tds exposure



Why and How to Manage TDS Data

FoodCASE - A TDS Data Management System

David Weber (ETH Zurich)

TDS Workshop, Brussels, 25th March 2015

Outline

- Why to manage TDS data
- What data needs to be managed
- How data needs to be managed
- Concepts for data management
- Use cases / Implementation in FoodCASE
- Data quality in FoodCASE

David Weber

- MSc in Computer Science with focus on information systems and business informatics
- Several years of work experience as software engineer in the financial industry
- PhD student at ETH Zurich
- 3 years experience with food composition and TDS data management system
- Research interest on data quality

Why to manage TDS data?

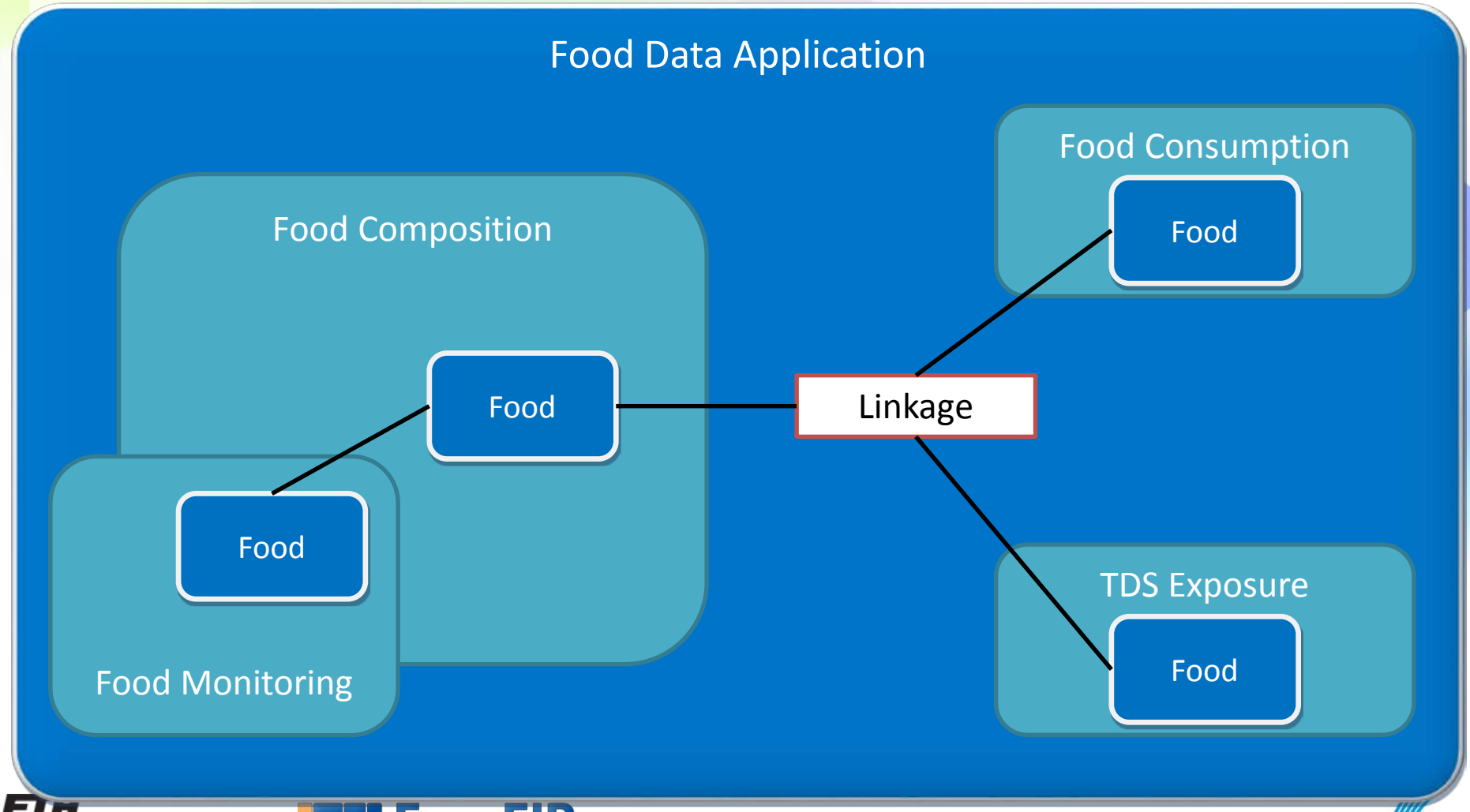
Why to Manage TDS Data

- Preserve data for long-term access
- Increase the visibility of your work
- Save time and resources in long run
- Enable other to access your data (in a unique way)
- Open data means that publicly funded data must be publicly available (EU directive)

Why to Manage TDS Data

- Enhance data security
- Minimise risk of data loss
- Meet grant requirements
- See data as a product for which funding were used
- Documentation of your data (evidence) ensures data integrity and replication
- Facilitate new discoveries

Linkage of Food Items



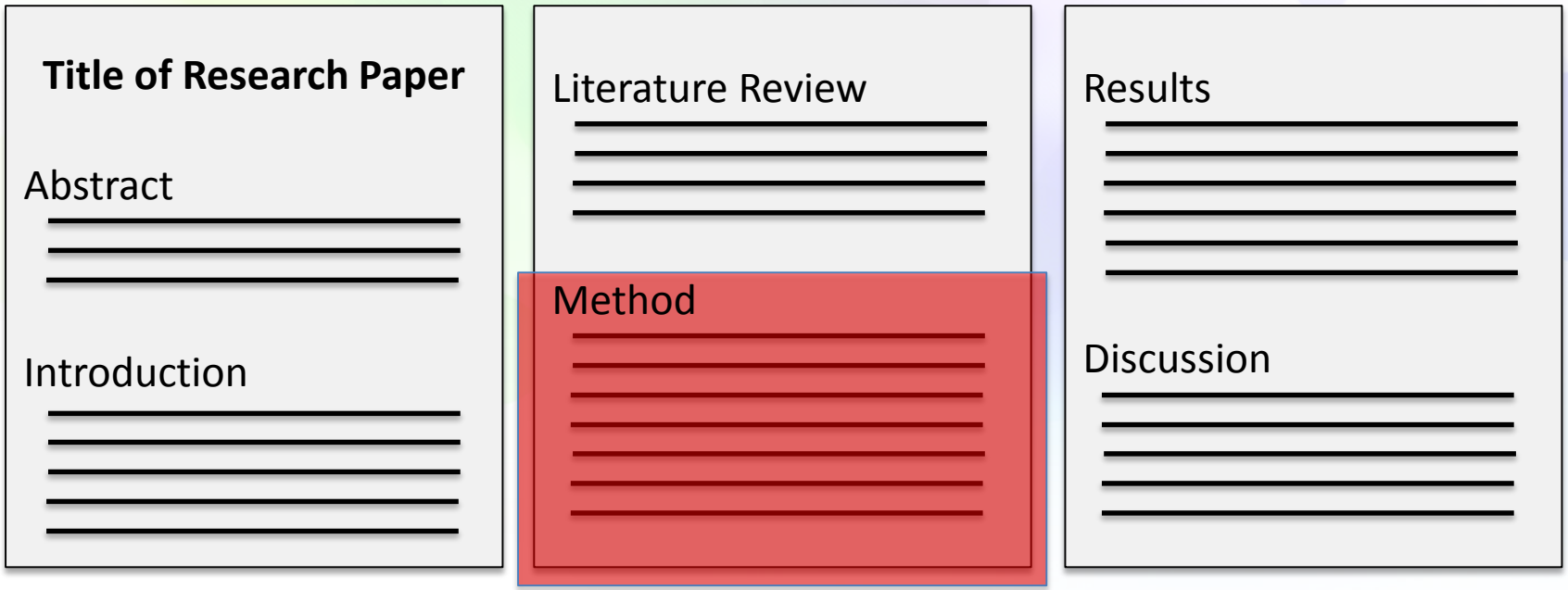
Take Home Message I

TDS data is the output of your work and your asset.
Take care of it.

What Data Needs to be Managed?

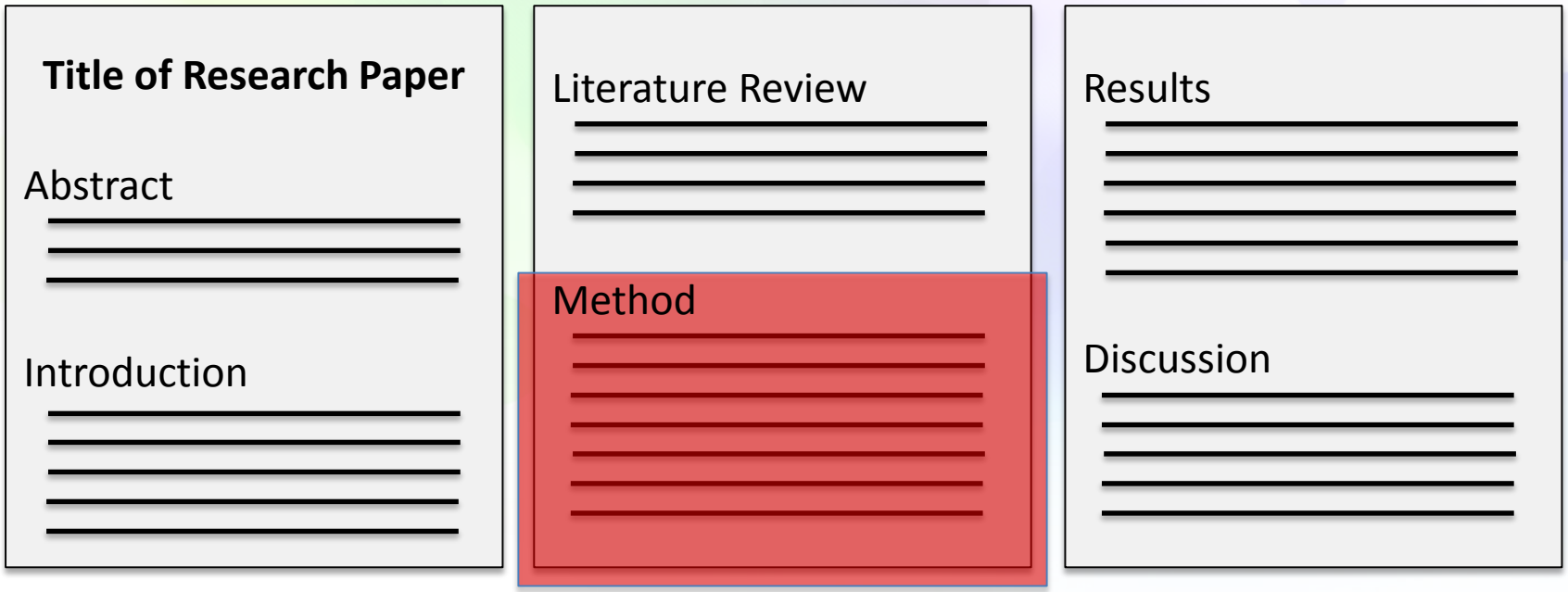
What Data Needs to be Managed?

Analogy with a research paper:



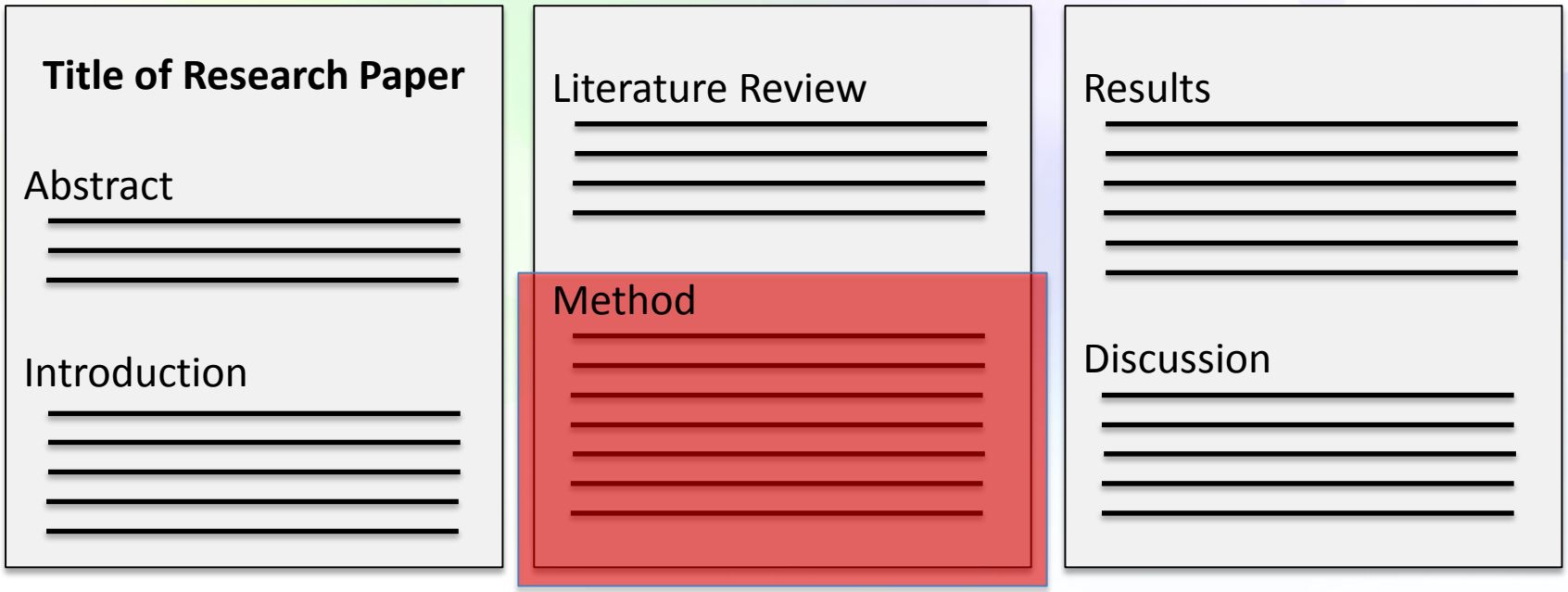
What Data Needs to be Managed?

