PPS annual report 2018

Approval secretary / consortium The secretary / consortium will review the annual report. The Topconsortia voor Kennis en Innovatie (TKI) (top consortium for knowledge and innovation) would like to be informed about any remarks on the annual report.	
The secretary on behalf of the consortium has (…) the annual report	V approved □ disapproved
Any remarks on the annual report:	-

General information	
PPS number	AF-16160
Title	1H4F-SaferFood-BIGdata
Top section and innovation theme	Smart Agri & Food, Food safety
Project leader (research)	Dr. Alex Bossers (formerly Paul van den Wijngaard)
PPS coordinator (private partner)	Vion Food NL - Dr. Bert Urlings (coordinator / penvoerder) - Dr. Martijn Bouwknegt (secretary)
Contact government	Cor Wever
Status (ongoing or finalized)	Ongoing
Type of research (F, T of V)	F, T
Official start date	2017 April 1 (cash payment association shifts approved at the start)
Official end date	2021 March 31
Organization / management structure	Representatives of each private partner (Vion, Thermo Fisher (TF), IBM) and the DLO/WUR-partners (WBVR and WU) consult quarterly with each other through teleconferencing. Vion organizes and IBM facilitates the plenary teleconferences. At least once each year a physical meetings take place, in which results are presented and plans for the coming period are outlined.

Counseling structure (sounding board etc.)	Partner WU (prof Marcel Zwietering) acts as an advisor for the project contents. The consortium as such outlines the project and its tasks, while the 1H4F advisory and steering group reviews the progress and content.
Brief description of the content (max 4 lines).	With big data and machine learning we aim to identify microbiome derived biomarkers for the early detection of product contamination in slaughterhouses. Through this early warning system interventions can take place to avoid further contamination of meat products with food pathogens. This monitoring will lead to improved food safety for consumers and reduced contamination related economical losses.

Planning and progression (please indicate changes with respect to the project plan)	
The PPS is going according to plan?	Yes
Are there changes in the consortium / project partners?	No
Are there any delays or postponing of completion date?	The project starting date was set at April 1st 2017. Budgets were agreed to shift one quarter in time accordingly. The Consortium Agreement was signed on 2017 July 24 by all parties.

Highlights: give a brief description of the most important results (this description will be posted as public summary on the websites of the TKI's/topsectoren)

The expertise of the various partners has been used to set up and perform several pilot experiments. In addition, 16S sequencing and preliminary statistical analysis has been performed on a subset of the collected samples and its meta data. A central data-sharing platform has been initiated through the Thermo Fischer cloud.

As proof-of-principle we aim *to identify microbiome related biomarkers that correlate with potential spill-over of pathogenic bacteria (i.e. Salmonella) that can be found on the carcasses in the pig slaughterline.* To achieve this, we defined a number of short-term sub questions to explore and to improve our knowledge in this subfield.

1) What is the scientific state-of-the-art literature regarding biofilm formation in the meat processing industry?
2) Are the salmonella present in the biofilms clonal or a reflection of the salmonella introduced by the transported animals?
3) Does full metagenome sequencing have an additional value for the detection of the primary microbiome composition of the carcass?
4) What insights can be gained from an initial analysis of the 16S sequencing results?
5) Will pooling of the samples have impact on the 16S sequencing results?
6) Are there differences in pattern in microbiome composition between days ahead a Salmonella positive and negative sample?

Point 1-4 have been presented at a physical meeting in April 2018 in Zurich, which was hosted by TF. Point 5 derived directly from this meeting. Point 6 is currently being investigated by IBM. As soon as we have a proof of concept to identify biomarkers using described methods the project can shift focus to the usability of such markers either directly or by developed novel diagnostics.

#1 Biofilm formation/composition

A literatures study has been started to update the current knowledge on biofilms in the meat processing industry. This study aims to answer questions such as which bacteria and composition of the microbiome play an important role in biofilm formation, which factors contribute to the formation and what are the intervention possibilities. Specific knowledge gaps on biofilms could potentially be bridge in laboratory experiments at a later stage. Primary involved are WBVR, WU and Vion. A first draft of this study has been written and revised by the partners. The next version is in progress where more specific targets for future research can be proposed.

