Near real-time large scale (sensor) data provisioning for PLF

M.R. Vonder[1], B.D. van der Waaij[1], E.J. Harmsma[1] and G. Donker[2]
[1]*TNO, Eemsgolaan 3, 9727 DW Groningen, The Netherlands*
[2]*Rovecom, Elbe 2, 7908 HB Hoogeveen, The Netherlands*
Matthijs.Vonder@tno.nl

Abstract

Think big, start small. With that thought in mind, Smart Dairy Farming (SDF) developed a platform to make real-time sensor data from different farms available, for model developers to support dairy farmers in Precision Livestock Farming. The data has been made available via a standard interface on an open platform in real-time at the individual animal level. The platform, known as InfoBroker, must furthermore be suitable for a large-scale roll-out. The concept of this InfoBroker is designed as a breakthrough when it comes to making data stored in diverse places available in an efficient manner. Data is not stored centrally, but remains at the source. The InfoBroker is capable of retrieving individual cow data from an infinite number of sources while at the same time serving a large number of models on-demand. For each farm it is specified which data may be released by the InfoBroker. This means that the farmer continues to be the owner of the data. In SDF 1.0 (2011 - 2014) the InfoBroker is developed and used for supporting 7 farmers by providing them 11 different cow-specific work instructions, generated by 3 expert models.

This paper describes the results from SDF 1.0 and also the next steps to be made in SDF 2.0 (2015 - 2017): the use of data-driven analyses of the cow centric data and the use of Linked (open) Data technics as a mean for standardizing the exchange of sensor data for the 60 farmers in SDF 2.0 (with an outlook to the 15.000+ farmers in the Netherlands).

Keywords: linked data, infobroker, sensor data, cow centric, big data, precision livestock farming

Introduction

The SDF program’s overall aim is to support dairy farmers in the care of individual animals, with the specific goal of a longer productive stay at the farm and with an increased lifetime milk production due to improvement of individual health (Knijn et al, 2013). Within this paper we focus on the IT architecture of the SDF program.

In order to quickly respond to changes in the cow’s situation, cows need to be monitored in near real-time and this data must be made available for near real-time models. In other words the program provides *cow specific advice to the farmer, based on near real-time analysis models using near real-time sensor data*.

Two design principles were formulated:

1. Be as flexible as possible: let the farmer decide which sensor devices he uses and from which analysis models he wants to make use of, and also the additional granted parties which are allowed to use his data.

2. Sensor data is not copied and or stored centrally, but remains at the original storage location (e.g. of the device supplier).

The first design principle involves the requirement that the farmer should be in control of which equipment he uses on his farm. Farmers have several devices in use, such as: yield measurement in the milking parlor, water dispensers, weighing scales. Several manufacturers supply such devices and ideally they all should be useable to provide sensor data for the analyses models. At the same time the farmer should be free in selecting a consultancy organization to provide him with insights and advice on his cattle. All consultancy organizations should be able to feed their analyses models with real-time sensor data coming from the farm of the customer. In both cases it is the farmer himself who determines what he wants to use and pay for.

The second design principle is about the location and control over the sensor data. The sensor data of a particular farm is owned by the farmer. He should be the only one in control over who is using his sensor data for what purpose. In order to emphasize this level of control, the data should remain at the source, under the responsibility of the device supplier. Although storage, access control, billing, etc. can be centralized (outsourced), the data itself should stay under control of the owner (the farmer).

Material and Methods

<u>Available Data</u>

Seven farms participated in the SDF 1.0 program, all located in the northern part of the Netherlands. On these farms eight categories of sensors were used in different combinations, see the table below. These sensors came from a diverse group of suppliers (in alphabetical order): Agis, DeLaval, Förster, Gallagher, Lely, Nedap, SCR and smaXtec.

Table 1: Indication of available sensor data categories at the farms

	Farm 1	Farm 2	Farm 3	Farm 4	Farm 5	Farm 6	Farm 7
# cows/calves[1]	459	186	315	239	706	202	351
Behaviour	x				x		
Temperature	x				x		
Activity	x	x	x	x	x	x	
Milk production	x	x			x	x	x
Food intake		x				x	x
Weight	x	x	x	x	x	x	x
Water intake			x	x			
Milk intake			x	x			

[1] This row gives the total numbers of cows and calves at the farms (February 2015). Not all the animals are monitored for SDF yet (e.g. at farm 3 and 4 only calves are monitored for Young Stock Rearing).

