14th Symposium

of the Scientific Committee of the Federal Agency for the Safety of the Food Chain

BIG DATA IN THE FOOD CHAIN: the un(der)explored goldmine?

Tuesday 4th December 2018

Pachéco Center - Finance Tower

Boulevard Pachéco 13

BRUSSELS

SciCom

Federal Agency
for the Safety
of the Food Chain

4th December 2018

Publication of the Scientific Committee and the Staff Direction for risk assessment of the Federal Agency for the Safety of the Food Chain (FASFC)

Editors
Etienne Thiry, Chair Scientific Committee FASFC
Xavier Van Huffel, Director Staff Direction for risk assessment FASFC
Herman Diricks, CEO FASFC

Federal Agency for the Safety of the Food Chain (FASFC)
CA-Botanique
Food Safety Center
Boulevard du Jardin botanique
B-1000 Brussels

Lay-out
Axel Mauroy, Expert Staff Direction for risk assessment FASFC
Gert Van Kerckhove, Communication service FASFC

Table of contents

Welcome address 2

Programme 4

Introduction - Big data definition and objectives of the symposium 7

Session 1: Big data sciences 13

Mining Spaghetti and Lasagna Processes: Bridging the Gap Between Data Science and Process Science 15

Blockchain: concept, critical success factors and possibilities in the food chain 25

When data science meets food safety 37

Session 2: Big data applications to the food chain safety 47

The potential of blockchain technologies in food safety 49

Data value chain in the dairy production: opportunities and challenges 51

Spread model: a forecasting and managing tool in microbiological safety 57

Smart imaging from space for crops management 65

Smart imaging for vector-borne diseases management 75

OMICs in food safety: contribution to chemical risk assessment 83

Session 3: Perspectives 91

Gotcha! Network analytics for Fraud Detection 93

Unlocking the power of citizen science 95

Welcome address

“Big data in the food chain: the un(der)explored goldmine?” is the topic of this 14th annual symposium of the Scientific Committee of the Belgian Food Safety Agency.

The increased amount of numerical data in the food chain opens a new area of research that has to be used in every step of the risk assessment process. The current challenge is to develop adapted methodologies to make the best use of these “Big Data”, especially in our domain, the safety of the food chain.

The objective of the symposium is to present and discuss concepts related to Big Data and the food chain: basic concepts of “Big Data” sciences; “Big Data” applications to the food safety in various domains as microbiological and chemical risks, animal health, vector borne diseases, plant health and crop management, chemical risks; as well as perspectives in fraud detection and citizen participation to data acquisition. Our aim is indeed to deliver to the participants of the symposium an information as exhaustive as possible on “Big Data” related to risk assessment and risk management of the safety of the food chain.

The safety of the food chain is precisely the mission of our Food Safety Agency and the Scientific Committee has the mandate to assist the Agency by its scientific opinions. Our Scientific Committee constitutes a board of experts who gives independent scientific advice to the Chief Executive Officer of the Agency and to the Minister in all matters related to risk assessment and risk management in the food chain; especially regarding food safety but including also animal and plant health. We enforce a policy of impartial and independent scientific consultation and transparent communication and management of conflicts of interest. We acknowledge the strong support of the Agency through its Chief Executive Officer, the General Director of the Directorate General Control policy and the Staff Direction for risk assessment.

I express very special thanks to my fellow members of the Scientific Committee, Prof. Annemie Geeraerd, Dr. Lieve Herman, (vice-chairwoman of the Scientific Committee), Prof. Jeroen Dewulf, Dr. Thierry van den Berg, Prof. Claude Saegerman, to Dr. Pierre Wattiau (former member of the Scientific Committee) and Peter Rakers (Smart Digital Farming Network Manager and invited expert of the Scientific Committee) and to Dr. Xavier Van Huffel (Director) and Dr. Axel Mauroy (expert) both from the Staff Direction for risk assessment, for their invaluable contributions in the working group in charge of the scientific organization of this symposium. These acknowledgements are extended to the Agency and the Staff Direction for risk assessment for providing the needed human and financial resources.

The very high scientific quality of the symposium is supported by prominent speakers coming from Belgium, Germany, Spain and the United Kingdom. They are warmly acknowledged.
Your presence as participants reveals that the selection of this topic by the Scientific Committee was more than appropriate and meets the concerns of scientists, risk assessors, risk managers and stakeholders involved in the analysis of risks in the food chain. Indeed, we all pursue the same objective: to provide safe food to our society.

I wish you a very fruitful symposium.

Prof. Etienne Thiry,
Chair of the Scientific Committee
of the Belgian Food Safety Agency.

Prof. Etienne Thiry graduated as doctor in veterinary medicine in 1980 and doctor in veterinary sciences in 1985. He was recognized in 2001 as diplomate of the European College for Veterinary Public Health. He is full professor and head of veterinary virology and animal viral diseases laboratory, Faculty of veterinary medicine, university of Liège University, Belgium. He is also part-time professor at the Free University of Brussel. He won the International Pfizer award by the international committee of the World Buiatrics Society in 1996 and the Gaston Ramon award by the French Academy of Veterinary Medicine in 2008. He was awarded with the prix de la Francophonie by the Fédération des Associations francophones des vétérinaires d'Animaux de Compagnie (FAFVAC) in 2011 and with the excellence price Jean-Jacques Tondreau by the French association Groupements techniques vétérinaires (GTV 52) in 2012.

Etienne Thiry is member of several Scientific Committees: the National Committee for Microbiology of the Belgian Royal Academy; chairman (2014-2021) of the scientific committee of the Belgian Food Safety Agency (AFSCA); chairman (2012-2018) of the expert committee for animal health and well-being at the French Food Safety Agency (ANSES); vice-chairman of the board of directors of Sciensano. He is also member, previously vice-chairman, of the European Advisory Board on Cat Diseases.

His research interests cover several aspects of animal virology, especially the study of animal virus-host interactions and the evolution of viral populations through genetic recombination and reassortment in herpesviruses, noroviruses, hepeviruses and orbiviruses. These scientific activities generated more than 480 papers in specialised scientific journals. He is the author of four books in the collection Clinical Virology at the Editions du Point Vétérinaire.

SciCom

14th Symposium of the Scientific Committee
of the Federal Agency for the Safety of the Food Chain

PROGRAMME

Big data in the food chain: the un(der)explored goldmine?

Tuesday, 4th December 2018
Pacheco auditorium - Finance Tower
Pachecolaan/Boulevard Pachéco 13
1000 Brussels - Belgium

9:00 REGISTRATION

9:30 ***Welcome*** *2*
Etienne Thiry (ULiège – Chairman SciCom)

9:40 ***Introduction – Big data definition and objectives of the symposium*** *7*
Jeroen Dewulf (UGent – SciCom member)

Session 1: Big data sciences

Chair: Lieve Herman (ILVO, Vice-chairman SciCom), Jeroen Dewulf (UGent, SciCom member)

9:55 ***Mining Spaghetti and Lasagna Processes: Bridging the Gap Between Data Science and Process Science*** *15*
Wil van der Aalst (Aachen University)

10:25 ***Blockchain: concept, critical success factors and possibilities in the food chain*** *25*
Frank Robben (Smals Research)

10:55 ***When data science meets food safety*** *37*
Guy Poppy (Food Standard Agency & University of Southampton)

11:25 Coffee break

Session 2: Big data applications to the food chain safety

Chair: Peter Rakers (ILVO),
Thierry van den Berg (Sciensano, SciCom member)

11:40 ***The potential of blockchain technologies in food safety*** ***49***
Christopher Brewster (TNO)

12:05 ***Data value chain in the dairy production: opportunities and challenges*** ***51***
Stephanie Van Weyenberg (ILVO)

12:30 LUNCH

13:40 ***Forecasting and managing tool in microbiological safety using network analysis*** ***57***
Pablo Fernandez (Universidad Politécnica de Cartagena)

14:05 ***Smart imaging from space for crops management*** ***65***
Pierre Defourny (UCLouvain)

14:30 ***Smart imaging for vector-borne diseases management*** ***75***
Els Ducheyne (AVIA-GIS)

14:55 ***OMICs in food safety: contribution to chemical risk assessment*** ***83***
Matthew Wright (University of Newcastle)

15:20 Coffee break

Session 3: Perspectives

Chair: Annemie Geeraerd (KU Leuven, SciCom member),
Xavier Van Huffel (FASFC)

15:35 ***Gotcha! Network analytics for Fraud Detection*** ***93***
Véronique Van Vlasselaer (SAS)

16:00 ***Unlocking the power of citizen science*** ***95***
Filip Meysman (UAntwerpen)

16:25 ***Closing remarks***
Etienne Thiry (ULiège, Chairman SciCom)

Introduction
Big data definition and objectives of the symposium

Jeroen Dewulf
University of Gent, Belgium
SciCom member

Prof. Jeroen Dewulf graduated in 1998 as a veterinarian from the Faculty of Veterinary Medicine of the Ghent University, Belgium. He immediately started working as researcher at the Herd health department of the faculty of Veterinary Medicine of Ghent. In 2002 he finished his PhD on the epidemiology and control of classical swine fever. In that same year he received a master of science degree in veterinary epidemiology from the University of Utrecht, the Netherlands (Cum Laude). He became diplomate in the European College of Veterinary Public Health in 2005.

Since 2006 he is associated professor in Veterinary Epidemiology at the Faculty of Veterinary Medicine of the Ghent University. In 2014 he was appointed as full professor in Veterinary Epidemiology. His main research interests are quantitative epidemiology and control of zoonoses with a specific emphasis on antimicrobial resistance and antimicrobial use in animal production as well as the prevention of epidemic and endemic diseases with a focus on the application of biosecurity measures.

His research is focused on prevention of endemic and epidemic diseases through biosecurity biosecurity measures and animal health aspects that may influence public health such as antimicrobial use and resistance in animals as well as zoonotic infections. He is the head of the Veterinary Epidemiology Unit and is supervising over 10 PhD students who are doing research in the field of veterinary epidemiology. He is (co-)author of over 270 A1 publications in the field of veterinary epidemiology with a H-index of 41.

He is the principal author of the annual Belgian report on antimicrobial consumption in animals (BelVetSac) and is member of the European Surveillance on Veterinary Antimicrobial Consumption (ESVAC) network and chair of the JPI-AMR network on quantification of antimicrobial consumption in animals at herd level. Since 2009 he is member of the Scientific Committee of the Belgian Food Safety Agency and is the founder and chair of board of the centre of expertise on antimicrobial use and resistance in animals (AMCRA) in Belgium. He is the senior vice president of the European College of Veterinary Public Health.

He is also the author of the book "Biosecurity in animal production and veterinary medicine" as well as the book "8 myths on antimicrobial resistance disproved, practical guide for reducing antibiotic use in animal husbandry".

'Big data' are considered to represent the black gold of the 21st century. A quick search in the PubMed repository for the terms "Big data" demonstrate the ever-increased interest of the scientific community on aspects associated to data sciences during the last decade and a field of sciences in constant evolution (Figure 1).

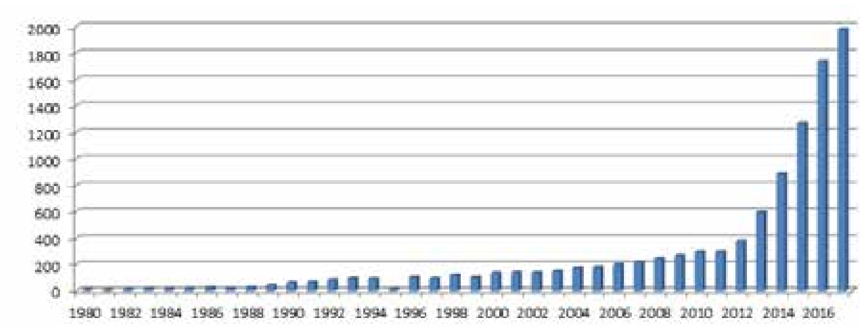

Figure 1: Number of PubMed entries for the terms "big data" (access on the 3rd November 2018).

However big data in the food chain remain largely unexplored. On one hand, they arrive in increasing volumes and with ever-higher velocity. Yet, a very small percentage of them are structured or organized in such a way they can be easily exploited. Data deposits are mostly unknown (12% of data are analyzed on average by companies), while their accessibility can be a factor of competitiveness (50% of decision-makers do not have access to the data they need).

On the other hand, lack of transparency in the use of big data, has recently alarmed society and has animated the public debate. New legislation has been put in place to better protect the privacy of the public (GDPR).

Big data is usually studied through the prism of its four pillars, the 4 'V' (for 'Volume', 'Velocity', 'Variety' and 'Value'), to which is juxtaposed a fifth one: 'Veracity' (representing the uncertainty in the target datasets). Big data analysis can also be split in two ways:

- business intelligence, using descriptive statistics on data with a high density of information in order to measure phenomena or to detect trends;
- Big data, properly, using inferential statistics on data with a low density of information whose large volume makes it possible to infer laws (regression) thus giving to big data (with the limits of the inferencing) predictive capacities.

New concepts associated to 'big data' are building now the relations of tomorrow between consumers, producers and authorities. These concepts can be summarized as following:

- Interoperability: the ability of machines, devices, sensors, and people to connect and communicate with each other via the Internet of Things (IoT) or the Internet of People (IoP);
- Information transparency: the ability of information systems to aggregate raw sensor data to higher-value context information. Among other opportunities, this can increase traceability and potentially consumer's confidence in the food chain;
- Technical assistance: First, the ability of assistance systems to support humans by aggregating and visualizing information comprehensibly for making informed decisions and solving urgent problems on short notice. Second, the ability of cyber physical systems to physically support humans by conducting a range of tasks that are unpleasant, too exhausting, or unsafe for their human co-workers;

- Decentralized decisions: The ability of cyber physical systems to make decisions on their own and to perform their tasks as autonomously as possible (or also artificial intelligence). Only in the case of exceptions, interferences, or conflicting goals, are tasks delegated to a higher level.

The advanced analysis of big data opens vast potential for different stakeholders in the food chain. Farmers rely on predictive analytics and other big data solutions to help them with a better yield at lower cost. Big data-driven analytics support food industry businesses with critical decision-making capabilities in the areas of pricing, product promotion, product development and demand forecasting. Food delivery companies may utilize data in diverse ways to estimate delivery times and price. Consumers demand better transparency into their food lifecycle - where is it coming from, how was it produced, how does distribution work, etc. This encourages stakeholders to adopt big data solutions to acquire/provide better information about the food production and distribution process.

The application of big data for food safety purposes remains up till now rather exceptional, certainly by national authorities which are confronted with scattered data across the food chain and lack of expertise and fit-for-purpose tools for data- and process analysis. During this 14th symposium of the Scientific Committee of the Belgian Food Safety Agency, various aspects of data science application in the food chain will be discussed by experts in this domain.

The organizers of this symposium have realized a great effort to set up an interesting and challenging programme for professionals who are involved in safeguarding the safety of the food chain by combining state-of-the-art presentations focused on the science of big data and the principles of new technologies such as blockchain with presentations showing more practical applications of big data in different domains of the food chain and the related challenges and opportunities.