#2 Clonality of Salmonella on the slaughterline

To determine whether Salmonella is derived from external sources or from biofilms present in the slaughterhouses, we set up a study to compare salmonella isolates. Little variation in the genetic sequences indicates an internal source of contamination. The samples were collected at different time points and from several pig slaughterhouses. In total 100 samples were sent to WBVR from which the whole genome was sequenced and analyzed. From the primary genetic analysis, samples cluster according to date of sampling and to a lesser extent on sampling location. To determine if strains are from the same source, publicly available data from known outbreaks strains have been investigated with the same clustering methods. If possible the findings will be published in a scientific journal in 2019. Salmonella WGS proved to be a good alternative to serotyping methods providing more detail (at single nucleotide level) if strains are clonal or not (high resolution).

#3 Primary microbiome composition of the carcass using metagenome sequencing

Samples for routinely monitoring were doubled to provide samples for this project. With little adjustment these samples are proven to be useful for assessing the microbiome composition. The adjustments needed concerned the processing of the primary sample in which the sample was washed (the microbiome consisting fraction) instead of complete homogenization (the standard procedure) to reduce the amount of pig DNA released in the buffer. These samples were subjected to 16S-rRNA gene sequencing (16S-barcoding) and meta genome sequencing to determine the microbiome composition and quantification. Tissue homogenized samples are usually not suitable for the direct metagenome sequencing method, as with this method all present DNA and RNA is sequenced and such samples will most likely contain > 99% host (pig) DNA. A pilot confirmed that the before assumption was indeed the case. For a subset of 12 samples, WBVR used the DNA full metagenome sequencing method to assess the microbiome composition. The results show that >98% of the sequences are derived from pig DNA and the remaining are from bacterial and viral genomes. At this stage it was decided not to use the metagenome sequencing method to assess the microbiome composition since the search space (2%) was very small and probably multiple deep-sequencing rounds would be necessary to obtain enough data. Nevertheless, IBM will also inspect the generated raw data and see whether it might deliver some signatures anyway. A subsequent RNAseq experiment was discussed and proposed on similar samples to assess the present RNA repertoire which ideally might allow the assessment of the microbiome metabolic state (dead or stressed primary bacteria still present from farm/transport or active biofilm spillovers). Details and formal approval of that study are pending until the next physical meeting in April 2019.
Secure data sharing large datasets has been initiated between WBVR and IBM.

#4 Insights from the initial analysis of the 16S-derived sequencing results

More than 2,000 samples have been collected from 2017 October till March 2018, 91 of these samples are currently being sequenced by TF using amplicon 16S RNA gene barcoding. Selection of these samples was based on classical monitoring results for enterobacteriacaea and Salmonella as proxy. Using the 16S data and collected metadata such as date of sampling, slaughter day and time, temperature, relative humidity, IBM performed preliminary statistical analysis. The purpose of the analysis was to explore the data, to find sameday and longitudinal

correlations between microbiome measurements and culture results, gain insights into the similarity and diversity of bacteria on carcasses. The initial results show that the collection and sequencing procedures are yielding good data. The microbiome compositions for different samples show a temporal coherence, meaning that samples on the same or adjacent days show more similarity than samples collected far apart in time. However, due to the limited Salmonella positive days in the 91 analyzed samples, it is too early to develop predictive models, for this the microbiomes from many more carcasses need to be sequenced and analyzed. To maximize the number of carcasses and days covered within the sequencing budget, a strategy of pooling samples across carcasses from a given collection day has been selected. A pilot study has been set up to test the effect of pooling samples on the 16S sequencing results, see point #5. This will be done to reduce DNA processing time and sequencing cost. If the results of the pooling pilot are as hoped, the next stage is to perform sequencing on a larger number of samples, with two pooled samples per collection day, so that a statistically meaningful number of days with both positive and negative results from classical monitoring may be analyzed for development of a predictive model.