Analogy with a research paper:



What Data Needs to be Managed?

Analogy with a research paper:



How does a TDS work?









Food 1
Food 2
Food 3
…

plan

Collect food according to consumption info

Food A
Food B
Food C
…

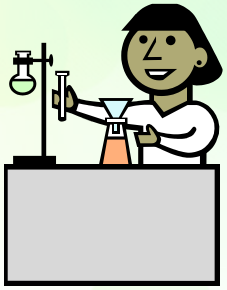
prepare as consumed

Food A
Food B
Food C
…

pool

Pooled food X

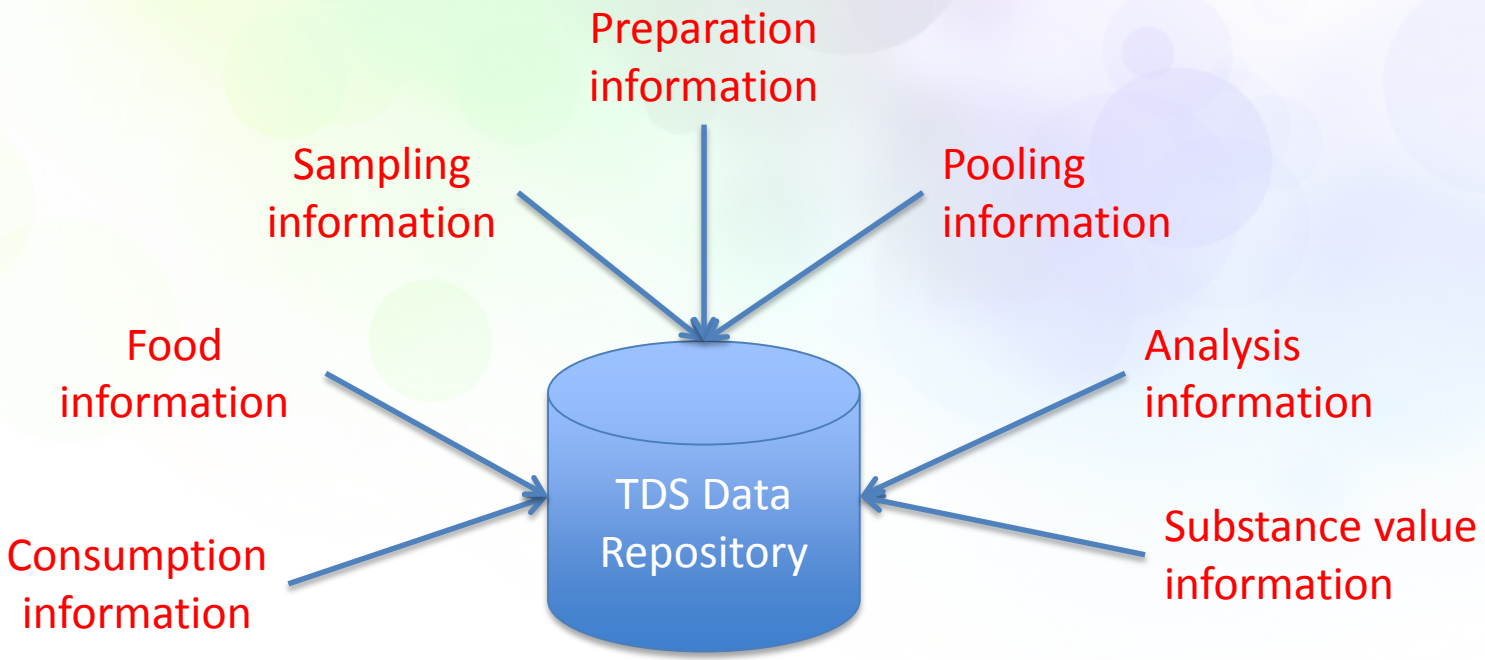
How does a TDS work?

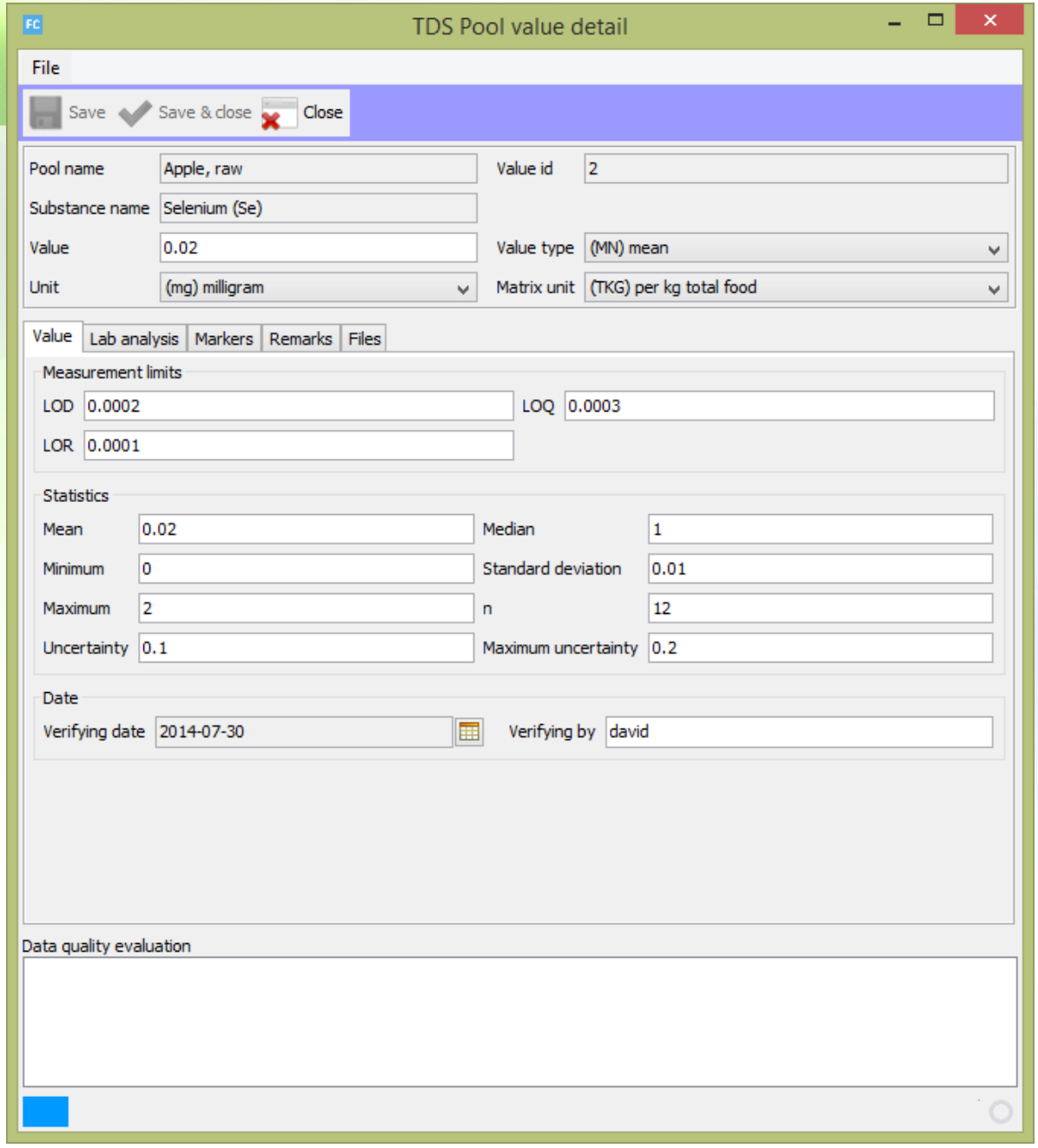




Pooled food X → analyse → Nutrient values, Contaminant values → Assess risk → Recommendation

Management of TDS Data




TDS Pool value detail
File
Save
Save & close
Close
Pool name
Apple, raw
Value id
2
Substance name
Selenium (Se)
Value
0.02
Value type
(MN) mean
Unit
(mg) milligram
Matrix unit
(TKG) per kg total food
Value
Lab analysis
Markers
Remarks
Files
Measurement limits
LOD
0.0002
LOQ
0.0003
LOR
0.0001
Statistics
Mean
0.02
Median
1
Minimum
0
Standard deviation
0.01
Maximum
2
n
12
Uncertainty
0.1
Maximum uncertainty
0.2
Date
Verifying date
2014-07-30
Verifying by
david
Data quality evaluation

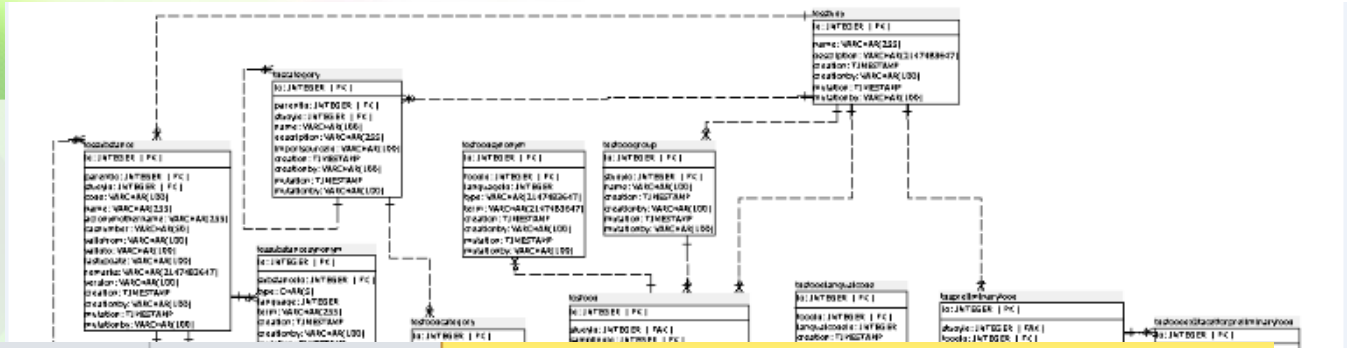
Take Home Message II

Value documentation is crucial for scientific evidence.

What Data Needs to be Managed?

- Until now every country did its own TDS and collected data according to their requirements.
- One goal in the TDS-Exposure project is to establish recommendations what data should be collected.

What Data



	Kategory	Name	Type	Comment
1				
2	ids	id	integer	
3		study id	integer	
4		sampling id	integer	
5		food group id	integer	
6	names	name	string	
7		english name	string	
8		scientific name	string	
9		other names	list of other names	
10	classification	FoodEx2	FoodEx2 code	
11		FoodEx2 facets	list of facets	--> tdsfoodex2facetforfood
12		LanguaL	list of LanguaL codes	--> tdsfoodlangualcode
13	consumption info	typicalconsumtion	numeric	typical consumed amount in g
14	Culinary factor	culinaryfactoredible	numeric	average culinary factor from raw and edible to consumed
15		culinaryfactorpurchased	numeric	average culinary factor from raw and purchased to consumed
16	varia	remarks	string	
17		images	list of images	--> tdsfilelink
18		marker	marker	



What Data Amount Needs to be Managed?