It is hoped that this symposium will trigger new developments and future projects around food safety risk assessment and risk control.

Session 1:
Big data sciences

Mining Spaghetti and Lasagna Processes: Bridging the Gap Between Data Science and Process Science

Wil van der Aalst

Process and Data Science, RWTH Aachen University, Germany

Prof.dr.ir. Wil van der Aalst is a full professor at RWTH Aachen University leading the Process and Data Science (PADS) group. He is also part-time affiliated with the Technische Universiteit Eindhoven (TU/e). Until December 2017, he was the scientific director of the Data Science Center Eindhoven (DSC/e) and led the Architecture of Information Systems group at TU/e. Since 2003, he holds a part-time position at Queensland University of Technology (QUT). Currently, he is also a visiting researcher at Fondazione Bruno Kessler (FBK) in Trento and a member of the Board of Governors of Tilburg University.

His research interests include process mining, Petri nets, business process management, workflow management, process modeling, and process analysis. Wil van der Aalst has published more than 200 journal papers, 20 books (as author or editor), 450 refereed conference/workshop publications, and 65 book chapters. Many of his papers are highly cited (he is one of the most cited computer scientists in the world and has an H-index of 138 according to Google Scholar with over 85,000 citations) and his ideas have influenced researchers, software developers, and standardization committees working on process support. He has been a co-chair of many conferences including the Business Process Management conference, the International Conference on Cooperative Information Systems, the International Conference on the Application and Theory of Petri Nets, and the IEEE International Conference on Services Computing. He is also editor/member of the editorial board of several journals, including Business & Information Systems Engineering, Computing, Distributed and Parallel Databases, Software and Systems Modeling, Computer Supported Cooperative Work, the International Journal of Business Process Integration and Management, the International Journal on Enterprise Modelling and Information Systems Architectures, Computers in

Industry, IEEE Transactions on Services Computing, Lecture Notes in Business Information Processing, and Transactions on Petri Nets and Other Models of Concurrency. He is also a member of the Council for Physics and Technical Sciences of the Royal Netherlands Academy of Arts and Sciences and serves on the advisory boards of several organizations, including Fluxicon, Celonis, Processgold, and Bright Cape. In 2012, he received the degree of doctor honoris causa from Hasselt University in Belgium. He also served as scientific director of the International Laboratory of Process-Aware Information Systems of the National Research University, Higher School of Economics in Moscow. In 2013, he was appointed as Distinguished University Professor of TU/e and was awarded an honorary guest professorship at Tsinghua University. In 2015, he was appointed as honorary professor at the National Research University, Higher School of Economics in Moscow. He is also a member of the Royal Netherlands Academy of Arts and Sciences (Koninklijke Nederlandse Akademie van Wetenschappen), Royal Holland Society of Sciences and Humanities (Koninklijke Hollandsche Maatschappij der Wetenschappen), and the Academy of Europe (Academia Europaea). In 2017 he was awarded a Humboldt Professorship, Germany's most valuable research award.

ABSTRACT

Process mining provides a generic collection of techniques to turn event data into valuable insights, improvement ideas, predictions, and recommendations. Process mining is the missing link between model-based process analysis and data-oriented analysis techniques. As such, it is an important bridge between data science (including Big data, machine learning, and artificial intelligence) and process science. Obviously, food chains are composed of collections of processes that increasingly generate events. Processes in the "farm to fork" supply chain (i.e., cultivation, processing, packaging, distribution, and preparation) can benefit from process-aware data analytics. Experienced process miners already talk about "Spaghetti and Lasagna processes" to describe process characteristics in terms of food. However, food safety and avoidance of waste and inefficiencies can benefit from process mining.

DATA SCIENCE: BEYOND THE HYPE

In recent years, data science emerged as a new and important discipline. It can be viewed as an amalgamation of classical disciplines like statistics, data mining, databases, and distributed systems (see Figure 1). The spectacular growth of the digital universe, summarized by the overhyped term "Big Data", makes it possible to record, derive, and analyze events. Events may take place inside a machine (e.g., an X-ray machine, an ATM, or baggage handling system), inside an enterprise information system (e.g., an order placed by a customer or the submission of a tax declaration), inside a hospital (e.g., the analysis of a blood sample), inside a social network (e.g., exchanging e-mails or twitter messages), inside a transportation system (e.g., checking in, buying a ticket, or passing through a toll booth), etc. Events may be "life events", "machine events", or "organization events". The term Internet of Events (IoE) refers to all event data available. Data scientists assist organizations in turning data into value. A data scientist can answer a variety of data-driven questions. These can be grouped into the following four main categories: (1) Reporting: What did it happen?, (2) Diagnosis: Why did it happen?, (3) Prediction: What will happen?, and (4) Recommendation: What is the best that can happen?

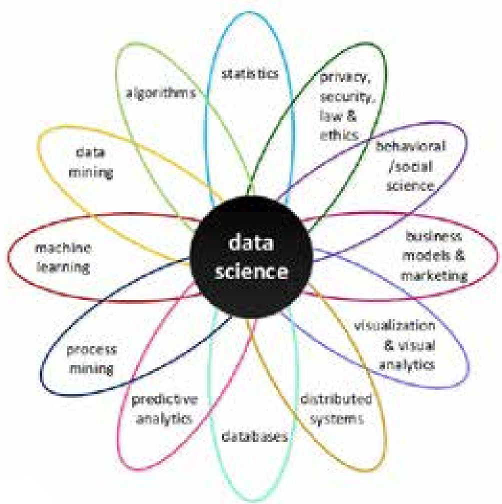

Figure 1: Data science combines different disciplines

The data science pipeline shown in in Figure 2 relates the different "ingredients" of any larger initiative that aims to "turn data into value". The main challenge in infrastructure is to make things scalable and instant, e.g., handling data that does not fit into a single computer or that cannot be stored. The main challenge in analysis is to provide answers to known and unknown unknowns (quoting Donald Rumsfeld). Analysis may be question- or problem-driven or one may simply explore data and processes to see improvement potential. The main challenge in effect is to do all of this in a responsible manner (looking at aspects such as fairness, accuracy, confidentiality, and transparency). One would like to use the data without negative side effects.

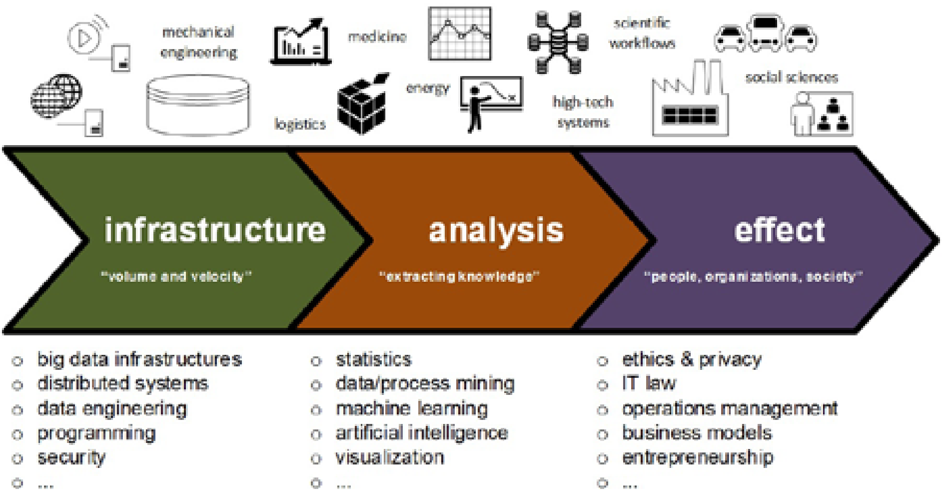

Figure 2: Data Science pipeline

Although the elements described in the two figures are broadly accepted as the key ingredients for data science, investments in Europe are lagging behind the US and Asia. The "platforms" on which our society and business run (Windows, iOS, Android, Amazon, Google, Oracle, etc.) are all developed in the US. Thinking about a data science solution as a "pizza", the focus of European organizations is on the "pizza toppings" and not on the "pizza base". Pizza toppings include elements like smart mobility, smart energy, personalized health, e-government, etc. However, these elements can only be successful when one is able to exploit powerful data science platforms (databases, distributed systems, analytics tools, machine learning algorithms, etc.). Unfortunately, the pizza base is often simply imported from the US. Organizations simply purchase a database system or a Big data solution without considering this dimension.

To remain competitive and innovate food chains, we need to bring together data science experts and domain experts. However, attention needs to be paid on the core technologies that make all of this possible. Often "the winner takes it all" (economy of scale), making it important to become leading in data science. The platforms controlled by US organizations on personal computers (Windows), phones and tablets (iOS and Android), shopping (Amazon), search (Google), and databases (Oracle) illustrate this.

PROCESS MINING

Process mining is one of the few data science areas where Europe is clearly leading. Process mining can be seen as a means to bridge the gap between data science and process science (see Figure 3). Data science approaches tend to be process-agnostic whereas process science approaches tend to be model-driven without considering the "evidence" hidden in the data. Process mining seeks the confrontation between event data (i.e., observed behavior) and process models (hand-made or discovered automatically). Mainstream data science approaches tend to be process agnostic. Data mining, statistics and machine learning techniques do not consider end-to-end process models. Process science approaches are process-centric, but often focused on modeling rather than learning from event data. The unique positioning of process mining makes it a powerful tool to exploit the growing availability of data for improving end-to-end processes.

Event logs can be used to conduct three types of process mining. The first type of process mining is discovery. A discovery technique takes an event log and produces a model without using any a-priori information. If the event log contains information about resources, we can also discover resource-related models, e.g. a social network showing how people are working together in an organization.

The second type of process mining is conformance. Here, an existing process model is compared with an event log of the same process. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa. For instance, there may be a process model indicating that purchase orders of more than one million Euro require two checks.

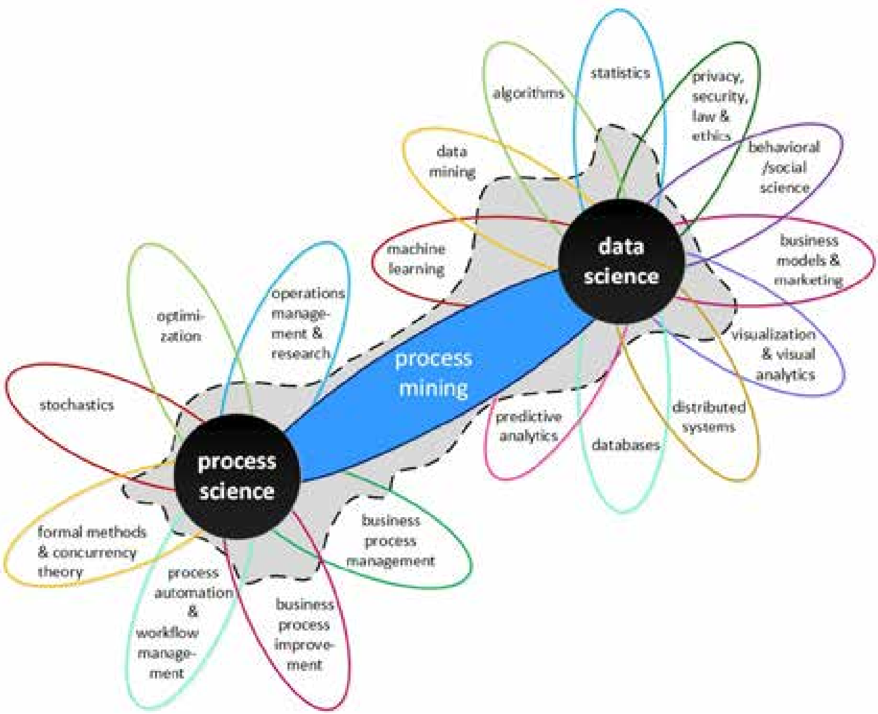

Figure 3: Process mining as the bridge between data science and process science

The third type of process mining is enhancement. Here, the idea is to extend or improve an existing process model using information about the actual process recorded in some event log. Whereas conformance checking measures the alignment between model and reality, this third type of process mining aims at changing or extending the a-priori model. One type of enhancement is repair, i.e., modifying the model to better reflect reality. For example, if two activities are modeled sequentially but in reality can happen in any order, then the model may be corrected to reflect this. Another type of enhancement is extension, i.e., adding a new perspective to the process model by cross-correlating it with the log. An example is the extension of a process model with performance data.

When analyzing event data one soon learns that there are two types of processes: Lasagna and Spaghetti processes. Unlike Spaghetti processes, Lasagna processes have a clear structure and most cases are handled in a prearranged manner. There are relatively few exceptions and stakeholders have a reasonable understanding of the flow of work. Spaghetti processes are more chaotic and many cases follow a unique path. Obviously, it is more difficult to analyze Spaghetti processes. Nevertheless, such processes are very interesting from the viewpoint of process mining as they often allow for various improvements. A highly-structured well-organized process is often less interesting in this respect; it is easy to apply process mining techniques but there is also little improvement potential. Therefore, one should not shy away from Spaghetti processes as these are often appealing from a process management perspective. Turning Spaghetti processes into Lasagna processes can be very beneficial for an organization.

RESPONSIBLE DATA SCIENCE

The importance of data science is widely acknowledged, but there are also great concerns about the use of data. Increasingly, customers, patients, and other stakeholders are concerned about irresponsible data use. Automated data decisions may be unfair or non-transparent. Confidential data may be shared unintentionally or abused by third parties. Each step in the "data science pipeline" (from raw data to conclusions, see figure) may create inaccuracies, e.g., if the data used to learn a model reflect existing social biases, the algorithm is likely to incorporate these biases. These concerns could lead to resistance against the large-scale use of data and make it impossible to reap the benefits of data science. Rather than avoiding the use of data altogether, we strongly believe that data science techniques, infrastructures and approaches can be made responsible by design.

Therefore, the author initiated the Responsible Data Science (RDS) consortium in 2015 in which leading Dutch research groups from multiple disciplines joined forces to develop novel solutions for problems related to Fairness, Accuracy, Confidentiality, and Accuracy (FACT) in data science.

In the meantime, the General Data Protection Regulation (GDPR) was established. However, creating stricter laws is only part of the solution. It is possible to create positive technological solutions. For example, one can do process mining on encrypted data that allows parties to get analysis results without having access to individual events or cases. Moreover, it is possible to create novel analysis approaches with predefined fairness criteria. The possibilities to create positive technological solutions ensuring the responsible use of data science are there but need to be developed further and become visible for policy makers and and-users.

FOOD FOR THOUGHT

Using metaphors like the data science pizza and Spaghetti and Lasagna processes, we tried to introduce process mining as an emerging new technology in food chains. The dynamic nature of food chains and the discrete nature of products fits well with the preconditions to apply process mining. Process mining can be used to address compliance and performance issues in food chains.

