BfR
Bundesinstitut für Risikobewertung

User requirements and concepts II

Stephan Worseck

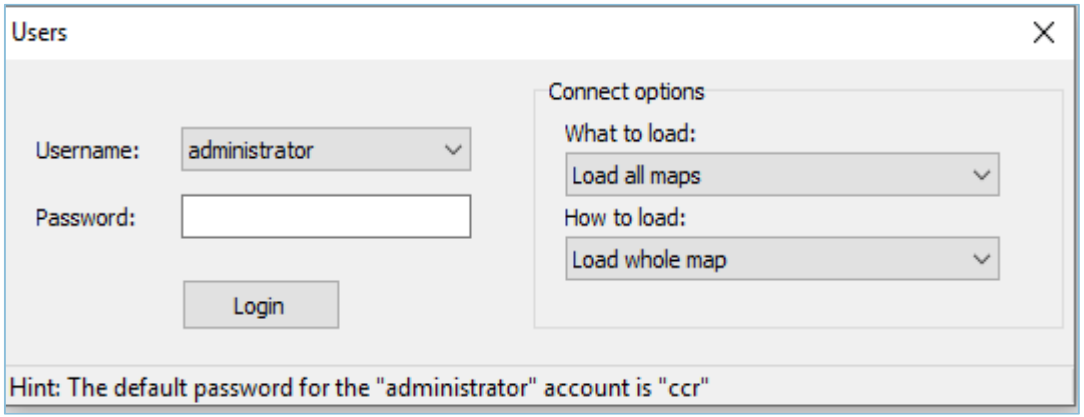
Content

Main messages and open questions aspects regarding

- the Database,
- the Tool,
- Data Management in the Reference Collection
- Management of (Q)SAR results
- Migration

The current MetaPath database / tool implementation

- „MetaPath" works with **Firebird** as the database management system.
- Firebird is on place 17th of the relational database management systems ➔ high risk of an investment.
- Has anybody experiences in using „MetaPath" in a **multi user** environment with a central remote database?
- **No interoperability** is foreseen to other IT-Tools.
- No documentation is available regarding a **role concept** of "MetaPath".



- There exists **no governance model** for the further development of „MetaPath".

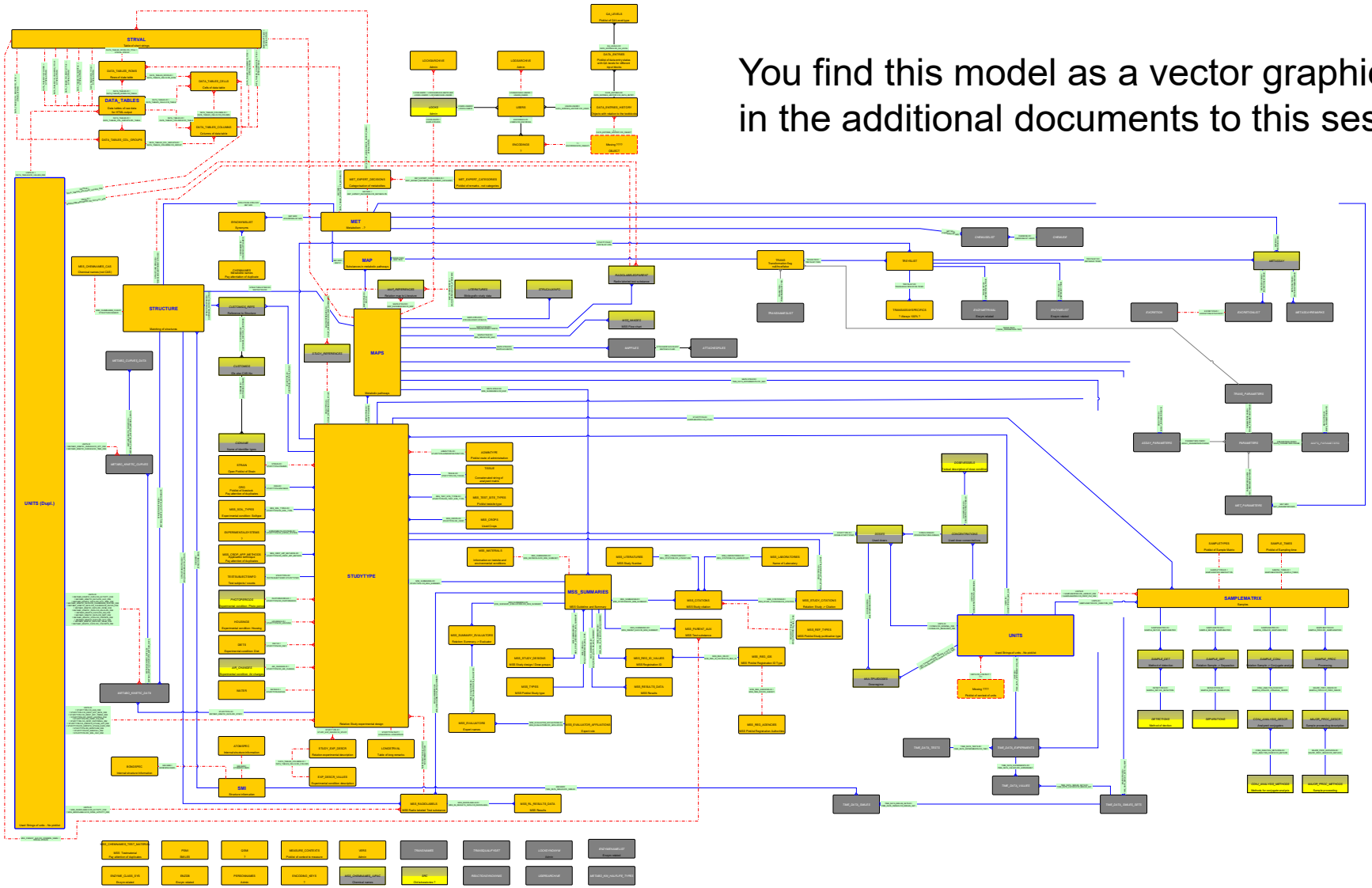
The current MetaPath database implementation

The following database files were used for the analysis:

- Public MetaPath Db_EFSA project.MTB with 341 maps
- RegulatoryDB_771_Sept_2018_v3.1.MTB with 771 maps

Important	Important table used in DER and MSS composers	Normal	Table used in DER and MSS composers
Only used in EFSA DB	Used only by MSS composers	Only used in Regulatory DB	Used by DER composer
Empty table, never used	Never used table	MAPS.STRUCID= RADIOLABELEDPARENT.MAPID	References between columns of two tables.
	Relation based on foreign keys between used tables		Relation based on foreign keys but between a used and an empty table
	Relation interpreted by sub-strings in the fieldname e.g. ATTACHEDFILES.FILEID= MAPFILES.FILEID		Relation interpreted by sub-strings in the fieldname but between a used and an empty table e.g. TRANS.TRID= TRANSNAMESLIST.TRID
	Relation interpreted on basis of an existing index		