The following numbers were identified:

- 200 foods
- 96 attributes per food and 49 for one compound value

Example:

200 food items with about 10 compounds -> at least 117’200 data points

Take Home Message III

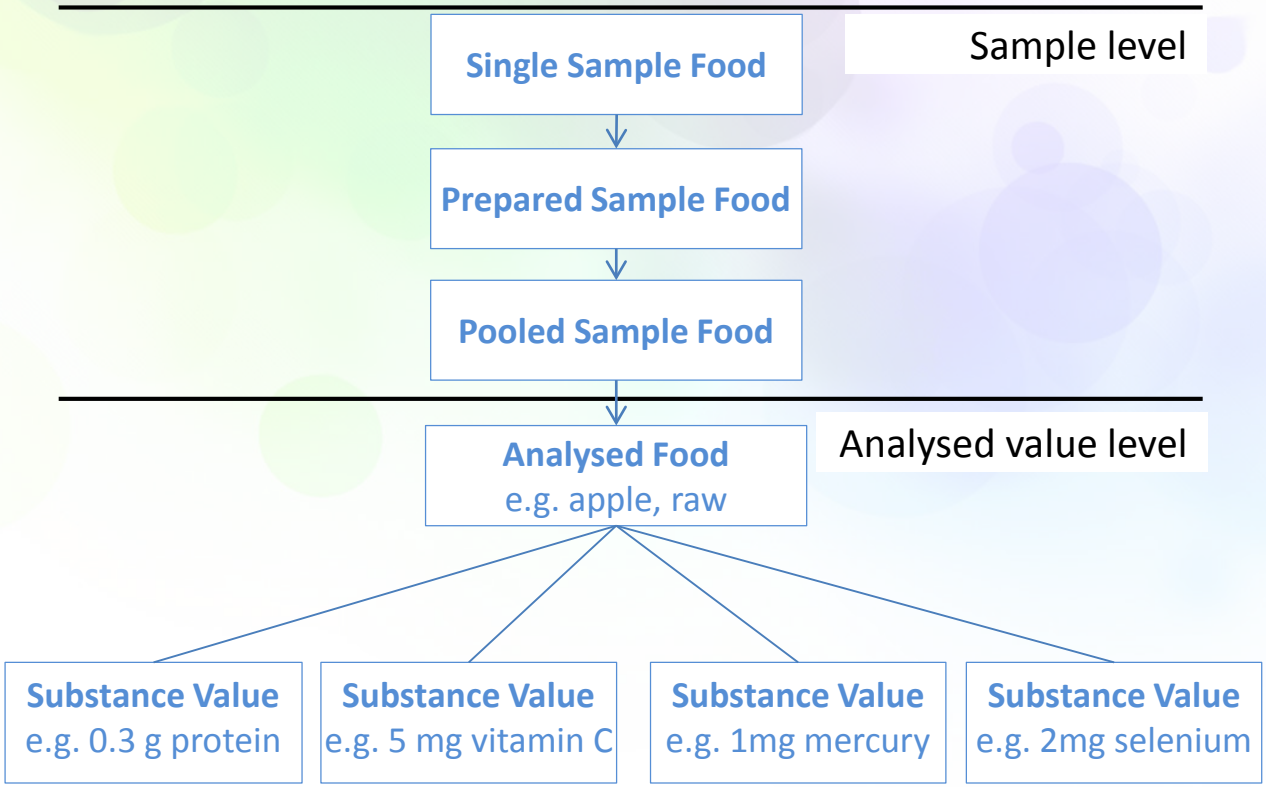
Use a feasible tool to manage your TDS and make your data sustainable.

How Data Needs to be Managed?

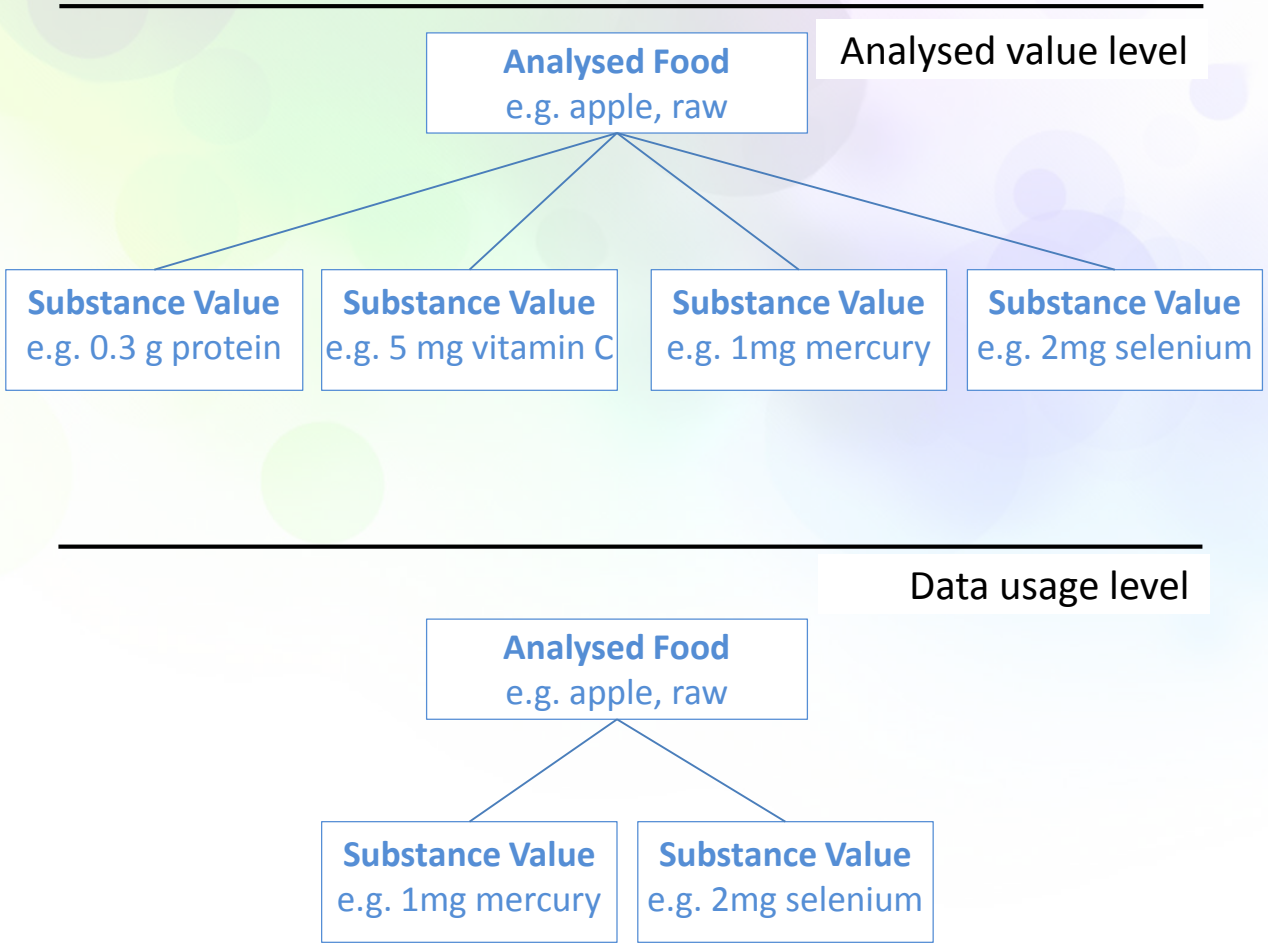
TDS Data Entities and Levels

- Entity = Real world object such as chair or table
- Data Levels = Different stages that data go through in a database

TDS Data Entities and Levels



TDS Data Entities and Levels



Take Home Message IV

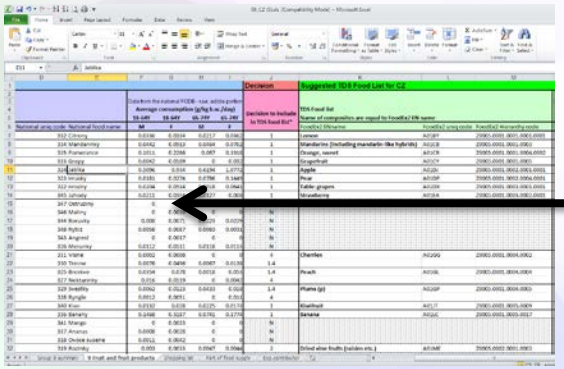
TDS data goes through a data live cycle.

How to Work with a TDS Data Management System?

Planning Phase

Assemble food list

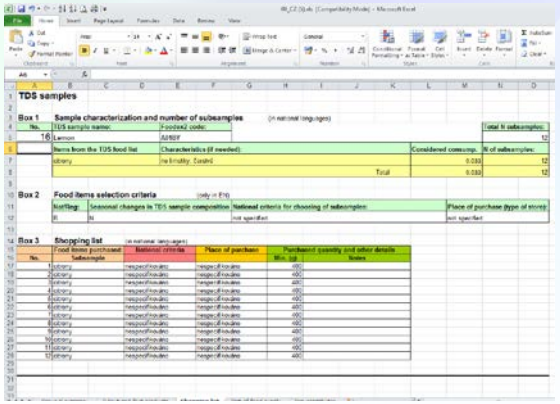
Different data sets are needed: food comp data, consumption data and former TDS data



FOODCASE

Assemble shopping list

EXCEL is a handy tool to manipulate multiple entries. Issue: 18 EXCEL files (one for each food category)



FOODCASE

Prepare samples

Form pool groups



FOODCASE

t

How to Work with a TDS Data Management System?

Execution Phase

Collect food items



Prepare samples



Form pool groups



Analyse food items



t

Only deviation from original plan must be documented.

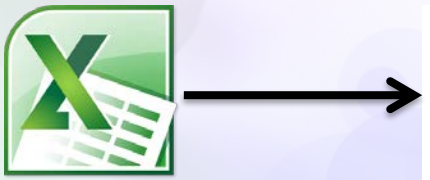
How to Work with a TDS Data Management System?

Lab Analysis Phase

t

Enter substance values

LIMS



Archive data



Export data for risk assessment

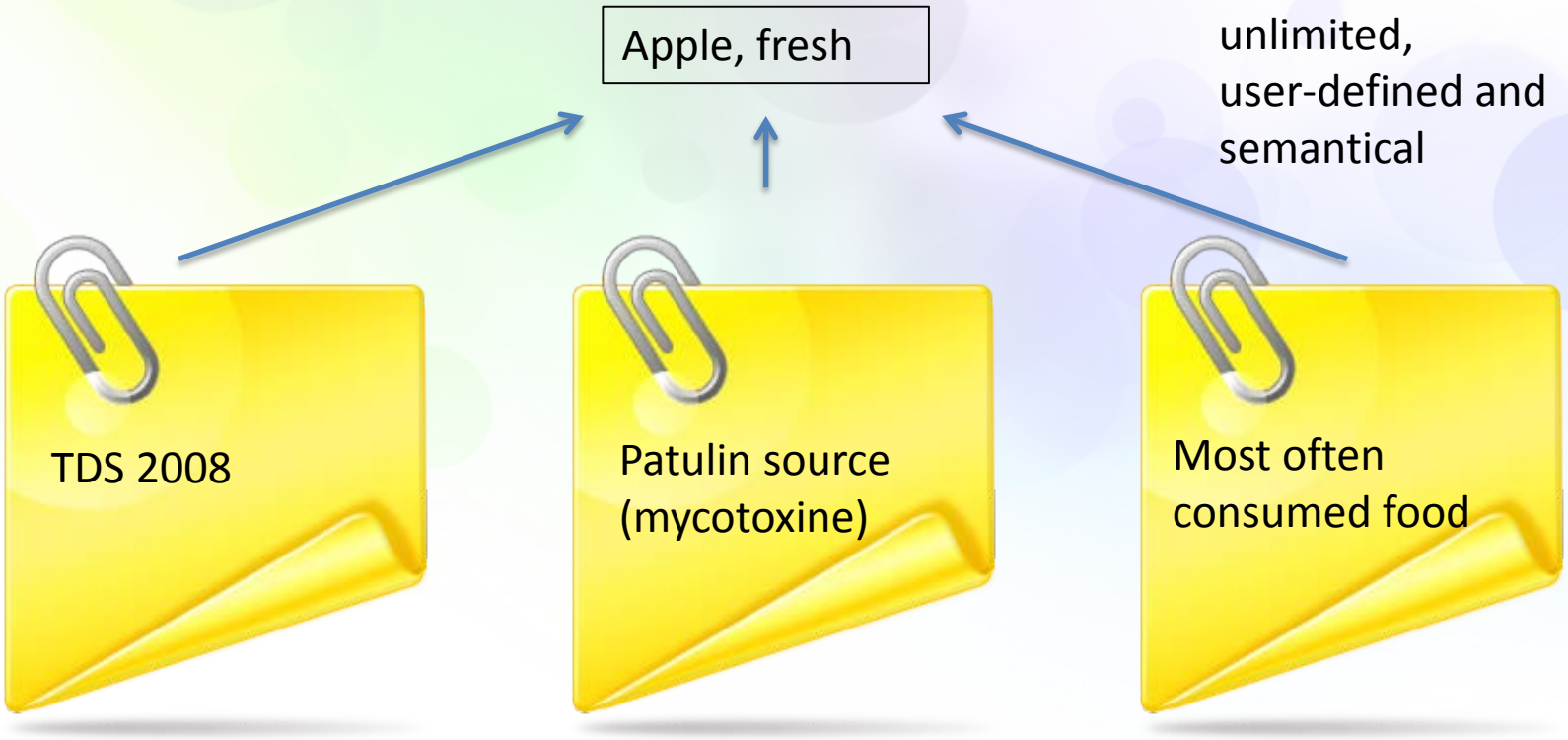
Take Home Message V

A good TDS data management system does not restrict you in your TDS approach. It rather supports your value documentation and risk assessment.