References

Kerremans M. 2018. Market Guide for Process Mining. Gartner Report.

van der Aalst W.M.P. 2016. Process Mining: Data Science in Action. Springer-Verlag, Berlin.

van der Aalst W.M.P. 2017. Responsible Data Science: Using Big Data in a "People Friendly" Manner. In Hammoudi S., Maciaszek L., Missikoff M., Camp O. & Cordiero J., editors, Enterprise Information Systems, volume 291 of Lecture Notes in Business Information Processing, pages 3-28. Springer-Verlag, Berlin.

van der Aalst W.M.P., Bichler M. & Heinzl A. 2017. Responsible Data Science. Business and Information Systems Engineering, 59(5):311-313.

van der Aalst W.M.P. 2013. Business Process Management: A Comprehensive Survey. ISRN Software Engineering, pages 1-37.

van der Aalst W.M.P., Adriansyah A. & van Dongen B. 2012. Replaying History on Process Models for Conformance Checking and Performance Analysis. WIREs Data Mining and Knowledge Discovery, 2(2):182-192.

Blockchain: concept, critical success factors and possibilities in the food chain

Frank Robben, Kristof Verslype

Smals Research, Belgium

Since 1991 Frank Robben is the general manager of the Crossroads Bank for Social Security, a public institution he conceived and founded, which ensures secure electronic data exchange between all actors in the social sector. Since 2008, he is also the general manager of the eHealth platform, a public institution responsible for organising and supporting secure electronic data exchange between all actors in the health sector with a view to improving the quality of care and patient safety, reducing the administrative burden on patients and health care providers and supporting scientific research in the field of health. He also conceived and founded this institution.

Frank Robben was also appointed managing director of Smals, the main ICT service provider in the Belgian social security sector and is a member of the Data Protection Authority. Frank was awarded the title of Flemish public manager of the year 2005.

ABSTRACT

Blockchain is a relatively new concept that has initially been applied in 2009 with the launch of Bitcoin. It allows for disintermediation; processes that traditionally require dependencies of a central and/or intermediary entity can now – at least conceptually – be organized without this entity. The technology is enormously hyped and expectations are huge. This article clarifies when a blockchain approach can be useful, draws lessons from projects worldwide and gives additional recommendations, based own experience, to increase chances to go life with a blockchain project.

WHAT IS BLOCKCHAIN?

Blockchain is a technology for disintermediation; processes that traditionally required dependencies is central and/or intermediary entity can now – at least conceptually – be organized without this entity, or at least the reliance on these entities can be reduced. The technology is enormously hyped and expectations are huge. The Harvard Business Review states (Iansiti & Lakhani, 2017):

> *"Blockchain is not a "disruptive" technology, which can attack a traditional business model with a lower-cost solution and overtake incumbent firms quickly. Blockchain is a foundational technology: it has the potential to create new foundations for our economic and social systems. But while the impact will be enormous, it will take decades for blockchain to seep into our economic and social infrastructure. The process of adoption will be gradual and steady, not sudden, as waves of technological and institutional change gain momentum."*

But what is a blockchain? It is an append-only data structure that is collectively maintained by a group of participants. Many keep a local copy of the blockchain and the consensus mechanism guarantees that all agree on the same version. Hence, a malicious subset of participants cannot tamper the blockchain or negatively impact its correct functioning.

Three categories of applications of the technology can be distinguished:

- **Registration of facts.** Once data is registered in a blockchain, it can no longer be removed or changed. Tampering with the registration time is equally hard. This does not imply that the data itself is stored on the blockchain. Often, a unique fingerprint (cryptographic hash) suffices. E.g. the different steps in a supply chain could be registered in a blockchain. Additionally, meat processing companies could be issued a license on the same blockchain. Combined, this informs the consumer not only about the trajectory of the food, but also about the different companies involved.

- **Transfer of assets.** Everything of value can be represented on a blockchain: virtual currencies (e.g. Bitcoin) licenses, diplomas, domain names, securities, gold, cars, real estate, … The participants (the network) enforce that the creation, registration, transfer and destruction of assets happens according to the rules. Each of these steps is registered on the blockchain.
- **Enforcement of rules.** The transfer of assets requires the enforcement of relatively simple rules. With smart contracts, any kind of rule expressible in computer code can be enforced collectively by the blockchain network. Often it comes down to: " If conditions A & B are met, transfer assets to X. " For instance, real estate can be transferred only if the requested amount of (virtual) money is paid, if there is a valid soil certificate and if the notary has confirmed that buyer and seller are correctly informed. As a second example, a shop could order meat using a smart contract. He transfers the money to the contract. During the meat transport, smart sensors in the refrigerated vehicle register the conditions under which the meat is transported - such as temperature and humidity - and register it on the blockchain. If the meat was transported under the agreed conditions, the smart contract transfers the money to the meat supplier and indicates that the shop owns the meat. In the other case, it sends the money back to the shop.

Blockchain offers transparency & real-time auditability; participants can 1) verify when what information has been registered, 2) see the history of an asset and 3) verify whether the rules have been correctly enforced by the network. This could be convenient for supervisory authorities.

BAD CASE, GOOD CASE

Blockchain is about trust and robustness. If none of the participants has an issue with the existence of a central, trusted entity on which they depend, a blockchain approach is most likely a bad idea. Other properties that blockchain can realize, can usually more also, and more easily, be realized with traditional and more mature technologies. Think of process automation, consistency of databases, streamlined processes, real-time updates and insight in the decision-making process. This section presents negative and positive blockchain business cases.

A first negative example is a typical flow of personal data in government context, as shown in the figure.

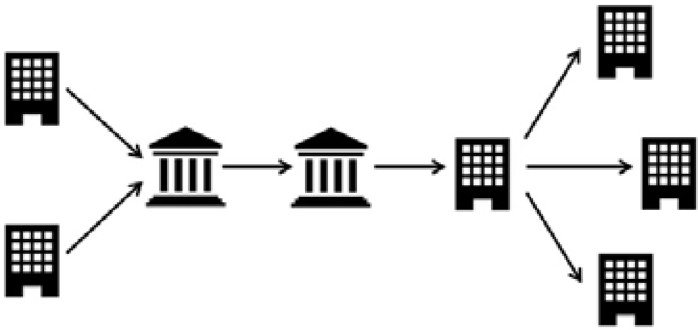

Figure 1: Typical flow of personal data in government context.

A blockchain approach Is not useful if the three central entities are maintained. The government interested in a blockchain approach was confronted not with a trust issue, but with a complexity issue. It was hoping that blockchain could reduce this complexity while maintaining the central entities. However, a blockchain approach requires a more complex infrastructure and extra cryptography to sufficiently protect the personal data. The blockchain approach has rightfully been abandoned.

A second negative example is found in the Netherlands (application for a wheelchair - Gemeente Stichtse Vecht - Rolstoelgebruik in Blockchain, Blockchain pilots Dutch Government, blockchainpilots.nl, Pilotronde 2, 2017). Currently, the citizen submits a request for a specific wheelchair to the municipality, which, in turn, contacts several care providers. In the proposal for a private blockchain, not only the care providers, but also the municipality would participate. The advantage is that the municipality cannot tamper and does not control the data. However, this approach requires a server for each participant, as well as a more complex implementation. It's unlikely that the reduction in dependency outweigh these extra costs.

A similar blockchain case is found in Flanders (Rousseau & Buytaert, 2017). Currently, a citizen applies at multiple social housing companies, which is cumbersome, and also transparency is lacking. The citizen does not know why she has not yet been assigned a social house. The social housing companies would have to participate in a blockchain network. It turned out that they prefer the cheaper and more convenient approach of a centralized system.
A first positive example are therapeutic relations in Belgium. They determine if a care provider has access to the patient's health data. Today, this information is dispersed over several entities that are to a certain extent competitors that are suspicious to hand over sensitive medical data to a central service. Also here, the citizens lacks overview. From a conceptual level, this is a good blockchain case.

A second positive example are cross-border processes, such as international container traffic, nuclear waste transport, food supply chains and the issuance and showing of diplomas in an international context. A centralized approach would raise the question about what country should be the primus inter pares and, hence, be responsible for the system.

A third positive example is the demonstrability service (Verslype, 2018). Within the government an eBox is used for exchanging documents between end users belonging to a specific sector. Different organizations each represent a different, non-overlapping part of the end users (see figure).

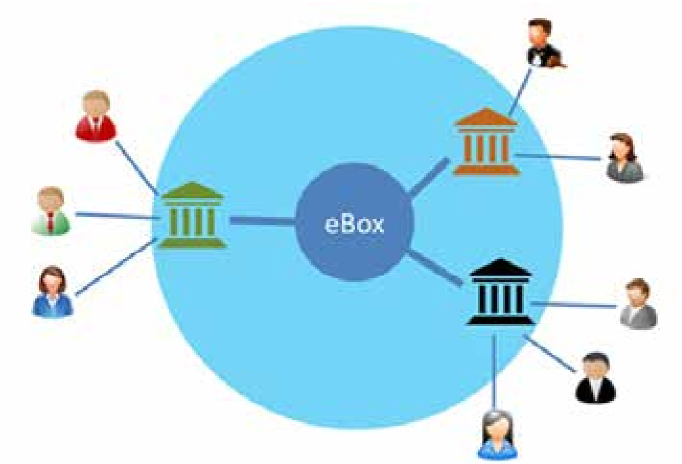

Figure 2: schematic representation of the use of an ebox in a governmental context

When Alice sends a message to Bob, Alice sends the message through her organization to the eBox and later Bob downloads it from the eBox through his organization. We need a proof that the document was sent by Alice at a certain moment, and that it was received by Bob at a certain moment. Unfortunately, the end users do not trust the eBox sufficiently, nor do they really trust each other. A blockchain approach is not only useful here, but also allows each participant to upgrade its role.

EXPERIENCES & LESSONS

Now that we can identify good business cases, let's draw some lessons from concrete projects.
The World Food Program of the United Nations offers a cash-for-food program for 100.000 Syrians in Jordan refugee camps. Thanks to blockchain technology no longer had to rely on local banks, which allowed a reduction of 98% of the transactions costs (Juskalian, 2018). Hence, a strong cost reduction is indeed possible by eliminating intermediary parties.

Several players in the financial sector are doing blockchain experiments. Swift has built with 34 banks a PoC for account settlement (Adoption of DLT presents significant operational challenges for Swift member banks, Finextra, 8 March 2018). Although the PoC was allegedly a resounding success, SWIFT acknowledges significant re-engineering costs and operational challenges. Ripple has built a PoC for international payments, but also had to acknowledge that blockchain technology faces significant scalability, confidentiality and flexibility challenges. Their new system, xCurrent, enables instant settlement, but is not based on a distributed ledger such as blockchain (Irrera, 2018). The World bank is issuing bonds with blockchain technology. This reduced the settlement time from five days to a few seconds, which is an immense improvement. On the other hand, the promise of transparency and reduced transaction costs has not yet been realized (Sender, 2018). The financial industry was the first to start experimenting with blockchain and are investing millions in it. The three examples illustrate the challenges and that not all promises are realizable today.

The previous section mentioned the diploma case as a good one. However, this does not necessarily result in an optimal solution, as can be observed from the blockchain PoC by the Flemish government (Daniels, 2018). The participants in the blockchain network would be schools and governments. All expected functionality was present. However, the PoC ignored security and privacy aspects. Every participant had full access to the diploma data of each citizen (personal data). Instead of distributing trust over the participants, it is multiplied by their number. Coming up with a blockchain solution that does take into account security and privacy would severely reduce the functionality. Blockchain not only enables transparency, but it also requires it. It can be hard – although not impossible – to reconcile this transparency with strong confidentiality and privacy requirements. Michèle Finck, research fellow at the Max Planck Institute for Innovation and lecturer at the University of Oxford, states: "There are many tensions and uncertainties between GDPR and blockchain and many blockchain projects are likely not compatible with GDPR." (EU Blockchain Observatory and Forum, 2018).

If a centralized approach is undesirable, blockchain might be a good approach, which does not imply that it necessarily is one. Often, we have more choices than either centralized or blockchain. In the diploma case, we can envisage a system where each country stores the diplomas issued by its own accredited educational institutions. When a citizen wants to manage or show her diplomas, she contacts the service provided by her country and maintains references to diplomas that the citizen obtained elsewhere. This results in a decentralized system without blockchain. We also see blockchain regularly in the context of self-sovereign identity, which enables the citizen to manage her identity and to selectively disclose personal data. She could just prove that she is an adult, without disclosing all information on her identity card. However, if we are only interested in this selective disclosure of personal attributes, another technology, called attribute-based credentials (Rannenberg et al., 2015), offers better privacy properties and does not require a full blockchain network. It's good to look at new technologies such as blockchain. Just don't forget that also alternative, less visible technologies exist.

The remainder of this section discusses blockchain experiments specifically in supply chain context.
Already in 2006, before blockchain existed, Walmart launched a project to trace the provenance of products by using RFID tags. The projects has been abandoned due to high investments and complexity at the side of the producers (Gaudin, 2008). Although meanwhile these costs might have lowered, we should realize that this will not be solved by using blockchain. Blockchain is not a full business solution, it is just one component in a bigger system.

Often over 30 parties involved in container transport from A to B. There is a low degree of digitization and a lot of paper works that constitutes over 50% of the total costs. During the transport, the same pin code is passed and reused, resulting in security risks. Therefore, Maersk, IBM and around hundred other companies form a joint venture. Their aim is more transparency and simplicity of cross-border transport of goods with an open blockchain platform for the sector. However, their blockchain-based solutions was rejected by its rivals (Andersen & Vogdrup-Schmidt, 2018). Why? Because within the sector, many similar projects are being developed, of which the one of Maersk is most the prominent one. Companies don't want to abandon theirs in favor of one ofa competitor. "We are going to waste a lot of money" says Hapag-Lloyd CEO Rolf Habben Jansen, who adds that common standards and a joint solution are necessary. This emphasizes the

need for a common project, in which all, or at least the most prominent, stakeholders are involved.
In conclusion, we see encouraging results, but not all blockchain projects result in a reduction of costs and blockchain is just one part of a complete solution. Additionally, today we are still facing several challenges. And we must not forget that there is more in the world than blockchain.

BEYOND PROOF OF CONCEPT

This section complements the guidance in the previous sections with concrete advice for your own blockchain project and stems from experiences of Belgian government agencies.
Clarify your ambitions before starting with a PoC. We distinguish four levels. 1) Maybe you just want to announce that you company is experimenting with the technology. As long as the hype lasts, this might have a positive effect on the company's prestige, media attention and share price. From a blockchain perspective, this is the least challenging. 2) It's more useful if you want to discover the new possibilities of the technology. The actual implementation of the PoC can be outsourced, and you focus on what the technology can do for you on a business level. 3) Maybe you want to acquire technical knowledge and competences in your own company. In that case, make sure to integrate a learning path in the process. 4) The most challenging, but also the most useful is building a PoC as a preparation towards a production-ready system. This requires a profound analysis to ensure that all possible issues, such as confidentiality and privacy, can be sufficiently covered.
The business model of many blockchain start-ups is selling PoCs. To avoid surprises, it is therefore paramount that you make clear agreements (on paper) beforehand if you are planning to go further. We heard start-ups saying: "We make abstraction of the GDPR", "Given the limited budget, we do not do a prior analysis" and even "The blockchain PoC was not meant to run in a distributed way".

