Big data in food safety

Hans Marvin & Yamine Bouzembrak
(RIKILT Wageningen UR)

WAGENINGEN
UNIVERSITY & RESEARCH

WDCC, 10 April 2018, Wageningen

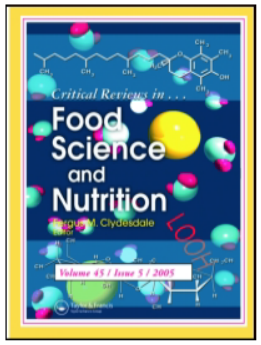
Outline

1. Literature study; What is Big data and is it used in food safety research

2. Stepwise testing & implementation of Big data elements at RIKILT

3. European development: DEMETER project

1) Literature study; What is Big data and is it used in food safety research



Taylor & Francis
Taylor & Francis Group

Critical Reviews in Food Science and Nutrition

ISSN: 1040-8398 (Print) 1549-7852 (Online) Journal homepage: http://www.tandfonline.com/loi/bfsn20

Big data in food safety: An overview

Hans J. P. Marvin, Esmée M. Janssen, Yamine Bouzembrak, Peter J. M. Hendriksen & Martijn Staats

To cite this article: Hans J. P. Marvin, Esmée M. Janssen, Yamine Bouzembrak, Peter J. M. Hendriksen & Martijn Staats (2017) Big data in food safety: An overview, Critical Reviews in Food Science and Nutrition, 57:11, 2286-2295, DOI: 10.1080/10408398.2016.1257481

To link to this article: http://dx.doi.org/10.1080/10408398.2016.1257481

Big Data, definition provided by EC

- 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.[7]

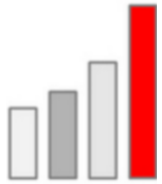
[7] Going beyond traditional "data mining" tools designed to handle mainly low-variety, small scale and static datasets, often manually.

WAGENINGEN
UNIVERSITY & RESEARCH

European Commission (2014) COM(2014) 442 final

Characteristics of Big Data

- **Volume**: refers to the vast amounts of data generated every second.
  - (Mega, Giga), Tera, Peta, Exa, Zetta, Yotta bytes,...



- **Velocity**: refers to the increasing speed of which data is created and the speed at which it can be stored, processed and analyzed.
  - Batch, Real Time

- **Variety**: refers to the different types of data including structured data, semi-structured data, and unstructured data.

- **Veracity** refers to the trustworthiness and accuracy of the data.

From: Prof. Bedir Tekinerdogan
Information Technology Group, Wageningen University