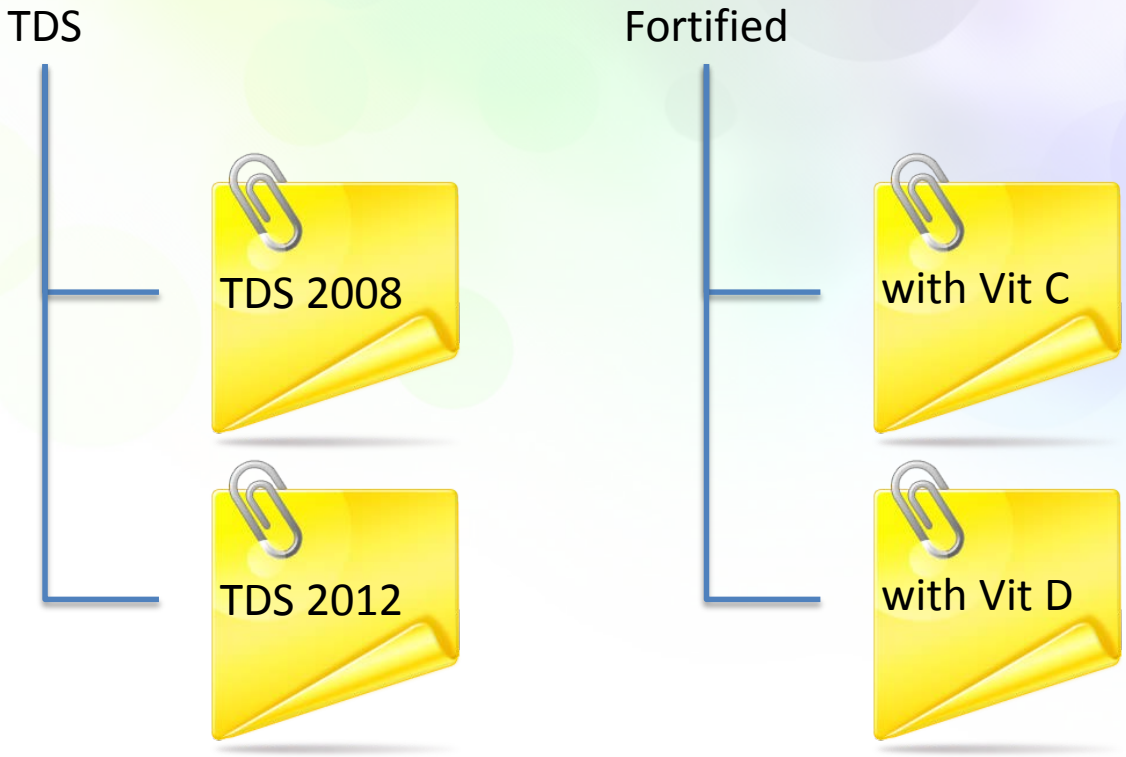
Concepts for Data Management

TDS Data Markers

Concept 1:
Add information to data which is unlimited, user-defined and semantical



TDS Data Markers



Concept 2:
Structure markers to simplify overview

TDS Data Markers

Concept 3:
Markers are searchable

Search food items which were not
used in the TDS 2008
but in TDS 2010
and are fortified with vitamin C

“TDS 2008” AND NOT “TDS 2010” AND “fortified with vit C”

Take Home Message VI

Markers is a flexible concept for semantical data tagging.

Reports and Exports of TDS Data

Reporting/Exporting means

- to search and prepare existing data,
- to bring it into a usable form such as table or graphic,
- and to put it into a **human or computer readable** file format such as PDF or EXCEL

Reports and Exports of TDS Data

Challenge: Reports are useful but exist in endless number of variations and implementation costs.

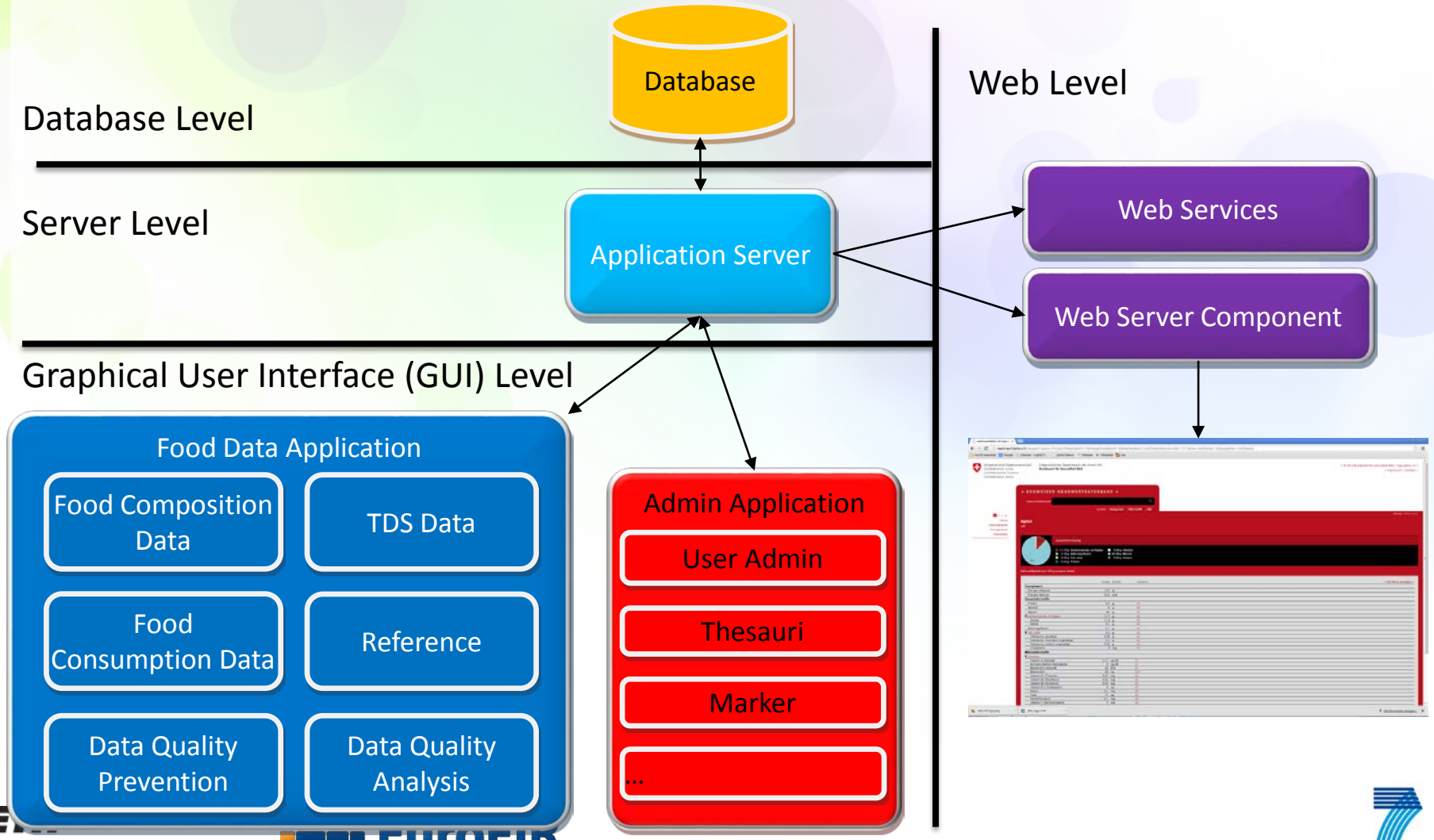
Solution:

- Data must be selectable for use in report/export
- User should be able to define their own reports in terms of form and format

Take Home Message VII

Report and Export functionalities should satisfy needs and in best case can be extended by users.

Architecture of a Food Information System



Use Cases in FoodCASE

Starting Mask - Overview

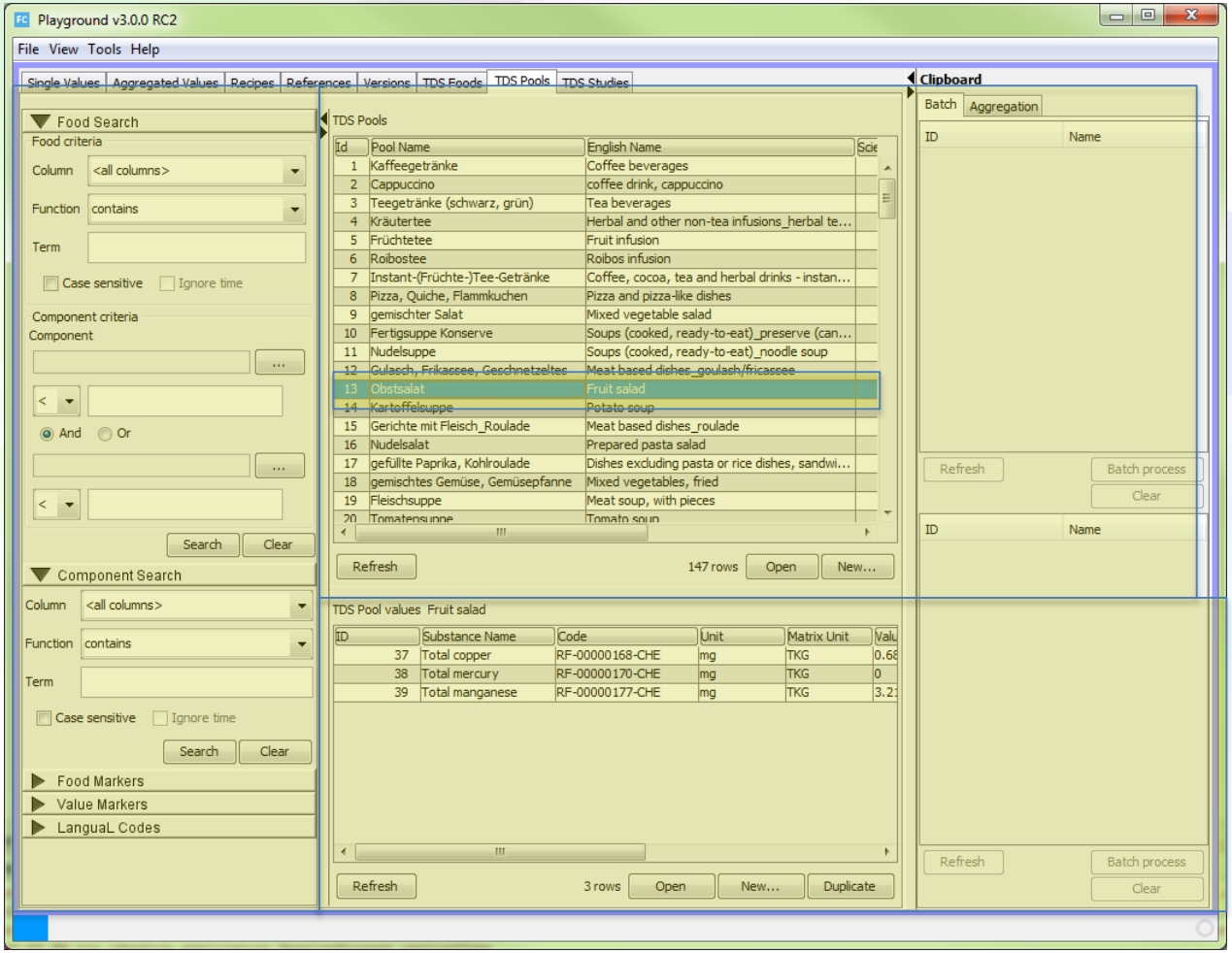
Search Area

Data Listing Area

- All or certain columns
- Content of substances
- Data markers
- LanguaL / FoodEx2 codes