If you want to go life, we also recommend to start small. This is what for instance Carrefour and Belfius did. Carrefour uses blockchain to control the supply chain of fresh products (Biesmans, 2018). They started with eggs from the Auvergne region and are gradually expanding. Belfius uses blockchain to incentivize children to go by bike or on foot to school. For now, they rolled it out in three Belgian towns. Carrefour and Belfius start small, but with the ambition to expand.
From our experience, we also recommend to initially avoid using blockchain technology that is too complex given the young and immature market, unless you have plenty of resources. And, finally, avoid technology lock-in by providing a migration path. Because today we cannot know what blockchain technology or technologies will become the de facto standards.

CONCLUSIONS

Blockchain is seen as a solution for many problems. This article clarified when it can have an added value, but also drew lessons from experiences in blockchain projects internationally and nationally, including in supply chain management. Additionally, recommendations are given for organizations, based on our own experiences.

References

Andersen O. & Vogdrup-Schmidt L. 2018. Rivals reject blockchain solution from Maersk and IBM. Shipping Watch.

Biesemans J. 2018. De blockchain van de kip en het ei, EOS Magazine.

Daniels K. 2018. Certified for Life—International exchange & authentication of diplomas via blockchain. Medium and Forum - GDPR Workshop Report.

Gaudin S. 2008. Some Suppliers Gain from Failed Wal-Mart RFID Edict, CIO.

Iansiti M. & Lakhani K. 2017. The Truth About Blockchain. Harvard Business Review.

Irrera A. 2018. Banks unlikely to process payments with distributed ledgers for now, says Ripple, Reuters.

Juskalian R. 2018. Inside the Jordan refugee camp that runs on blockchain. MIT Technology Review.

Rannenberg K., Camenisch J., Sabouri A. 2015. Attribute-based Credentials for Trust: Identity in the Information Society. Identity in the Information Society, Springer.

Rousseau P. & Buytaert P. 2017. Sociale woningen toewijzen via blockchain, Infosessie blockchain. Informatie Vlaanderen.

Sender H. 2018. World Bank breaks ground with blockchain bond sale. Financial Times.

Verslype K. 2018. BeSure - A realistic blockchain case for the Belgian government. Medium.

When data science meets food safety

Guy Poppy [1], **Ben Goodall** [2]

1Chief Scientific Adviser, Food Standards Agency's & University of Southampton, The United Kingdom

2Food Standard Agency, The United Kingdom

Prof. Guy Poppy took up his role as the FSA's Chief Scientific Adviser in August 2014. He also continues with his research in global food security at the University of Southampton, where he is Professor of Ecology and previously directed Interdisciplinary Research across 11 research themes and 4 institutes.

Professor Poppy has significant research experience in food systems and food security and has advised governments around the world on these issues. He has published over 100 peer-reviewed papers including a number of highly cited articles on risk assessment, risk analysis and risk communication. He is currently a member of the Research Excellence Framework (REF2021) panel assessing the quality of agriculture, food and veterinary science in the UK, having previously served on the REF2014 panel.

A graduate of Imperial College and Oxford University, Professor Poppy previously worked at Rothamsted Research, becoming Principal Scientific Officer. He left in 2001 to join the University of Southampton where he has been Head of Biodiversity and Ecology and, more recently, Head of Biological Sciences.

As the FSA's Chief Scientific Adviser, Professor Poppy provides expert scientific advice to the UK government and plays a critical role in helping to understand how scientific developments will shape the work of the FSA as well as the strategic implications of any possible changes. His series of CSA reports have reached a very wide audience and have had impact on issues ranging from AMR to Big data and Whole Genome Sequencing through to the Food Hygiene Rating Scheme (FHRS). He has focused on connecting science to those using it and has pushed for scientists to be intelligent providers to intelligent customers of science within the FSA and beyond.

ABSTRACT

Data and data science present both opportunities and challenges to a regulator working in an ecosystem as diverse and complex as the UK Food and Drink Sector. Here, some of the Food Standards Agency's recent efforts in applying data science to support UK food safety and authenticity are introduced, along with our wider data aspirations and how we collaborate with others for horizon scanning and assurance.

UK FOOD STANDARDS AGENCY USE OF DATA SCIENCE

Data and data science continue to have a transformative impact on our lives and interaction with the world around us. This is no different with respect to food safety. However, data on its own is of little to no value. Many within the food chain, including the FSA, have historically sat on vast data collections, doing little with them, regarding the practice as some sort of safety deposit, with disproportional intrinsic value.
Data becomes more valuable as it moves up a data pyramid, firstly generating information. Information is intended to grant knowledge or insight, for the delivery of impact through action (Figure 1). This is most effectively done collaboratively, and in our work to protect UK food safety, FSA data science collaborates with other Government Departments, Local Authorities, Trade bodies and Industry, and Academia.

Much of what is presented here was first outlined in the 2017 CSA Science Report on Data Science1.

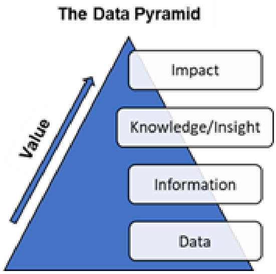

Figure 1: The data pyramid shows how we use data to create impact. Raw data lays the foundations. It is essential that the 'right' data is included. Ideally, this data already exists as collecting data is often expensive and time consuming. The value of data is most readily realised high in the pyramid through the conversion of knowledge to impact. By making our data open and making effective use of others' data we strive to maximise value and impact.

The UK Agri-Food Sector generates some $113bn for the UK economy each year [2]. It is our biggest Manufacturing Sector yet considered to be of low productivity; employing nearly 4 million, often low payed individuals, frequently delivering small profit margins.

As the UK's regulator for food safety and authenticity, the scale of our challenge is huge. As a snapshot; between 2016/2017 we investigated 2,265 food, feed and environmental contamination incidents. The UK produces some 50% of its own food [based on the farm-gate value of unprocessed food in 2017] with 30% coming from our European neighbours [3]. In 2017 we managed compliance for some 822,000 export- and 796,000 import- consignments respectively.

The FSA is an independent Non-Ministerial Department, science led at its core. We are a UK Governmental leader in our data application and policy, with a publicly available data strategy [4], where we commit to being 'open by default' in publishing data collected in pursuit of our statutory functions. We currently maintain approximately 70% of all our data as open [5]. Within this, data from the Food Hygiene Rating Scheme (FHRS) is the most commonly accessed asset in our collection.

We find that our datasets are now being used for applications beyond their original purpose. This is in part because some of our open datasets are well established (FHRS since 2012 for instance), but utilisation in further applications is also practically supported by making our data available through Application Programming Interfaces (APIs). These provide easy accessibility to coders, for the provision of ratings on 3rd party websites [6], as well as some more unexpected innovations such as helping emergency services in the West Midlands better plan, model and respond to incidents [7]. APIs are also used to provide a live data-stream on product withdrawals or recalls, supporting those with food allergies by ensuring that individuals who need to know about food alerts have a means of doing so in a timely and efficient way, however they choose to access this information [8].

The overall success of the FHRS must be attributed to the extensive preparatory work of our data team with respect to data standards; a low quality of baseline data is persistent throughout the entire lifecycle, resulting in low trust, more management effort being required and therefore incurring greater costs. Data standards is a top FSA priority, relevant to major programmes of activity such as the 'Regulating our Future' [9] workstream.

THE FOOD HYGIENE RATING SCHEM [10]

In the UK, there are an estimated 500,000 cases of foodborne illness per year from known pathogens [11].

In 2010, the Food Standards Agency introduced the FHRS. By combining compliance scores from inspection results on factors including safety procedures, food handling practices and temperature control, an overall food hygiene rating from '0' (Urgent Improvement Necessary) to '5' (Very Good) is generated.

These scores are conveyed to proprietors and then made public to help consumers make informed choices about where to eat based on hygiene ratings. Evidence shows that this has had a positive impact on compliance, largely through economic impact, but also on public health. For every 1% change in Food Businesses moving from not broadly compliant (ratings 0,1,2) to broadly complaint (ratings 3,4,5), we predict a reduction of between 2,000 and 3,900 cases foodborne illnesses [12]. This work is world leading in demonstrating public health impacts of a food policy intervention[13].

In England, display of the stickers is voluntary, while mandatory display was introduced in Wales in 2013 and in Northern Ireland in 2016. The FSA has recently reaffirmed its ambition to have a statutory requirement for the display of hygiene ratings in England.

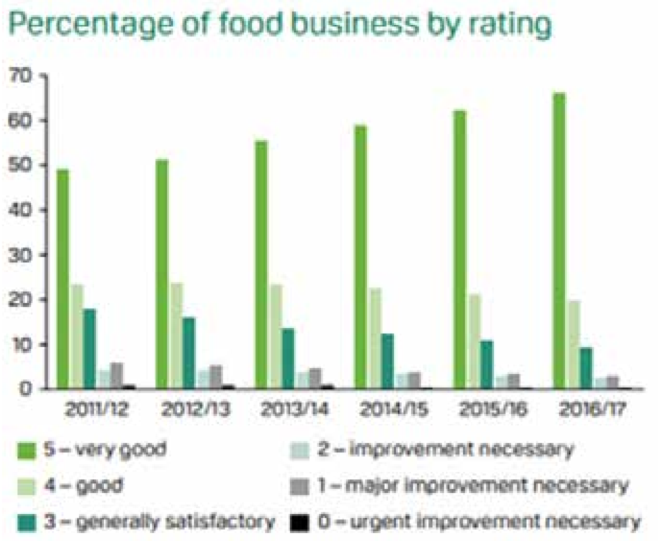

Figure 2: Percentage of food business in the UK following the Food Hygiene Rating Schem (FHRS)

Other innovative FSA data projects include:

- Improving the collection and communication of meat inspection results at abattoirs through Blockchain. The UK Government has given significant attention to Distributed Ledger Technology [14] but the FSA is leading the way in real world application; we have a multi-site pilot live. Blockchain is no magic bullet but does rebalance the reporting burden for our official veterinary inspectors and provides an opportunity to add value; for the first time we have been able to geographically map inspected conditions such as Liver Flukes. This will not only improve our understanding of such conditions but provides an opportunity for intervention; inspection results are accessible to the producers of a given animal, who may adapt their husbandry practices accordingly. This work has been more linked to developing the right data standard as opposed the use of DLT per se, illustrating again the importance of data standards.

- Using Twitter as an early warning tool to predict Norovirus outbreaks. Norovirus is the most common cause of gastroenteritis in the UK. It is highly contagious, and outbreaks frequently cause major disruption in places where a lot of people gather such as schools and hospitals. There is a lag of about two weeks between illness and lab results being published. Historic data allowed us to create a predictive model, inspired by initial success of Google's Flu Trends [15]. When a significant rise in the number of cases over three consecutive weeks is predicted, confidence that the data is real is such that an intervention may be triggered. This means action can be taken earlier than would be the case if laboratory reports were the initiator.

- An 'internet of things' sensor trial with operators in Cambridge, the Local Authority and industry that saw business replace their paper-based food safety management processes with Checkit's [16] digital solution. Daily compliance checks were recorded via wireless sensors and the results automatically uploaded to cloud storage [digitally time-stamped and tamper-proof]. Users could access records remotely, receive alerts on any anomalies and track performance in real-time. Such data sharing would increase efficiency of food safety inspections, support better regulatory judgements, improve transparency and provide businesses with regulated assurance.

Such projects run parallel to the overarching ambitions of the FSA's 'Regulating our Future' programme of modernisation [17]. This promotes a target operating model individualised to each business's needs, and a shift towards greater regulated self-assurance and earned recognition. It includes a new digitally-enabled approach to the registration of food businesses, capturing more information about its intended activities. This will better determine proportionate, risk-based regulatory intervention and provide the FSA with a clearer oversight of the UK food industry. Compared across the UK Government, public trust that the FSA does its job well and that we tell the truth, run very high; 69% and 72% respectively. Trust in the values of the food industry are much lower; 45%. 'Regulating our Future' will help to maintain or even improve public trust in the FSA and the industry that we are responsible for. Data and data standards, assurance and science are all central to this work.

Regardless of what has been introduced here, our data environment is a continuously evolving landscape and it is difficult for a regulator to remain on the front foot without support of external partners and independent expert advice. The FSA co-funds a research Fellowship at the University College London's Big Data Institute [18] for instance and we have recently asked our advisory Science Council [19] to take a refreshed, independent look at our data exploitation, governance and horizon scanning to help assure that we maintain a strong position.

Harnessing disruptive/transformational data science is a challenge for Government but the consequences of not keeping pace, or at least being data aware are poor. The FSA will continue with its collaborative and open data science approach, endeavouring to deliver efficient, effective food safety and authenticity to the UK public. The ambition of the FSA to be a modern, accountable and excellent regulator requires excellent data usage to ensure that the very best science and evidence is at the core of our work, as was the intention when we were established by an Act of Parliament in 1999 [20].

References

[1] *https://www.food.gov.uk/sites/default/files/media/document/chiefscientificadvisersssciencereport%20%281%29.pdf*

[2] *https://www.gov.uk/government/publications/food-statistics-pocketbook-2017/food-statistics-in-your-pocket-2017-food-chain*

[3] *https://www.gov.uk/government/publications/food-statistics-pocketbook-2017/food-statistics-in-your-pocket-2017-global-and-uk-supply*

[4] *https://www.food.gov.uk/our-data*

[5] *https://data.food.gov.uk/catalog*

[6] *https://www.food.gov.uk/business-guidance/making-the-most-of-your-business-rating*

7 *http://bluelightcamp.org.uk/bringing-data-loop-to-bluelightcamp/*

8 *http://data.food.gov.uk/food-alerts/ui/reference*

9 *https://www.food.gov.uk/about-us/regulating-our-future*

10 *https://www.food.gov.uk/safety-hygiene/food-hygiene-rating-scheme*

11 *https://www.food.gov.uk/sites/default/files/media/document/IID2%20extension%20report%20-%20FINAL%2025%20March%202014.pdf*

12 *https://www.food.gov.uk/sites/default/files/media/document/csareportfhrsfinal.pdf*

13 *Fleetwood, J., Rahman, S., Holland, D., Millson, D., Thomson, L. & Poppy G. 2018. As clean as they look? Food hygiene inspection scores, microbiological contamination, and foodborne illness. Food Control. DOI: 10.1016/j.foodcont.2018.08.034*

14 *https://www.gov.uk/government/news/distributed-ledger-technology-beyond-block-chain*

15 *https://www.google.org/flutrends/about/*

16 *https://www.checkit.net/*

17 *https://www.food.gov.uk/sites/default/files/media/document/changing-food-regulation-what-weve-done-where-we-go-next.pdf*

18 *https://www.ucl.ac.uk/big-data/bdi*

19 *https://science-council.food.gov.uk/*

20 *http://www.legislation.gov.uk/ukpga/1999/28/contents*

Session 2:
Big data applications to the food chain safety

The potential of blockchain technologies in food safety

Christopher Brewster

Data Science Group, TNO, The Netherlands

Dr. Christopher Brewster is a Senior Scientist at TNO, and Visiting Professor in the application of emerging technologies at the Institute of Data Science, University of Maastricht. He specializes in semantics and interoperability architectures with an interest in emerging technologies such as blockchains. He was scientific coordinator for the Dutch public-private funded Techruption Blockchain project and co-led the Blockchain in agrifood project funded by Dutch Ministry of Economic Affairs.

