A GIS-based Data Analysis Platform for Analyzing the Time-Varying Quality of Animal Feed and its Impact on the Environment

A. Taliun

M. Bohlen

A. Bracher

F. Cafagna

International Environmental Modelling and Software Society (iEMSs)
2012 International Congress on Environmental Modelling and Software
Managing Resources of a Limited Planet, Sixth Biennial Meeting, Leipzig, Germany
R. Seppelt, A.A. Voinov, S. Lange, D. Bankamp (Eds.)
http://www.iemss.org/society/index.php/iemss-2012-proceedings

A GIS-based Data Analysis Platform for Analyzing the Time-Varying Quality of Animal Feed and its Impact on the Environment

A. Taliun [a], **M. Böhlen**[a], **A. Bracher**[b] **and F. Cafagna**[a]

[a]*University of Zürich, Department of Informatics, Binzmühlestrasse 14, CH-8050 Zürich (taliun@ifi.uzh.ch, boehlen@ifi.uzh.ch, cafagna@ifi.uzh.ch)*

[b]*Agroscope Liebefeld-Posieux ALP, Rte de la Tioleyre 4, Case postale 64, CH-1725 Posieux (annelies.bracher@alp.admin.ch)*

Abstract: The Swiss Feed Data Warehouse is a public service for companies, farmers and research institutions that provides detailed and up-to-date information about the concentration of nutrients in animal feed from all across Switzerland. The core of the Swiss Feed Data Warehouse is a carefully curated data warehouse with more than 2 million chemical analyses of 600 feed types and 400 nutrients. The nutrient measurements are enriched with geographical (as postal code and altitude), temporal (as harvesting and analyses dates), biological (as variety and botanical composition) and technical information (as conservation and production methods). We propose a solution that offers to users a fast, effective and intuitive approach to query and analyze large amounts of high dimensional feed data. An interactive web-application enables dynamic query construction and offers multiple charts to visualize feed data. Techniques such as radius search and charts allow non-expert users to detect local correlations and trends in the feed data. Historical information makes it possible for advanced users to analyze changes in the feed quality and determine the balance of nutrients in feed mixes that minimizes the load of severe nutrients from animal excreta into ecosystems.

Keywords: GIS, Data Warehouse, Data Analysis

1 INTRODUCTION

An extensive knowledge of the properties of feedstuffs provides the basis for the understanding of the interdependencies between animal husbandry, animal nutrition, ecology, nutrient cycles, and the influence of feeding on the quality of food of animal origin. Feed science is closely linked to plant production since crops largely contribute to the supply of feed for farm animals. For climatic and topographic reasons, grassland products and herbage based feeding systems are of particular importance in Switzerland. Due to very diverse climatic, topographic and pedologic conditions that vary within short distances, the local and regional influence on feed quality can be quite substantial. The Swiss Feed Data Warehouse employs state-of-the-art open source GIS techniques to provide information that allows to fine tune animal feeding down to the farm level. A precise diet formulation that matches the nutrient requirements of farm animals as close as possible prevents nutrient oversupply. This results in an efficient use of limited resources and low emissions, which is particularly relevant for nitrogen (N) and phosphorous (P) cycling (Bracher et al. [2011]).

In this paper we propose a rich dimensional data warehouse model for the enhanced analyses of sparse and multi-granular feed data. We describe a GIS-based web application that provides different views on the data to satisfy the needs of the three main user groups. The effectiveness of our solution is illustrated by three example analyses of feed data collected in Switzerland.

Feed databases of other European countries are limited with respect to geographic information, data views and query reports. Charts containing historical evolution (Germany: `http://datenbank.futtermittel.net`) or histograms displaying frequency distribution (France: `http://www.feedbase.com`) of feed nutrients are available. The list of individual feed samples can be viewed but the corresponding grouping filters for the origin of the samples are restricted to country, province and district as the lowest resolution. The linking of the sample origin with a map and visualization of the spatial data is a new approach that does not exist so far.