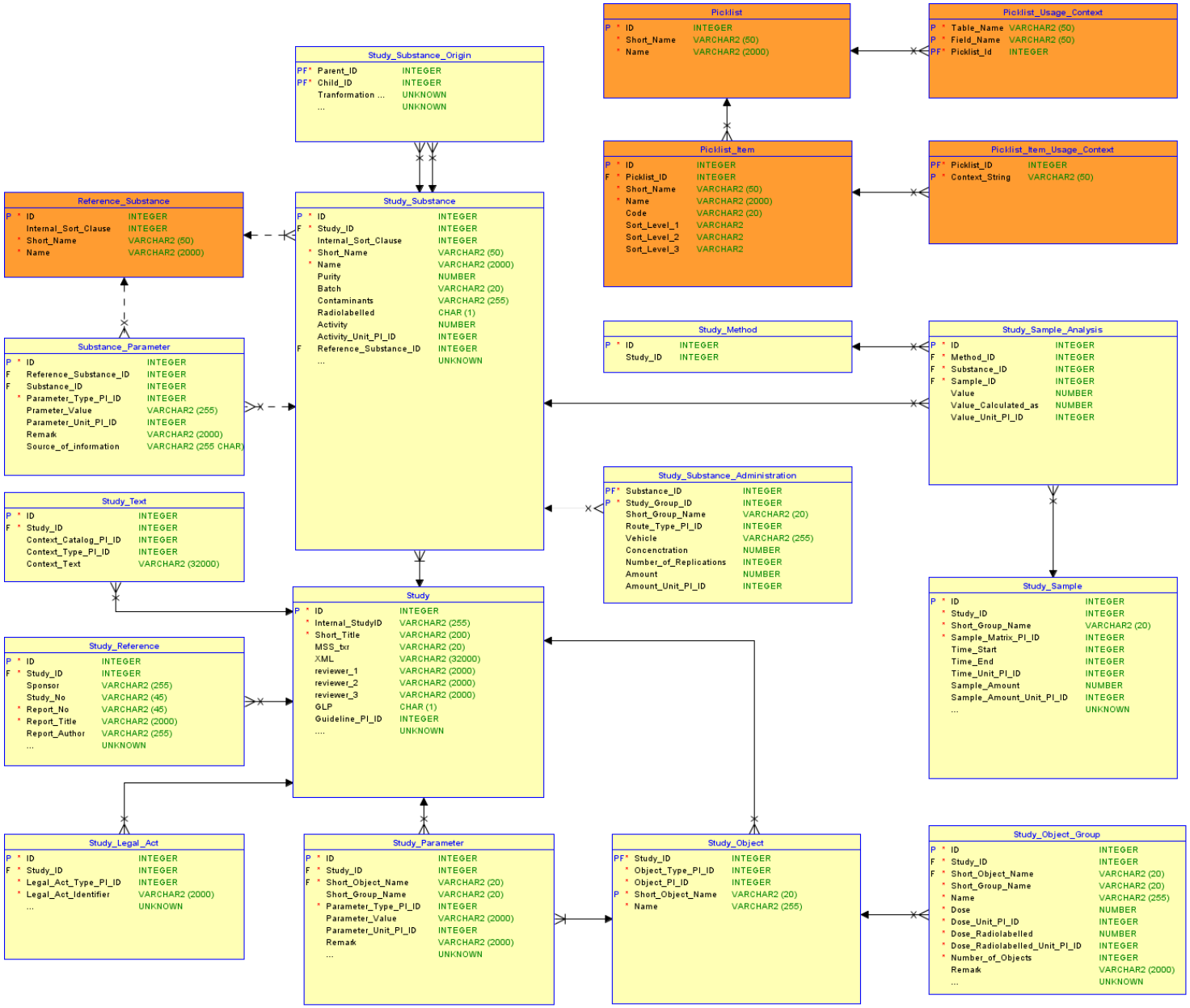
The current MetaPath database implementation



You find this model as a vector graphic
in the additional documents to this session.

An improvement of the information flow should be combined with a redesign of the used database model!

The proposed core structure



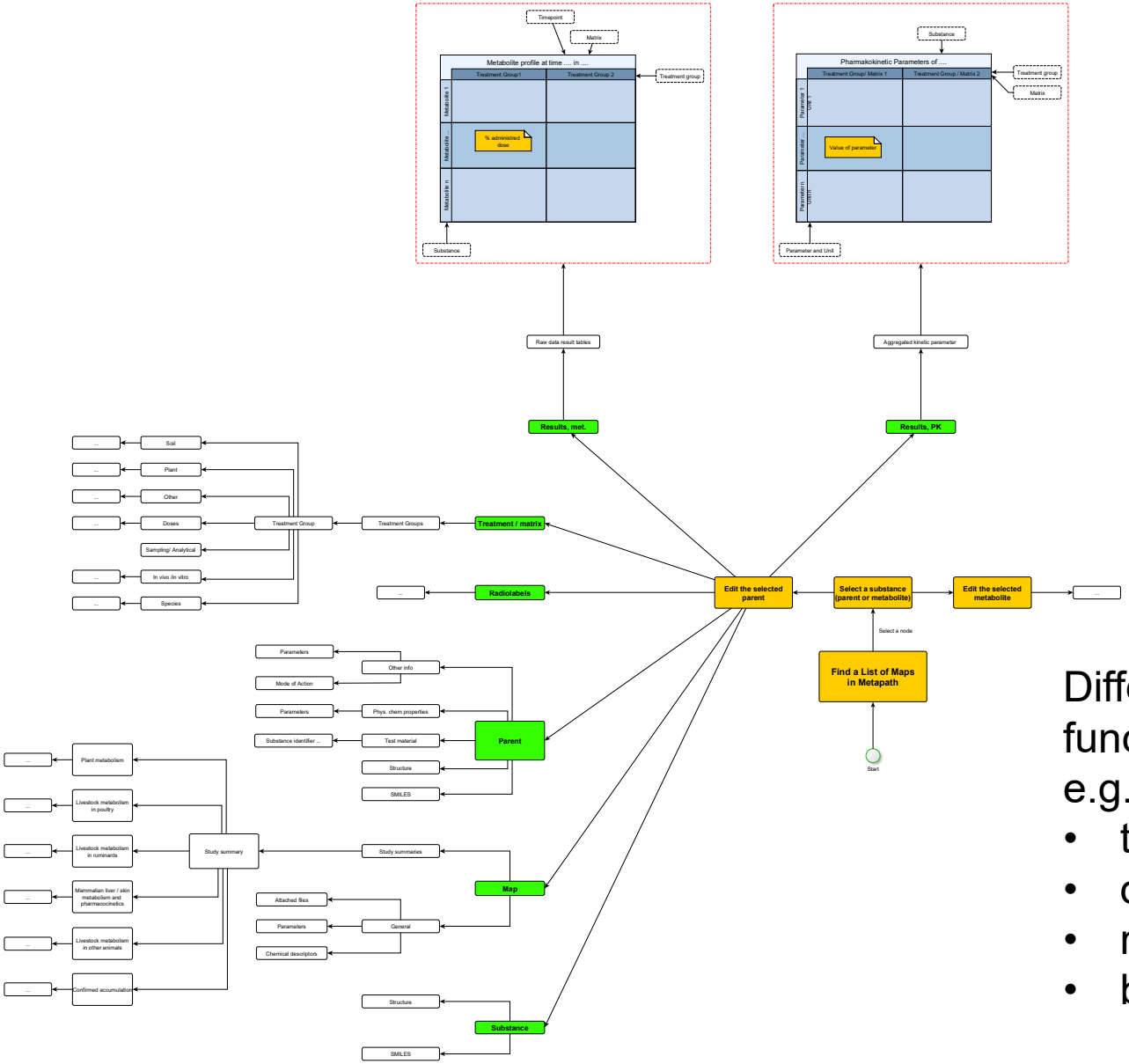
You find this model as a vector graphic in the additional documents to this session.