Each sensor category consists of one or more sensor types. The milk production measured by Lely consists next to the milk yield, also of the milk duration, visit duration, protein, fat, lactose, milk temperature and more. Other manufacturers of milking equipment might use a different set of parameters. In average there are about 25 sensor types per cow, the exact number of sensor types per cow might vary, depending on the farm.

In addition to the near real-time available sensor data we make use of other cow related data, e.g. birth date, gender, last in heat date, last insemination date, expected calving date, number of inseminations, etc.

Methodology

Based on the two design principles mentioned in the introduction, the following methodology is developed:

1. Convert device centric sensors into *cow centric sensors*
2. *Decouple* direct link between sensor sources and analyses models, by making a broker which routes data requests towards the correct source(s).
3. *Semantically integrate* sensor data from different sources:
   a. Describe semantically sensor data and bring sensor types closer together.
   b. Standardize sensor data in order to avoid inconsistencies between two sensor sources.

Step1: Why cow centric sensors?

The cows are passing along most devices on the farm. Each cow visits the device at certain moments in time. The sensor in the device produces a list of measurements containing data for all visits. Often a farm has several devices of the same type, multiple milking stands, multiple weighing scales, etc. Each device is producing measurements for all cows.

A cow specific analyses model would like to have sensor data on a *per cow* and *per device type* basis. For each cow and for a requested period all its weights measurements, unrespectable which weighing scale produces the measurement, as long as semantically the measurements are the same, for instance tare weight in kg. If the measurements are semantically not the same, step 3 is necessary.

In order to facilitate these cow specific analyses models, the device centric sensor data should be converted into cow specific sensor data.

Step 2: Why decouple?

A cow specific analyses model needs input from one or more real-time sensors. There are several reasons to decouple the direct link between sensors and analyses models:

- Provide a one-stop-shop for analyses vendors for retrieving data to feed their analyses models, avoiding each model to implement all available device specific interfaces to collect the necessary real-time sensor data.
- To be able to process sensor data requests about data which is divided over several locations (farms).

- To be easily able to add access control, in order to limit which analyses models (or any other interested party) may use which part of the sensor data of a certain source.
- To be easily able to add billing, in order to bill analyses model vendors (or other using parties) for the usage of the sensor data of a specific farmer.
- To be able to outsource the hassle of managing all these points and make it available as a service.

Step 3: Why semantic integration?
Adding new sensors instantly, without involved parties having to implement a new interface, would require also semantic integration. As mentioned before, possibly not all measurements of the same device type, such as a weighing scale, a water dispenser, etc., have indeed the same semantics. Some weighing scales produce tare weight, others don't. Some water dispensers produce measurements in ml (milliliters) others in l (liters). Analyses models have to understand the semantics of a certain measurement if they want to use that sensor data. To reduce the burden on the analyses model developers it would be convenient if these differences can be automatically converted into a concept (liters for instance) that the analyses model wants to use. Some of these differences can be solved by describing them all semantically and also describe their relation towards one another (a 1000 ml is a liter), step 3a. Ideally there is a commonly agreed standard which describes the concept in a uniform manner, step 3b. Then all devices can internally convert the output of their sensors into this standardized concept and all measurements are directly comparable and usable by an analyses model without any necessary conversions.

Results and Discussion

To successfully develop a platform that makes real-time sensor data available to model developers, from different farms, the data must be made available via a standard interface on an open platform in real-time at the individual animal level. The dairy farmer remains the owner of his data. He must be able to decide which parties can have access to this data. The platform must furthermore be suitable for a large-scale roll-out in order to be prepared for the majority of the Dutch cattle farms. This includes farmers, suppliers of sensors, model developers and service suppliers who currently are not participating in the program. The ambition is to extend this also to other countries. Interesting work concerning ontologies and semantics has already been done by the project agriOpenLink (Tomic et al, 2014).

Based on the methodology, the InfoBroker is being developed as a generic platform for making cow-centric data available. The suppliers of sensor data, as well as the model developers, will subsequently have to adopt a cow-centric approach. The real-time models translate the retrieved sensor data into work instructions (also called Standard Operating Procedures, SOP's) that can be applied instantly in practice by the dairy farmer or his employees. In figure 1 a schematic overview of the architecture is depicted. The platform is currently being used by 7 farms, but has been designed to at least support the more than 15000+ Dutch dairy farms in real-time.