Typical Big Data workflow

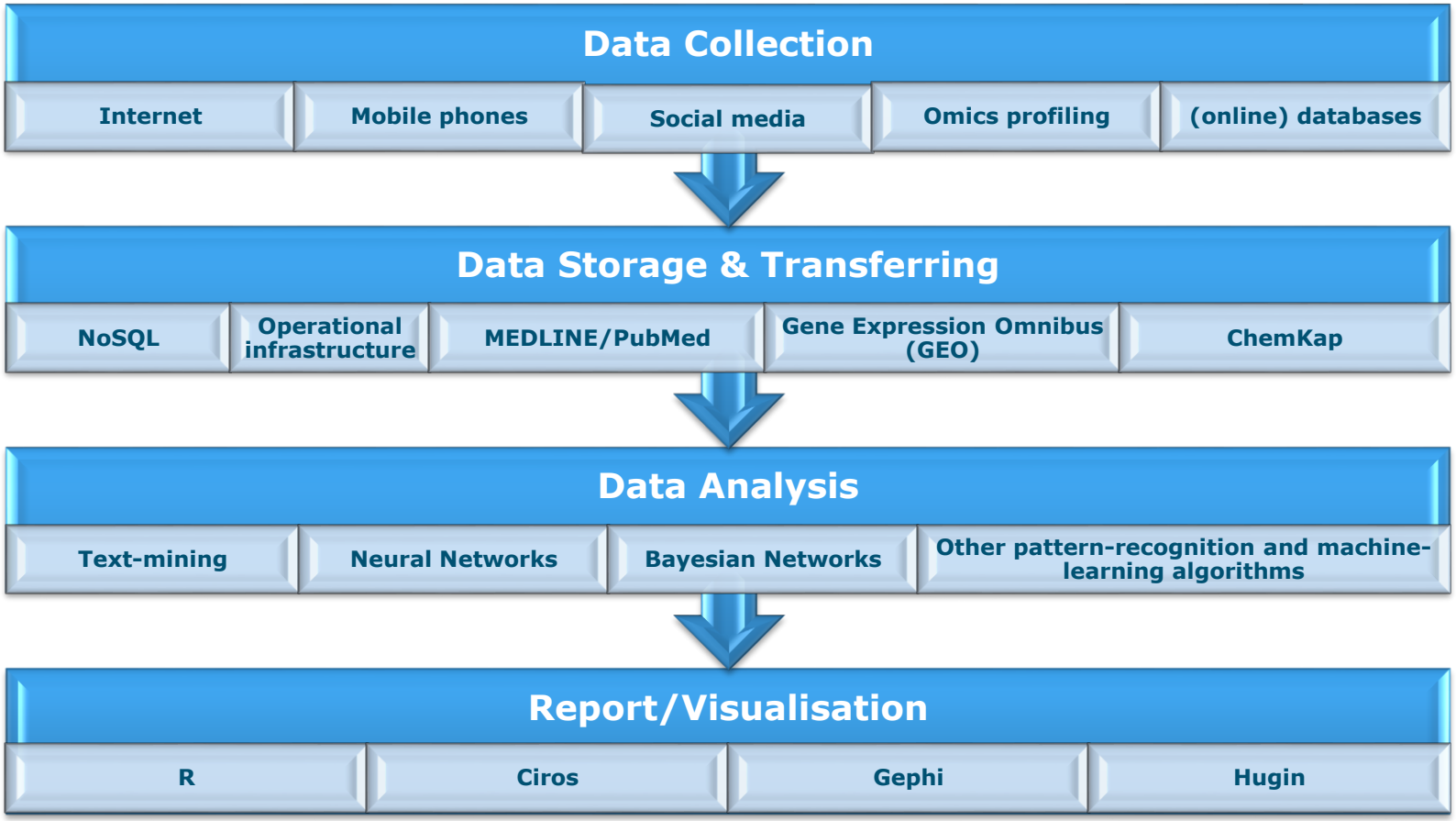


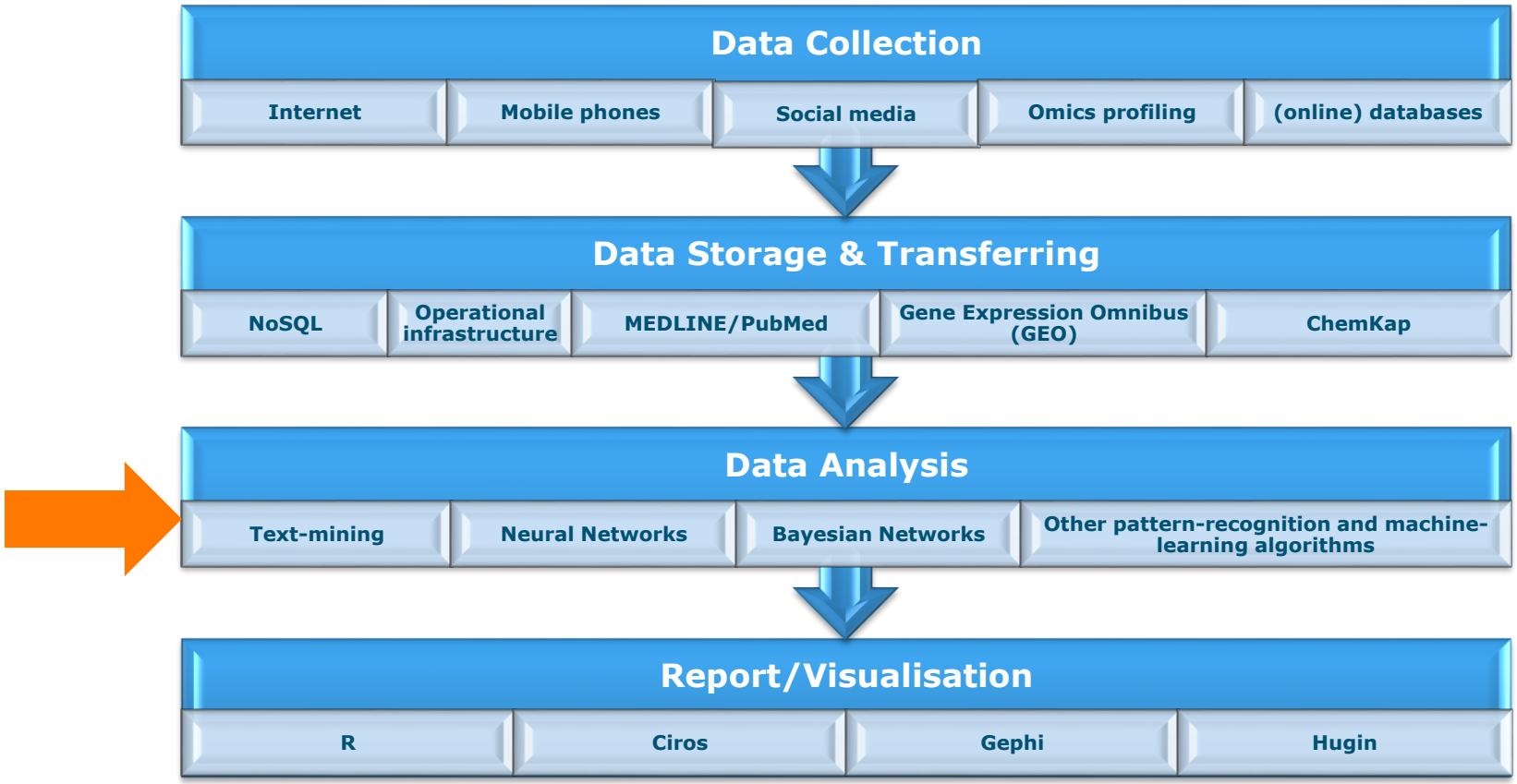
Figure adapted from Huang et al. (2015)

