or import declaration	45	5%
Expired Date	Expired Date	24	3%
Mislabelling	Mislabelling	13	2%
Grand Total		834	100%

Application of BNs to predict food fraud

Modelling of Fraud in RASFF

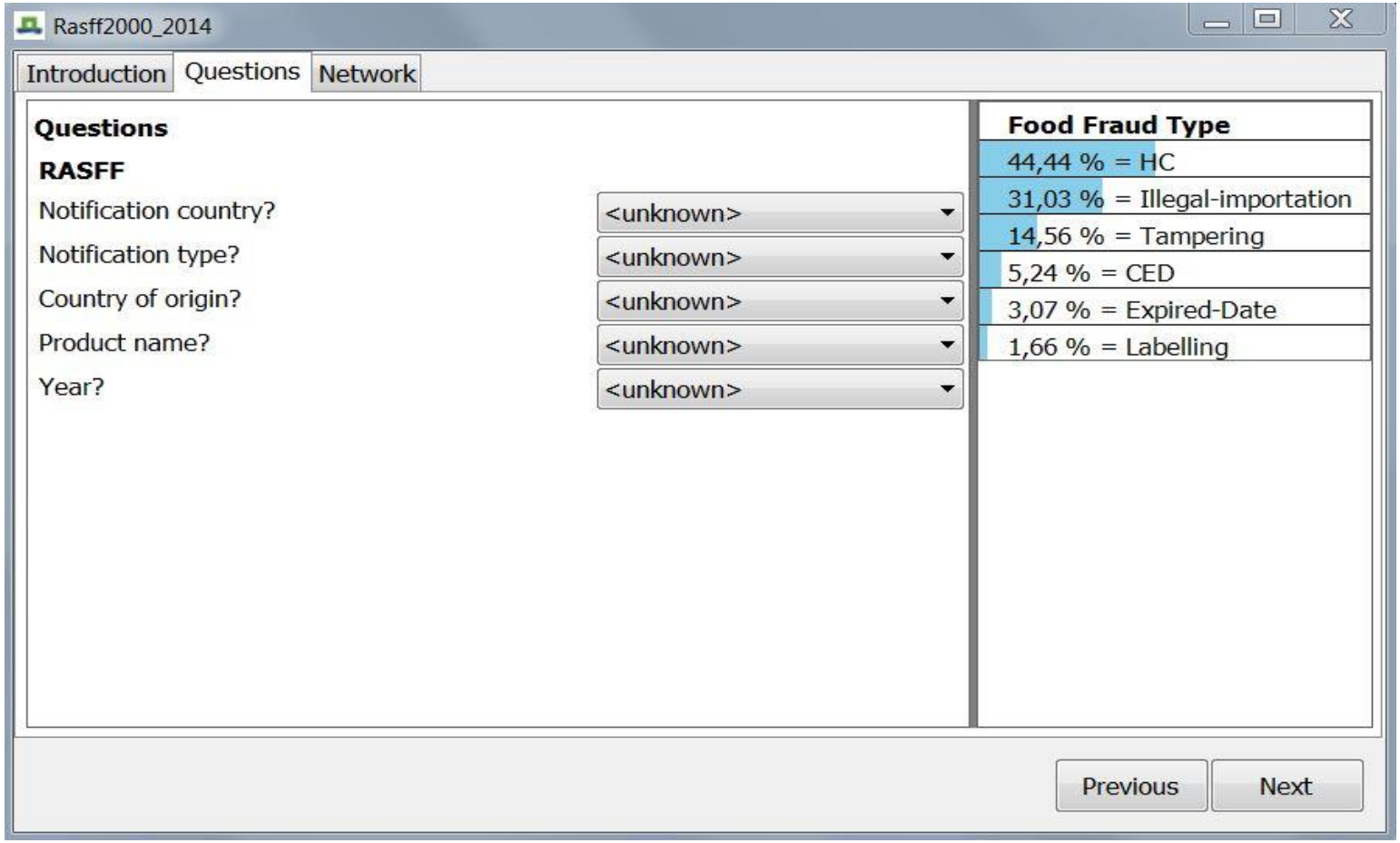
Model developed with RASFF data up to 2013