Content

Main messages and open questions aspects regarding

- the Database,
- the Tool,
- Data Management in the Reference Collection
- Management of (Q)SAR results
- Migration

The hierarchy of some MetaPath functions

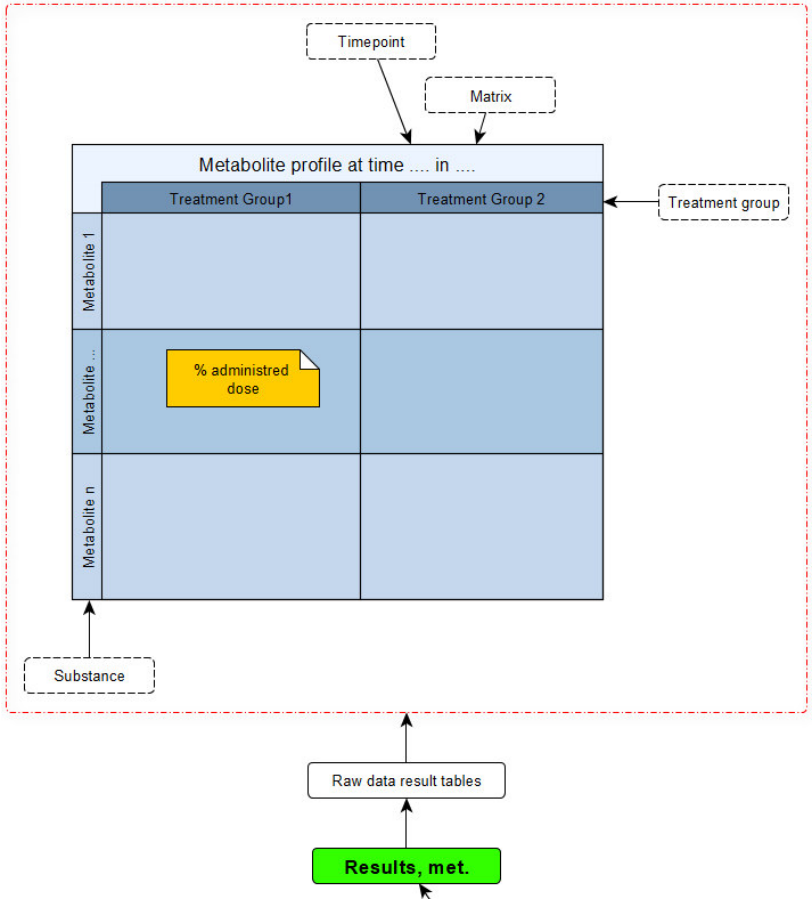


You find this model as a vector graphic in the additional documents to this session.

Different way to activate the functions could confuse the users e.g. by

- the main menu
- different additional tab menus,
- right click context menus and
- buttons

Manage HTML tables from the aggregated raw data in MetaPath

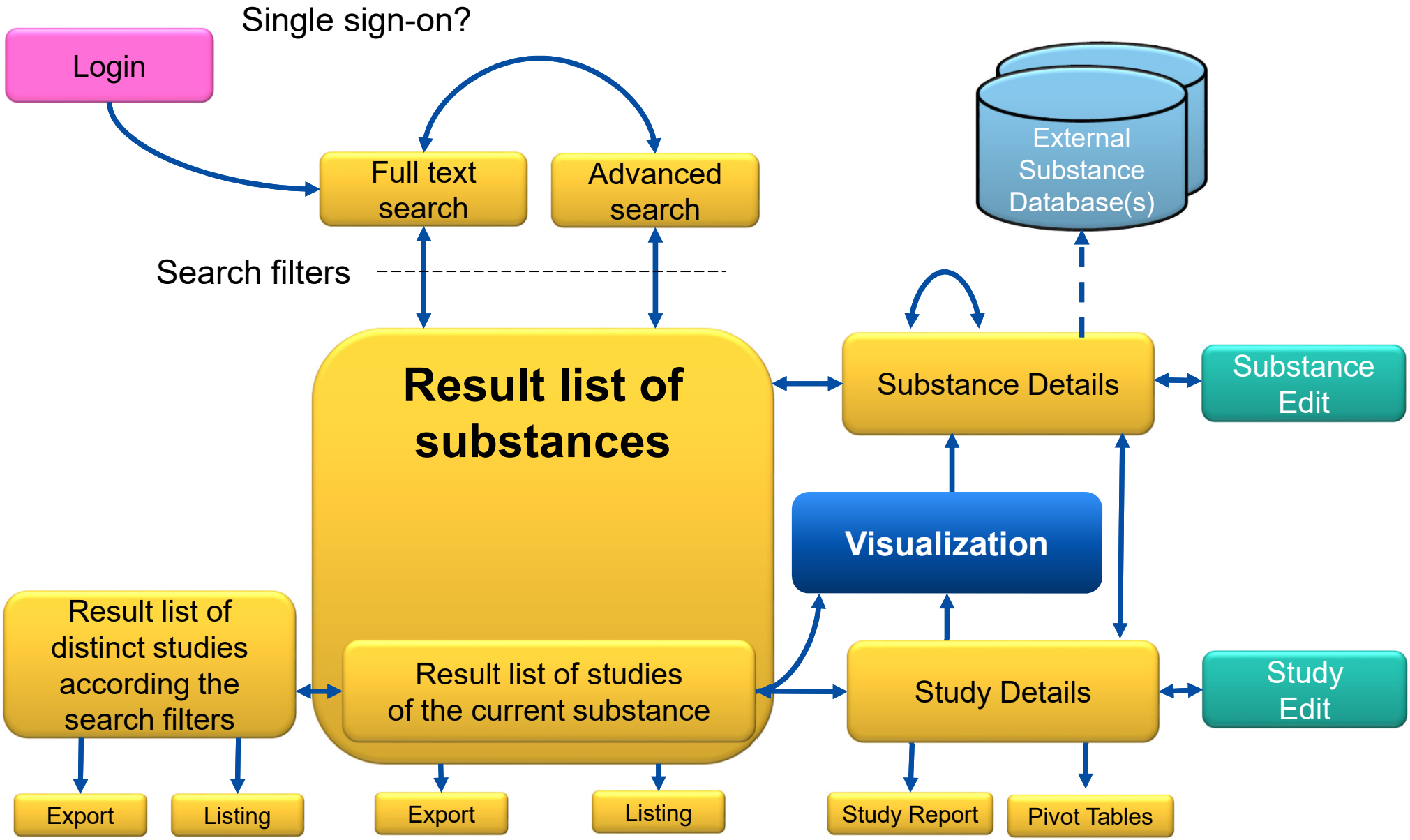



```
<table5 name="" title="Metabolite profile of excreta collected fr
  <footer>"Compounds were not resolved.  This peak included Unkno
  <column group="10 mg/kg male urine">5a</column>
  <column group="10 mg/kg female urine">6a</column>
  <column group="1000 mg/kg male urine">7a</column>
  <column group="1000 mg/kg female urine">8a</column>
  <column group="10 mg/kg male feces">5b</column>
  <column group="10 mg/kg female feces">6b</column>
  <column group="1000 mg/kg male feces">7b</column>
  <column group="1000 mg/kg female feces">8b</column>
  <tableRow>
    <compound>URINE</compound>
    <value>12.56</value>
    <value>13.10</value>
    <value>0.81</value>
    <value>1.14</value>
    <value/>
    <value/>
    <value/>
    <value/>
  </tableRow>
  <tableRow>
    <compound>Amisulbrom (NC-224)</compound>
    <value/>
    <value/>
    <value/>
    <value/>
    <value>40.5</value>
```