2 THE DATA MODEL

Feed quality is determined by the properties of feed samples: first, feed samples are collected from various locations and, next, they are prepared, analyzed and cleaned by the laboratories at agricultural research stations. During the analyses the containment of nutrients in feed samples is measured and the achieved values are stored in a single table with hundreds of columns together with other detailed information (cf. Figure 1).

sample_number	**city**	**harvest**	**feed**	...	**RA**	**RF**	**GLU**	**MG**	...
302026-9	Pensier	2006	Milocorn, Körner $<$ 4 % RF (Sorghum, Hirse)	...	17.2	19.5	0.09		...
259532-4	Oltingen	2005-08	Sorghum, Körner 4 - 6 % RF (Milocorn, Hirse)	...		100.7			...
264521-2	Oltingen	2005-08	Tapiokamehl (Maniok, Cassava)	...	65.0	48.0		1.1	...
...	...	...	...	...	...	...	...	...	...

Figure 1. Feed Data in an Excel File Provided by the Laboratories

In order to provide added value to expert and non-expert users it is necessary to model this data such that ad-hoc data analyses are supported. We propose a dimensional model for the Swiss Feed Data Warehouse (cf. Figures 2 and 3), which supports enhanced analyses of the feed data. The dimensions categorize the available properties in a consistent way to support different types of analyses. For example, all properties that are required for temporal analyses are contained in dimension *time* and all biological properties of the feed sample are in dimension *sample*.

Another specific part of our design is a vertically partitioned fact table (Abadi et al. [2009]), i.e., *raw_measurements* that store individual nutrient measurements rather than individual samples. Such a design approach addresses two critical properties of feed data. First, there are more than 600 nutrients and only small part of them (5-20) are measured per sample. Therefore, a fact table that stores rows of nutrient measurements would be extremely sparse. Second, it is often the case that nutrients are different because of the analyses methodology (for example, two methods x and y to measure fat content RL result in two nutrients RL-x and RL-y). Because of this the list of nutrients changes over time as new methodologies become available. Our design allows to add new nutrients without changes in the data model or a redesign of the queries. Figure 3 illustrates a small extract of the data warehouse, which is further explained below.

Feed Types and Feed Properties. Dimension *feed* contains about 1000 feed types most of which are grouped into feed categories. Feed categories form a hierarchy. For example in Figure 3, feeds with ids 1 and 2 belong to feed category *Getreidekörner*

