

Excel is not a database



QUESTIONS

Ulf
Månsson



Mikael
Månsson



Agenda

- Horror stories
- What is a real database?
- Excel and databases can live together!
- Excel data checks – *Tips & Trix*





POLL

Horror stories - background



Horror stories



<http://www.eusprig.org>

Identifier:	POB2001
Title:	Data not controlled, 16000 UK Covid-19 test results lost for a week
Source:	https://www.bbc.co.uk/news/technology-54423988
Release Date:	08 October 2020
Risk:	Lives put at risk because the contact-tracing process had been delayed
Discrepancy:	16,000 test cases in a week

THE ROLE OF SPREADSHEETS

"...spreadsheets will always fill the void between what a business needs today and the formal installed systems..." Mel Glass et al [<http://arxiv.org/abs/0908.1584>]

Ungefär:

"...kalkylark kommer alltid att fylla tomrummet mellan vad en organisation behöver idag och de officiella IT-systemen som nu är i drift!"

What is a "real" database then?

In terms of a "real database", I think of systems like PostgreSQL, SQL Server, Oracle, or even more lightweight systems like SQLite. These systems focus on storing large amounts of data — with integrity. Meaning you can set up rules regarding your precious data. For the case of COVID-19, this could mean ensuring that the same test is not registered twice.

SQL Constraints

SQL constraints are used to specify rules for the data in a table.

Constraints are used to limit the type of data that can go into a table. This ensures the accuracy and reliability of the data in the table. If there is any violation between the constraint and the data action, the action is aborted.

Constraints can be column level or table level. Column level constraints apply to a column, and table level constraints apply to the whole table.

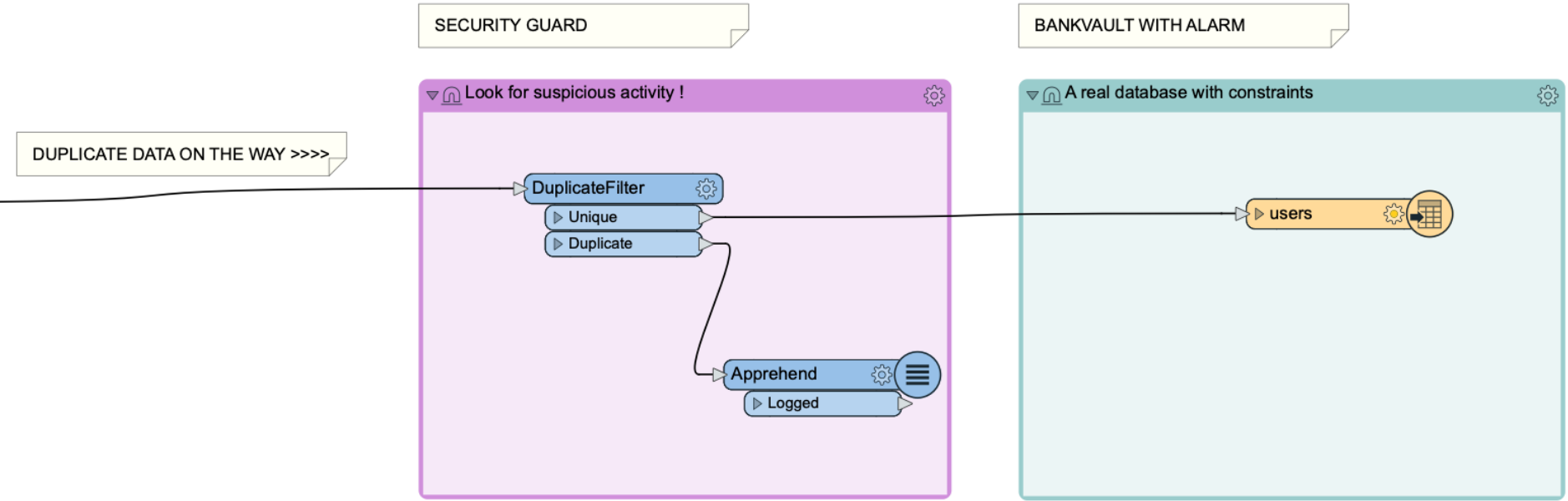
The following constraints are commonly used in SQL:

- **NOT NULL** - Ensures that a column cannot have a NULL value
- **UNIQUE** - Ensures that all values in a column are different
- **PRIMARY KEY** - A combination of a NOT NULL and UNIQUE. Uniquely identifies each row in a table
- **FOREIGN KEY** - Uniquely identifies a row/record in another table
- **CHECK** - Ensures that all values in a column satisfies a specific condition
- **DEFAULT** - Sets a default value for a column when no value is specified
- **INDEX** - Used to create and retrieve data from the database very quickly



DEMO

There are layers of "data-security"



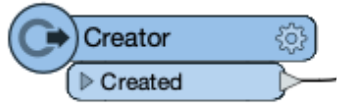
Excel and Databases can live together

- Excel – easy tool for input!
- Use templates for fractions of data (Pro tip: Use checksums!)
- ETL Tool for gathering and verifying.
- One single database!

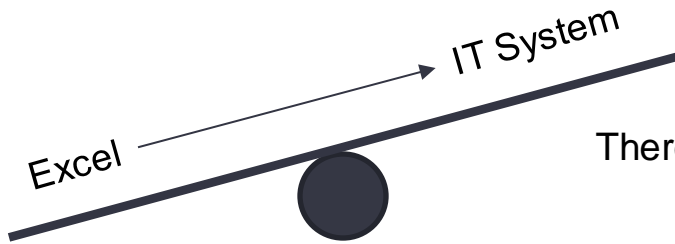
Mr Jan Bliki at EEA:

“Excel is still the easiest input tool among non-database users. It’s hard to replace if users are that used to that format. In our case, we ask to deliver multiple files using the same template: Excel files that only store a fraction of the data (for example, by hospital and by day). Then, we use a central service that brings all these daily results together into one single database using a single ETL tool. To make easy use of such a database, a BI tool could be linked so those same users would still have the ability to produce an overview of all results.”

Excel data checks – Tips & Trix



Look out!



There is a risk that your *"just a quick spreadsheet"* assumes the role of a formal IT System.

(This **without** the Systems Development process that should follow a formal IT System)

Vital data? At least consider a database!

THE ROLE OF SPREADSHEETS





































"...spreadsheets will always fill the void between what a business needs today and the formal installed systems..." Mel Glass et al [<http://arxiv.org/abs/0908.1584>]

Summary how FME can help

- Move data in/out of databases!
- Checks – before and after!
- Automation
 - Self Service validation on FME Server/Cloud
 - Triggers and alarms
- And much more



Upcoming Events

MARCH 2021				
 	11/3 9.00-9.30 Online	 	Webinar - Excel is not a database FREE	Register →
 	16-18/3 Online	 	FME Desktop Basic SEK 5 900	Register →
 	18/3 9.00-9.30 Online	 	Webinar - E-mails & FME FREE	Register →
 	25/3 9.00-9.30 Online	 	Webinar - FME goes BI (Swedish) FREE	Register →
APRIL 2021				
 	13-15/4 Online	 	FME Server Authoring SEK 5 900	Register →
MAY 2021				
 	18-20/5 Online	 	FME Desktop Advanced SEK 5 900	FULLBOKAT
 	18-19/5 08:30 – 12:30 Online	 	FME Point Cloud Mastery SEK 5 900	Register →
 	25-27/5 Online	 	FME Desktop Basic SEK 5 900	FULLBOKAT
 	26/5 08:30 – 12:30 Online	 	FME Master that Raster SEK 4 000	Register →

<https://dataflow.center/training-events/>

Contact



fmesupport@sweco.se

ulf.mansson@sweco.se

mikael.mansson@sweco.se

<https://dataflow.center/contact>