He has extensive experience of EC funded research, having been PI on the FI PPP Smartagrifood project, the FI PPP FIspace project, and the CIPS funded Disaster 2.0 project. He has a PhD in Computer Science from the University of Sheffield on the topic of automated ontology learning from texts and was previous Senior Lecturer at Aston University. His research focus is Semantic Technologies, Open and Linked Data, interoperability architectures and Data Governance, with a special interest in the roles of Blockchain technology in solving real world problems, the ethical implications of blockchain technology, and the interaction of semantics with smart contacts. He has focused on the agrifood domain as an application domain focusing on the use of Semantic Technologies to the food supply chain and logistics. He has published over 60 papers in conferences and journals, and organised many workshops.

Data value chain in the dairy production: opportunities and challenges

Stephanie Van Weyenberg

Instituut voor Landbouw- en Visserijonderzoek (ILVO), Belgium

Dr. Stephanie Van Weyenberg is a veterinarian with a PhD on animal nutrition. Her current research at ILVO (Flanders research institute for agriculture, fisheries and food) focuses on the use of sensor technology to support dairy management, to improve cow health and to ensure the milk quality. Examples of research projects include cubicle design with innovative bedding material, studies on teat conformation and condition and the relation with udder health, early detection of lameness and CONTROL, which coordinates the maintenance of all milking machines, milking robots, cooling tanks and milk meters in Belgium. The last years, a strong competence in statistical data analysis was developed. She is involved in various national and international data projects, such as ICT-AGRI era-net, IOF2020, and she is WP leader within H2020 4D4F. She supports companies to explore data driven business opportunities (Smart Digital Farming network) and is coordinator of the DataHub for Agrofood.

DAIRY FARMING IN BELGIUM AND FLANDERS

Sustainability, international competition, transparency, upscaling, profitability, climate and environmental challenges ... Dairy farming in the 21st century is a very complex production environment. In 2017, 2,616 million liters of milk were produced by 4,397 producers in Flanders. Compared to 2016, there were 5 % less dairy farms, while 5.1 % more milk was delivered (MCC Vlaanderen, 2017). In 2014, the production value of the dairy sector was 844 million Euro, which makes dairy the most important livestock sector after pig and beef cattle (Van der Straeten, 2015).

The Flemish milk quality is excellent. In 2017, significantly fewer penalty points were awarded: 3,082 (4,325 in 2016). The number of delivery bans decreased from 107 in 2016 to 60 in 2017. The percentage of companies without penalty points rose to more than 97% (MCC Vlaanderen, 2017).

The milking techniques on these modern farms has evolved in two different directions: farmers either opted for a milking system with high capacity with a higher number of cows milked per hour, or they chose for full automation with an automatic milking system (AMS). In Flanders we see an increasing popularity of the AMS. Nowadays, 10 % of dairy farms in Belgium are equipped with an AMS (Control, 2018), while 45 % of the new milking parlors installed in 2016 in Flanders were AMS (MCC Vlaanderen, 2017).

PRECISION LIVESTOCK FARMING

Dairy farming has become strongly technologized over the past years and applications that evaluate elements within the production system, such as rumination time, milk yield etc., are widely available. Large quantities of data are hereby generated in a fully- or semi-automated way and give indications on improvement potentials within farm management. Following the principle "What you can measure, you can manage", these data is integrated into farm management information systems aiming to provide farmers with easy-to-use decision support tools.

That dairy farmers are enthusiastic about the new precision techniques and applications is shown in a survey conducted by the Agriculture and Fisheries department, where 68 % of dairy farmers indicate to use these techniques or are planning to within the next 5 years (Van Bogaert et al, 2017).
When farmers were asked what holds them back from investing (extra), answers such as: "too expensive", "insufficient information on benefit", "too complex", "privacy insufficiently protected", "does not match my needs", "unclear what happens to my data", were received.

In order to increase adoption among all dairy farmers, additional awareness campaigns and training are required. Dairy farmers are convinced of the usefulness and even the necessity of digitization, but they ask for easy to use applications that meet their needs, conduct individual or group benefits, and with a 100 % transparency about data ownership and privacy.

DATA PRIVACY AND DATA OWNERSHIP

Farmers are not the only stakeholders for whom data ownership and data privacy is crucial. The data revolution
in the agrifood sector is a common economic opportunity for all stakeholders along the entire food chain: sensor- and IoT developers, companies who focus on data analytics and algorithm development, manufacturers of milking machines and barn equipment, veterinarians, animal nutrition companies and advisors, pharmaceutical industry, dairy producers, retail, licensing authorities, audit organizations, paying agencies, food safety agencies, etc.

The enormous and ever-increasing amount of data collected at dairy farms, enriched with data further in the chain, offers a potential for new data driven business models. In addition, European requirements and legislative trends show a thrive for greater transparency. Production circumstances such as pasturing, medicine use, etc., become more important for retail and consumers. But this information is also useful to optimize delivery forecasts and milk processing. Data will only become valuable if interlinked and integrated throughout the entire production chain. Data integration is the way to fully explore the potential of digitization in agriculture. Yet, all stakeholders along the entire food supply chain depend on the data integration and the willingness of data sharing. And trust and recognition of data ownership is critical for all players.

Therefore, a Code of Conduct was constructed by COPA-COGECA, Fertilizers Europe, CEETTAR, CEJA, EFFAB, FEFAC, ECPA, ESA and CEMA. This code is a non-binding document providing guidance on the use of agricultural data, looking at general principles and contractual relations where the farmer remains at the heart of the collection, processing and management of agricultural data. This set of fair and transparent rules makes it possible for all relevant parties to have the necessary access to data and benefit from its value.

DATAHUB4AGROFOOD

In this context of transparency of data sharing and transfer, it is important to involve the farmers in this data value construction. Sustainable digital innovation ecosystems with respect for privacy, ownership and added values for all stakeholders need to be developed.

DataHub4Agrofood aims at the development of such an ecosystem for data innovation in the agrifood sector. This project is an initiative of ILVO, the Innovative Business Network (IBN) Smart Digital Farming (SDF) and a number of leading companies from the Belgian dairy industry (AVEVE, Boerenbond, CRV, DGZ, Innovatiesteunpunt, Milcobel). The project is jointly invested by the companies and the European Regional Development Fund (Europees Fonds voor Regionale Ontwikkeling – EFRO).

The aim of the project is to build a digital cloud platform with main components (or functional entities) that are able to support data sharing between the participants having multiple business roles. The data hub wants to achieve horizontal stakeholder integration along the supply chain to secure data exchange for the future and stimulate innovation by creating trust among partners. Clear agreements about data access, use, management and ownership are made. All parties must be able to rely on the agreements being kept so that they will continue to share their data. The farmer is recognized as the owner of the raw data. He determines which other parties may have access to his data. Only authorized data can be further exchanged and combined.

The potential benefits of data sharing in the agricultural sector are demonstrated through use cases in dairy farming. The different cases will illustrate how integrated data will enable the diverse stakeholders (farmers, veterinarians, transport, dairy processors, retail, pharmaceutical companies, sensor and algorithm developers, audit organizations, etc.) to optimize their activities (farm management, administrative workload, disease detection and prevention, planning and delivery forecasts, feed and medicine traceability, product traceability (origin, supplier identification) on different levels. Ongoing data feedback between the project partners will allow improved "cause-effect" management at dairy farms, but also in companies further down the supply chain. The effort of the whole dairy sector in terms of sustainability and animal welfare can be shared transparently with retail, the consumer and the government.

The administrative unburdening of farmers is another use case topic. Administrative burden is a complaint from all farmers. This time-consuming activity often has to take place in the evening. A digital processing of data not only reduces the workload, it also diminishes the possibility of errors and it increases transparency throughout the whole chain. All these initiatives contain a great deal of information that, without any doubt, is also interesting for other stakeholders within and outside dairy farming. Therefore, the data hub is an open initiative and in the future use case beyond dairy will be possible as well.

References

Control, Kwaliteitszorg onderhoud melkwinningsapparatuur. 2017. MCC Vlaanderen, jaarverslag 2017.

Van Bogaert, Janssens & Maertens. 2018. Toepassing van precisielandbouwtechnieken.

Van der Straeten. 2015. Rentabiliteits- en kostprijsanalyse melkvee op basis van het Landbouwmonitoringsnetwerk. Boekjaren 2007-2012, Departement Landbouw en Visserij, afdeling Monitoring en Studie, Brussel.

Spread model: a forecasting and managing tool in microbiological safety

Pablo Fernandez

Universidad Politécnica de Cartagena, Spain

Prof. Pablo S. Fernandez is MsC (1990) and PhD (1994) in Veterinary Science by the University of Murcia (Spain). He is full time professor in Polytechnic University of Cartagena (Spain) since 2009, where he leads the group of "Food Safety and Preservation". He has published over 100 scientific papers in high impact, peer-reviewed journals, book chapters and the EFSA Journal with a total of 1.410 citations (WOS all databases) and an 'H' index of 24. He has led several research proposals from public and private competitive calls. His research has been focused on quantitative microbial risk assessment, mathematical modelling and preservation technologies of foods. He has been an expert of the BIOHAZ Panel of EFSA (2012-2018), chairing working groups on Qualified Presumption of Safety and Evaluation of new processes for animal by-products.

He leads the Spanish Network of Excellence "Development of the structure required to perform risk rankings and quantitative risk assessment in Spain", with 10 participating institutions and has provided training in QMRA in Latin American countries. He co-authors a patent, "Development of thermoresistometer Mastia", equipment to perform dynamic heat inactivation. He has supervised successfully four PhD students over the last 10 years and over 20 Master and Degree students He has extensive experience in management of research, as Vice-rector of Research and Innovation (2012-2015) and Director of the Doctoral School of his University (2015-2019).

ABSTRACT

Food trade has grown enormously over the last years and its impact on the food available in the markets of developed countries is very significant. Such a global market represents an ideal scenario for the spread of foodborne diseases, where contaminated products can quickly propagate through the network. Consequently, it is essential to understand the implications that the structure of the food network has on food safety. This would allow to design efficient control systems and to quickly respond to foodborne outbreaks, benefitting both producers and consumers. In order to do so, the use of big data represents a key factor to develop new rapid control or decision-making systems. In this work, a mathematical model is proposed to simulate the spread of food products within the trade network, based on the FAO database on international imports, exports, consumption and production. It is applied to investigate the origin of real foodborne disease that took place in the EU in 2017.

INTRODUCTION

Food production has become global, therefore producers of primary sectors can be located from a few to thousands of kilometres away from end consumers (Gilbert & Morgan, 2010). Global food trade is therefore essential to feed the population of developed countries. The increasing complexity of the food trade network faces a lack of control of factors that have emerged over the last years. An optimization of the structure of the network could contribute to meet the future food demands of the population while ensuring its food safety (Alamar et al., 2018), or to implement most effective food control systems. This target can only be attained based on scientifically sound knowledge of the food distributions channels and the structure of the food trade network.

MATERIALS AND METHODS

Mathematical model proposed.

A mathematical model to describe the spread of food products within the food network was constructed using network analysis, with the nodes representing countries and the edges trade flows. The following variables were considered: (1) the national production, (2) the national consumption, (3) the national amount of food wasted and (4) the international trade (imports/exports). The structure of the model is exemplified in Figure 1 for two nodes of the network.

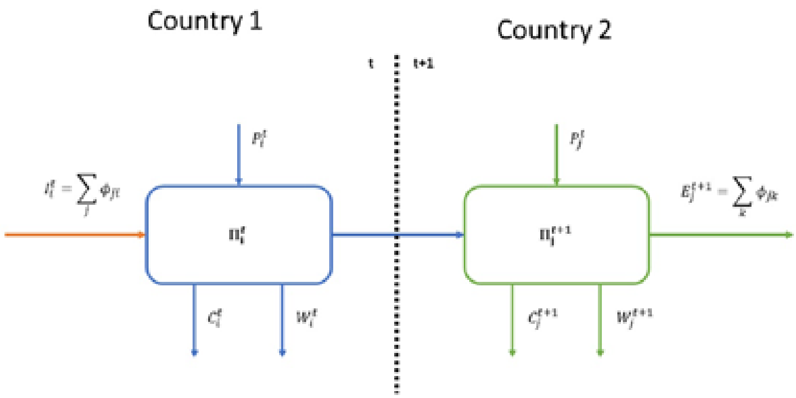

Figure 1: Steps in Exchange of foods between countries.
Production: P; Consumption: C; Exports: E; Waste: W; Imports: ϕ.

Under the assumption of mass conservation (imports and national production should equal consumption, exports and food waste), the model can be used to simulate the spread of food products within the network. This hypothesis is exemplified in Equation 1, where P_i^t, C_i^t, I_i^t, E_i^t and W_i^t are, respectively the national production, consumption, imports, exports and amount of food waste of the relevant item for country i at time t.

$$P_i^t + I_i^t = C_i^t + W_i^t + E_i^t \text{ (Equation 1)}$$

It is, thus, possible to simulate the spread of contaminated (or adulterated) products, as well as identify the most likely origin of items consumed in a selected country using Equation 2, where α_i^t represents the fraction of contaminated (or adulterated) food available, τ_i is the fraction of the national production that is contaminated (or adulterated) and ϕ_{ji} is the trade flux from country j to country i.

$$\alpha_i^{t+\Delta t} = \frac{\tau_i^{t+\Delta t} P_i + \sum_j \alpha_j^t \phi_{ji}}{C_i + E_i + W_i} \text{ (Equation 2)}$$

The model proposed is written in R (R Core Team, 2016) and accessible in open access.

DATABASE USED

Big datasets describing global trade and consumption of different food commodities that were generated by FAO are available in Open Access. In particular, the Detailed Trade matrix and the Food Balance Sheets (FAO, 2014) were used to build the model.
Food Balance Sheets provide the information of the food for consumption of the population and production and waste of certain items at national level, as well as the amount of food available for human consumption. The information of Detailed Trade matrix has also been taken considered (Figure 2).