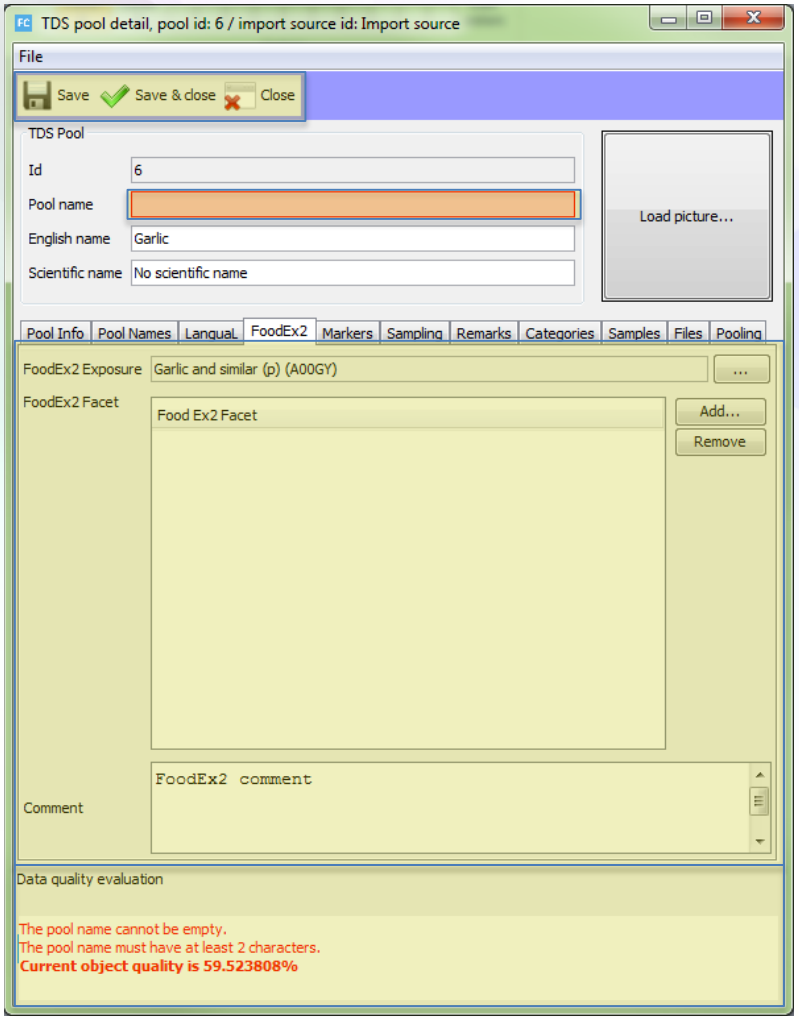
Food items

Chemical occurrences (beneficial and harmful substances)



Managing Food

- Mandatory fields
- Many other attributes for a single food
- Data quality panel

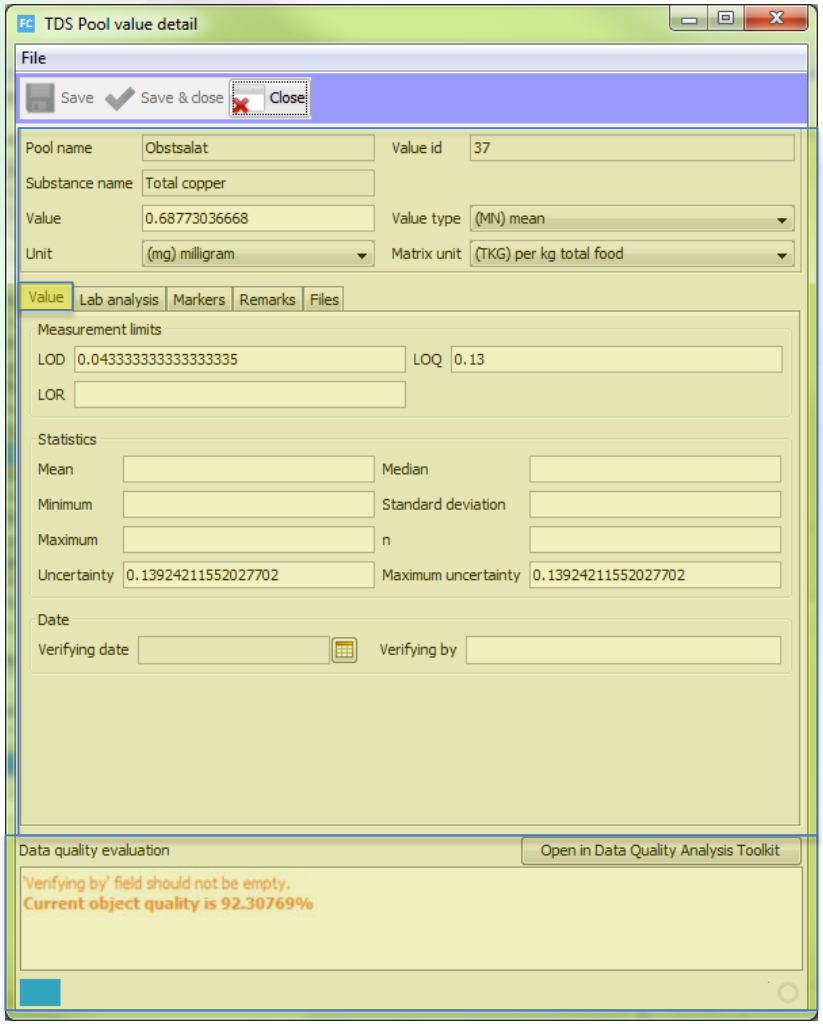


Managing Analysed Values

- Laboratory report
- Other scientific publication
- Basic idea:
  Provide as much value documentation as possible
- Attach original published document

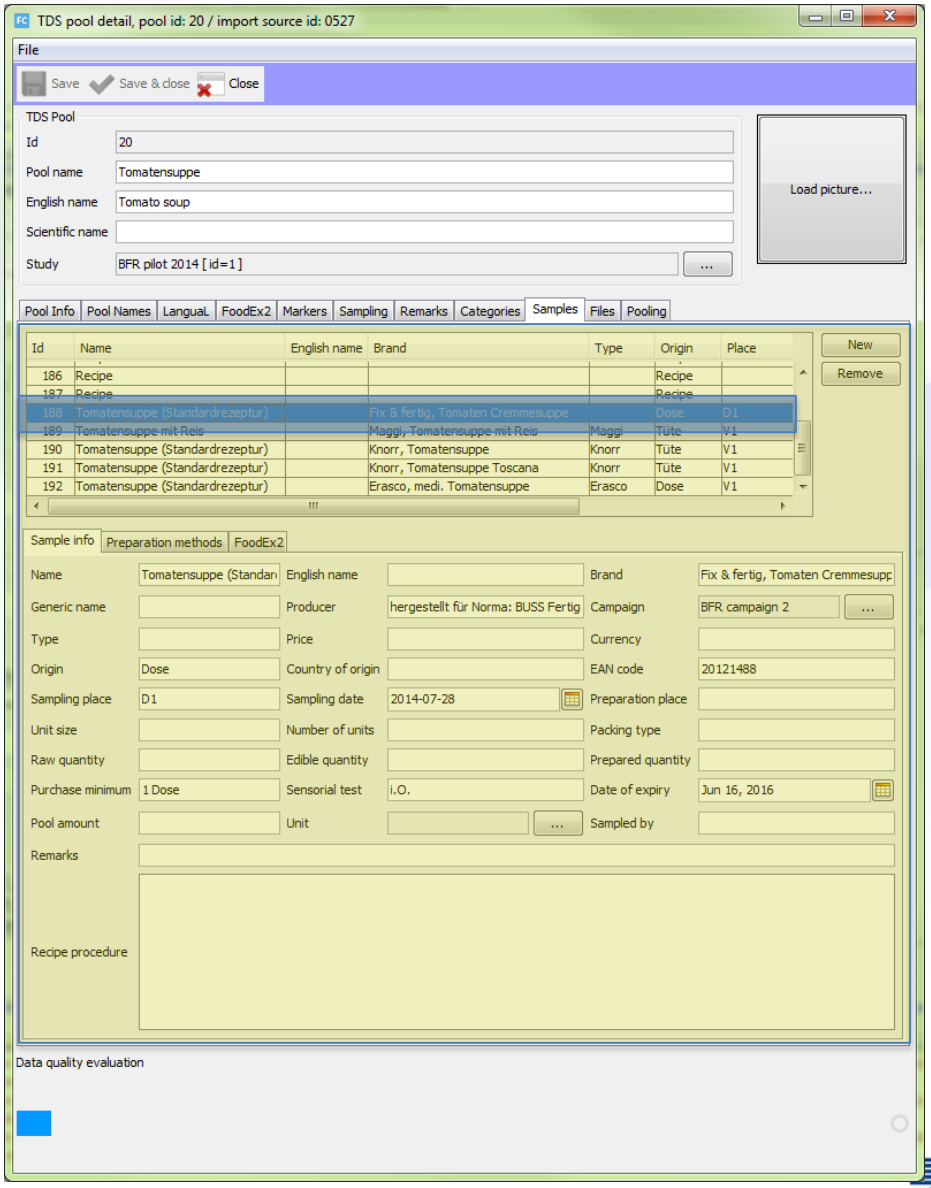
Managing Analysed Values

- Many available attributes
- Data quality panel



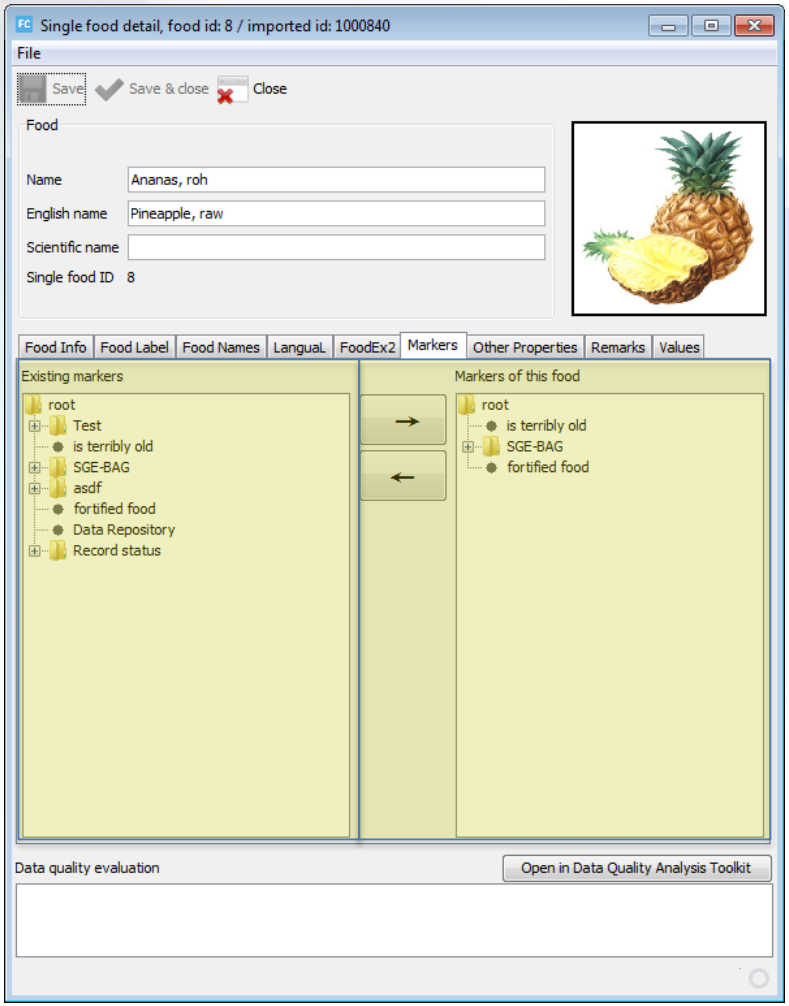
Managing Samples

- Attribute values depending on the table selection



Markers

- Available markers
- Assign or remove from current food item
- Definition of markers in the Admin client
- Custom naming
- Multilevel markers for TDS



Take Home Message VIII

FoodCASE contains all necessary concepts and user Interfaces are simple structured and intuitive.

