University of Zurich[UZH]

Department of Informatics

University of Zürich
Department of Informatics
Binzmühlestr. 14
CH-8050 Zürich
Phone. +41 44 635 43 11
Fax +41 44 635 68 09
www.ifi.uzh.ch/dbtg

UZH, Dept. of Informatics, Binzmühlestr. 14, CH-8050 Zürich

Nathalie Torrent

Prof. Dr. Michael Böhlen
Professor
Phone +41 44 635 43 33
Fax +41 44 635 68 09
boehlen@ifi.uzh.ch

Zürich, January 17, 2019

Master's Basic Module
Datenbanktechnologie

Topic: Data Curation in the Swiss Feed Database

The Swiss Feed Database is a data warehouse with a star schema [**?**] that models the concentration of nutrients in feed samples collected across Switzerland. The Swiss Feed Database is used by Swiss farmers and research institutions, and it is actively being developed and maintained by UZH and Agroscope. The development version of the online interface can be accessed at `http://vm-164.s3it.uzh.ch/`.

The goal of this Master's basic module is to extend the Swiss Feed Database with an effective and efficient solution for data import and maintenance. Specifically a solution that can be used to add, update and delete measurements, formulas, feeds, nutrients and analyses methods together with associated mappings shall be designed and implemented. The requirement elicitation shall be done in close collaboration with Agroscope to end up with a deployable solution.

The work includes the following tasks, outcomes and deliverables:

- **T1: getting started**
  install system (PostgreSQL, web server, Feedbase, web application, science cloud). Execute the following tasks without breaking existing functionality to get familiar with the Feedbase: Rename summary_data to reference_data; eliminate tables not needed (e.g., table names ending on _to_be_dropped); eliminate attributes not needed; eliminate functions not needed; eliminate users not needed (e.g., steger).
  Outcome: documentation of installation steps, database schema

- **T2: requirement analysis**
  Analysis of the part of Agroscope's data import and maintenance process that is related to the Swiss Feed Database. The starting point for the requirement elicitation are the existing solution, the feedbase documentation, and input from domain experts. The analysis part shall be done in close collaboration, including physical meetings, with domain experts from Agroscope.
  Outcome: documentation of requirements.
- **T3: definition of input data and operations**
  Definition of Excel templates for the data import and curation. The data in the Excel files must be mapped uniquely to fact and dimension tables in the data warehouse. A precise mapping shall be defined that supports all the necessary data manipulations.
  Outcome: definition of Excel input format and semantics of operations.
- **T4: Implementation of data import and maintenance**
  Implementation of a data import solution based on SheetJS js-xlsx. Important criteria of the solution are robustness and ease of use.
  Outcome: import and maintenance application, documentation.

Supervisor: Michael Böhlen

Start date: 3.1.2019

End date: 3.4.2019

Exam: 22.2.2019 11:30-12:30 BIN 2.E.13

Department of Informatics, University of Zurich

Prof. Dr. Michael Böhlen