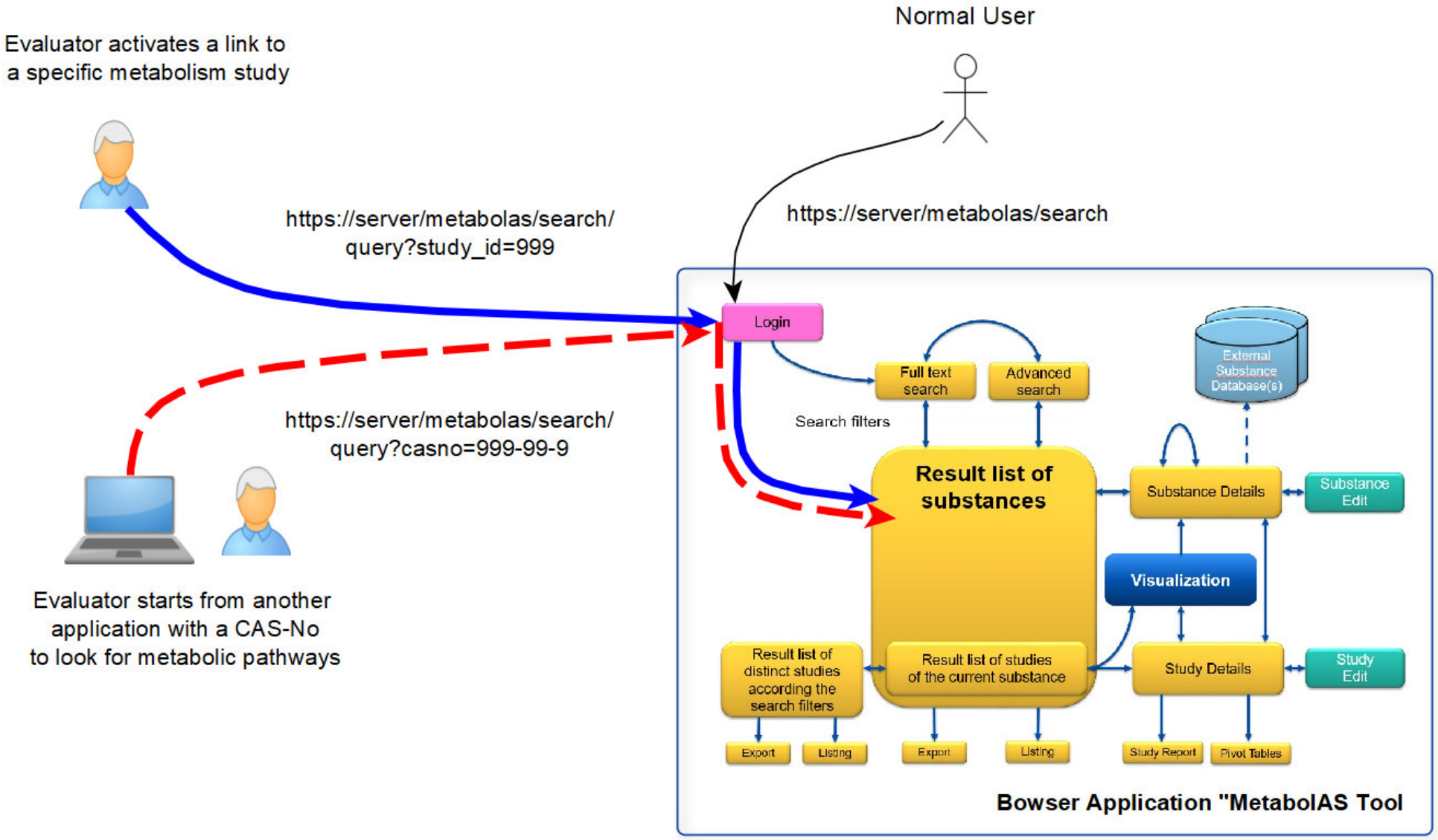
MetaPath user can define different table designs at the time point of data input. After data input, these data are frozen into HTML tables. No further possibility to calculate with the values and to create other flexible pivot reports.

But the data model is open for a migration!