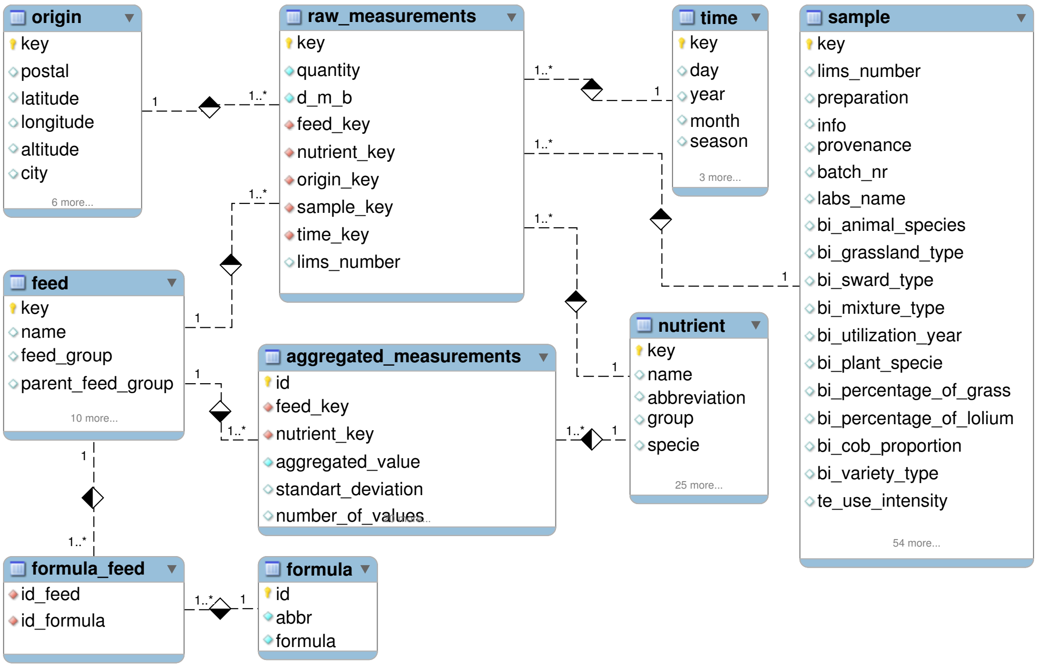


Figure 2. The Swiss Feed Data Warehouse Design

raw measurements

key	quantity	d_m_b	feed_key	nutrient_key	origin_key	time_key	sample_key
1	17.2	true	1	1	1	1	1
1	19.5	true	1	2	1	1	1
1	0.09	true	1	4	1	1	1
1	100.07	true	2	2	2	2	2
1	65.0	true	3	1	2	2	3
1	48.0	true	3	2	2	2	3
1	1.1	true	3	5	2	2	3

feed

key	name	feed_group	parent_feed_group	...
1	Milocorn, Körner < 4 % RF (Sorghum, Hirse)	Getreidekörner	Einzelfutter	...
2	Sorghum, Körner 4 - 6 % RF (Milocorn, Hirse)	Getreidekörner	Einzelfutter	...
3	Tapiokamehl (Maniok, Cassava)	Knollen und Wurzeln	Einzelfutter	...

nutrient

key	name	abbreviation	group	...
1	Rohasche	RA	Rohnährstoffe	...
2	Rohfaser	RF	Rohnährstoffe	...
4	Glutaminsäure	GLU	Aminosäuren	...
5	Magnesium	MG	Mineralstoffe	...

sample

key	sample_number	...
1	302026-9	...
2	259532-4	...
3	264521-2	...

origin

key	postal	latitude	longitude	altitude	...
1	1783	46.856928	7.152509	600-799	...
2	4494	47.434896	7.936865	600-799	...

time

key	day	year	month	...
1		2006		...
2		2005	8	...

formula

id	abbr	formula
1	OS	1000 - RA
2	NfE	#OS - RP - RL - RF
3	UEG	(12.98 * RP + 32.23 * RL + 15.76 * #NfE / 1000
4	UEG	(16380 - (16.38 * RA) - (34.64 * RF)) / 1000

formula_feed

id_feed	id_formula
1	3
2	3
3	4

Figure 3. An Extract of the Feed Data Warehouse

which, in turn, is a subcategory of *Einzelfutter*. Dimension *sample* provides information about the categorical properties of individual feed samples. Even for the same feed type or category, feed samples may differ in up to 30 unique properties. The general class contains description of the sample, project code, sample preparation method, etc. The second class are biological properties that include grassland characteristics, sward type, variety and etc. The third class are technical properties that describe the production system used to produce the feed such as grassland use intensity, conservation and storage methods. All this data is highly demanded by the agriculture research to observe old and to determine new factors that influence the feed quality.

Geographical and Temporal Properties. Temporal information is required to detect trends and seasonal effects in the feed quality. The most relevant geographical factors that impact feed quality are regional climate, altitude, soil type, production intensity and animal density. Processing and storage of temporal and geographical information must deal with different levels of detail at which data is provided. For example, harvesting and analyses dates are difficult to track down and are often coarse or approximate, whereas sampling and arrival dates are usually precise. For each feed sample we store complete geographical information including city names, altitude, latitude and longitude (cf. dimension *origin* in Figure 2). However, it is often the case that only the postal code or community name are given. We retrieve the missing information by investigating the provided description for the trial reports, names or abbreviations, and, by querying Google Web Services for geographic coordinates.