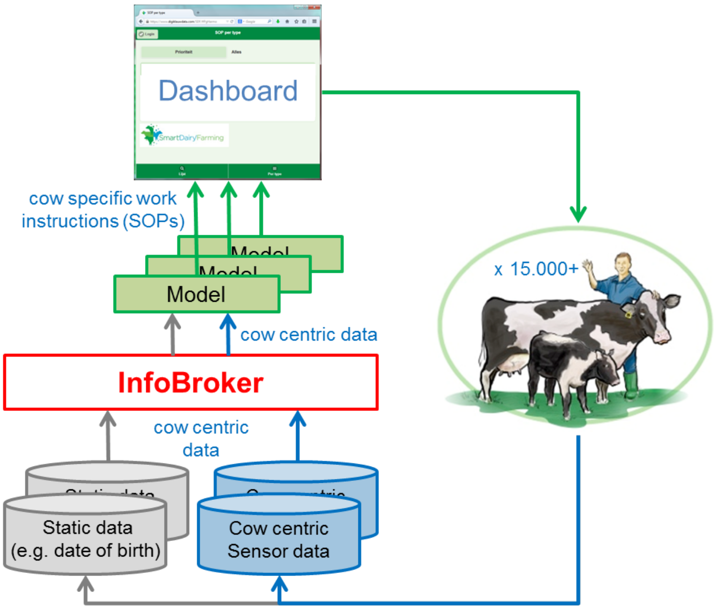

Figure 1: Schematic architecture

Below a number of graphs which give an indication of the amount of data that has been made available via the InfoBroker (November 2014 up to January 2015).

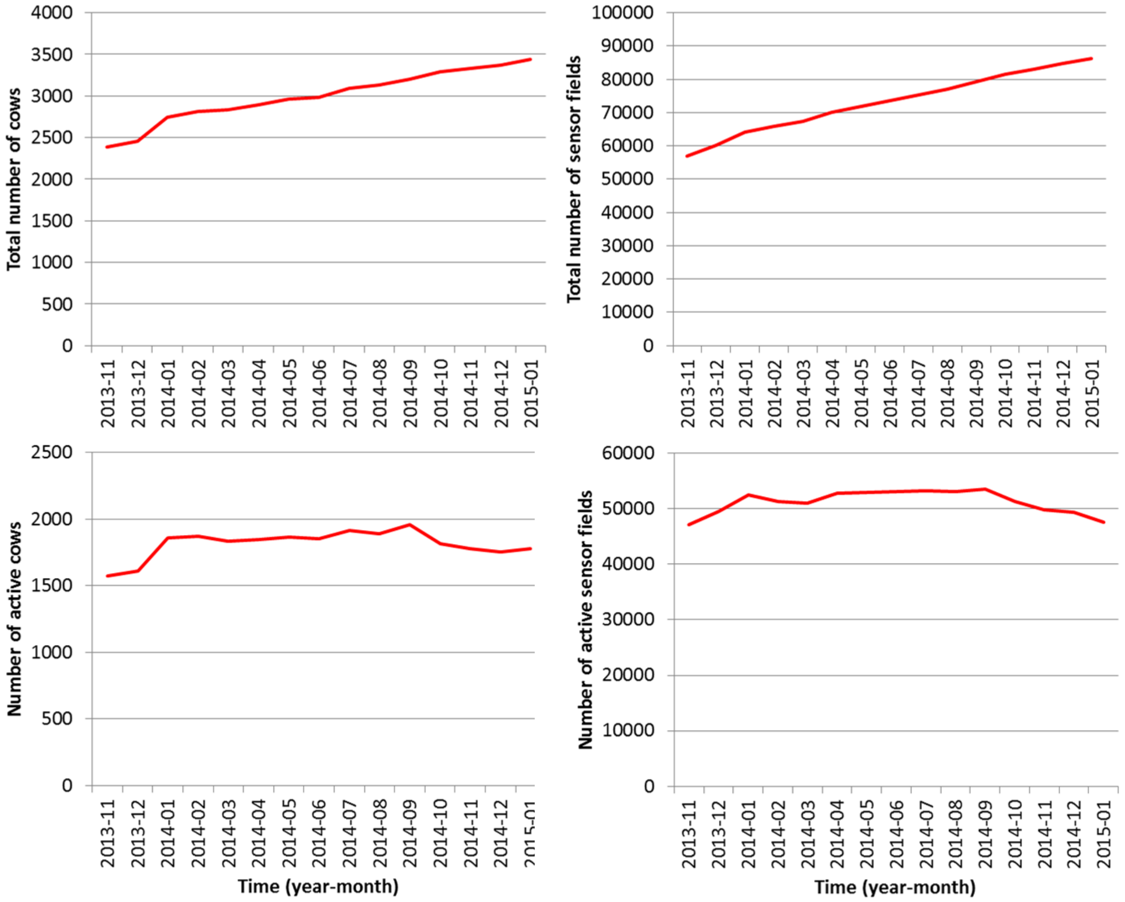

Figure 2: Number of *cows* (left column) and number of *sensor fields* (right column), with the *total numbers* (top row) and the *active ones* (bottom row)