The proposed MetabolAS Tool web flow

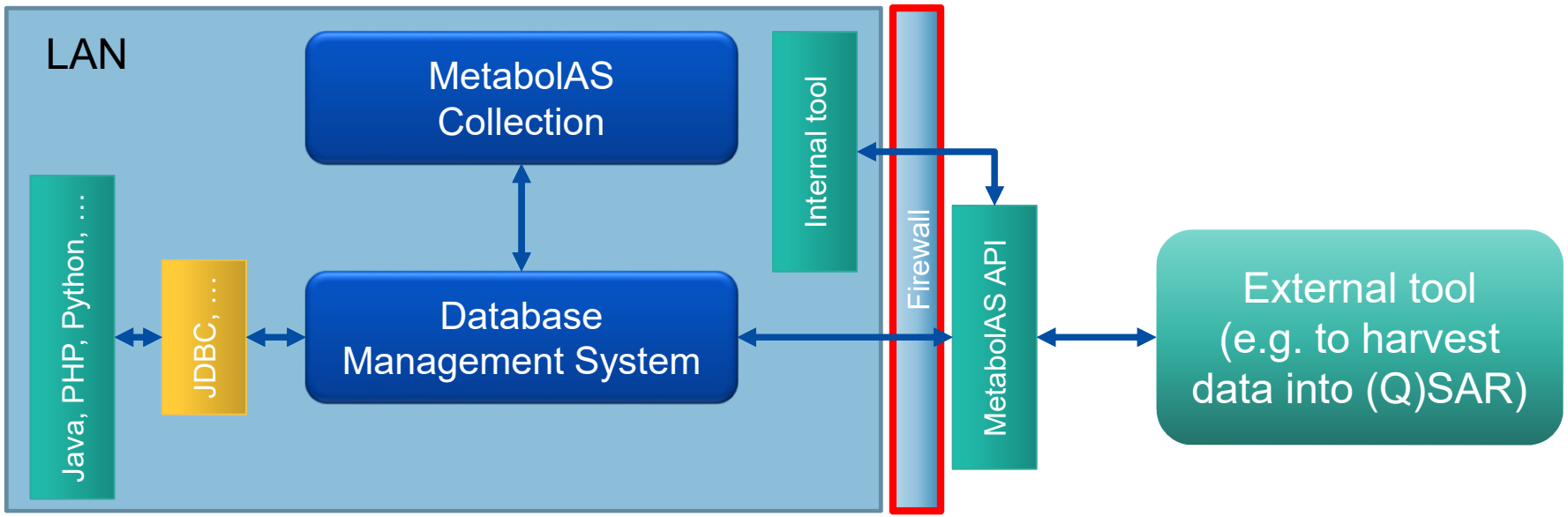


The concept of different start options into the MetabolAS Tool



The API concept of the MetabolAS Database / Tool

- The database management support common used access methods (e.g. JDBC, ODBC) and programming languages (e.g. Java, PHP, Python)
- The “API” provides functions for reading and writing data from / into a “MetabolAS collection” on element and record level.
  The “API” should provide a data interface that (Q)SAR models can “harvest” validated data sets.

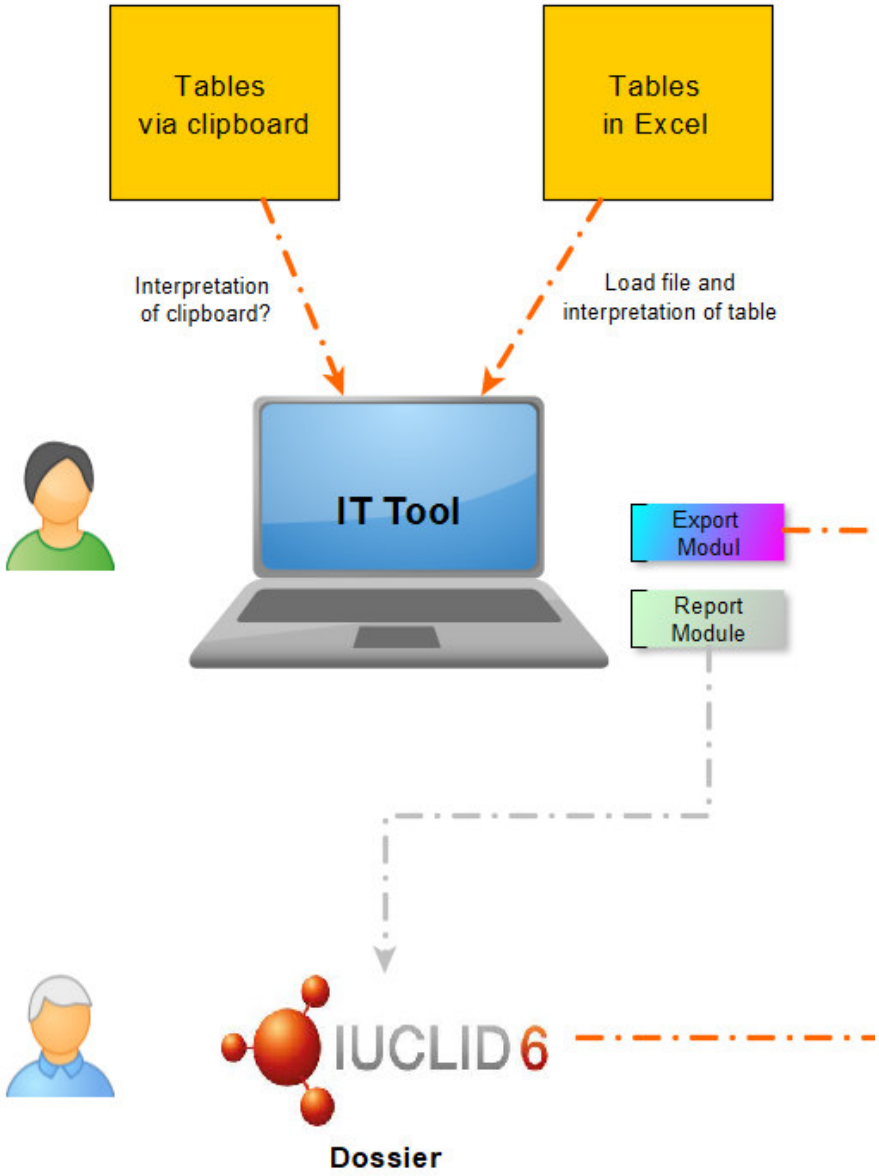


Switch to the voting system now
regarding create the “aggregated raw data” sets

Are there any questions? Please use the hand raise in the TEAMS environment.
For statements you could use also the TEAMS chat. The chat will be recorded. **So no idea is lost.**

Is it realistic / purposeful to plan a data transfer from a LIMS systems into a local MetabolAS collection via APIs and / or common used access methods ?

How realistic is the following scenario?



What happens if the required information is not available in the LIMS, but is part of the text of a study report?

Switch to the voting system now regarding create the “aggregated raw data” sets

Are there any questions? Please use the hand raise in the TEAMS environment.
For statements you could use also the TEAMS chat. The chat will be recorded. **So no idea is lost.**

Questions to applicants / laboratories:

How often do you have aggregated raw data in a simple electronical format (outside of LIMS)?