Nutrients and Nutritive Values. More than 600 nutrients and nutritive values (cf. table *nutrient* in Figure 2), characterize the feed quality in terms of proteins, minerals, vitamins and energy containment. The nutrients are grouped according to the chemical category (as amino acids or minerals) and nutritive values are specific to the animal species (as pigs or poultry). This categorization is necessary for joint comparison of feeds on specific properties and for optimizing the feeding of each animal specie individually.

We provide a detailed and aggregated view of nutrient content (cf. *raw_measurement* and *aggregated_measurements* in Figure 2). The detailed view allows to investigate nutrients of individual feed samples. In the aggregated view for each feed type and nutrient we provide an aggregated value that accurately quantifies representative nutrient content of a given feed type at the current date. The aggregated values are necessary for determining the best diet for the animals based on the feeds that are currently present on the market. Since the feed data is collected irregularly, a simple averaging of measurements with a fixed time interval results in biased values. Therefore, domain experts manually choose measurements from the recent history.

Nutritive values, such as digestibility and energy value, must be computed from the nutrients measurements based on predefined formulas. As an example, consider tables *formula* and *formula_feed* in Figure 3. Rows 3 and 4 of table *formula* contain the formula definitions for UEG (apparent metabolisable energy N-corrected of poultry) for three different feeds. Formula with id 4 is related to the feed *Tapiokamehl* while formula with id 3 is related to *Sorghum* and *Milocorn*. Note that the formula with id 3 is recursive, i.e, the computation of UEG for Sorghum requires to evaluate NfE which depends on nutritive value OS. In case, a feed sample contains measurements of all required nutrients, the computation of nutritive values in SQL is straightforward. Since the feed data is sparse and not all required nutrient measurements are available for a feed sample nearest neighbor joins must be used (Yao et al. [2010]). For each component of the formula the temporal closest feed sample with analogues characteristics that contains the missing nutrients must be retrieved.

3 The Application

The crucial functionality for the investigation of the feed quality are views tailored for specific user groups. Our web application supports three major user groups. First, individual farmers and the feed industry use the feed data to ensure the optimal balance of feed types in compound feeds. Thus, the main target of this user group are up-to-date aggregated values on the nutrient contents, which must be displayed in a sortable table for better comparison of individual values. Second, agricultural research is interested in detailed information that includes raw measurements, time, geographical and various biological properties of the feed samples. The user interface must provide the ability to refine the search and provide an enhanced graphical visualization with statistical information, colored charts and maps. Third, education and government organizations use the feed database mainly for data explorations and the most desirable property is an intuitive and simple navigation with additional information about feeds (feed catalog, statistics).

