F|J L MINISTERIUM FÜR EIN LEBENSWERTES ÖSTERREICH

HBLFA FRANCISCO JOSEPHINUM WIESELBURG

DATA INTEGRATION AND ANALYSIS IN PRECISION DAIRY FARMING: A SEMANTIC DATA WAREHOUSING APPROACH

Martin Wischenbart[1], Simon Schausberger[2],

Christoph G. Schuetz[2], S. Dana Tomic[3]

1. Josephinum Research, 2. Johannes Kepler University Linz, 3. Smartbow GmbH

Dipl.-Ing. Martin WISCHENBART, M.Sc.

Josephinum Research, Wieselburg, Austria

Linked Open Data in Agriculture – Applications in Livestock Farming

28th of September, 2017, H4 Hotel Berlin Alexanderplatz, Berlin, Germany

*josephinum research*

J|R wieselburg

Situation:

- Precision dairy farming **sensors** deliver **large amounts of diverse data**
- Many **heterogeneities** between data **sources**, **farms** and **tools**
- Need for **interoperable interfaces**
- **Data integration** tends to require enormous **efforts**
- **Analyses** often **limited** to **regional** reports in **non-standardized** formats

Goals:

- Create a **reference data structure** usable across different systems and farms
- **Facilitate integration** of available dairy farm data
- Enable **systematic analyses within and across** farms
- Throughout all steps: consider **individual access policies and privacy needs**

PROJECT AGRIPROKNOW: ARCHITECTURE
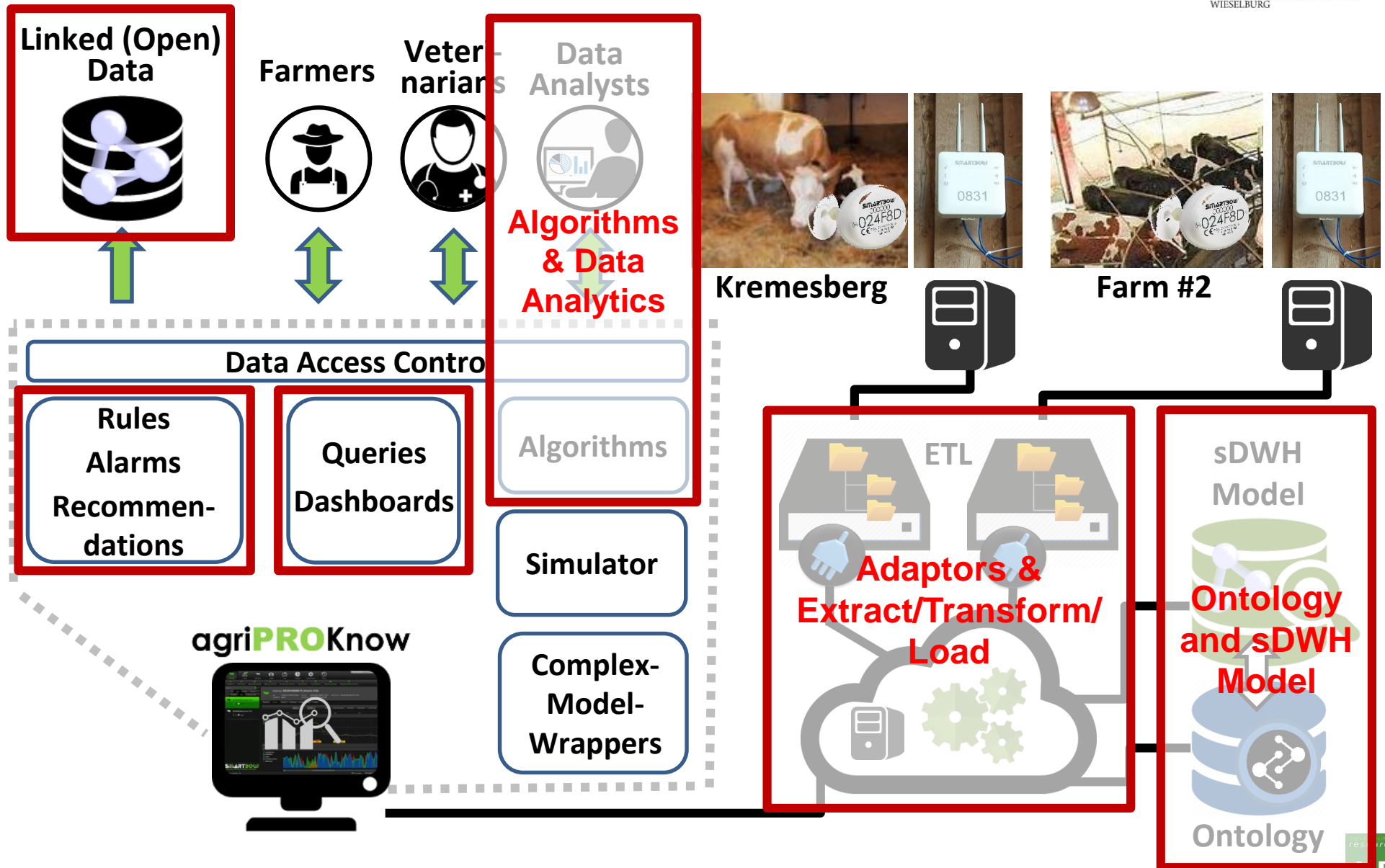
- **1. Ontology and sDWH Model**
- **2. Adaptors & Extract/Transform/Load**
- **3. Algorithms & Data Analytics**
- **4. Queries & Dashboards**
- **5. Rules, Alarms & Recommendations**
- **6. Exporting Linked (Open) Data**
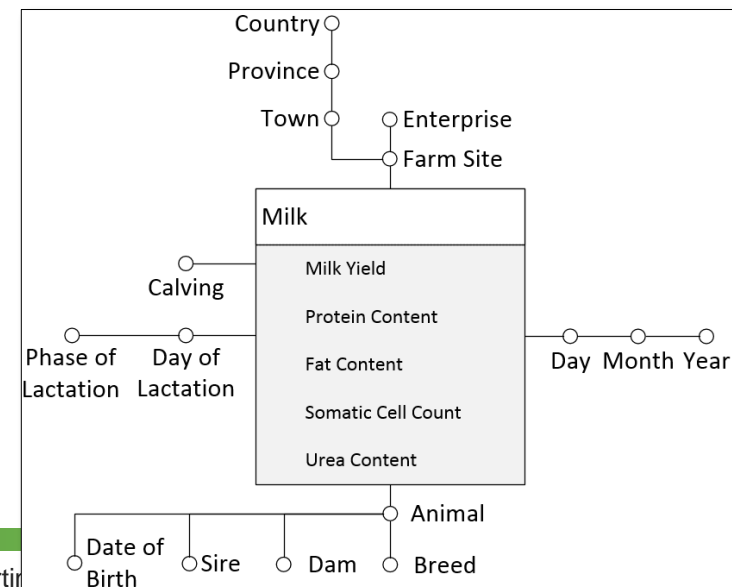- **Conclusions & Outlook**