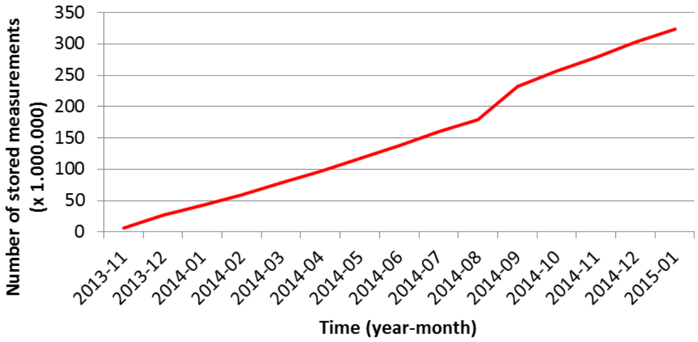

Figure 3: Number of stored measurements over time

The setup of the SDF 1.0 program resulted in the following metrics:

- 7 farms were participating, with 180 – 700 cows per farm (see also table 1).
- In average there were circa 1800 actively monitored[2] cows, with circa 51000 active sensor fields (see figure 2).
- This is an average of 28 different sensor fields per cow.
- More than 1 year of sensor data is gathered, resulting in 325 million measurements (figure 3) which equals to 23 Gb of stored measurement data.

The developed architecture and InfoBroker platform of figure 1 is used by several real-time models, like *young stock rearing* (Ipema et al, 2014), *heat detection* (Kamphuis et al, 2014) and *transition* (not yet published).

Conclusions and Future work

This paper recommends a platform (architecture and approach) for near real-time large scale (sensor) data to provide Precision Livestock Farming. The platform gives access to the sensor data of approximately 1800 actively monitored cows (and a history of an additional 1700 cows), with an average of 28 sensor fields per cow. The platform is used by several near real-time models to create work instructions (SOP's) for farmers on individual cow level in the field of young stock rearing, heat detection and transition.

<u>Future work</u>

In the follow-up program SDF 2.0 the transition is made from a "Proof of Concept" (SDF 1.0) to a "Proof op Practice". That means the developed real-time models using the InfoBroker will be used in a 50 - 60 additional farms (with an outlook to the 15000+ farmers in the Netherlands).

In order to implement the semantical integration, Linked (Open) Data technics will be used as a mean for standardizing the exchange of sensor data.

Next to this, the use of data-driven analyses of the cow centric data will be explored, e.g. to what extend can this support the farmer with concrete work instructions in Precision Livestock Farming?

---

[2] If sensor data for a specific cow in a specific month is available via the InfoBroker, this cow is considered to be actively monitored.

Acknowledgements

This research was supported by the Dutch Smart Dairy Farming research program, which is financed by Friesland Campina (Amersfoort, the Netherlands), CRV (Arnhem, the Netherlands), Agrifirm (Apeldoorn, the Netherlands), Dairy Campus (Leeuwarden, the Netherlands), Investment and Development Agency for the Northern Netherlands (Groningen, the Netherlands), the Dutch Dairy Board (Zoetermeer, the Netherlands), Mesdag fund and the Ministry of Economic Affairs, Agriculture and Innovation, Pieken in de Delta (Den Haag, the Netherlands).

References

Ipema, A.H., de Mol, R.M. and Hogewerf, P.H. 2014. Young Stock Rearing using automatically recorded data. In: *Int. Conference of Agricultural Engineering*.

Kamphuis, C., Huijps, K. and Hogewerf, P.H. 2014. Smart Dairy Farming in The Netherlands: case heat detection. *Abstract for Smart Agrimatics 2014*, Paris.

Knijn, H., Taweel, H., van Wichen, H., Wulfse, B.J. and Vonder, M. 2013. Smart Dairy Farming program in the Netherlands. In: *Proceedings of the precision dairy conference and expo "Precision Dairy 2013"*, 141-142.

Tomic, S.D.K., Drenjanac, D., Lazendic, G., Hörmann, S., Handler, D., Wöber, W., Schulmeister, K., Otte, M. and Auer, W. 2014. agriOpenLink: Semantic Services for Adaptive Processes in Livestock Farming, In: *Int. Conference of Agricultural Engineering.*