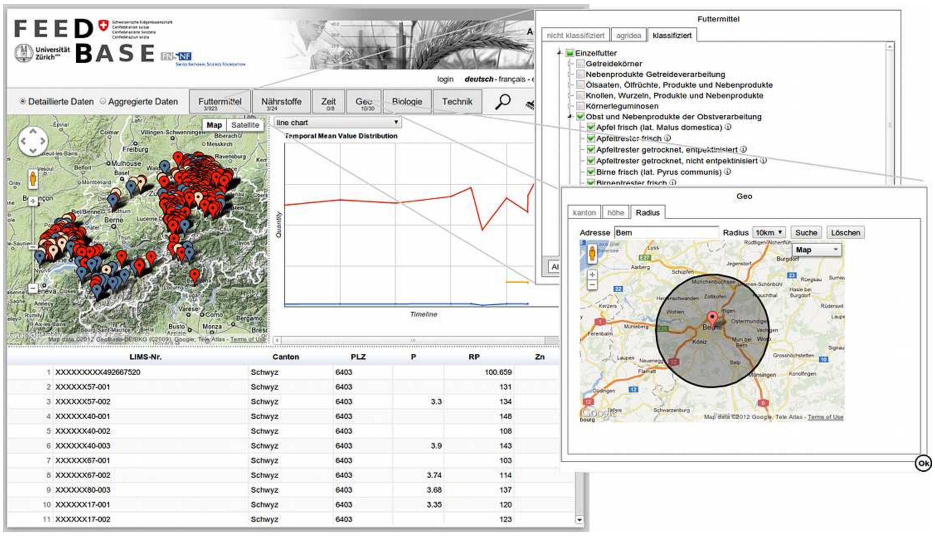


Figure 4. Web Application of the Swiss Feed Data Warehouse

Figure 4 shows the main window of the application. The design is optimized for the control of the search parameters and the visual investigation of the results. Search parameters are organized into a row of five buttons where each button corresponds to a search category. A click on a button expands an overlay window where the relevant search options are represented either by a list, a tree of check-boxes or a map. The search options are loaded dynamically based on the already chosen options. For example, once a feed is selected only the nutrients that are contained in the analyses of the selected feed are displayed. The query results are displayed in the three windows that occupy the main part of the screen. Each window provides a specific view: plotting the data on the map, charts with statistical information, and raw measurements in a sortable table.

Figure 5 illustrates the system architecture. On the server side we use a PostgreSQL database and Apache Web Server with PhP support. On the client side a web browser must support DHTML and JavaScript. All logic is pushed to the client, while the server is responsible for translating search options into SQL and passing back the query results with XML or JSON. Ajax enables asynchronous requests. jQuery library is used to manipulate DOM objects: check boxes, trees, buttons and etc. jQueryTools provides a cross-browser approach to create overlay windows, tabs and tooltips. Google Maps API [2011] is used to retrieve an excerpt of the map and display it together with feed samples. Google jSAPI provides scripts to display large amounts of values.

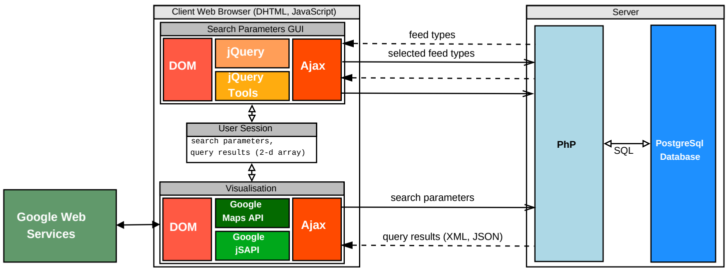


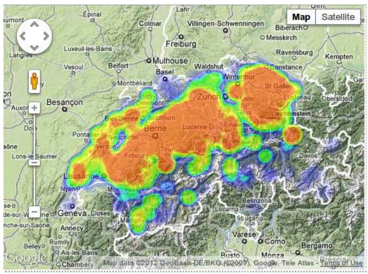
Figure 5. System Architecture of the Swiss Feed Data Warehouse

4 Analyses and Graphical Reports

The first on line version of the Swiss Feed Database contained only aggregates of the most recent nutrient measurements. The aggregated data is visualized in a table (cf. Figure 6(a)) without further possibilities for data analyses. Starting from 2010 the data warehouse was enhanced with data about individual samples including geographical and temporal information. The new application supports spatial queries and offers various charts for analyses. The current development aims to enrich visualization of the spatial data with statistical information. For example, Figure 6(b) uses a heatmap to highlight the regions where the occurrence of the feed samples is high.

	TS g/kg	OS g/kg TS	RA g/kg TS	RP g/kg TS	RL g/kg TS	RF g/kg TS	NfE g/kg TS	Stärke g/kg TS
Ackerbohne, Samen {Pferdebohne, Saubohne} (lat. Vicia faba)	870	958.9	41.1	296.0	17.8	94.8	550.3	420.7
Kristallines Fett für Milchvieh {Pulverfett, pansenstabiles Fett, Alikon}	990	999.6	0.4	0.5	981.5	0.0	17.60	0.0
Ameisensäure	1000	1000.0	0.0	0.0	0.0	0.0	1000.00	0.0
Apfel frisch (lat. Malus domestica)	165	978.0	22.0	23.0	19.0	104.0	832.00	NA
Apfeltrester frisch	253	983.4	16.6	51.0	19.4	191.1	721.90	0.0
Apfeltrester getrocknet, entpektinisiert	900	935.5	64.5	81.3	37.1	317.5	499.60	0.0