Data Quality in FoodCASE

Data Quality (DQ)

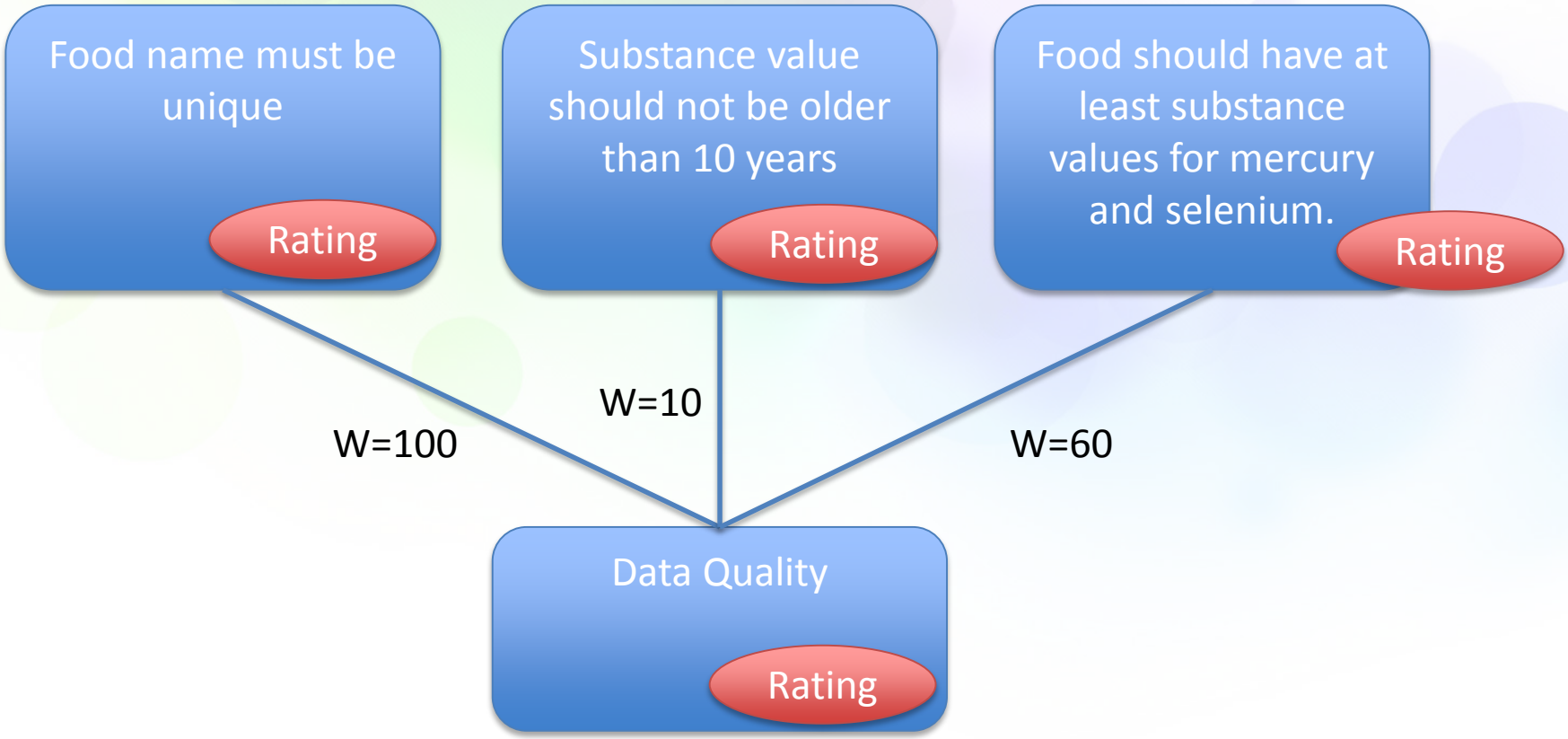
Definitions of data quality:

1. Data or information has quality if it is appropriate for the intended use
2. The quality of data or information depends on users' expectations and requirements

Data Quality Requirement (DQR)

- A data quality requirement is a concrete intrinsic or contextual constraint to data that helps to satisfy an aspect or construct of data quality.

Data Quality Requirements



Data Quality Framework (DQF) concept

- Managing and measuring data quality (DQ)
- Constraints as DQ indicators
- Unique definition and validation
- Flexible and highly configurable DQF
  - Importance weights for constraints
  - Grouping of constraints
  - DQ contracts for different users
  - Various visualisations

DQF Weights & Contracts

- User can define weights and threshold
  - 0: Not important, 100: Most important
  - Threshold between warning and error
  - Default weight
- A contract defines the weight (importance) for each of the constraints
- Default contract exists as fallback
- Users can define their (multiple) own contracts
- Contracts can easily be imported and shared

Types of DQ Requirements

Food name must be unique

Hard constraints → Error

Determines if data is deficient or not

Food should have at least substance values for mercury and selenium.

Soft constraints → Warning

The more requirements are satisfied, the better is data quality

Substance value should not be older than 10 years

Indicators → Warning

Indicate better quality but quality improvement is not evident

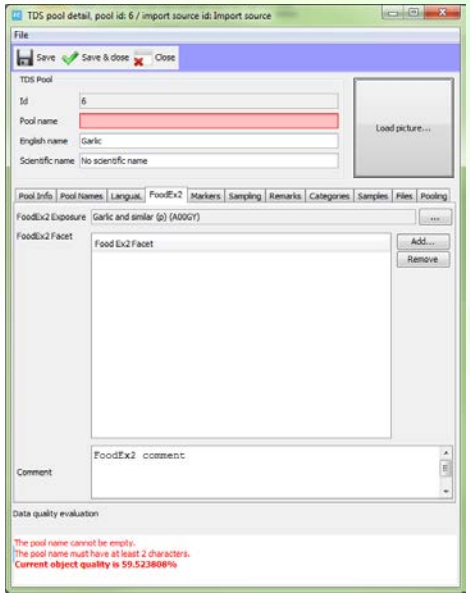
Data Quality Rating

- Can be simple as ‘satisfied’ or ‘not satisfied’ or more complex, for instance 0 to 100
- Can be defined by user

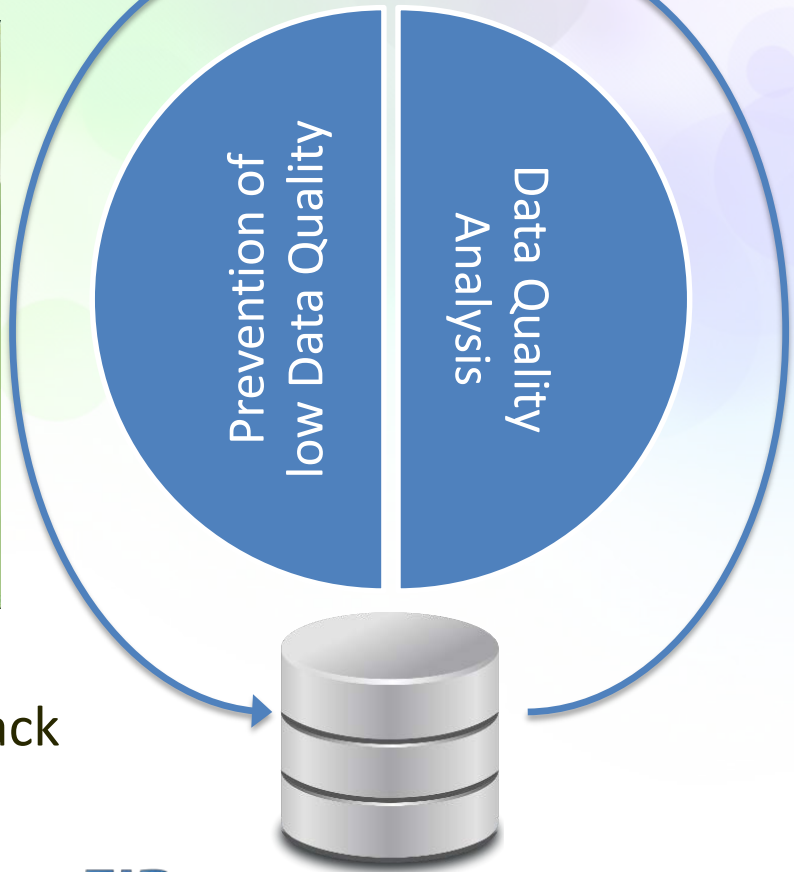


Data Quality Areas





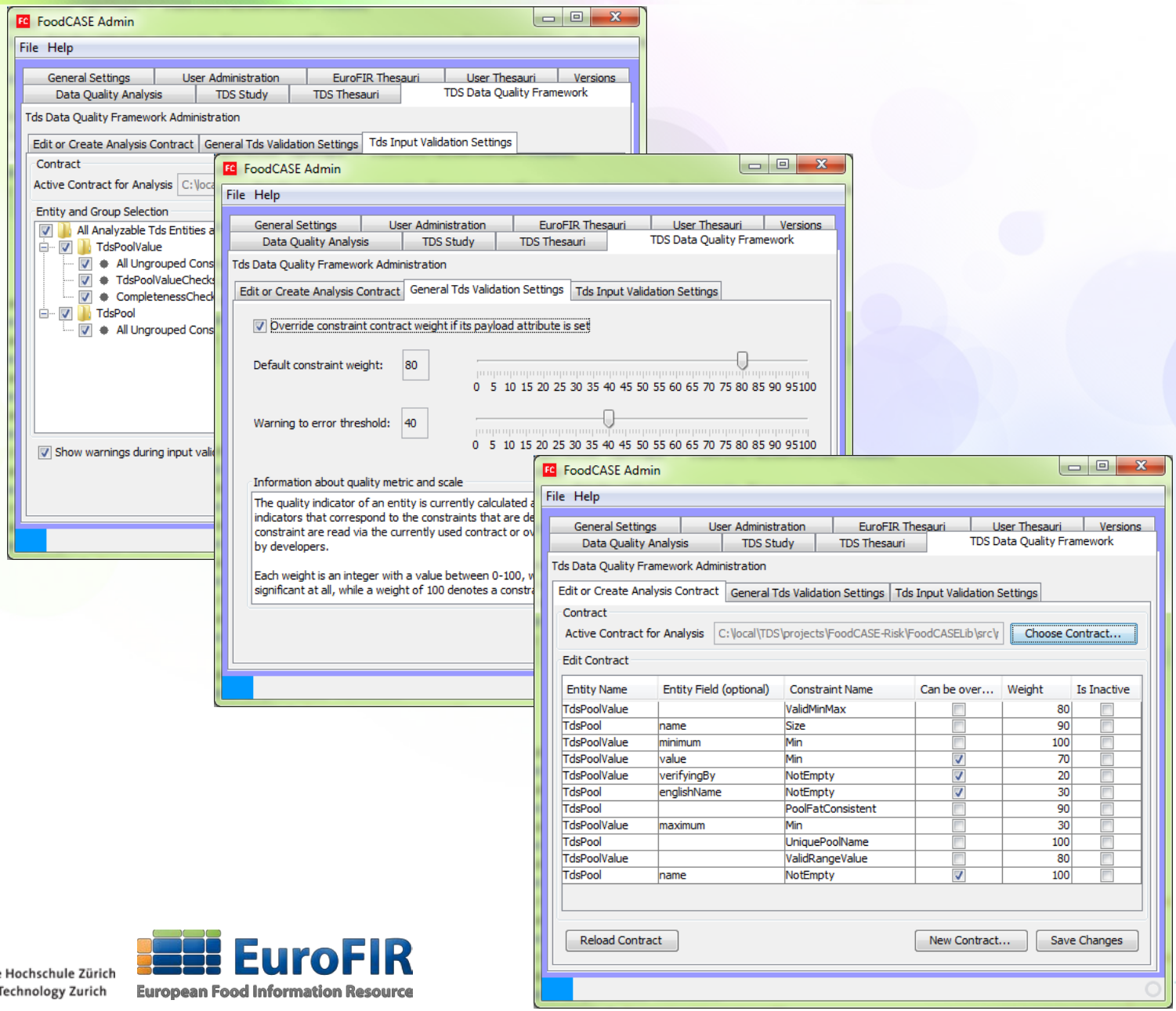
Data quality input validation & feedback