Switch to the voting system now
regarding create the “aggregated raw data” sets

Are there any questions? Please use the hand raise in the TEAMS environment.
For statements you could use also the TEAMS chat. The chat will be recorded. **So no idea is lost.**

Questions to all:

Would you say a bulk import of substances / dosegroups or result tables is possible most of the time?

If a bulk import of files is possible, is an interpretation of tables via the clipboard still necessary?

Content

Main messages and open questions aspects regarding

the Database,

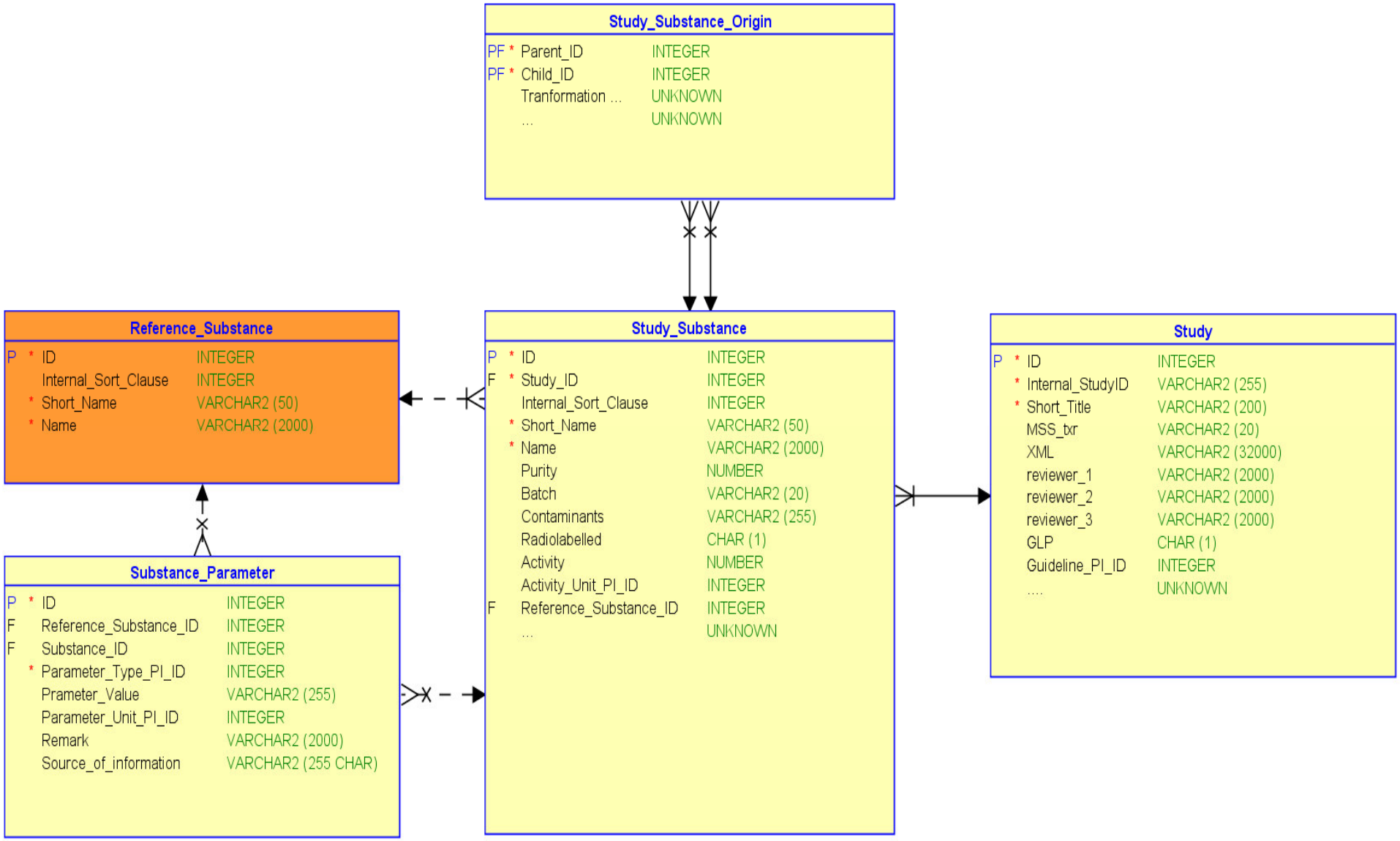
the Tool,

Data Management in the Reference Collection

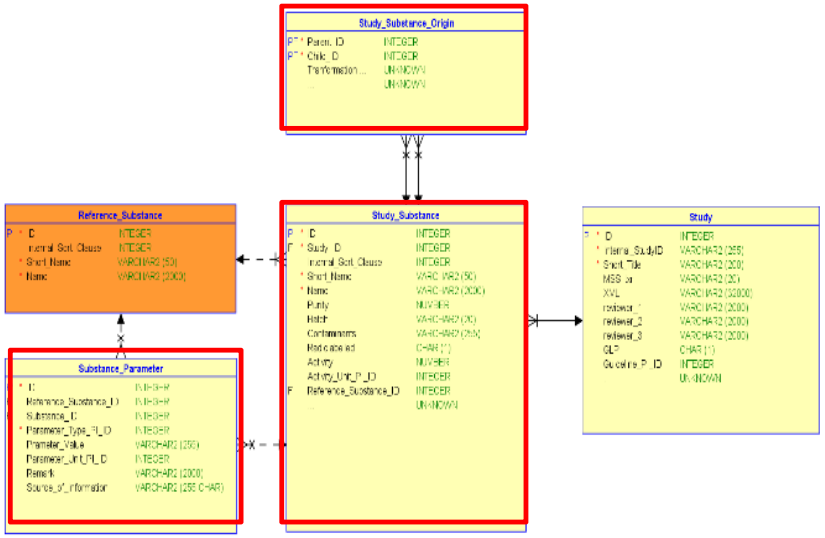
Management of (Q)SAR results

Migration

Detail: The substance model

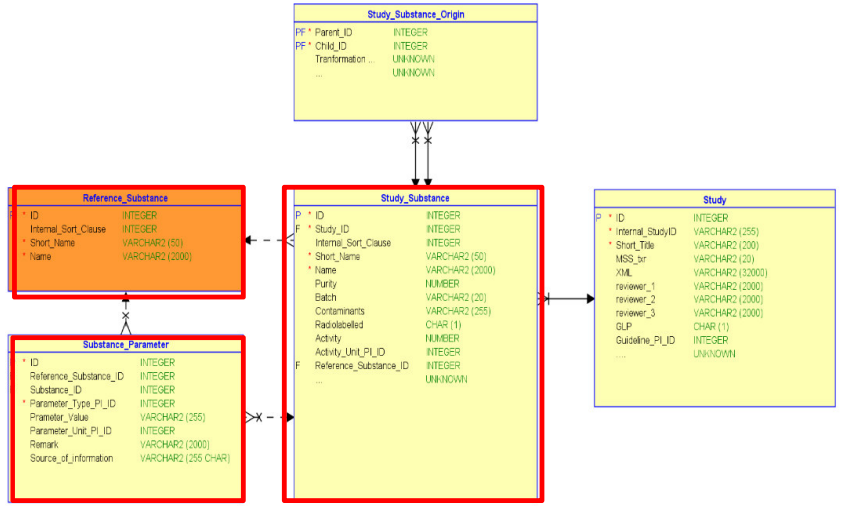


Detail: The substance model



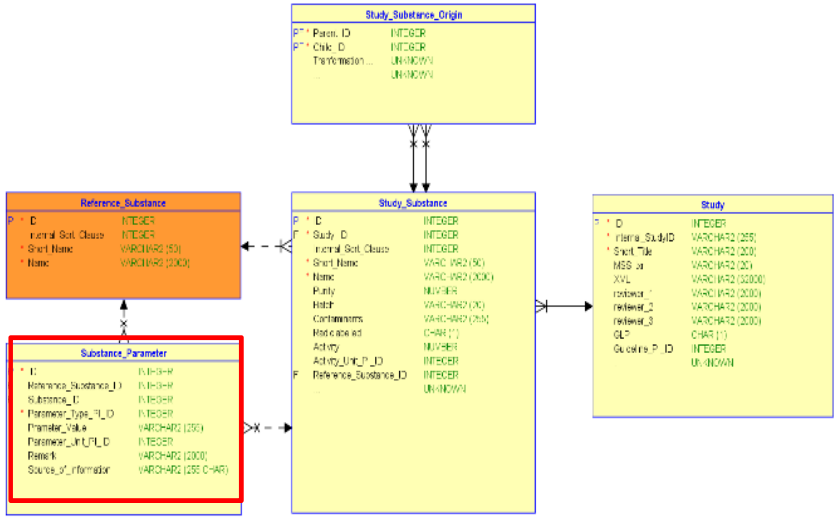
- All study substances are stored in context of the **study**:
  - Inactive parent substance,
  - Radiolabelled substance(s),
  - Metabolites
- The “Metabolic Pathway” is stored in the table **study substance origin**. Each record of this table has to have references between the parent and the child.
- Each of the substances could have a set of **substances parameter**.
- All of these substance information could be stored independently!

Detail: The substance model



- A study substance should have no or one reference to a **reference substance** of an internal inventory.
- The inventory of **reference substances** should be imported initially.
- Each of the reference substances has a minimal set of **substances parameter.**
- The **reference substance** is always the non radiolabelled substance.
- Unknown metabolites have no reference to a **reference substance**.
- Reference substances should be managed in a specialized administrator module

Detail: Substances parameter



The substance parameter types are defined by a picklist.