(a) Aggregated Data

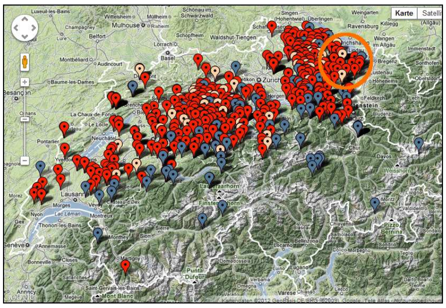


(b) Statistical Information

Figure 6. Development of the Swiss Feed Database

Regional Influence on Hay Quality. We provide a graphical interface to compare the content of nutrients across regions. Protein (RP) and phosphorous (P) are of particular interest due to their environmental impact. Protein contains nitrogen which, if fed in surplus, is directly linked to high ammonia emissions and nitrate leakage that pollutes surface and ground water. Phosphorous is a non-renewable resource that contributes to the eutrophication of lakes. Having detailed information about the regional, or even better local, nutrient content helps to choose the correct feed components to adjust the diet to the animals requirements and thus prevent oversupply. This is of particular importance in regions of high animal density, which largely contribute to air and water pollution.

Figure 7 illustrates regional patterns in the quality of hay, which is a staple feed component in dairy cow feeding. The data was collected between 2005 and 2009 (Boessinger and Python [2012]). The coloring helps to detect regional patterns. In Figure 7(a), the blue and yellow colors denote hay samples in which the P content is below and above the average by one standard deviation respecitively. As can be seen, blue flags prevail in mountain regions. Figure 7(b) summarizes the difference in the quality of hay of different altitude classes. Lowland hay has a higher protein (RP), phosphorous (P) and sugar content than mountain hay. This divergence is influenced by the botanical composition of the



(a) Hay Samples with Above and Below Average Phosphorous Content (all Altitude Classes)

	$>$ 1000m	$<$ 800m	Uzwil
P g/kg DM	2.75	3.53	3.57
RP g/kg DM	119.12	129.16	132.72
Mn g/kg DM	120.21	68.32	74.7
NEL MJ/kg DM	5.41	5.48	5.64
Sugar g/kg DM	100.19	124.96	135.49

(b) Nutrient Values

Figure 7. Regional Influence on the Nutrient Content in Hay

sward reflecting the difference between ley farming and permanent grassland, climate, soil type and the production intensity. The most striking difference concerns manganese (Mn) . Based on the P requirements of milking cows (between 3 and 4 g of P/kg of dry matter intake), mountain hay requires in most cases a P supplementation in a hay based winter diet whereas with lowland hay, P requirements are easily met if not surpassed.

The user can zoom into a region of intensive feed production such as Uzwil (cf. circle in Figure 7(a)). Because of a higher cut number per year and, thus, of a younger phenological stage, the hay is comparatively rich in energy, protein, P and sugar. The same region coincides with high ammonia emissions. The high emission level arises from high animal density, inadequate diet formulation and concentrate feeding. Particularly during summer feeding with herbage containing more than 180 g protein/kg of DM, protein intake is exceeding requirements of cows producing less than 25 kg of milk.

Temporal and Regional Influence on Hay Quality. 2007 was characterized by a cool and rainy summer for western Switzerland. The protein content in hay samples drops on average to 113 g/kg of DM in 2007 compared to above 130 g/kg of DM of the neighboring years (cf. Figure 8). No such jump was observed in eastern Switzerland. This underlines the relevance of geographic and temporal information that prove to be important factors with respect to varying feed quality. The monitoring of the hay quality by the yearly hay survey helps farmers to plan and adjust feeding on a yearly and regional basis.