Literature study: conclusions

- Big data is not yet fully applied in food safety. Volume and Velocity, is generally no issue but applications dealing with Variety & Veracity were found.
- Several examples were found that uses some of the Big data tools.
- Future trends on smartphone (online, onsite analysis), block chain (sharing data in the food supply chain), using social media for food safety identification will stimulate Big data approach.
- The trend to make data from public funded research projects available on internet opens new opportunities but requires presence of infrastructures (open cloud) and knowledge about the Big data tools.

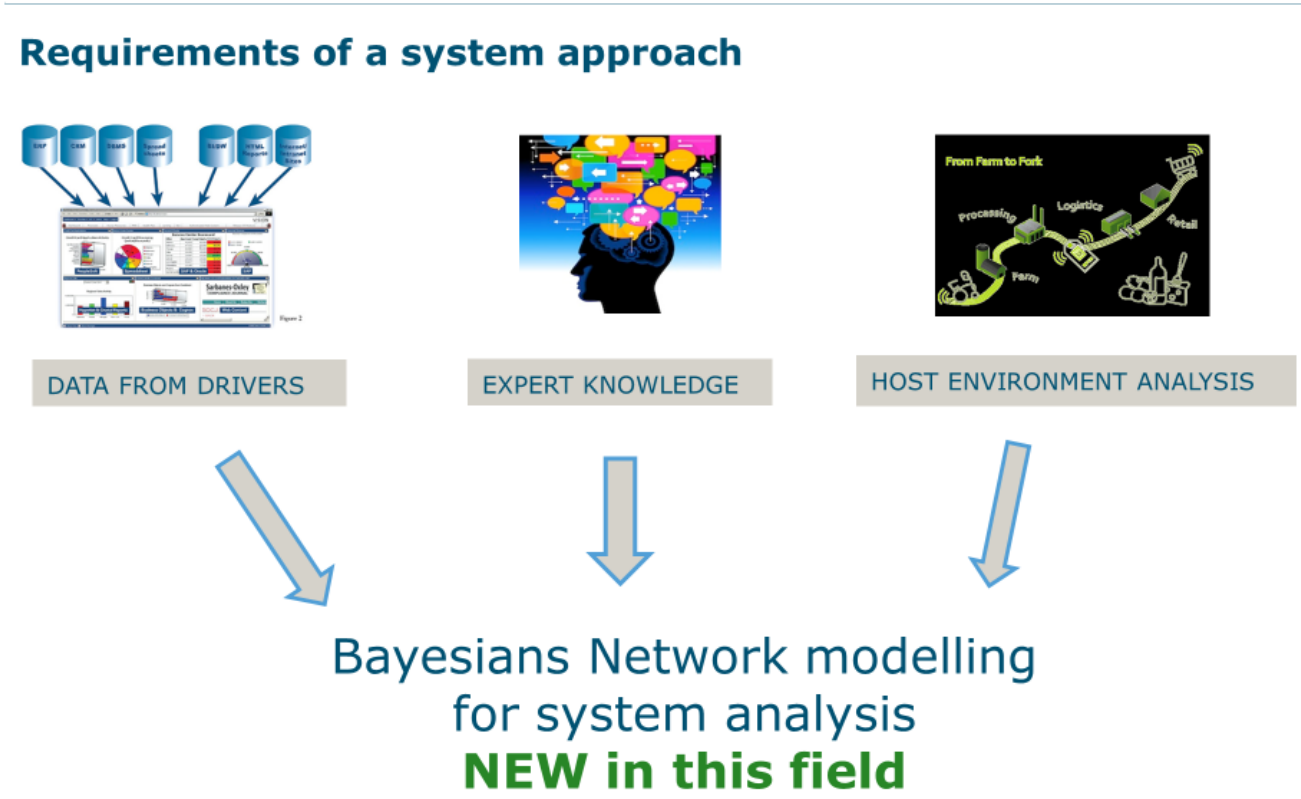
RIKILT started considering Big data in 2015