#5 The effect of sample pooling strategies on the 16S-barcoding results

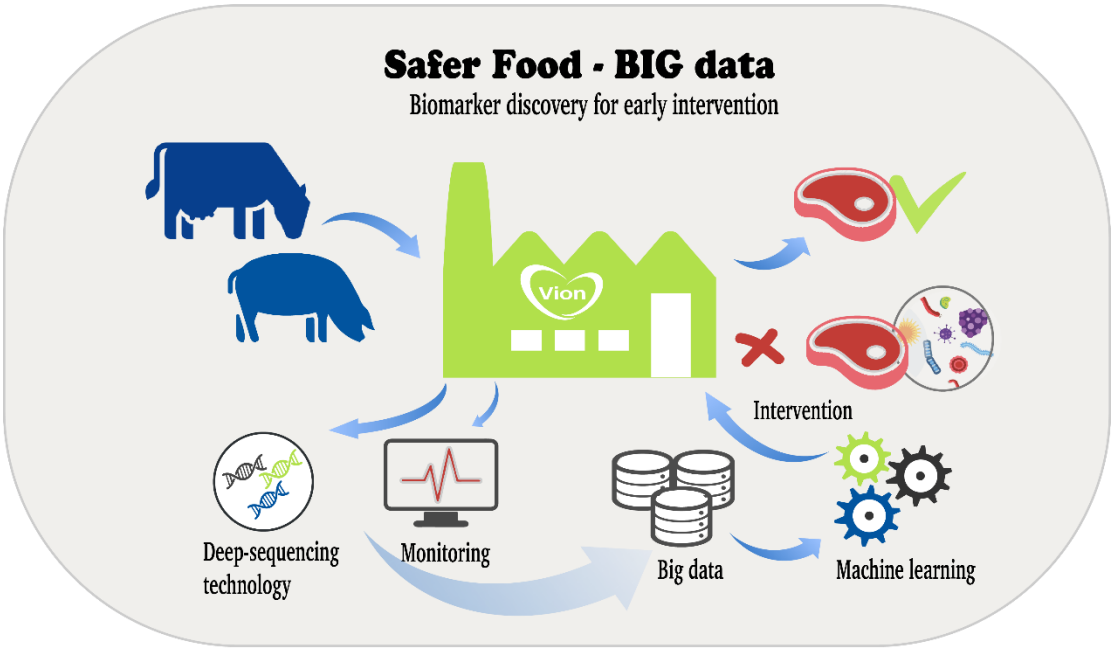
To test whether there will be a significant data loss by pooling of samples (where pooling would decrease the cost-per-sample), a pilot study was initiated in which various pools of 5 and 10 samples were 16S sequenced and the results were compared to that of the individually sequenced samples. Samples were selected such that it potentially included the preliminary signals picked up from the data by IBM in the data from the pilot-month of sampling (i.e., October 2017). The pooling experiment was completed end 2018 and the results showed that pooling reduced the detection sensitivity of low abundant species. However, this is balanced by: 1) still an as-good (or at least not-worse) capture of the overall microbial diversity compared to analysing individual samples, and 2) the lower costs of the analyses. Furthermore, we hypothesized that biofilm-species supporting growth of foodborne pathogens in the slaughterhouse possibly have an ecological advantage over other microbial species in the biofilm and therefore possibly are not present in undetectably low numbers. Therefore, it was decided to continue with this pooling strategy for the remaining samples.

#6 Patterns in microbiome composition ahead a Salmonella positive or negative day

This question is currently being investigated by IBM. The results will be presented in a physical meeting held by Vion in the Netherlands April 2019.

Social application: give a brief description of the social application / payoff

There are strict food safety regulations for the production of pork and beef to guarantee consumers safety. Despite these regulations contamination with food pathogens such as *Salmonella*, *Listeria monocytogenes* and STEC during the production chain cannot be fully eliminated. Contaminated products can lead to food poisoning, wasting of food products, substantial economical losses and brand damaging in case of recalls. This project aims to further reduce the small percentage of potential contamination using early detection signals (biomarkers) developed with innovative methods. These biomarkers derived from the existing microbiome in the production chain will act as early predictors when contamination occurs with food pathogens. We will use BIG-data and machine learning algorithms to identify such biomarkers which subsequently can be used to develop low cost rapid routine diagnostics.



Bottlenecks: give a brief description of the potential bottlenecks

Not applicable

Number of delivered products in 2018 (specify in the appendix the titles and/or descriptions of the products or a link to the products on public websites)					
Scientific articles (*)	Reports	Articles in trade journals	workshops/ invited lectures	Patent applications /first filings	Spin-offs (**)
Not yet (2 planned)	None	none	Consortium meetings	none	none

(*) A reminder: in case of financial support from the TKI allowance, the TKIs should be acknowledged in the publications. In case of publications in high impact journals the TKIs would like to be informed in advance.
(**) This indicates: financial support from other parties resulting from this project, additional subsidies and spin-offs activities.

Do you expect patent applications this year?	NO

Planned publications:

Publications in scientific journals

- Review 'Biofilms in the food industry with the focus on meat processing plants'. Planned.
- Genome announcement paper (Salmonella genome analysis for clonality in meat processing plants). Planned.

Publication in trade journals/websites:

- 1H4F website: project description