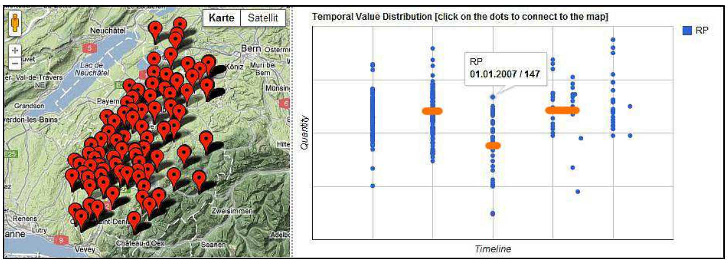
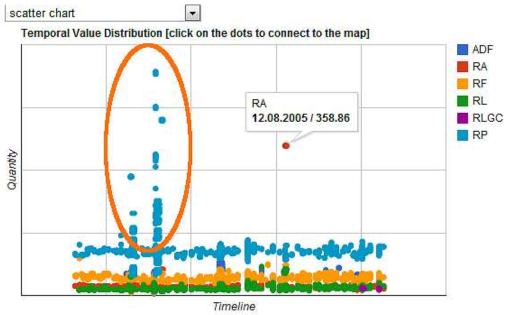


Figure 8. Year Influence on the Protein Content in Hay from the Kanton Fribourg.

Outlier Detection in Raw Data. Graphical and statistical reports are also used to detect outliers and clean raw data. Figure 9 illustrates examples of outlying barley samples. Based on expected values for barley, some samples have a clearly incorrect ash, fat and protein content. By checking the detailed sample description it becomes evident that a wrong feed code was used at the lab leading to misclassified feed samples. This



statistical information of nutrients

Nutrient	Minimum	Maximum	Average	2σ*	Count
ADF	4.000	91.800	55.914	15.068	310
RA	12.600	358.860	22.242	13.312	2716
RF	6.000	89.021	41.177	7.226	2724
RL	9.100	69.100	19.894	5.792	3254
RLGC	13.186	19.400	16.550	1.536	52
RP	48.200	535.500	10.213	35.227	2942

Figure 9. Detection of Misclassified Barley Samples.

underlines the importance of plausibility checks in cleaning raw data.

5 Conclusions and Future Work

Compared to the first on-line version of the Swiss Feed Database, the newly developed Data Warehouse offers enhanced search, reports and analyses functions. The most relevant factors that influence feed quality are an integral part of the underlying data model. The web application supports spatial queries on time-varying data sets by flagging individual samples on a map that can be further differentiated according to biological and technical properties. As part of the GIS-based data analyses visualizing below and above average values, regional patterns in feed quality can be detected. As shown with the hay survey, such a functionality helps to monitor and interpret existing regional influences, which provide the information needed for ecologic and economic feeding. The same features are applicable for other environmental issues linked to GIS-data.

Acknowledgments

This work was funded by the Swiss National Science Foundation (SNSF) through the Tameus project (proposal no 200021_135361) and by the Department of Computer Science, ETH Zurich.

References

Abadi, D. J., A. Marcus, S. R. Madden, and K. Hollenbach. Sw-store: a vertically partitioned dbms for semantic web data management. *The VLDB Journal*, pages 385–406, 2009.

Boessinger, M. and P. Python. Faktoren mit einfluss auf die nähr- und mineralstoffgehalte von belüftetem dürrfutter. *Agrarforschung Schweiz 3(1)*, pages 36–43, 2012.

Bracher, A., W. Münger, Stoll, P. Schlegel, and H. Menzi. The relationship between diet characteristics, milk urea, nitrogen excretion and ammonia emissions in dairy cows. *EAAP – 62nd Annual Meeting, Stavanger*, page 127, 2011.

Google Maps API. Google. `http://code.google.com/apis/maps/`, 2011.

Yao, B., F. Li, and P. Kumar. K nearest neighbor queries and knn-joins in large relational databases (almost) for free. In *ICDE*, pages 4–15, 2010.