1. First focus on data analysis



RIKILT started considering Big data in 2015

1. **Data analysis:** research question: can we combine data from different sources and structures to predict (food safety or fraud); projects on nanosafety, fruit and vegetables, food fraud, dairy)

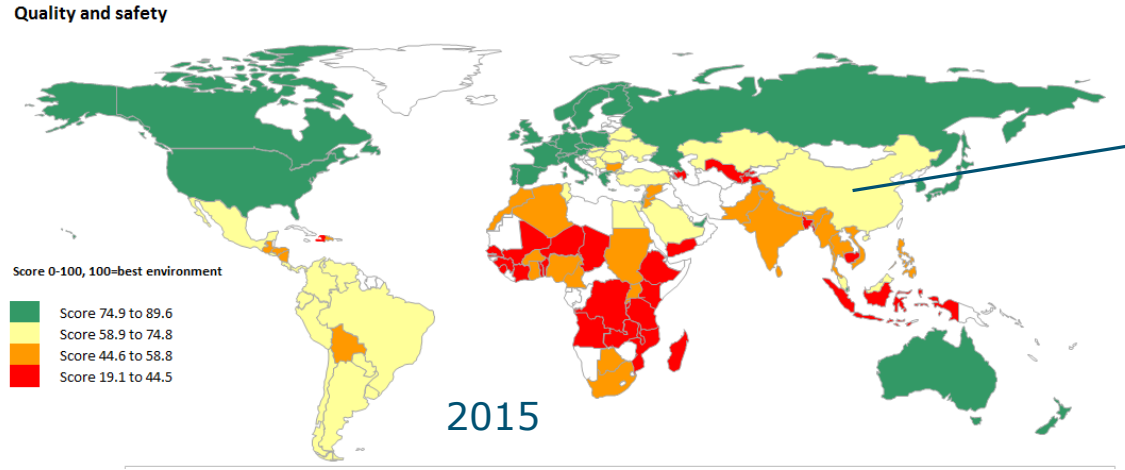


Preliminary BN model for predicting fraud

