Lay Summary
PIG DATA: Health analytics for Swiss pig farming

Project team

- Dr John Berezowski, Veterinary Public Health Institute Vetsuisse-Fakultät, Universität Bern

- Dr Rolf Grütter, Eidg. Forschungsanstalt für Wald, Schnee und Landschaft WSL

- Prof. Abraham Bernstein, Institut für Informatik, Universität Zürich

- Prof. Heiko Nathues, Clinic for swine, Department of Veterinary Clinical Medicine, Universität Bern

- Dr Christina Nathues, Veterinary Public Health Institute, Vetsuisse-Fakultät, Universität Bern

Contact address
Dr John Berezowski
Veterinary Public Health Institute Vetsuisse-Fakultät, Universität Bern
Schwarzenburgstrasse 161
3097 Liebefeld
john.berezowski@vetsuisse.unibe.ch

Date
January 11, 2021

1. Background
With the continued improvement of computer and IT technologies over the past few decades, many (maybe most) of human endeavors in high income countries are managed to some degree through the collection of large volumes of data. These data, called Big Data are created at a rapid pace, are highly variable, and are produced in large volumes. They have the potential to be a valuable resource because they may contain new information that may be useful in a variety of settings. However, the complexity of Big Data means that these data can no longer be analyzed using conventional methods and new methods must be developed to extract useful information from them.

Big Data methods have been developed and are being used in many parts of society including business, health, information technologies, computer science, engineering and others. So far, these methods have not had much impact on animal production systems, especially the Swiss pig and pork production industry, even though they would be of great interest. Reasons for this are unclear but could be related to the unique and very complex nature of pig production in Switzerland. In most industrialized countries, pork production systems are supply chains made up of small numbers of very large production units integrated into one single or multiple owner organization. The Swiss pig production system is quite different. It is made up of a very large number of small independently owned businesses (pig producers) and smaller numbers of larger businesses (breeding farms,

marketers, feed suppliers, slaughter plants, distributers and others) that have self-organized into very complex pig and pork supply chains. Most of the small and even some of the large business would not have the resources to combine and analyze these data.

Much data is currently being used to manage different parts of the Swiss pig and pork production system. Similar to other industries, these data have the potential to be analyzed to create useful information that may improve industry efficiency, profitability, sustainability and pig health and welfare. However, these data are not being used in a way that brings together all the various data from all stages of production. If this information were suitably prepared and analyzed, it will be possible to recognize new links, causes and risk factors in relation to diseases and/or a drop in performance, and to identify the best strategies for combating them. Information extracted from these data could be beneficial for Swiss pig production, the Swiss economy and society in general. Failing to use this resource will result in the Swiss pork production system falling behind competitors in neighboring countries.

2. Goals of the project
The overarching goal of the project is to develop new methods aimed at gaining a better understanding of, and optimizing, the structure and complexity of the pig farming and production network in Switzerland. This will not only utilize existing data but also tap into potential sources of new data in order to improve animal health, boost animal welfare and make pig production more sustainable

Specific goals of this project are to see if it is possible to:

1. demonstrate that new, useful and valuable information can be created by combining many disparate data sets from across a complete pig supply chain;

2. develop trusting relationships between the researchers in the project and the data collectors in a complete Swiss pig production system so that they are comfortable enough to share their data and the results of data analyses with each other;

3. process and join the many different data sets from a pig supply chain together in a single data store that is suitable for data analyses;

4. use existing methods or develop new methods to analyze the data in order to produce new information that is useful and valued by the swine supply chain partners in the project;

5. successfully conduct a research project involving researchers from different disciplines who have never worked together before;

6. develop new methods and create new information that is valued by the scientific community and therefore accepted for publication in scientific journals and presentation at scientific conferences;

7. attract and mentor graduate students to complete their research thesis within the project.

3. Methods

We began working on this project well before the official project start date. The first step was to assemble a team of researchers (senior research team) with the necessary expertise and research backgrounds to ensure the successful completion of the project. The senior research team consisted of specialists in swine veterinary medicine and production, computer science and Big Data, veterinary epidemiology, and geography. The second step was to identify members of a pig supply chain who would be willing to contribute to the project by providing their data and collaborating with the senior research team. In order to ensure that the project would create useful information, pig supply chain partners were asked to identify information that they didn't have but needed to solve important problems (called Dream Queries). After completion of the first two steps, the senior research team developed the project plan and submitted the research grant application.

Once the project started, a research team consisting of a project manager, Post-Doctoral students and other graduate students (PhD, MSc and Bachelor) was created. Key partners from all areas of pig production (farmers, veterinary surgeons, marketing organizations, animal feed manufacturers and slaughter companies) were re-contacted and confirmed their willingness to participate in the project and contribute their data. Data sharing and use agreements were signed with all data providers. These agreements, which included, anonymizing the data, confidentiality of data, privacy of data suppliers, control of which data could be used for research, and which information could be shared were essential for building trust among all members of the project.