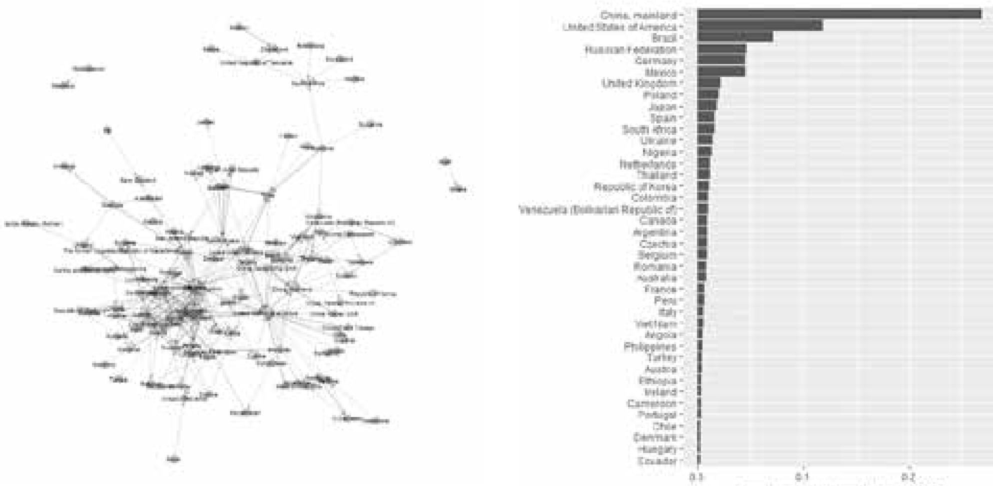

Figure 2: Food Spread Model: FAO dataset for Trade flows (left); production and consumption (right).

GENERAL CONTENT

The production and trade patterns of different food products depends on factors such as the geographical needs for its production and the capability to transport it safely through long distances. Consequently, different food products have a different structure of the trade matrix. To assess the impact that the network structure has on the spread of food products, simulations were performed using two reference networks frequently used in the literature: a scale-free network and a small-world network (Barabási & Albert, 1999, Watts & Strogatz, 1998). The results demonstrate that the structure affects the spread of food products and is a factor to be considered in risk assessment.

A case-study based on a recent multi-country outbreak of Salmonella Enteritidis in the EU was analyzed. It was assumed that the outbreak had already been linked to a food, but its origin was unknown. Here, the model proposed that it can simulate how food products spread within the trade network. It can be used to estimate the most likely origin of a product consumed in a country.

The model tracks the spread of selected food products through the network. Therefore, it can be used to estimate the destination of the products produced in a specific country. Once the likely source of the outbreak was narrowed down to just three countries, a similar simulation could be performed to identify other countries whose consumption was largely originated in these sources and could be at risk of consuming the contaminated eggs. This tool has proved to be effective to investigate multi-country outbreaks in order to identify the most likely source and take specific measures both in countries affected and in those with a high probability of being at risk. The availability and use of large databases provide new possibilities to generate powerful tools to ensure food safety at a multicountry level and can contribute to maintain consumer confidence in international trade.

ACKNOWLEDGEMENTS

The financial support of this research work was provided by the Ministry of Economy and Competitiveness (MINECO) of the Spanish Government and European Regional Development Fund (ERDF) through project AGL2017-86840-C2-1-R and by Fera Science. Alberto Garre (BES-2014-070946) is grateful to the MINECO for awarding him a pre-doctoral grant.

References

Alamar M. del C., Falagán N., Aktas E. & Terry, L. A. 2018. Minimising food waste: a call for multidisciplinary research. Journal of the Science of Food and Agriculture, 98(1), 8–11.

Barabási A.-L. & Albert, R. 1999. Emergence of scaling in random networks. Science, 286(5439), 509–512.

Gilbert C. L. & Morgan C. W. 2010. Food price volatility. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1554), 3023–3034.

Watts D. J. & Strogatz, S. H. 1998. Collective dynamics of 'small-world'networks. Nature, 393(6684), 440–442.

Smart imaging from space for crops management

Pierre Defourny

Catholic University of Louvain, Belgium

Prof. Pierre Defourny is professor of geomatics, Earth remote sensing and land use planning at the Faculty of Bioscience Engineering of the Université catholique de Louvain. He is the founding president (2010-2015) of the UCL Earth and Life Institute gathering more than 300 researchers in Earth and biological sciences to address major challenges facing our planet (food production, biodiversity, climate, water). He leads a research lab in satellite remote sensing and GIS modelling for land applications, in particular for land cover, agriculture and forest ecosystem monitoring. As the co-lead of the global network for a Joint Experiment for Crop Assessment and Monitoring (JECAM), he is an executive committee member of the GEOGLAM initiative for global agriculture monitoring supported by the G20.

He is also an Earth Observation expert for space and UN agencies including the European Space Agency and the NASA where he worked as a visiting scientist for a year. Currently leading several flagship projects of the European Space Agency, his research team joint forces with many international partners to produce from multiple satellites the most precise time series of global annual land cover maps from 1992 to 2015 recently adopted by FAO, OEDC and some IPCC modelers. Similarly, the Sentinel-2 for Agriculture open source system capable to process massive volume of satellite observation for crop monitoring in near real time at 10 m resolution was designed and demonstrated in many countries around the world. Most recently, he is invited as expert to various committees related to the EU Common Agriculture Policy to support the DG-AGRI in the exploitation of the new Sentinel satellite fleet for the CAP reform post 2020.

INTRODUCTION

A tsunami of Earth Observations downloaded from space is changing the game for land surface monitoring in any place on Earth. For 40 years, the Earth Observation community has developed concepts, methods and algorithms using experimental or commercial satellites. Since a few years, the convergence of the performances of cloud computing infrastructures, the success of artificial intelligence and the new European Copernicus program have opened the avenue for the operational implementation of many new space applications supporting crop management, food traceability, and agriculture practices monitoring, amongst many others. Some are already mature applications currently implemented as operational services at full scale, while others correspond to emerging technologies that are often very promising but have still to be confirmed in an operational context.

AN INFORMATION TECHNOLOGY AND EARTH OBSERVATION (R)EVOLUTION

Satellite remote sensing is an undisputed source of regular land information for a vast range of users at all geographical scales with pixel size now varying from 30 cm to 10 -20m on the ground. After many years of satellite observation driven by NASA, the European Copernicus program has made a breakthrough for operational applications by committing to open, free, systematic and long term imaging capacities over the whole surface of the planet using a fleet of Sentinel satellites. Earth Observation is progressively becoming a common good, like meteorological observation with long-term policies and system redundancy.

Agricultural applications are directly supported by the new Sentinel-2 mission, including two satellites in near-polar sun-synchronous orbits, with a revisit cycle of 5 days over each location of any continent. They measure a very well calibrated surface reflectance in 13 different spectral bands (from 443 nm to 2190 nm) specifically designed for vegetation monitoring at 10 and 20 m resolution on the ground and for correcting atmospheric perturbations. The Sentinel-1 mission also consists of two satellites in orbit providing a revisit of 6 days at the equator (3 days over Northern Europe) and imaging of the land surface by sending a C-band signal (central frequency of 5.404 GHz) and recording the backscattered signal, regardless of cloudiness and environmental conditions. These all-weather and day-and-night radar imaging capabilities largely compensate the fact that the images recorded by these Synthetic Aperture Radar (SAR) instruments are much less familiar. The first one of these four satellites was launched in 2014 and the last one has been fully operational since July 2017, with the following Sentinel-2C and D spacecraft already under development. With the systematic coverage of the satellite constellation, the never-ending waterfall of Sentinel-2 images results ultimately into over 4 Terabytes of fresh products, published every day on the Copernicus portals. This represents a real challenge for the ground systems to cope with and likely also for users themselves to process in their applications. Only the fast development of cloud computing infrastructures like Amazon, Alibaba and Google Earth Engine allows the handling of such a considerable amount of observation time series over large areas in a timely manner.

MAPPING CULTIVATED AREAS THROUGHOUT THE SEASON AND MONITORING CROP DEVELOPMENT

Estimating planted areas at an early stage, detecting plant emergence date at field level and identifying crop type along the season are the current challenges that are being successfully tackled by some open source software such as the Sen2Agri system, which has been specifically developed to streamline the management of massive data volumes in a seamless way. Such a system automatically downloads the data from the satellite catalog and processes it as it arrives. Biophysical variables like the Leave Area Index (LAI), proportion of vegetation cover and fraction of the Absorbed Photosynthetically Active Radiation (fAPAR) are retrieved for each pixel of 10 x 10 meters from each cloud free image acquired over the country or the region of interest (figure 1).

The monitoring of these biophysical variables for a given crop is relevant for vegetation development assessment, field heterogeneity characterization and in the best case, yield forecasting. For instance, in a context of a food unsecure country like Mali, crop yield indicatorss for the main crops was related to the peak of LAI providing early crop production estimate one month before harvest. Current research activities try to depict disease impacts from the evolution of these EO-derived variables throughout the season in order to support the modelling of the spread of disease such as wheat rust or maize lethal necrosis.

In Belgium, several commercial applications are emerging for crop monitoring, sometimes led by the food industry (e.g. potato sector), to deliver satellite information to the industry for planning purposes or to the farmers for crop management.

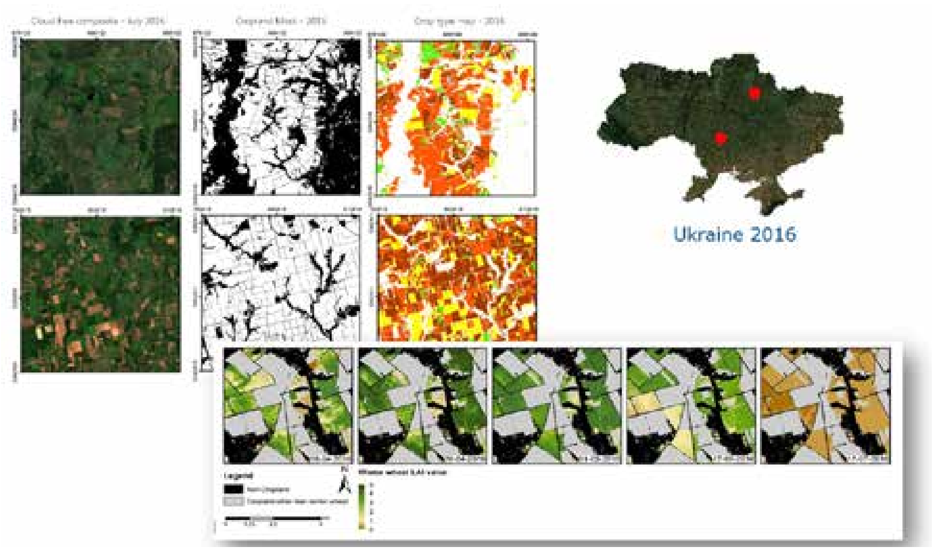

Figure 1 – First crop type map and Leaf Area Index (LAI) for the wheat fields produced at 10 m resolution from the Sentinel-2 mission in 2016 over Ukraine. This highlights the potential of near real time monitoring of each winter wheat field in Ukraine from the mid-season. The Leaf Area Index corresponds to a biophysical variable measuring the surface of all photosynthetically active leaves per ground unit and is closely correlated with the green biomass at early crop stages (Defourny et al., 2018).

SUPPORTING THE CROP MANAGEMENT

Exploiting the potential of the red-edge bands of Sentinel-2 has also led to the estimation of the crop canopy chlorophyll content from the images. Such information can be combined with eco-physiological knowledge to deliver the so-called Nitrogen Nutrition Index (NNI). In the context of the BELCAM platform supported by the STEREO program of the Belgian Science Office, this information is provided to farmers to deliver the current nitrogen plant status and advice on the fertilizer input (figure 2). In conventional, as in precision farming agriculture, the timeliness of such an information delivery is critical for fertilizer applications. The development of ecologically intensive agriculture could be directly related to such a capacity. Commercial services like Farmstar were initially developed many years ago with costly satellite images, but have become much more accessible today as they have been bought by fertilizer or seed companies.

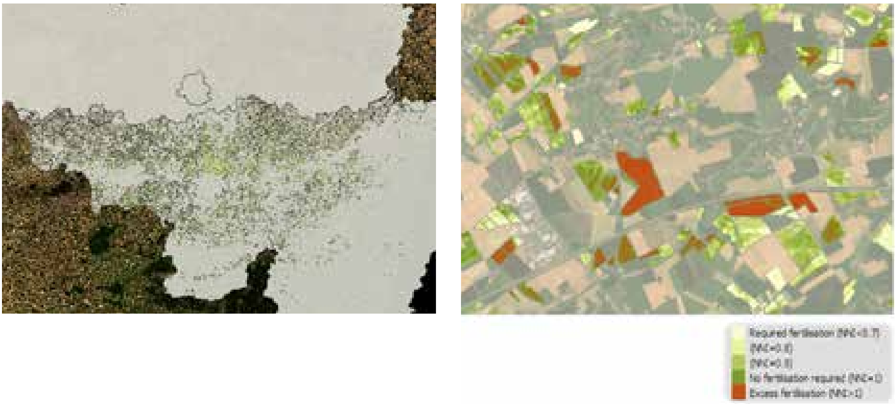

Figure 2 – First Nitrogen Nutrition Index map based on the canopy chlorophyll content measured for each 10 x10 m pixel. The Nitrogen Nutrition Index measures the capacity of a plant to take advantage of additional Nitrogen input (Delloye et al., submitted in 2018).

DRIVING THE REFORM OF THE 2020+ COMMON AGRICULTURE POLICY

On the request of the EU Commissioner Phil Hogan, the forthcoming reform of the 2020+ Common Agriculture Policy will have to rely on the potential of the Sentinel satellites in order to simplify its implementation over the whole European Union. Monitoring agriculture practices along the season and their compliancy with regards to the new CAP regulations are expected to drive the new CAP paradigm voted by the European Parliament late May 2018, namely the Monitoring approach. This will allow the support of farmers in their crop management and the associated eligibility at field and at farm level. In this context, the Sen4CAP project, supervised by the European Space Agency and supported the European Commission, is currently investigating new methodological developments and the Paying Agencies protocols, in order to best take advantage of the various metrics derived from these satellite signals.

As illustrated in figure 3, the unprecedented density of satellite observation allows detecting crop management practices like tillage or no tillage, harvesting date, or green manure management. Similarly for grasslands, the detection of mowing dates by satellites should inform the farmer in due time about the management needed to improve its productivity in a sustainable perspective and to maintain its eligibility for subsidies. The overall 2020+ CAP objective is to move from a control perspective to a monitoring perspective, supporting the farmers in their crop and grassland management for ecologically intensive agriculture.