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

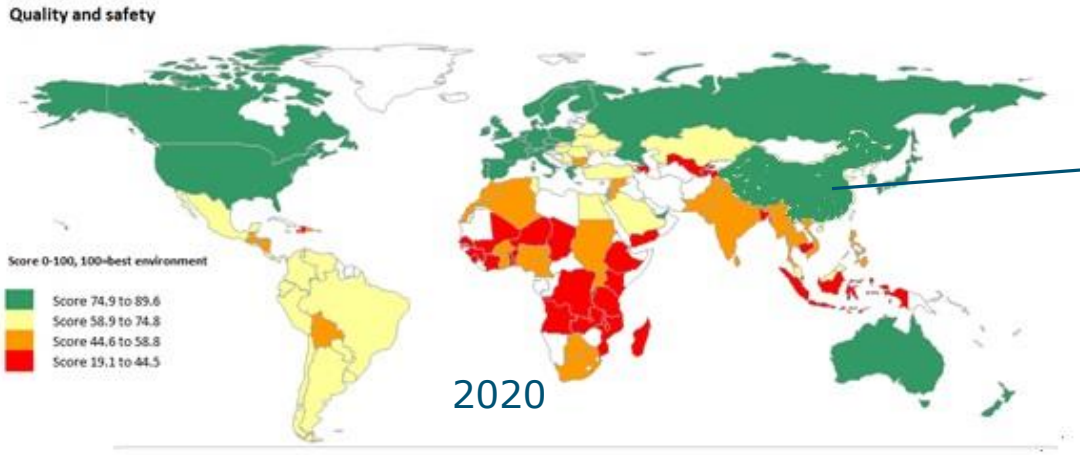
UNIVERSITY & RESEARCH

Prediction generally > 90% using BN. Forecasting also possible



2015: China in class 2 with a score of 69.3
Type of fraud:

1. **Tampering:89%**
2. HC: 1.2%
3. II: 1.2%
4. OL: 8.5%
5. CED: 0%



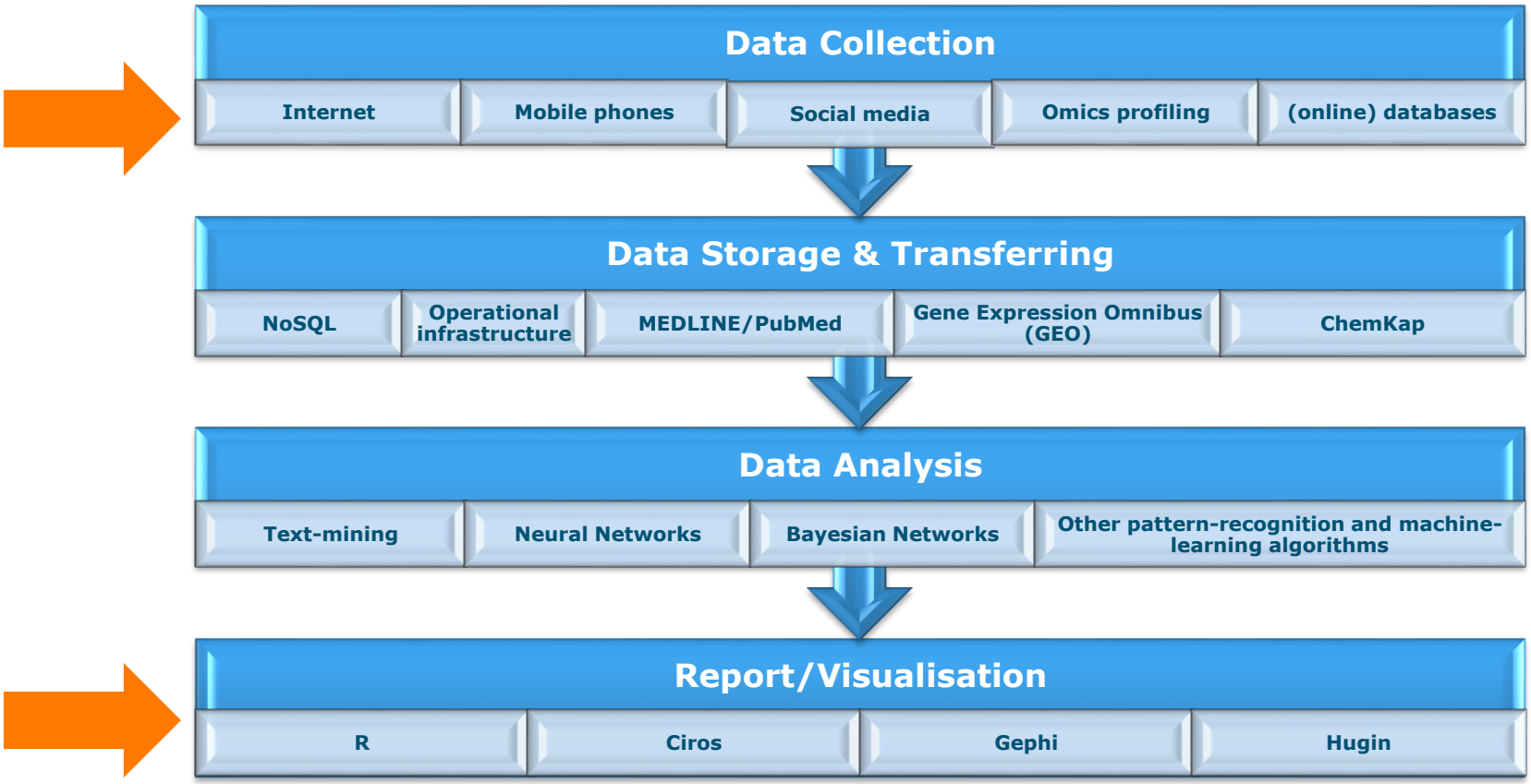
2020: China in class 1 with a score of 80
Type of fraud:

1. Tampering:2%
2. HC: 39%
3. **II: 55%**
4. OL: 0%
5. CED: 4%

WAGENINGEN
UNIVERSITY & RESEARCH

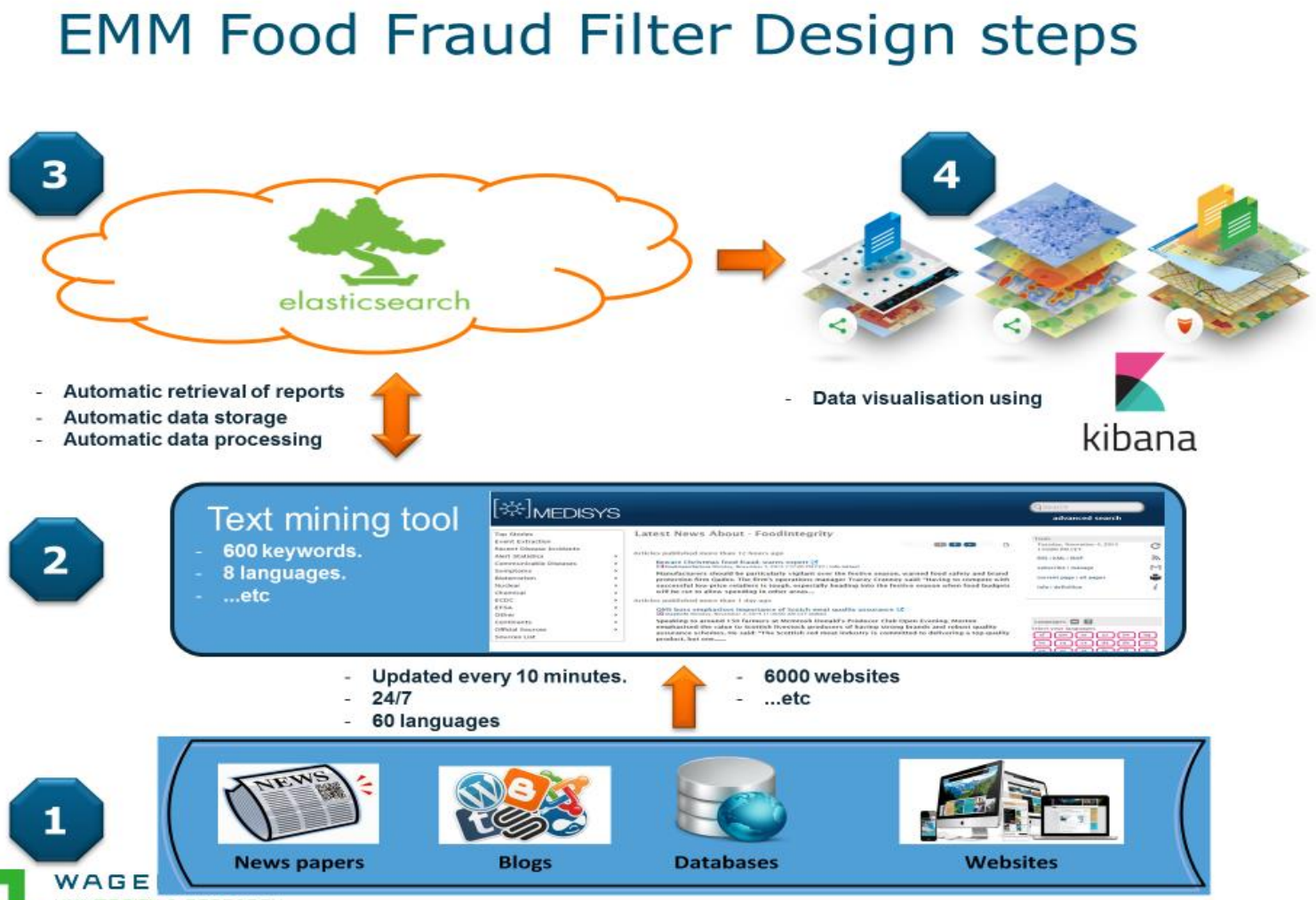
RIKILT started considering Big data in 2015