1. ONTOLOGY AND sDWH MODEL

Ontology:

- Serves as a **schema** and **defines generic terms** for precision dairy farming
- Includes **references** to other **ontologies**, **standards** and **literature**

Semantic Data Warehouse Model:

- Exensible **reference schema** suitable for **various kinds of dairy farms**
- Defined in **RDF** using **QB and QB4O vocabularies** for description of **multidimensional OLAP data cubes**
- **Redundancy** to **simplify queries** and for **faster responses**

Manually implemented adaptors as individual **plug-ins** for each type of source → **deployable** on **multiple farms** with equal equipment

Data **extraction**, **transformation** and **loading** (ETL):

- **Extraction** from **different source** systems & farms

  resolving **technical** and **syntactic heterogeneities**: e.g., access to database vs. parsing CSV retrieved via an API; deal with formats of animals' national IDs

- **Transformation** to **RDF format** and into **sDWH schema**

  resolving **structural** and **semantic heterogeneities**: e.g., plain text/CSV vs. relational vs. name-value-pairs & lists (JSON); different abbreviations and languages (DE/EN/SK/…); overlapping or related term definitions ("D.M. Weight / Cow" vs. "Futtermenge" in fresh weight for a group)

- **Loading into** the **sDWH**

  transfer and **processing** including **aggregation** to a unified level: e.g., plausibility checks; handling of missing inputs

Integrated data as **basis for statistic** and **stochasic analytics**

Publish aggregrated and **anonymous data,** e.g.:

- Average **daily milk yield** in liters, **rumination/lying durations** in minutes, for each breed across all farms in the sDWH

MainBreed	SecondBreed	Milkyield	RuminationDuration	LyingDuration
Holstein_Friesian	German_Red_Pied	22,1	516,1	928,8
Holstein_Friesian	NONE	25,0	491,3	895,2

Algorithms to **analyze relations** and **correlations** between different **parameters** and **production factors**