As parameter are foreseen

- CAS-Name,
- IUPAC-Name,
- Synonyms,
- SMILES,
- extended SMILES (CXSMILES),
- InChI,
- CAS-No,
- EINCS,
- IUCLID-Reference Substance UUID,
- $logP_{ow}$
- …

Switch to the voting system now regarding the need of substance parameter

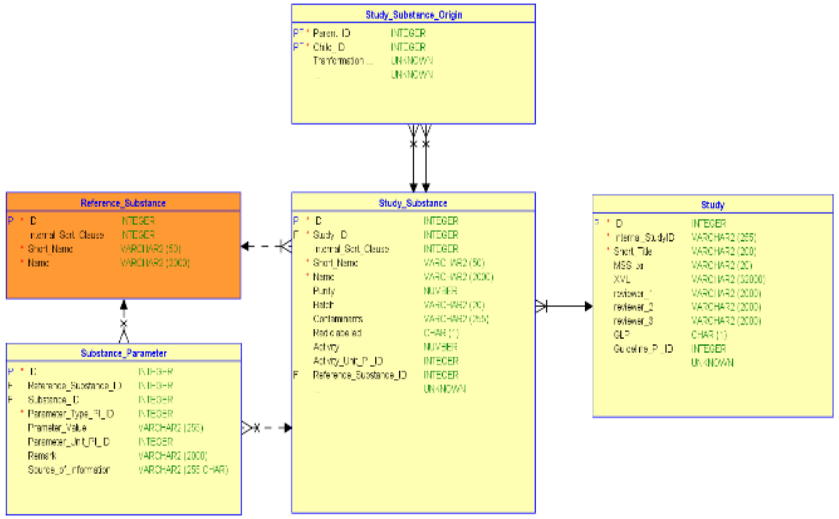
Are there any questions? Please use the hand raise in the TEAMS environment.
For statements you could use also the TEAMS chat. The chat will be recorded. **So no idea is lost.**

The $logP_{ow}$ (partition coefficient n-octanol/water) is often used in the risk assessment. This value could be measured or calculated from the structure.

How often do you measure this partition coefficient / do you get measured values from the applicants?

Is there a need to store additional phys-chem properties (beside logPow) or toxicological data as structured meta data needed in evaluation process?

Detail: The substance administration



- Before inserting / importing new reference substances a duplicate check has to be made.
  - If the reference substances already exists in the target MetaboIAS collection, all internal study references should be modified to reference to the existing row.
- Mechanisms should be implemented if reference substance duplicates are identified later. ➔
  “Merge (or pool) two substances and their references, because these substances are duplicates”.
- How to proceed with metadata of at first unknown metabolites, for which the structure was identified later.

Switch to the voting system now regarding life cycle of substance information

Are there any questions? Please use the hand raise in the TEAMS environment.
For statements you could use also the TEAMS chat. The chat will be recorded. **So no idea is lost.**

How often do you get later additional structure (or name) information on initially "unknown" metabolites reported in the original GLP report?