2. Second data collection and visualisation



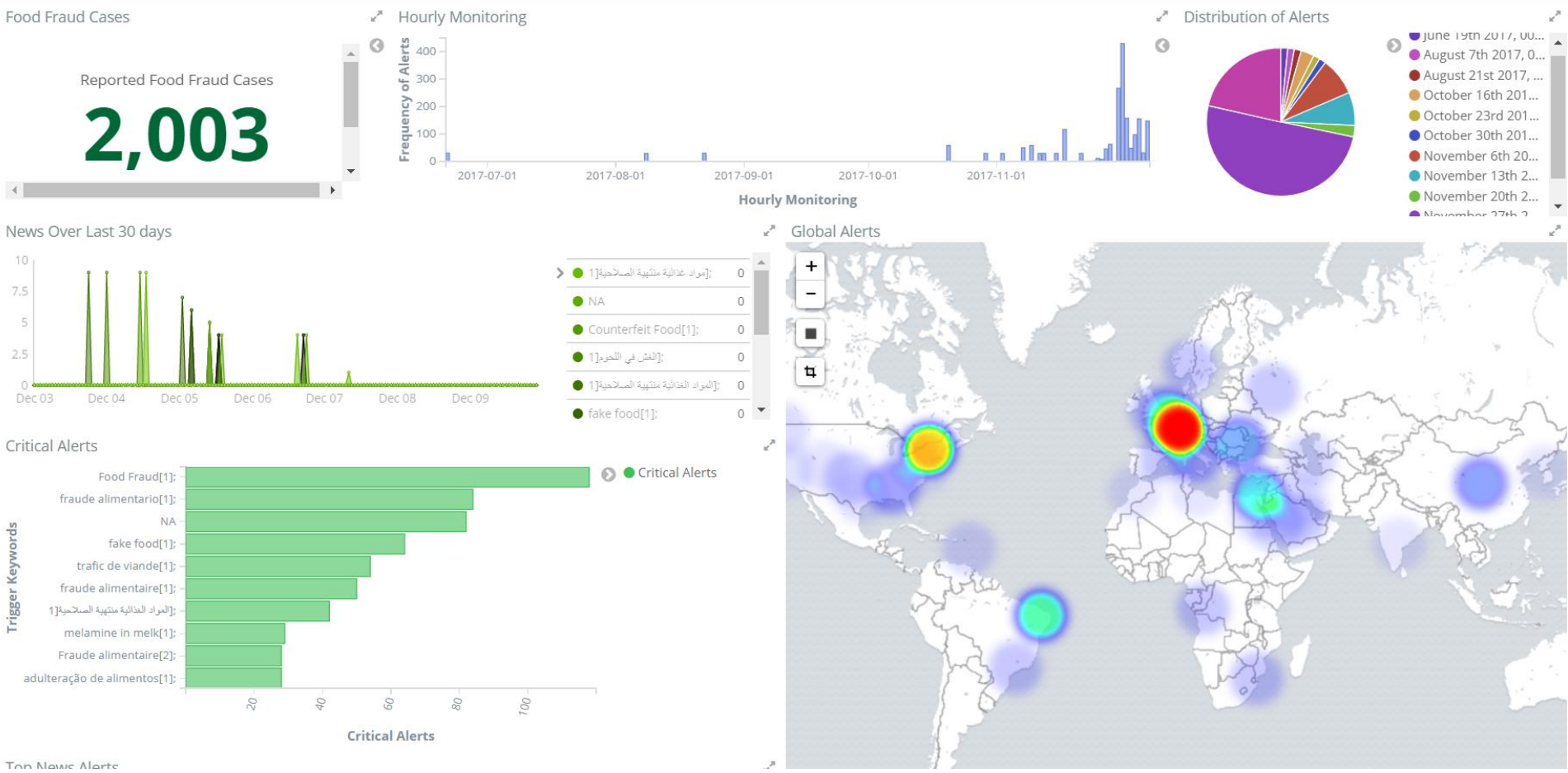
RIKILT started considering Big data in 2015