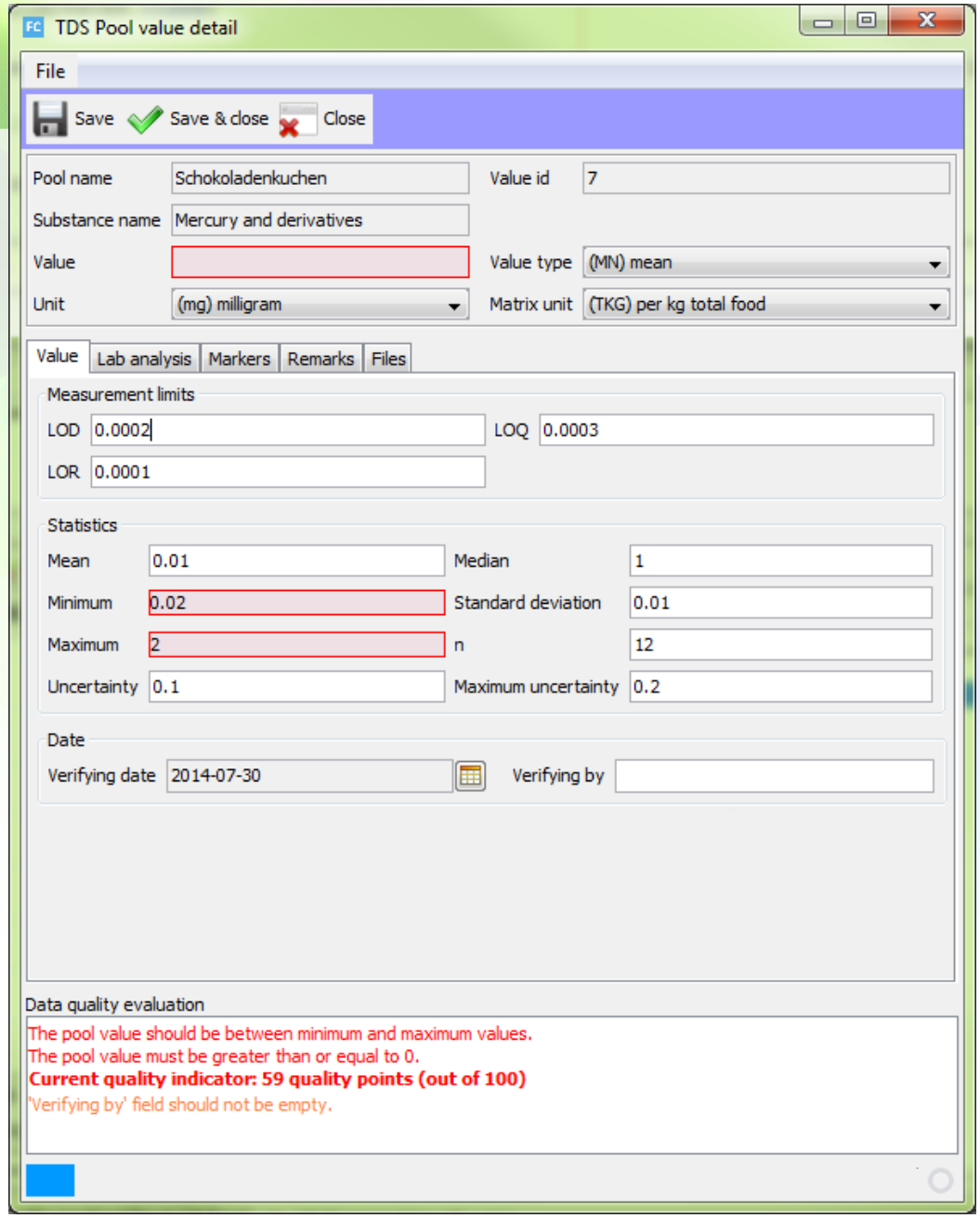


Data quality analysis & feedback

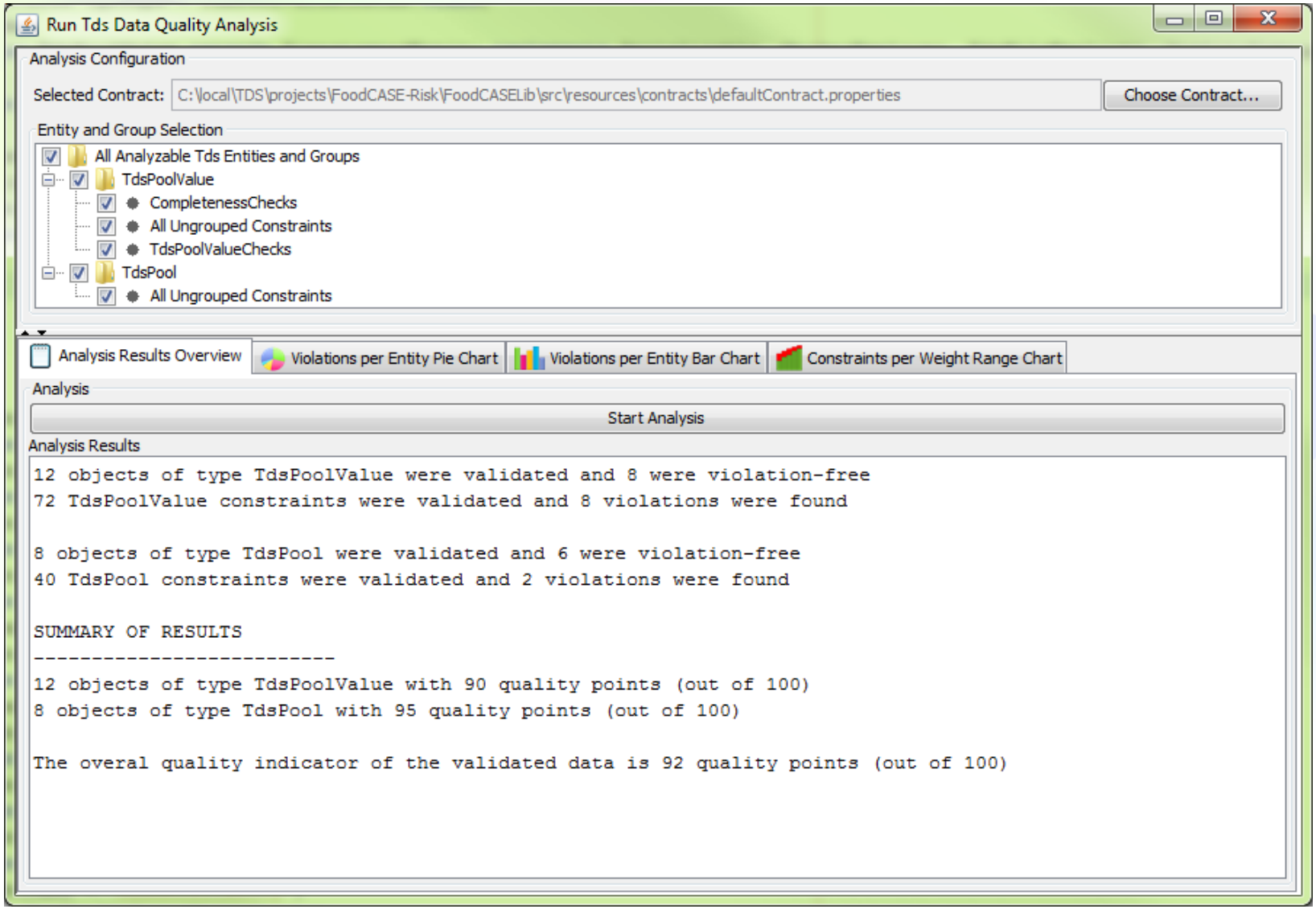
Data quality administration



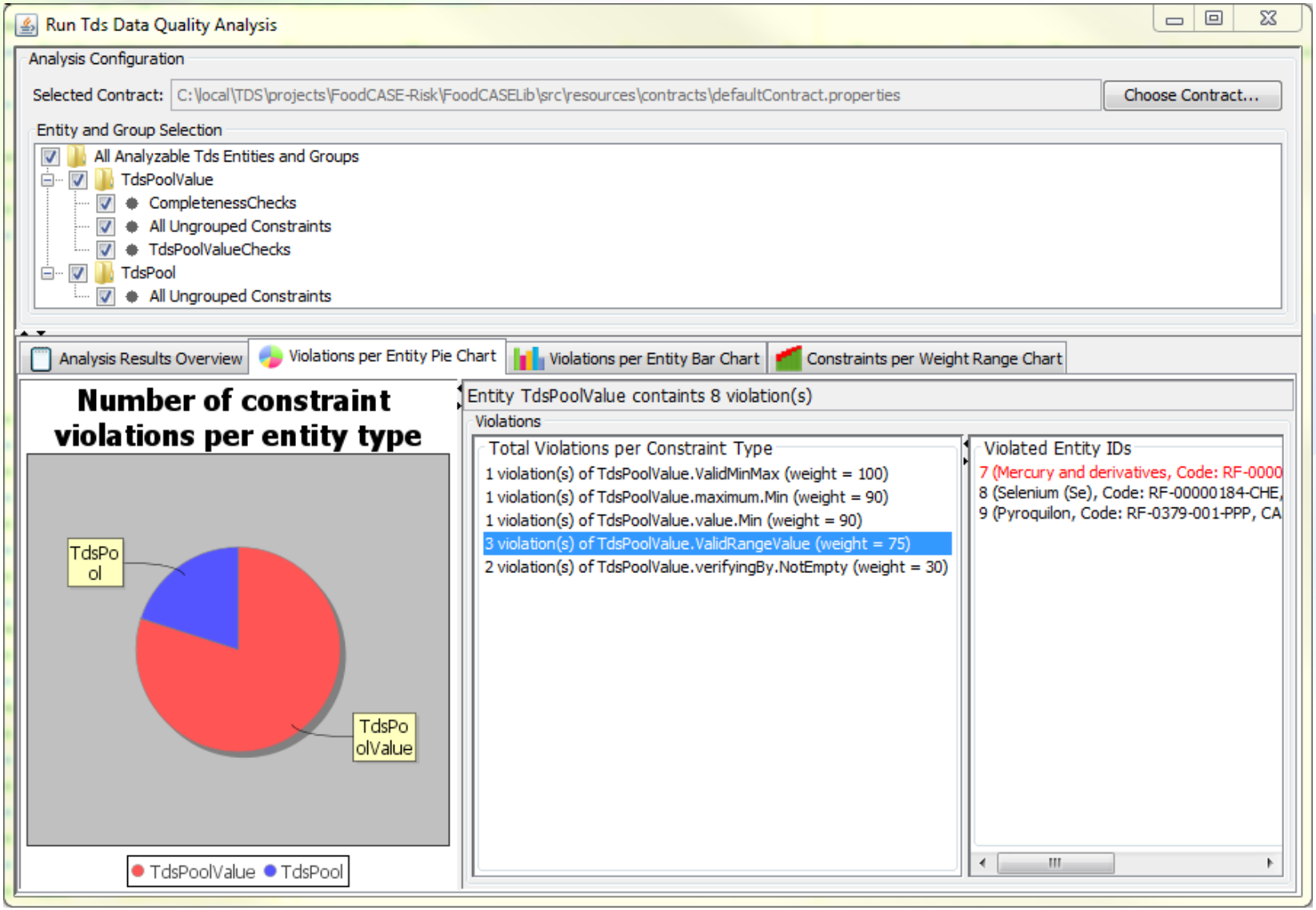
Data quality input validation



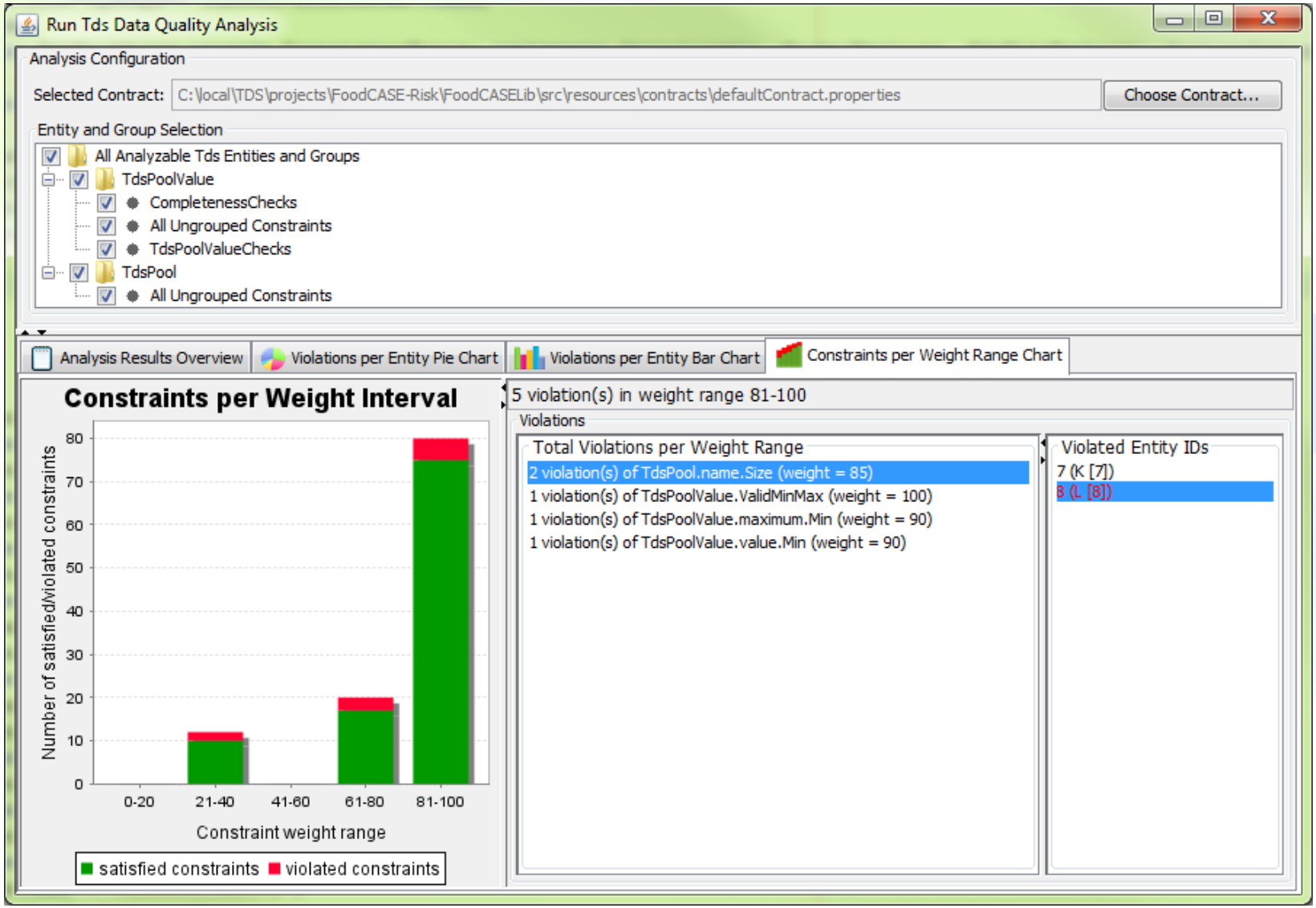
Data quality analysis