RIKILT
WAGENINGEN UR

Application of BNs to predict food fraud

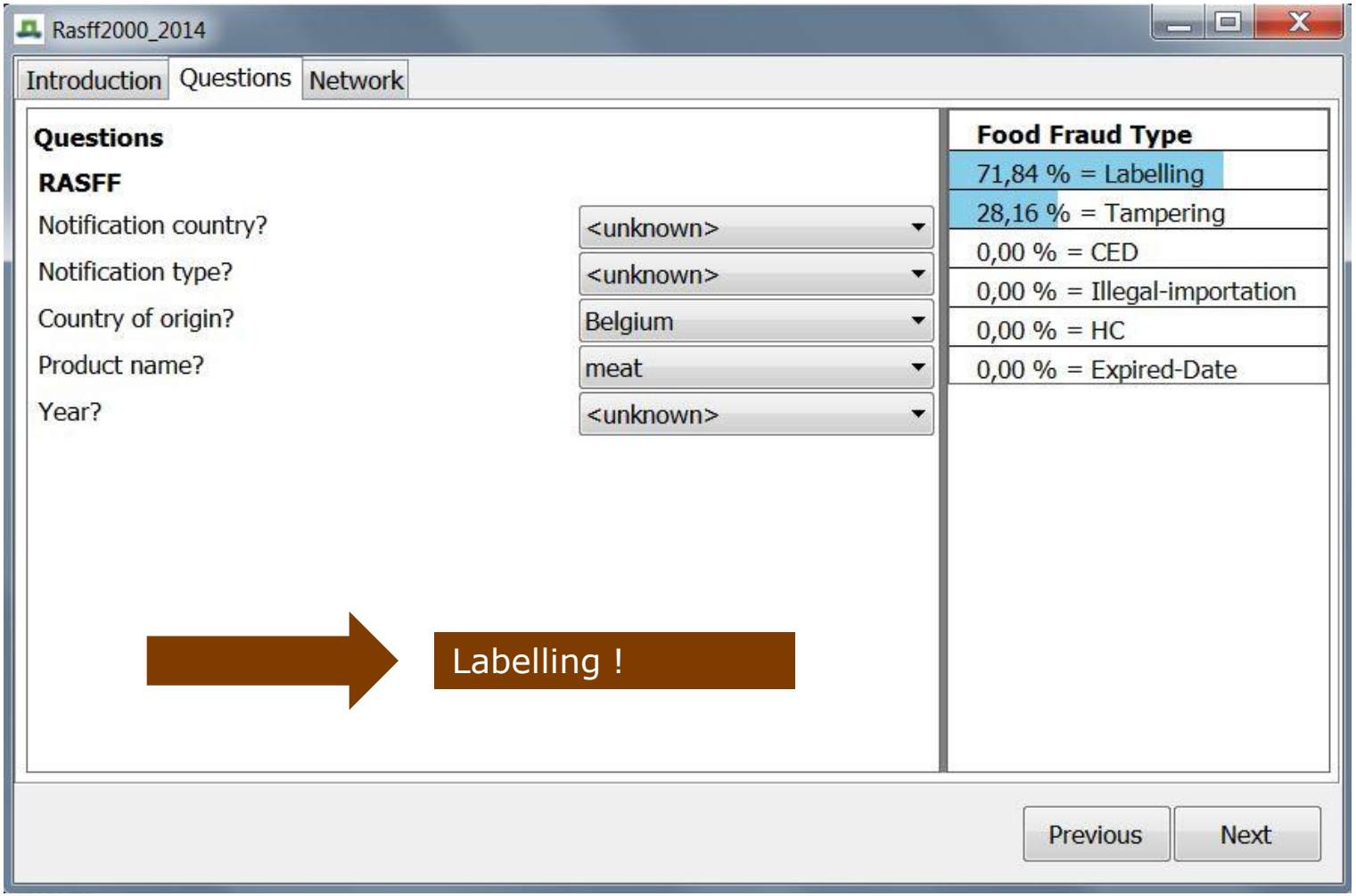
USER-INTERFACE



RIKILT
WAGENINGEN UR

Application of BNs to predict food fraud

CASE: What type of fraud can we expect from meat imported from Belgium



WAGENINGEN UR

Application of BNs to predict food fraud

Validation of the Model;

Method:

Predict the type of fraud for all reports in RASFF in 2014.