- Data-based **detection of diseases**, e.g., to improve **methods** for early prognosis and **diagnosis of ketosis** through *pattern recognition, ANOVA, regression and prinicial component analysis* to discover **influencing factors** for **beta-hydroxy-butyrate (BHB)** blood measurements

4. QUERIES & DASHBOARDS (1)

Interface for advanced **farmers, veterinarians** and **consultants**

Interactive construction and **execution** of **queries** on the basis of various **pre-defined query patterns**; further **structuring** by:

- Definition of **common business terms** ("predicates")
  e.g., IncreasedSCC := SomaticCellCount > 200000

- Calculation of **additional parameters** ("calculated measures")
  e.g., FatProteinRatio := FatContent / ProteinContent

Dashboards to visualize query results

e.g.,   showing **developement of body condition score (BCS) and BHB**

distribution of all BHB measurements within 30 days after calving

delta BCS from 30 days before to calving

delta BCS until 30 days after calving

5. Rules, Alarms & Recommendations
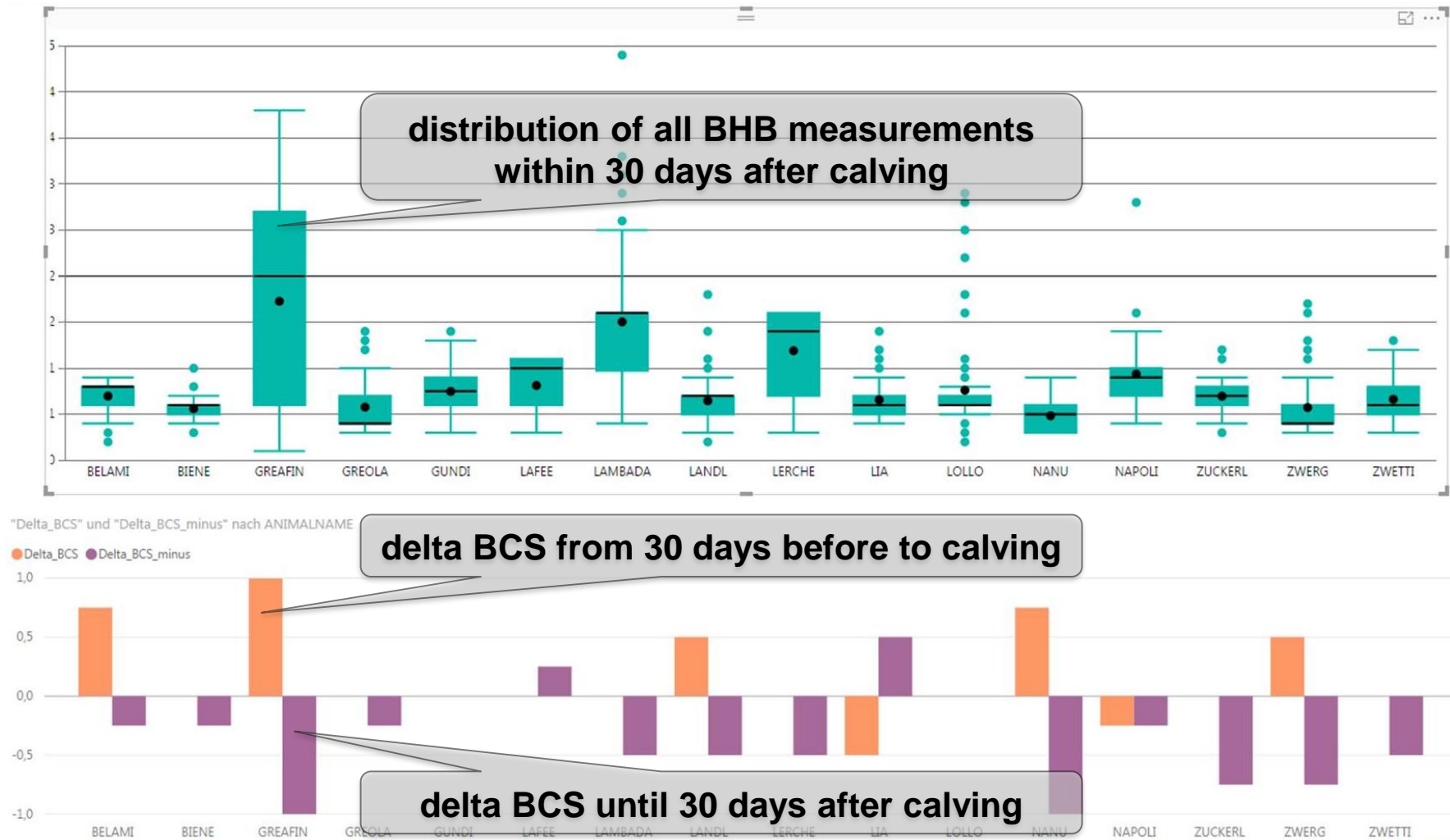
Specification of **rules**:

- Definition by domain **experts**
- Basing on **terms** from the **sDWH model**

Execution of **rules**:

- Creation of **alarms/reports** to **farmers** or **veterinarians**
- **Recommend actions** (or **trigger** them directly) on **machines**

Examples:

- Alarm **'laming cows'**: analyse animals' daily **walking** distance, time, **variation** of these parameters → **alarm if** a pre-defined no. of animals exhibits a certain **decrease** in walking activity (maybe suggest pedicure)
- Report **'insemnation due'**: report about cows' history after calving: milk yield and parameters, feed intake, ketosis diagnoses, lameness, … → help farmer decide about insemnation
- Action **'ketosis prevention'**: **prophylactic** feeding of **propylene glycol** to animals having an increased **risk for ketosis**: **automatically** or in **report**

1. Export linked data

– **Query results**, **alarms**, **reports** or recommended **actions** in **RDF** format

– Facilitate data **exchange** and **integration** with **devices** within the farm or with authorized personnel


```
rdfs:label   [language: en]
RiskOfKetosis

rdfs:comment   [language: en]
A FPR (fat-protein-ratio in milk) above 1.5
within the first 100 days of lactation is a
strong indicator for ketosis. Therefore, there
is a risk of ketosis.

expression
FatProteinRatio > 1.5 && DaysOfLactation <=
100

uses
DaysOfLactation

uses
FatProteinRatio
```

```
###  http://agriproknow.com/instances#
SK0008███████████
inst:SK0008███████████
    rdf:type owl:NamedIndividual ,
             apk:Cow ;
    apk:hasPredicate
    inst:riskOfKetosis_BN4aRVqzFf .

###  http://agriproknow.com/instances#
riskOfKetosis_BN4aRVqzFf
inst:riskOfKetosis_BN4aRVqzFf
    rdf:type owl:NamedIndividual ,
             apk:RiskOfKetosis ;
    apk:hasDate "2016-08-26 00:00:00.0" ;
    apk:hasDaysOfLactation "39" ;
    apk:hasFatProteinRatio "1.6" .
```

1. **Export linked data**

2. **Export** and **publish linked open data**
   - **Aggregated data** and **statistics**
   - No disclosure of **private operational farm data**

```
###   http://agriproknow.com/instances#breed0
inst:breed0 rdf:type owl:NamedIndividual ,
                     query:Result ;
            apk:MainBreed inst:Holstein_Friesian ;
            apk:SecondBreed inst:German_Red_Pied ;
            apk:LyingDuration 928.8 ;
            apk:Milkyield 22.1 ;
            apk:RuminationDuration 516.1 .

###   http://agriproknow.com/instances#breed1
inst:breed1 rdf:type owl:NamedIndividual ,
                     query:Result ;
            apk:MainBreed inst:Holstein_Friesian ;
            apk:LyingDuration 895.2 ;
            apk:Milkyield 25 ;
            apk:RuminationDuration 491.3 .
```

CONCLUSIONS & OUTLOOK (1)

- Built a **semantic data warehouse** (sDWH) for **cross-farm integration** of precision dairy farming data

- **Adaptors** to **resolve heterogeneities** in **schemas** and **data** from different sources

- Incorporated **data** on **1700 dairy cows** from **two farms**

- The sDWH **facilitates analyses, queries** and **rules** on integrated data, aiming to allow **new kinds of applications** and **interfaces** for different **stakeholders**

- Use of semantic technologies facilitates **exports** and **data exchange** in **RDF format** and **linked data**

- Motivate **vendors** to **create open interfaces** for providing data by **lowering obstacles** of data **integration**

- Advance **provision** of **cross-farm comparable data** and **decision support** (operational **farm data** for **private** use; **aggregated** data and **statistics** for **publication**)

- Project **agriProKnow** is still **ongoing**

- Further **development** and **evaluation** with different **stakeholders** planned

- A final **public agriProKnow workshop** will take place on **November 27th, 2017** (including **excursion** to farm)
  - contact martin.wischenbart@josephninum.at for an invitation

THANK YOU!

QUESTIONS ?

Dipl.-Ing. Martin Wischenbart, M.Sc.
Data and Information Sciences

Josephinum Research, Wieselburg, Austria
E-Mail: martin.wischenbart@josephinum.at
Phone: +43 7416 52175 630
Web: http://research.josephinum.at