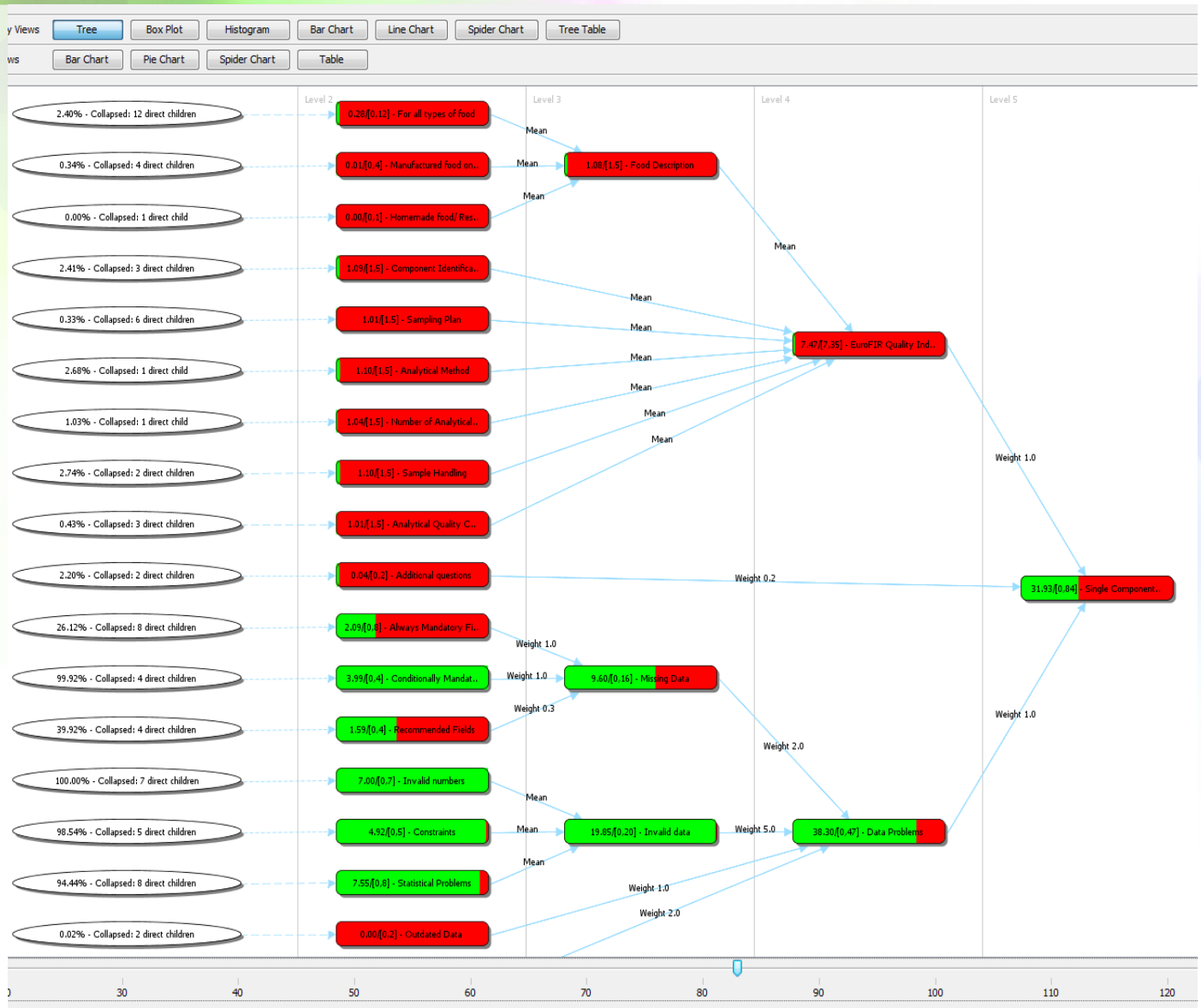
Data quality analysis



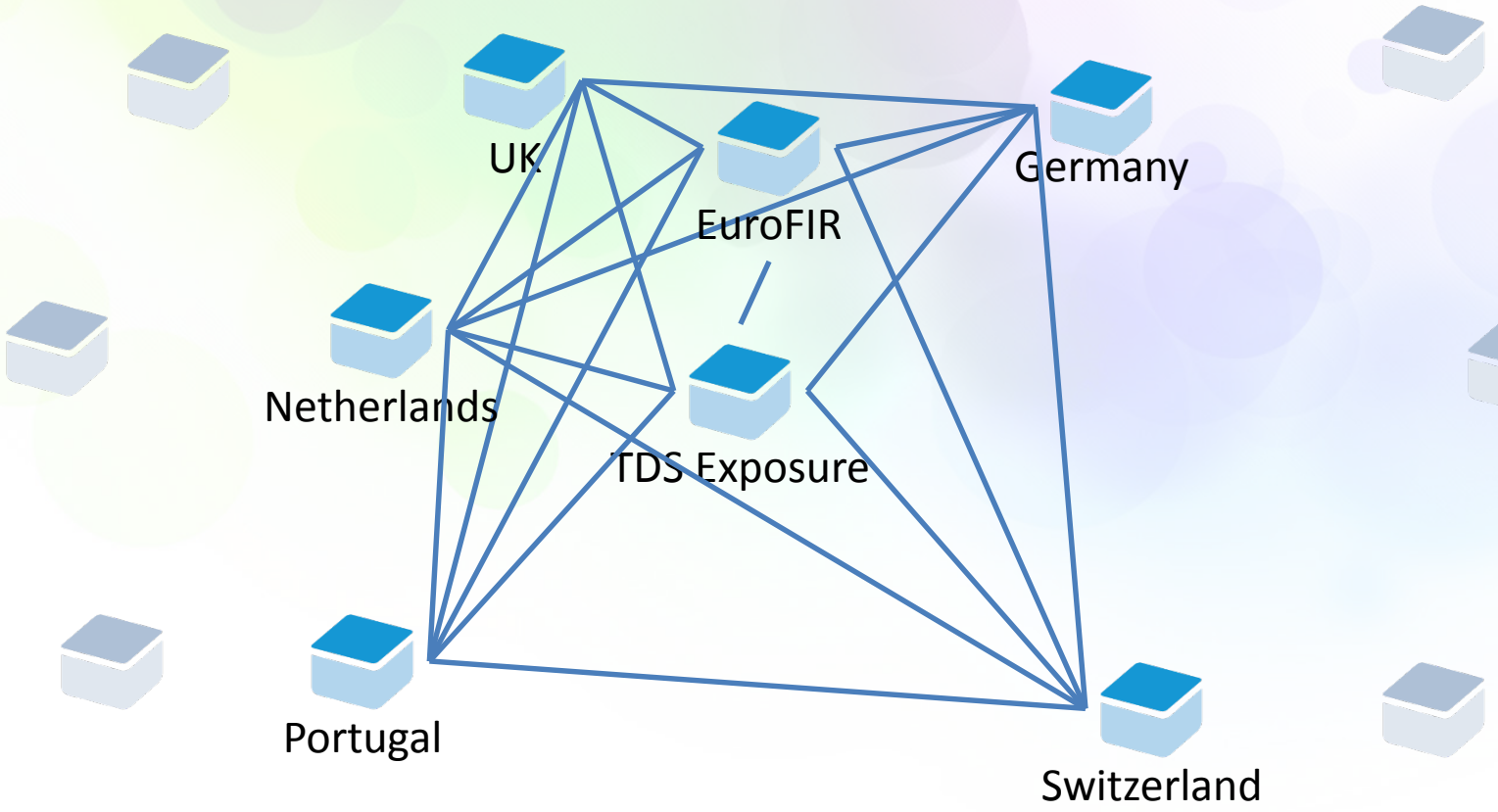
Data quality analysis



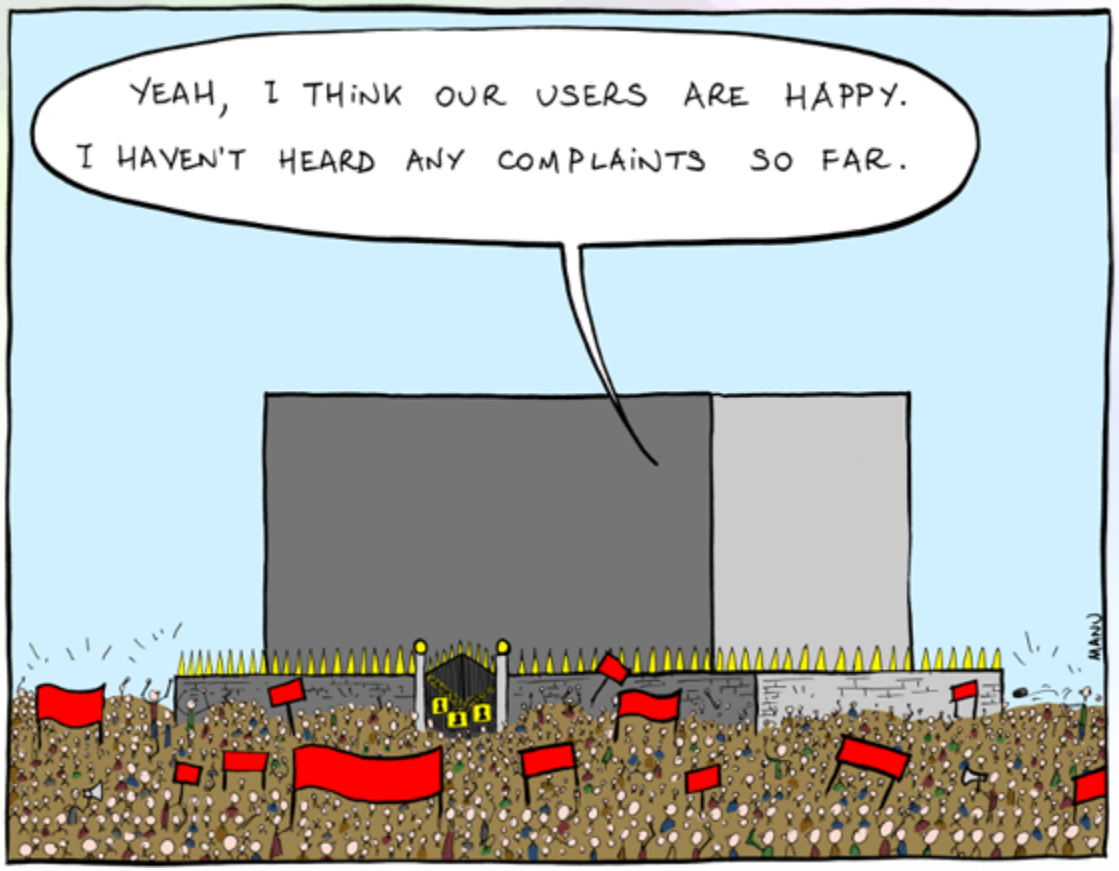
Data quality analysis



Outlook: European FoodCASE Infrastructure



Your feedback / discussion



Thank you for your attention