2. Second data collection and visualisation



WAGE
UNIVERSITY & RESEARCH

4 Data visualisation in a dashboard using Kibana (1)



WAGENINGEN
UNIVERSITY & RESEARCH

Worldwide coverage of media reports on food fraud; 24/7, refreshed every hour

European development: DEMETER project

Objective: to support current (and future) EFSA procedures for emerging risks identification by providing a set of integrated, open-source solutions that will allow EFSA and EU Member State authorities to share data, knowledge and methods in a rapid and effective manner

Duration: 3 years (2017-2020)

WP Objectives (1)

1. To create automated data retrieval, data validation and data mining pipelines that provide input to EFSA staff and responsible national stakeholders in their judgements on identification of possible emerging issues.

2. To develop a methodology to systematically integrate information and data from social sciences or derived from social science methodologies, into the emerging risk identification framework.

3. To build on and further develop EREN member state usage of the EFSA emerging risk identification framework currently in place. Current EREN membership is limited, potentially because barriers to utilising the current system are perceived by potential end-users.

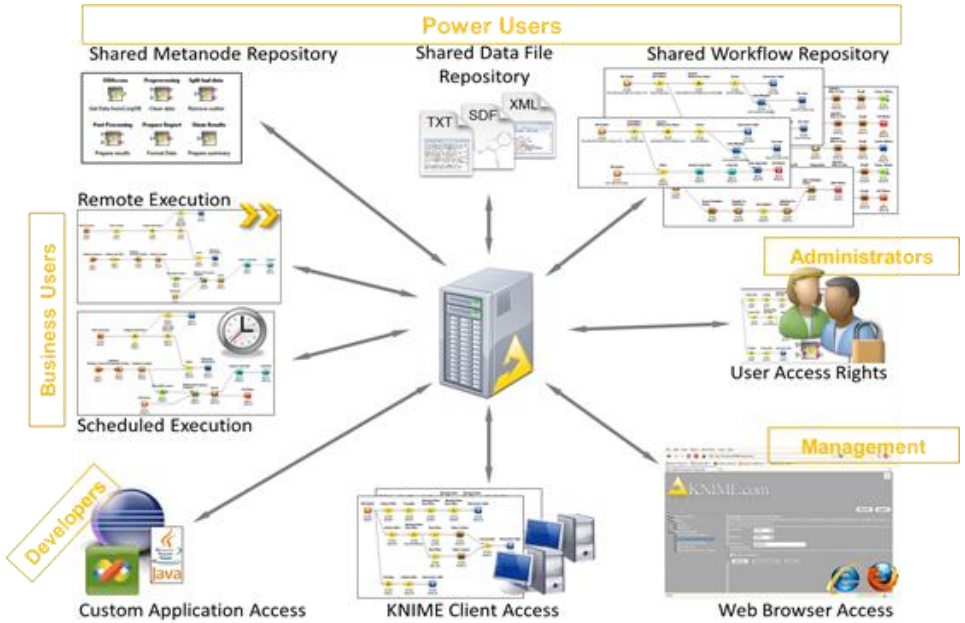
WP Objectives (2)

4. To facilitate software-technical implementation of the desired collaborative platform (ERKEP) supporting current EFSA procedures for emerging risks identification. Specifically this platform will allow EU Member State authorities and EFSA to share knowledge, data and methods for the identification of emerging food-related risks in a rapid and effective manner.

   This task is also responsible for providing necessary internal training on the open source KNIME data mining solution that will be used (among others) within WP1 and WP2

5. To ensure proper coordination, integration and execution of the DEMETER activities on all levels regarding project management & financial matters, communication and dissemination.

DEMETER collaborative platform (ERKEP)



https://www.farcrycms.org/content-management-system-static-site-best/

https://qph.fs.quoracdn.net/main-qimg-16dbf583d24564523d946ecd378d19c1-c

Thank you

Acknowledgement

The following RIKILT WUR colleagues were involved:

Yamine Bouzembrak

Esther van Asselt

Paulien Adamse

Esmée Janssen

Ine van der Fels-Klerx

Peter Hendriksen

Gijs A. Kleter

Nathan Meijer

Bram Steen

Anand Gavai

Martijn Staats