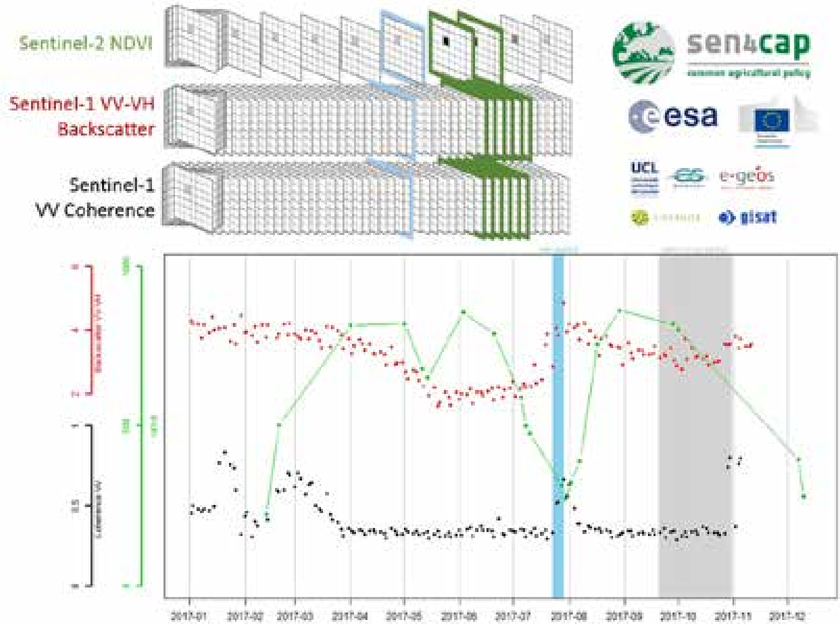

Figure 3 – Three different metrics respectively derived from Sentinel-1 and Sentinel-2 satellite constellations are computed from very dense time series of observations, in particular from the all-weather, day-and-night radar imaging system on board Sentinel-1 providing a useable image every 3 days over Belgium. The sharp coherence increase, matching the large drop of the vegetation index (NDVI) derived from Sentinel-2 depicts the harvest date of this field (source: Sen4CAP project, from GISAT – UCLouvain, 2018).

SHARING THE INFORMATION TO FARMERS AND TO THE GLOBAL LEVEL

The successful uptake of these new information sources is also much related to the communication tools for the information dissemination and its actual use by the targeted stakeholders. Numerous apps on smartphones, developed to display the field information along with meteorological forecasts, seems key to reach out to the farmer for crop management in near real time. Attractive web mapping interfaces are more required for analysts, who would prefer to browse a series of websites and to download fresh data in their own analytical tools. No standard is established yet and the innovation pace is currently particularly impressive. Beyond attractive and fancy interfaces, the service quality necessary to build user confidence will be key in the coming years as well as the integration of various services.

On the other hand, for international markets, the G20 request to enhance market transparency by widely sharing crop production estimates has been taken up by the GEOGLAM initiative, to derive such information partly from satellite imagery and crop modelling, in order to reduce food price volatility. More specifically the GEOGLAM crop monitor, compiling on a monthly basis all the available information from the main food producers, feeds the global, well-recognized Agriculture Market Information System hosted at FAO in Rome.

FURTHER PERSPECTIVES ON INFORMATION SOURCES

Next to the space agencies, Earth Observation by satellites and drones is also developing new low cost pathways. The Planets company currently claims to image the entire planet every day at 5 m resolution thanks to hundreds of nanosatellites built from cheap on-the-shelf components. Drones are sometimes introduced as robots to the farmers, imaging in an automated way all the fields of a farm whenever required. Last but not least, the Internet of Things linking drones, tractors, harvesters and satellites is another level of integration which is today tested at full scale. These big data flows have yet to be turned into relevant and timely information, which would have taken decades without the machine learning and artificial intelligence development. The societal challenge remains the same, making the best use of these information and communication technologies to move towards more sustainable healthy food production pathways.

References

Defourny, Pierre; Bontemps, Sophie, et al.. Sentinel-2 near real-time agriculture monitoring over large countries at field level exploiting Sen2-Agri automated system: performance assessment in various cropping systems around the world. In: Remote Sensing of Environment, in press (2018).

Delloye, Cindy ; Weiss, Marie ; Defourny, Pierre. Retrieval of the canopy chlorophyll content from Sentinel-2 spectral bands to estimate nitrogen uptake in intensive winter wheat cropping systems. In: Remote Sensing of Environment, Vol. 216, p. 245-261 (2018).

Lambert, Marie-Julie ; Traoré, Pierre C. Sibiry ; Blaes, Xavier ; Baret, Philippe ; Defourny, Pierre. Estimating smallholder crops production at village level from Sentinel-2 time series in Mali's cotton belt. In: Remote Sensing of Environment, Vol. 216, p. 647-657 (2018).

European Space Agency Sen2Agri system: an open source system - www.esa-sen2agri.org

EU – ESA joint research initiative : Sen4CAP : Sentinels for the EU Common Agriculture Policy - esa-sen4cap.org

FAO Handbook on remote sensing for agricultural statistics.Global Strategy for Improving Agriculture and Rural Statistics , ed. Delincé Jacques, 258 pages, 2017.

GEOGLAM : a crop monitor to support transparency on international markets - cropmonitor.org

Minet, Julien ; Curnel, Yannick ; Gobin, Anne ; Goffart, Jean-Pierre ; Mélard, François ; Tychon, Bernard ; Wellens, Joost ; Defourny, Pierre. Crowdsourcing for agricultural applications: A review of uses and opportunities for a farmsourcing approach. In: Computers and Electronics in Agriculture, Vol. 142, p. 126-138 (2017).

Rodrigues, Francelino A. ; Blasch, Gerald ; Defourny, Pierre ; Ortiz-Monasterio, J. Ivan ; Schulthess, Urs; Zarco-Tejada, Pablo J. ; Taylor, James A. ; Gérard, Bruno. Multi-Temporal and Spectral Analysis of High-Resolution Hyperspectral Airborne Imagery for Precision Agriculture: Assessment of Wheat Grain Yield and Grain Protein Content. In: Remote Sensing, Vol. 10, no.6, p. 930 (2018).

Smart imaging for vector-borne diseases management

Els Ducheyne

AVIA-GIS, Belgium

Dr. Els Ducheyne (1976) graduated as doctor in the agricultural and applied biological sciences in 2003 and has over 15 years of experience in the field of geographical information systems, spatial modelling using statistical and machine learning techniques and in the set-up of spatial decision support systems. During her PhD (1999-2003), she focused on the integration of GIS and machine learning techniques to optimize forest resource management in the framework of multiple conflicting objectives. After her PhD, she was employed by the Institute of Tropical Medicine in Antwerp (Belgium) to work on a collaborative project between ITM and Avia-GIS to model the spread of vectors of disease, in particular Culicoides in the Mediterranean basin. Since 2005, she joined Avia-GIS where she contributes to a wide range of projects which include spatio-temporal modelling and analysis, database set up and development, with a particular focus on data that can be used for spatial modelling sensu lato: presence/ absence and abundance data and development of spatial decision support systems.

A particular attention is given to data quality assessment, spatial model output validation and uncertainty assessment. As part of her current tasks as Scientific Director of Avia-GIS, Els coordinates the research activities to improve the current products and she oversees all necessary phases to achieve a marketable product: identification of user needs and requirements, designing system requirements and system architecture and development of the final system. Els is member of the editorial board of Geospatial Health, a leading peer-reviewed journal focusing on Health Applications in Geospatial Science.

Fasciola hepatica is a trematode parasite with a high economic impact on livestock production worldwide (Mas-Coma et al., 2005). For example, in northern Belgium (Flanders), the yearly cost of infections with this parasite in dairy cattle has been estimated at €8.2 million or €30 per adult dairy cow present in the population (Charlier et al., 2007, 2009). Similarly, Schweizer et al. (2005) estimated, using Monte Carlo simulation, that the median financial loss due to bovine fasciolosis in Switzerland amounts to approximately €52 million representing a median loss of €299 per infected animal. Most of these losses arise from reduced milk yield and reduced fertility, with smaller losses due to reduced meat production and liver damage.

In order to identify high-risk zones, several spatial modelling approaches have been suggested, either directly by modelling of the presence/absence of the disease on farms in Europe (Rapsch et al., 2008; McCann et al., 2010; Bennema et al., 2011; Fox et al., 2011; Kantzoura et al., 2011) and worldwide (Fuentes 2004, 2006; Asrat et al., 2007; Valencia-Lopez et al., 2012), or indirectly through species distribution modelling of the intermediate host Galba truncatula (De Roeck et al., 2014). These models usually focus on a region within a single country and have often a low spatial resolution. McCann et al. (2010), for example, described the prevalence of fasciolosis in England and Wales at the postcode level with each polygon having a mean surface of 20,000 km2.

Few studies focus on the development of a pan- European model, most likely due to the lack of comparable epidemiological surveys of the disease between different countries and study sites. Within GLOWORM, a harmonised sampling strategy generated in VECMAP version 1 and diagnostic method (Charlier et al., 2007) enabled the creation of a baseline dataset that can be used to model the probability of exposure to liver fluke at the European level (Ducheyne et al., 2015). The use of this harmonised sampling approach in five countries allowed the creation of two pan-European ensemble models using VECMAP that delineate the probability of exposure to F. hepatica with high accuracy (Figure 1). The two machine-learning models were capable of distinguishing areas with high probability of exposure from those with intermediate and low probability. Rainfall and temperature were the most important drivers for the probability of exposure consistent with previous modelling efforts. During the model generation, areas outside the eco-climatic range of the samples were masked out. These areas require further field study to increase the spatial extent of the model output.

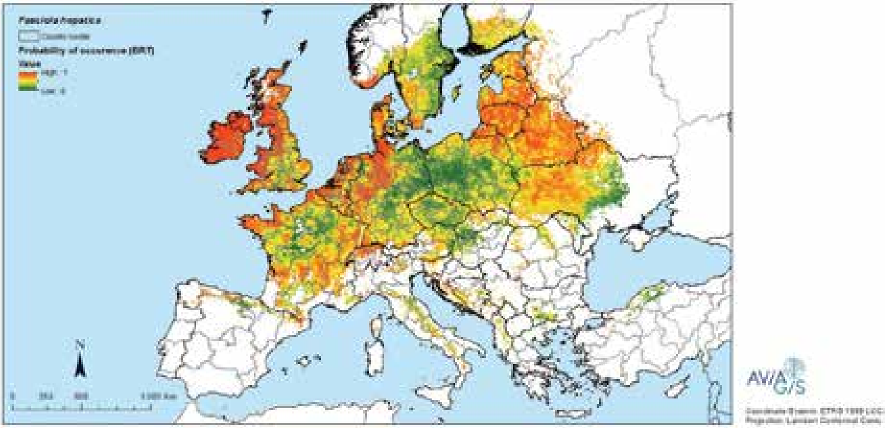

Figure 1: Probability of occurrence of F. hepatica in Europe modelled using Boosted Regression Trees in VECMAP.

Baseline models such as the European model described above or country-wide models such as Selemetas et al. (2015) developed for Ireland can be used as a first input to temporal models, to detect deviations in the baseline probability and to assess the economic cost of fasciolosis in Europe.

The University of Naples has collated a spatio-temporal database between 2000-2017 which contains presence and absence data of gastro-intestinal parasites including F. hepatica. In an ongoing study the objective is to generate dynamic yearly models of the GI nematodes. The first step is to first determine the minimum dataset on a yearly basis required to obtain reliable models; this minimum set is determined by number of observations in the different eco-zones in Italy and the accuracy of each model output.

Then the years which adhere to the minimum set are used to obtain a baseline model using a similar ensemble technique as for the continental models (Ducheyne et al. 2015). Yearly models will then be developed and contrasted against the 'normal' situation. The anomalies will then highlight years and locations with higher/lower risk than expected. This will impact the management planning of each individual farmer and will pave the way towards farm level models. Indeed, adaptive on-farm management under global change scenarios will become more and more important and tool development to support farmers in their decision-making process is therefore relevant.

Such on-farm models have already been developed on a prototype scale. Monitoring vector habitats from the finest scales up to farm level is of key importance to refine currently existing broad-scale infection risk models. Using Fasciola hepatica, a parasite liver fluke as a case in point, this study illustrates the potential of very high resolution (VHR) optical satellite imagery to efficiently and semi-automatically detect detailed vector habitats (De Roeck et al., 2014). A WorldView2 satellite image capable of <5m resolution was acquired in the spring of 2013 for the area around Bruges, Belgium, a region where dairy farms suffer from liver fluke infections transmitted by freshwater snails. Given that the intermediate host thrives in small water bodies (SWBs), such as ponds, ditches and other humid areas consisting of open water, aquatic vegetation and/or inundated grass, identification of the SWB is paramount in disease management. The water bodies can be as small as a few m2 and are most often not present on existing land cover maps because of their small size. Within an innovation project in collaboration with the University of Gent (Veterinary Faculty and Bio-engineering Faculty) a classification procedure based on object-based image analysis (OBIA) that proved valuable to detect SWBs at a fine scale in an operational and semi-automated way. The classification results were compared to field and other reference data such as existing broad-scale maps and expert knowledge. Overall, the SWB detection accuracy reached up to 87%. The resulting fine-scale SWB map was be used as input for spatial distribution modelling of the liver fluke snail vector to enable development of improved infection risk mapping and management advice adapted to specific, local farm situations. The risk models developped using machine learning techniques from VECMAP indicate pastures which may require additional management adaptations. Figure 2 shows a dairy farm with the pastures. The pasture on the right has a very low probability of G. truncatula presence due to lack of the SWBs. This pasture can be used throughout the year. The pasture on the left of this has many higher risk SWBs that will permit the intermediate host to be present.

Adaptive management may be required on this particular pasture. Future funding will permit to make these proof of concept tools operational and bring them to the farmer.

Figure 2: Very high resolution on-farm risk model for Fasciola hepatica.

References

Bennema S.C., Ducheyne E., Vercruysse J., Claerebout E., Hendrickx G., Charlier J. 2011. Relative importance of man- agement, meteorological and environmental factors in the spatial distribution of Fasciola hepatica in dairy cattle in a tem- perate climate zone. Int. J. Parasitol. 41, 225-233.

Charlier J., De Cat A., Forbes A., Vercruysse J. 2009. Measurement of antibodies to gastrointestinal nematodes and liver fluke in meat juice of beef cattle and associations with carcass parameters. Vet. Parasitol. 166, 235-240.

Charlier J., Duchateau L., Claerebout E., Williams D., Vercruysse J. 2007. Associations between anti-Fasciola hepatica antibody levels in BTM samples and production parameters in dairy herds. Prev. Vet. Med. 78, 57-66.

De Roeck E., Van Coillie F., De Wulf R., Soenen K., Charlier J., Vercruysse J., Hantson W., Ducheyne E., Hendrickx G. 2014. Fine-scale mapping of vector habitats using very high-resolution satellite imagery: a liver fluke case-study. Geospa.t Health 8, S671-S683.

Ducheyne E., Charlier J., Vercruysse J., Rinaldi L., Biggeri A., Demeler J., Brandt C., de Waal T., Selemetas N., Höglund J., Kaba J., Kowalczyk S.J., Hendrickx G. 2015. Modelling the spatial distribution of Faciola hepatica in dairy cattle in Europe. Geospat. Health 9, 261-270

Fox N.J., White P.C., McClean C.J., Marion G., Evans A., Hutchings M.R. 2011. Predicting impacts of climate change on Fasciola hepatica risk. PLoS One 6, e16126.

Mas-Coma S., Bargues M.D., Valero M.A. 2005. Fascioliasis and other plant-borne trematode zoonoses. Int. J. Parasitol. 35, 1255-1278.