Input values: type of product and country of origin

Results:

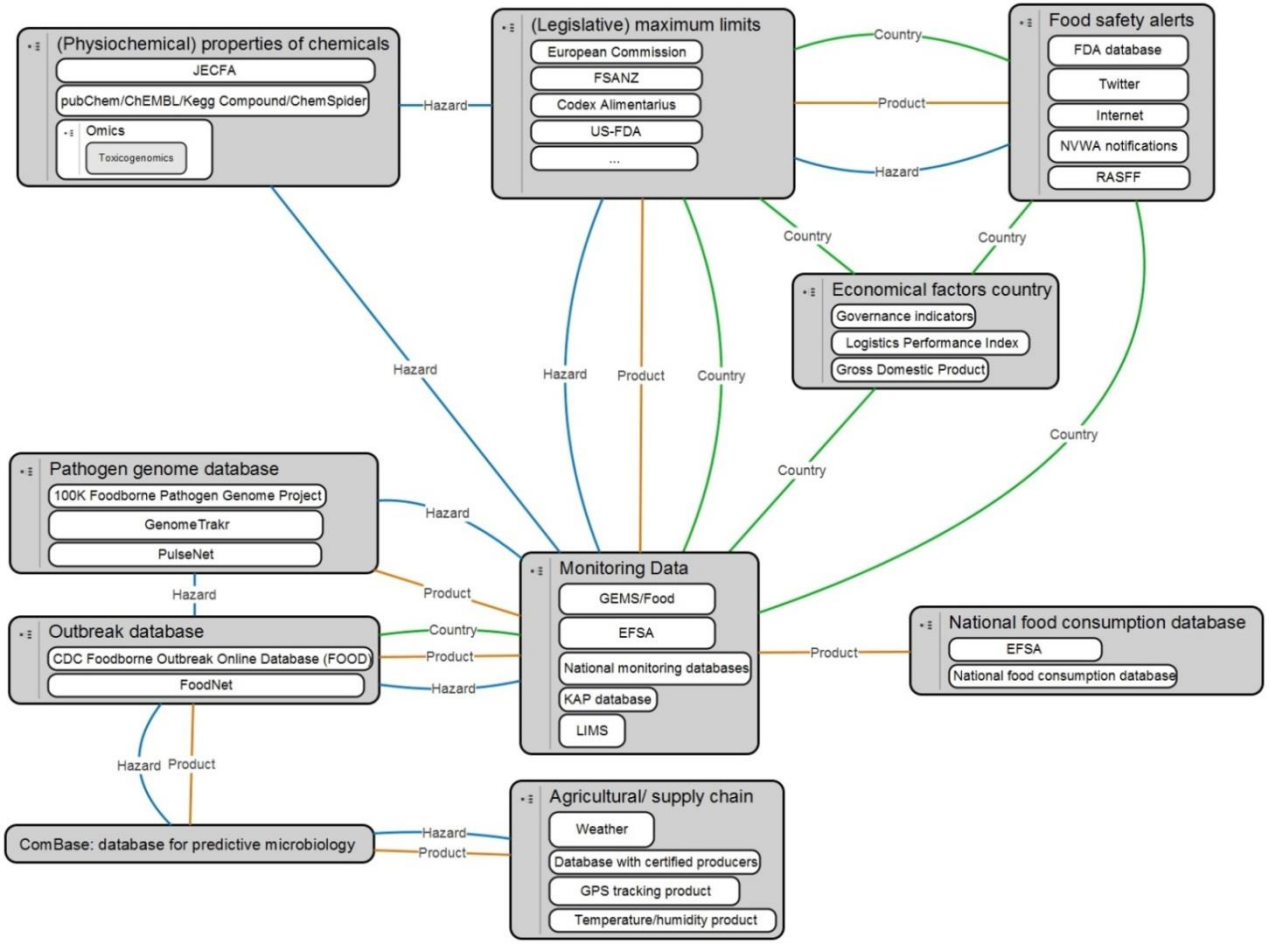
- 82% correct
- 51% prediction of new fraud combinations

Higher performance with dynamic model

RIKILT
WAGENINGEN UR

Bayesian Networks (BNs) modelling in Big Data research

BN is able to combine data from different nature and location



RIKILT
WAGENINGEN UR

Application of BNs to predict food fraud

Prediction of food fraud using multiple data sources and expert knowledge

- Identify parameters relevant for food fraud (=> nodes in BN model)
- Determine the relationships between these parameters (expert knowledge, literature)
- Identify data sources of these parameters
- Develop the BN model
- Validate the BN model

Application of BNs to predict food fraud

Parameters relevant for food fraud (some examples)

Economic parameters:

- Prices of the fraudulent product at the time of detection
- Price spike around the period of detection
- Trade volumes of the product between the country of detection and country of origin
- Complexity of the food chain

Parameters of the country of origin & detection

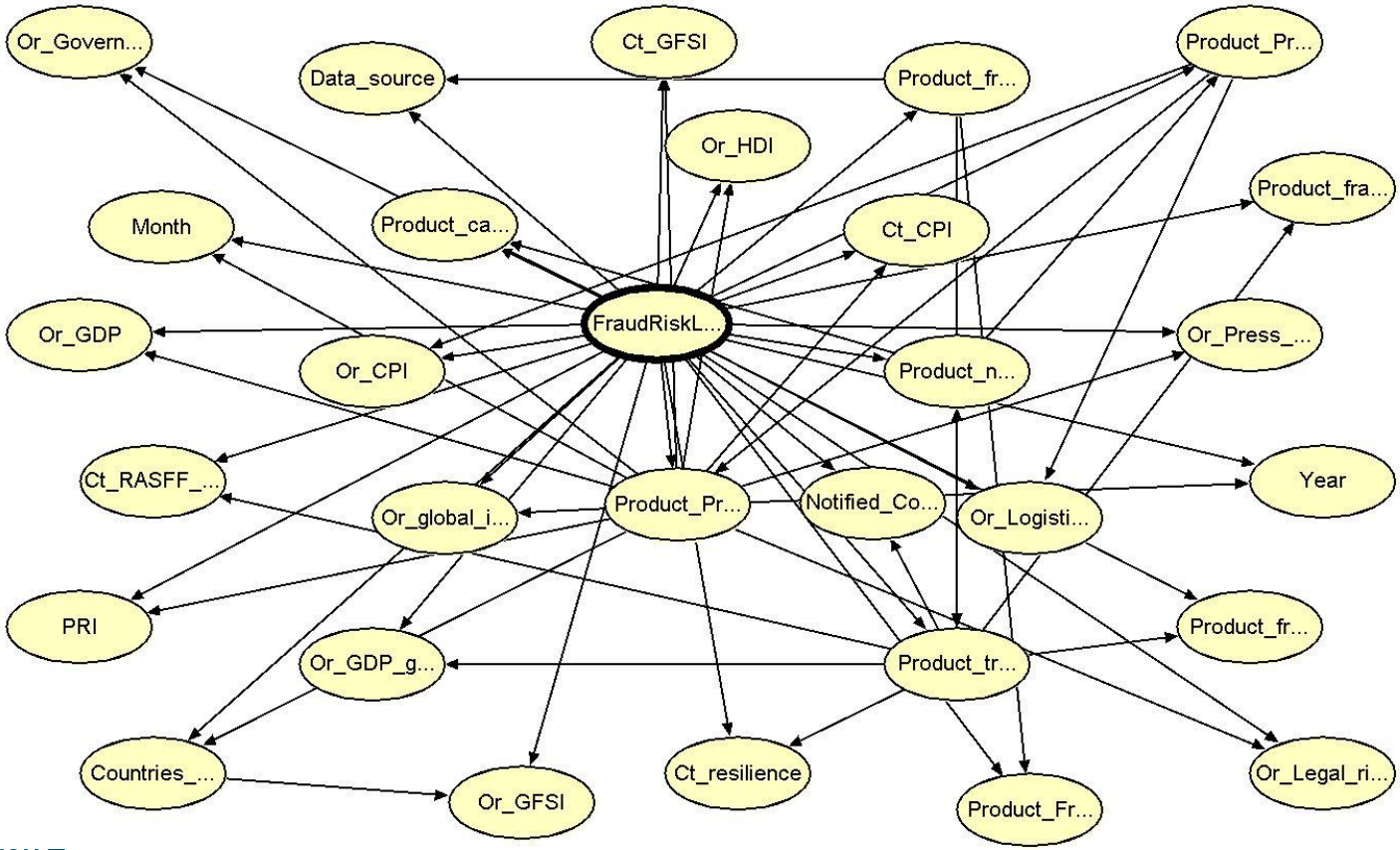
- Indices: corruption index, food safety index, governance index, legal system index, press index, human development index and technology index.