A central data storage space (Pig Data Space) was created and methods were developed to allow data providers to submit their data. Considerable time and effort were required to transport, clean, join and format all of the highly variable data so that it could be analyzed within the Pig Data Space. This required many discussions between individual data providers and the research team in order for the research team to understand the data. Because the research team members had different backgrounds, methods were developed to allow each researcher to use the analytical platform and methods they were comfortable with. This stimulated a new research focus aimed at understanding the ways in which researchers from different background clean, process and analyze the data. Since the research team came from 4 different institutions, methods that allowed researchers to conduct their research from different locations were implemented.

Data analysis focused on answering the Dream Queries that were provided by the pig supply chain partners. The first step was to examine the different datasets in order to determine whether the combined data were of sufficient quantity and good enough quality to answer the Dream Queries. Some Dream Queries could not be answered for reasons relating to data quality and quantity. Answers to the Dream Queries were communicated to supply chain partners in scientific reports and in regular meetings. Supply-chain partners provided feedback about the quality and usefulness of the Dream Query answers. Regular in person meetings were held with the research team and supply-chain partners in order to discuss the research, build trust and develop a strong relationship between all.

4. Results

The project was an overall success in many ways. Here we present the results for each of the specific goals of the project:

1. The project demonstrated that new, useful and valuable information could be created by combining many disparate data sets from across a complete pig supply chain. The pig data supply-chain partners reported that the information created to answer the Dream Queries was interesting and useful. In addition, the supply-chain partners recognized that the information could not have been created without combining data from many data providers who normally would not have combined their data.

2. The project successfully developed trusting relationships between the researchers in the project and the members of a Swiss pig production system. The pig supply-chain partners shared their data with the research team, they collaborated with the research team throughout the project, they allowed the results of the analyses of their data to be shared among all supply-chain partners and they approve the publication of these results in scientific journals and at scientific and industry conferences.

3. The research team successfully developed a central data store and developed methods to clean, process and join the many different data sets from a pig supply chain. The cleaned and joined data were made available to the research team who used them for analysis to answer the Dream Queries.

4. The research team used existing analytical methods and developed new methods to successfully analyze the data to create new information that was valued by the supply-chain partners in the project. New methods were developed by the graduate students in the research team and made publicly available through publication in scientific journals.

5. The research team was very diverse which created a challenge for agreement on the best research strategies. Developing consensus on the best approaches was achieved through frequent online and in person meetings. The team was able to jell into working unit that achieved the goals of the project.

6. The project was successful at developing new methods and creating new information that was accepted by the scientific community. The project team has published 3 studies in scientific journals (5 are currently being written), 2 studies in scientific conference proceedings, and presented at 9 scientific conferences.

7. The project successfully attached attracted graduate students who have completed or are in the process of completing their thesis within the Pig Data project. Furthermore, several residents from the European College of Veterinary Public Health and the European College of Porcine Health Management contributed to the Pig Data project, especially on the analysis of Dream Queries.

5. Significance of the results for science and practice

The adoption of Big Data approaches by the Swiss swine industry has been relatively slow, especially when compared to other industries and countries. This was one of the main drivers for initiating this project. It was well known that the structure of the swine and pork production system in Switzerland differs greatly from the intensive production systems in other countries. For that reason, the methods developed and applied elsewhere were of limited value for Switzerland. It required that we create a new strategy for the adoption of Big Data approaches for the Swiss swine industry. This project demonstrated that it is possible to combine data from various actors in the pig supply chain. It also firmly established that this approach can generate new and useful information. The Pig Data project can be seen as a pilot experience about the application of Big Data approaches in the Swiss swine industry. Its success was key to drawing the attention of key stakeholders (such as the Federal Food Safety and Veterinary Office) to the importance of this topic and for the planning of new projects that will further explore the field. This is evidenced by the official support of a new project that aims at establishing a 'competence & information centre for pig health in Switzerland' which will collect, analyze and publish health data from various projects and practitioners in the field in order to obtain a real-time overview of the population's health status and the occurrence of emerging and re-emerging diseases.

One of the Pig Data project components aimed at answering questions from the industry partners (Dream Queries). As agreed with the industry partners, some of these Dream Queries were answered using a scientific approach. The results of these Dream Queries provided the industry partners direct input that can be used to improve their processes. It also helped the industry partners to evaluate their policies. For example: the slaughterhouse company enrolled in the project applies economic penalties if carcasses present meat quality parameters that are not within a given range. When analyzing data related to these meat parameters, we verified that the vast majority of pigs are within the range set by the slaughterhouse company.

The results of this project will enable livestock owners and veterinary surgeons to take steps to improve animal health and prevent disease in their pig farming operations. Optimized production processes will lead to greater efficiency, sustainability and value creation in Swiss animal production. Ultimately, end users will also benefit from improvements in animal health, reduced antibiotic use and the resulting foodstuffs. We would also like to make the prototype of the Pig Data Space accessible to marketing organizations, veterinary surgeons and farmers, so that they can make beneficial use of the system to help them with future questions and decisions.