McCann C.M., Baylis M., Williams D.J. 2010. Seroprevalence and spatial distribution of Fasciola hepatica-infected dairy herds in England and Wales. Vet. Rec. 166, 612-617.

Selemetas N., Ducheyne E., Phelan P., O'Kiely P., Hendrickx G., de Waal T. 2015. Spatial analysis and risk mapping of Fasciola hepatica infection in dairy herds in Ireland. Geospat. Health 9, 281-291.

Kantzoura V., Kouam M.K., Feidas H., Teofanova D., Theodoropoulos G. 2011. Geographic distribution modelling for ruminant liver flukes (Fasciola hepatica) in south-eastern Europe. Int. J. Parasitol. 41, 747-753.

Rapsch C., Dahinden T., Heinzmann D., Torgerson P.R., Braun U., Deplazes P., Hurni L., Hansruedi B.., Knubben-Schweizer G 2008. An interactive map to assess the potential spread of Lymnaea truncatula and the free-living stages of Fasciola hepatica in Switzerland. Vet. Parasitol. 154, 242-249.

OMICs in food safety: contribution to chemical risk assessment

Matthew Wright

University of Newcastle, The United Kingdom

Prof. Matthew Wright has over 28 years research experience with liver-related drug /chemical metabolism and toxicity. He was the first to demonstrate that stimulating liver myofibroblast apoptosis reversed fibrosis; developed a human single chain antibody to target these cells (licensed to Pharma) and demonstrated an anti-fibrogenic, and anti-inflammatory role for the PXR in the liver. He has published in excess of 240 full research papers and scientific opinions related to toxicology; was an expert for the European Food Safety Authority ANS Panel (2011-2018) and Chair for the Working Group pertaining to Article 8 – rapid response opinions on member state/EC concerns. He currently acts as an Associate Editor for the journal, Toxicology and is a member of the UK Committee on Toxicity.

A number of in silico, in vitro and in vivo approaches have been developed to investigate toxicokinetic and toxicodynamic processes of chemicals. However, a pivotal toxicological study in the appropriate test species, using endpoints from pathological or biochemical investigations, is most often used to identify a reference point (RP) which is then used to derive either a health-based guidance value (e.g. ADI), a margin of safety (MOS) or a margin of exposure (MOE). In this respect, the EFSA ANS Panel - between 2011 and 2018 - re-evaluated 57 food additives (covering 96 separate chemicals, some were evaluated in groups to generate a group ADI). Eighteen had insufficient data to determine an ADI, 1 confirmed a derivation of SCF and 38 concluded on endpoints in animal studies to derive an ADI. No re-evaluations used OMICs data to derive an ADI.

OMICS technologies encompass genomics (structure and function of the genome); transcriptomics (changes in RNA expression); proteomics (protein expression and post-translational modifications) and metabolomics (metabolite profiling) in tissues and cells. Examples of the above will be given and their potential use in future regulatory decision-making process will be briefly discussed. Over the last decades, OMICS technologies have increasing matured, and may currently complement traditional toxicology studies. However, their translation to regulatory decision-making has been slow.

As part of the EFSA Scientific Colloquium 24 – OMICS in risk assessment: state-of-the-art and next steps, a discussion group considered whether OMICS data can be integrated into chemical risk assessment by supporting the identification of RP for hazard characterisation. Four major themes were discussed as outlined below. With the exception of selected examples, the author has chosen for this presentation, the consensus view of the group will be presented.

Do OMICS data provide molecular readouts that are suitable and sufficiently robust to derive Reference Points (RP) (or Points of Departure) for the derivation of health-based guidance values?

Approaches for the generation, analysis and storage of OMICS data was considered important. However, the development of appropriate analysis techniques was widely considered to be lagging.

The group considered that OMICS data has been used more successfully for discovery and hypothesis generation. OMICS has generated a large amount of data on the biological impacts of chemical exposures; has supported the development of adverse outcome pathways (AOPs) and facilitated the grouping of chemicals (e.g. genotoxic vs non-genotoxic carcinogens). Overall, the findings from OMICs datasets have generally been used to support discovery/development questions or regulatory submissions, with the potential advantage that OMICs approaches have greater sensitivities in detecting responses. However, OMICs datasets may also identify associated changes that are not relevant to those causing toxicity and/or reflect adaptive changes. Including at least one dose/time that elicits a classical toxicological response was considered imperative at this stage.

One example where OMICS approaches were being considered as tools having an application in the decision-making process, is via an initiative by Health Canada, where they are assessing health-based guidance values (HBGV) generated through OMICS data (Farmahin et al, 2017). The discussion group noted that OMICS investigations can lead to the identification of novel biomarkers. Few biomarkers had made the translation from research to use in regulatory spheres, however, this may change with time as novel biomarkers become more widely used and accepted.

However, guidelines are likely needed to support and standardise OMICS technology implementation, reporting and analyses so that the data generated is robust and reproducible. In the context of identifying a reference dose, both time and dose-response data will be required. However, many legacy datasets incorporating dose- and time-effects are not available due to cost/technical issues. The 'normal' biological OMICS profiles, encompassing inter-individual variability, is required to maximise the utility OMICS data. This could be of particular value in providing confidence around negative responses to chemical challenges. Ensuring these data are made publicly available will promote the use of OMICs data in risk assessment.

In summary, the discussion group concluded that OMICS technologies could be viewed as a complementary technique, similar to in vitro or QSAR. As a set of approaches to provide additional data that can support the setting of HBGVs, OMICS data was seen as moving from promise to (near) reality. A number of international initiatives may help to establish the role of OMICs in regulatory frameworks.

Which types of OMICS data are most suitable for use in risk assessment?

Overall, the discussion group considered genomics and transcriptomics data easiest to assess, though also their functional impact is furthest from the biological phenotype. Proteomics and metabolomics are near to the biological phenotype but analytical approaches are more limited in scope and may lack critical readouts pertinent to adverse effects. The balance between coverage and biological relevance is a dilemma in OMICs analyses, and one solution could be a move toward multi-OMICs approaches. The discussion group considered that two areas of research are receiving significant attention, allowing for their potential use in the regulatory arena: Whole-genome sequencing and epigenomic signatures. Together, these technologies have the potential to inform our understanding of susceptibility factors, and potentially signatures of prior chemical exposures.

OMICs data have contributed to the identification of novel prognostic, effect and exposure biomarkers (e.g. miR122 or Kim1). However, the discussion group considered that few biomarkers have been validated.

Overall, the discussion group considered that OMICs data sets are potentially suitable for use in risk assessment. Since each technology has their own limitations and advantages, one strategy may be to use a multi-OMICS approach in studies.

Are OMICS data helpful to identify the human relevance of an adverse effect observed in animal bioassays?

Extrapolating from animal models to humans was discussed as there are limitations around the role genetic susceptibilities in influencing adverse events. Under normal circumstances, a single animal strain is used. Even if the selected strain is outbred, the genetic variability within the strain remains limited. It was noted that the Jackson labs have recently introduced a diversity outbred (DO) mouse colony (https://www.jax.org/strain/009376). The colony was generated by crossing eight unique and genetically diverse inbred mouse strains (A/J, C57BL/6J, 129S1/SvImJ, NOD/ShiLtJ, NZO/HlLtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ), followed by subsequent inbreeding to produce 160 new and unique recombinant incipient inbred lines. Random outcross mating from these lines provides the DO colony, which may represent a new approach to studying human susceptibility in animal models.

The group considered if OMICS approaches added value for current challenges in toxicology and risk assessment at this point in time. The consensus view of the group is shown in the table below.

The group considered if OMICS approaches added value for current challenges in toxicology and risk assessment at this point in time. The consensus view of the group is shown in Table I.

Table I: OMICS approaches added value for current challenges in toxicology and risk assessment as determined by a workgroup during the EFSA Scientific Colloquium 24.

	TOXICITY	RISK ASSESSMENT
Species-specific toxicity	YES	YES
Mixture toxicity	YES	NO
Low-dose effects	YES1	NO
Endocrine effects	YES	NO
Nanotoxicology	YES	NO
Epidemiology	YES	YES
Grouping of Chemicals	YES	YES

1OMICs approaches were felt to be of use when there was an absence of overt phenotypic changes at low-dose.

Overall, the discussion group concluded that OMICs data had added significant value to our understanding of chemical mode of action (both desired and adverse), and human relevance of adverse effects identified in animal bioassays. These contributions have not fully translated to the risk assessment arena, but the discussion group concluded that as OMICs approaches mature, they were likely to add value in the future.

What are the challenges for collecting, processing, interpretation, storing and curating large-scale OMICs data?

The discussion group considered each of the major steps in an OMICs experiment, from data collection, through processing and interpretation to long-term storage and curation. It was agreed there are established standards for the collection, storage and curation of OMICs data. However, both data processing and interpretation were seen as challenging at the present time. This was due to a lack of standardisation in protocols, as well as the lag in biological information for contextualisation. While there are standards around data collection, processing and interpretation was a challenge as there are divergent views on these latter aspects. One possible approach considered by the discussion group would be to routinely operate a dual track system: all OMICs data would be processed (and interpreted?) using a reference protocol so that the generated data will be readily comparable with other datasets. However, this would not prevent bespoke processing and interpretation of the data to fit the specific requirements of the experiment. The discussion group noted that there were a number of activities being undertaken by OECD and ECETOC to support standardised curation and reporting processes (Buesen et al., 2017). However, these were at different stages for different OMICS technologies with genomics and transcriptomics being further developed than proteomics and metabolomics. The discussion group considered that performance and system performance checks on OMICs approaches were important.

Overall, while several aspects concerning OMICs experiments have agreed/developing standards, data processing and interpretation remain areas where considerable inter-lab variability exists. This has the potential to slow translation of OMICs approaches into risk assessment.

References

Buesen R., Chorley B.N., da Silva Lima B., Daston G., Deferme L., Ebbels T., Gant T.W., Goetz A., Greally J., Gribaldo L., Hackermüller J., Hubesch B., Jennen D., Johnson K., Kanno J., Kauffmann H.M., Laffont M., McMullen P., Meehan R., Pemberton M., Perdichizzi S., Piersma A.H., Sauer U.G., Schmidt K., Seitz H., Sumida K., Tollefsen K.E., Tong W., Tralau T., van Ravenzwaay B., Weber R.J.M., Worth A., Yauk C., Poole A. 2017. Applying 'omics technologies in chemicals rik assessment: Report of an ECETOC workshop. Regul. Toxicol. Pharmacol., 91 Suppl 1:S3-S13.

Farmahin R., Williams A., Kuo B., Chepelev N.L, Thomas R.S., Barton-Maclaren T.S., Curran I.H., Nong A., Wade M.G., Yauk C.L. 2017. Recommended approaches in the application of toxicogenomics to derive points of departure for chemical risk assessment. Arch. Toxicol., 91:2045-2065.

Session 3:
Perspectives

Gotcha! Network analytics for Fraud Detection

Véronique Van Vlasselaer
SAS Institute Inc., Belgium

Dr. Véronique Van Vlasselaer is a Consultant at SAS, and a true data science enthusiast. In her job, she passionately helps companies to envision and prepare for an AI-driven future, embrace the power of data science, and discover the real value in their data. Before she joined SAS, she graduated as Doctor in Business Economics at the KU Leuven with the department of Information Management and Decision Sciences under Prof. Dr. Bart Baesens. Her Ph.D is oriented towards the development of fraud detection frameworks and solutions from a data science perspective. She is co-author of the book "Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection".

Fraud is an uncommon, well-considered, imperceptibly concealed, time-evolving and often carefully organized crime, making it a challenge for the data scientist to create high-performing, automated detection algorithms.

Although many detection algorithms are often merely based on so-called intrinsic or local features, research has shown that networked data – e.g., how people relate to each other - offers a rich source of additional information to help the fraud investigator. In this presentation, Véronique tackles how such networked data can be easily visualized, mined and integrated in fraud analyses. She claims that effectively fighting fraud is finding the right balance between analytical models that are (1) powerful and (2) easily interpretable.

Unlocking the power of citizen science

Filip Meysman
University of Antwerpen, Belgium

Prof. Dr. Ir Filip Meysman (1970) is full professor within the Department of Biology at University of Antwerp (Belgium). Within the Global Change Ecology Excellence Centre, his research focuses on biogeochemical cycling, the large-scale interactions between biology, chemistry and geology. An important aspect of his research is to investigate how a range of complex environmental problems – like climate change, ocean acidification and urban air quality – can be investigated with the help of complex numerical models, and how such models can be groundtruthed by empirical data. In this light, he has recently pioneered how citizen science can be successfully used to collect powerful "big datasets" in order to constrain and improve air quality models.

Presently, he is coordinating the citizen science project CurieuzeNeuzen, which mobilizes 20,000 citizens across the whole region of Flanders to measure ambient air quality at their front door. In 2017, he received the science communication award from the National Academy of Sciences and Arts for his efforts in innovative citizen science.

CurieuzeNeuzen Vlaanderen is a citizen science project in which 20.000 citizens measure the air quality near their own house during May 2018. The aim is to acquire a detailed map of air quality in Flanders (the northern region of Belgium), both inside urban areas as well as on the countryside.
CurieuzeNeuzen Vlaanderen is the largest citizen science project on air quality to date.

Participants have installed a simple, standardized measurement device on a street-facing window of their house, apartment or building. Two diffusion tubes determined the mean concentration of nitrogen dioxide (NO2) in the ambient air over one month (May 2018).

The samplers are attached to a V-shaped window sign commonly used in advertising real estate in order to establish a standardized measurement setup. NO2 is an important indicator for traffic pollution. The data collected from the diffusion samplers are quality controlled and calibrated with NO2 measurements at reference monitoring stations operated by the Flemish Environment Agency (VMM).

The large dataset collected by CurieuzeNeuzen Vlaanderen will be used to test the state-of-the-art ATMOSYS computer model (developed by VITO for the Flemish Environment Agency VMM), which is currently used to assess air quality in Flanders. By improving the predictive capabilities of this model, we will arrive at a better estimation of the population exposure to NO2 and its effects on public health, allowing to provide better information and recommendations to policy makers.

Air quality can vary significantly over short distances, especially due to the street canyon effect (pollutants accumulate to higher concentrations in narrow, poorly ventilated streets with intense traffic density). Because air quality is so spatially variable, many measurement locations are required to properly assess the predictive capacity of the air quality model. This is why help from citizens is extremely valuable to gather "big data" on the spatial distribution of air quality.

Citizen science projects don't just target the collection of "hard "datasets to advance science, they also have a "softer" side, as they can raise awareness by providing information to the general public . In the latter regard, CurieuzeNeuzen Vlaanderen aims to increase public awareness of the importance of air quality for a healthy environment, and wants to stress the need and importance of performing reliable air quality measurements.

Federal Agency for the Safety of the Food Chain
CA-Botanique - Food Safety Center
Bd du Jardin Botanique 55 - B-1000 Brussels
www.fasfc.be

SciCom