Detected food fraud cases

- RASFF, UPS and EMA

Application of BNs to predict food fraud

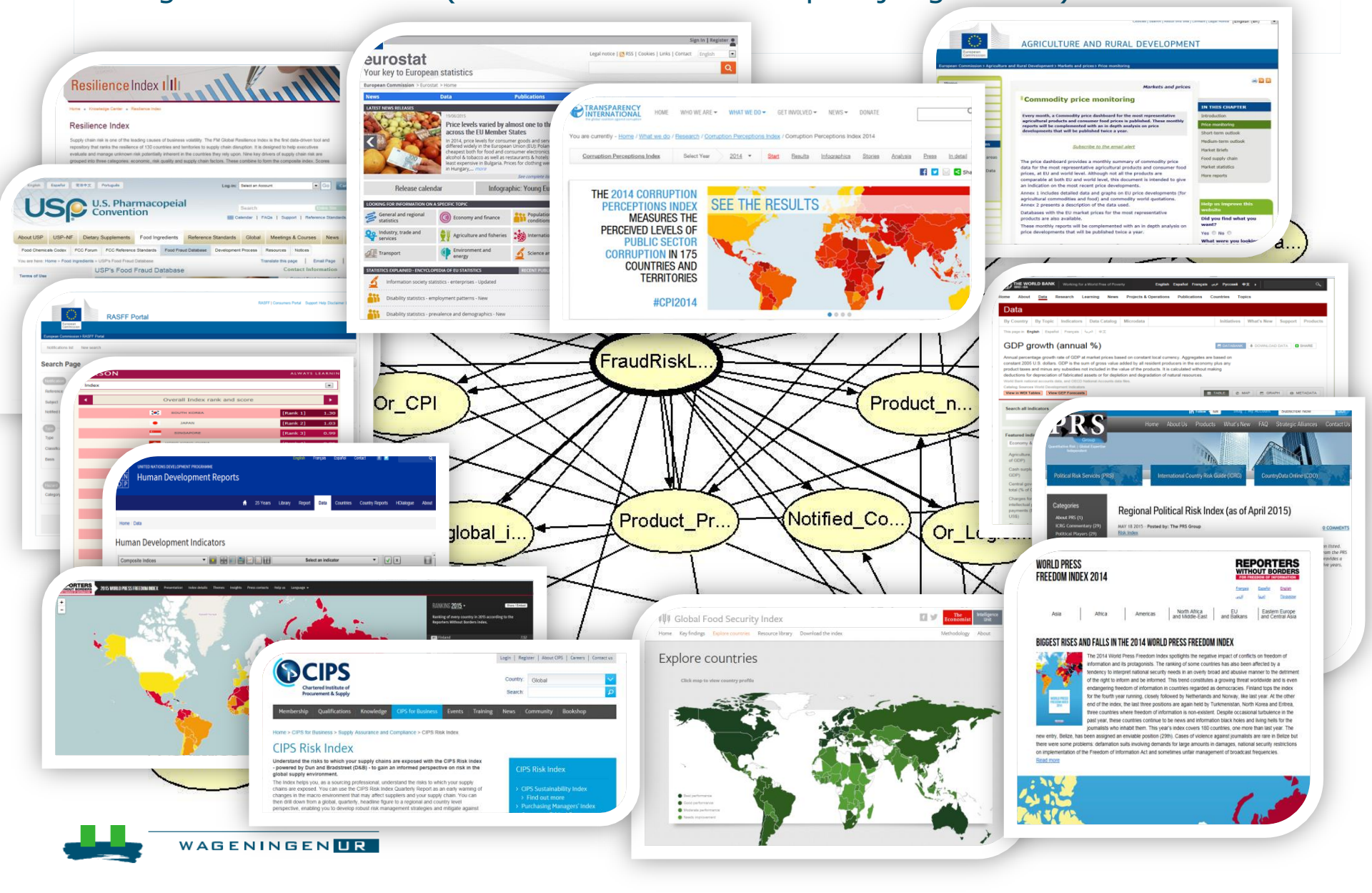
Preliminary BN model for predicting food fraud; relationships between all parameters



RIKILT
WAGENINGEN UR

Preliminary BN model for predicting fraud

linking 36 data sources (18 databases and 8 expert judgements)



Conclusions

- Big data developments show great potential in food safety research, food safety risk assessment and food safety risk management
- European Media Monitor MedISys is suitable to collect publications on food safety and food fraud on global level.
- Bayesians Network approach is useful to combine information from a variety of data sources
- Bayesians network models are useful to predict food fraud

Thank you

谢谢

RIKILT WAGENINGEN UR