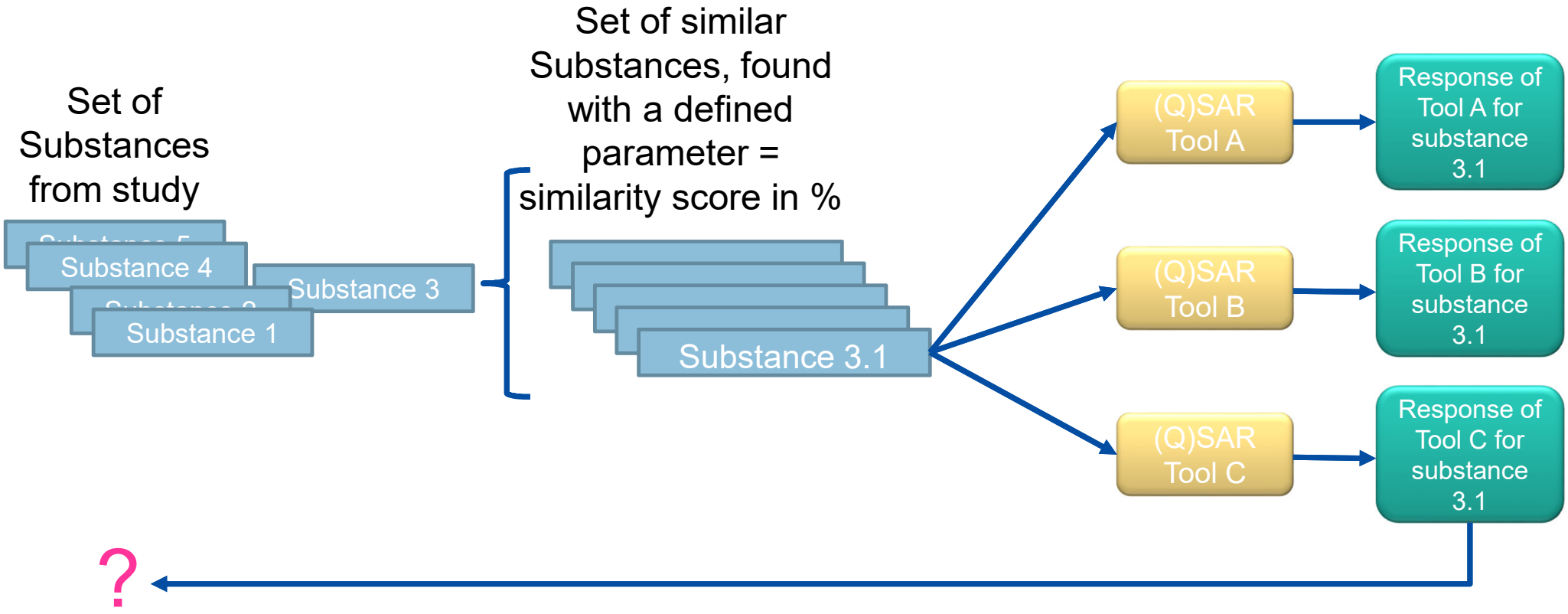
Content

Main messages and open questions aspects regarding

- the Database,
- the Tool,
- Data Management in the Reference Collection
- **Management of (Q)SAR results**
- Migration

The Problem

Evaluators have to loop over the “Set of Substances” and to start (Q)SAR analysis in different (Q)SAR Tools for each study substance.



Switch to the voting system now regarding management of (Q)SAR results

Are there any questions? Please use the hand raise in the TEAMS environment.
For statements you could use also the TEAMS chat. The chat will be recorded. **So no idea is lost.**

Should the MetabolAS Tool be able to manage (Q)SAR responses for each substance from different (Q)SAR Tools according the ECHA guide?

Content

Main messages and open questions aspects regarding

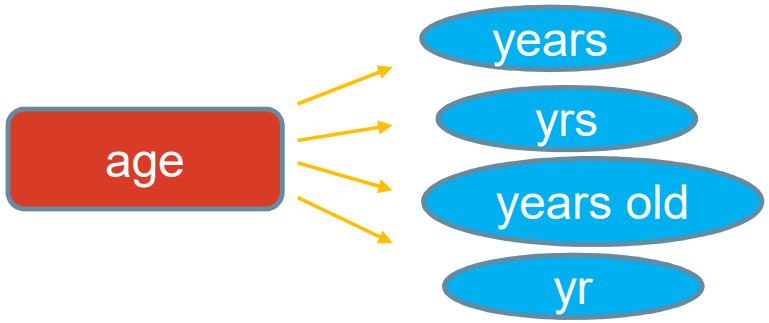
- the Database,
- the Tool,
- Data Management in the Reference Collection
- Management of (Q)SAR results
- Migration

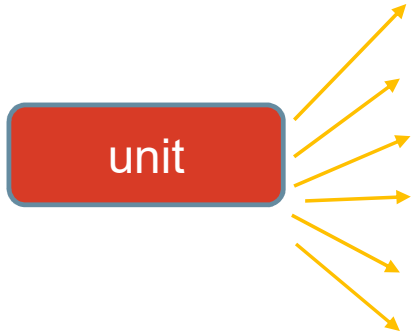
Migration should start from ….?

- BfR prefer the XML files, because the data entry was done by the MSS / DER composer. Nobody has done the data input into Metapath without the MSS /DER Composer.
- The imports into Metapath were incomplete a long time.
- Metapath hadn’t generated an additional “benefit” for the data sets while the import process.
- It is easier to analyse XML files than to understand the internal logic of Metapath database

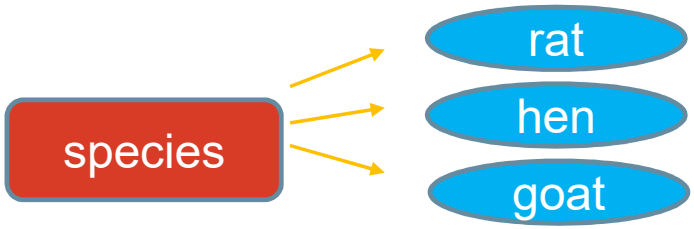
Tasks for the migration (I):

XML files:	341
MSSLivestock:	61
MSS Plants:	175
DER:	105





mCi/mmol	59
mCi/mmole	18
mCi/mMol	5
mCi/mmo	2
mCi/mM	2
mCi/mMole	2
mCi/ mmol	1
mCI/mmole	1



<testAnimals><species>**Goat**</species><strain>Sprague-Dawley **male rats** were used in the study</strain>

Conversion of free text fields into picklist elements

Minimize content related errors for depending picklist elements

Tasks for the migration (II):

Separation of the concatenated string separated by comma
e.g. relation between parents / child's or the test groups

```
</appendix2row>
<appendix2row>
  <id>6</id>
  <code>IT-6</code>
  <name>IT-6</name>
  <smiles>OCC1=C(Br)c2cc(O)c(F)cc2N1S(=O)(=O)C1=NNC=N1</smiles>
  <parents>3,5</parents>
  <expertise type="none" toleranceExpression="false" residueOfConcern="false"/>
</appendix2row>
<appendix2row>
  ...
</appendix2row>
<table2row>
  <testgroup>1b,2b,3b,4b,5b,6b,7b,8b</testgroup>
  <matrix>Feces</matrix>
  <sampletime>0-96</sampletime>
  <majormethod>Reflux with Acetonitrile and Water</majormethod>
  <conjugateanalysis>16 reference compounds</conjugateanalysis>
  <separation>TLC</separation>
  <detection>HPLC, LC-MS, and MS/MS</detection>
  <remarks/>
</table2row>
```

Switch to the voting system now regarding the migration concept

Are there any questions? Please use the hand raise in the TEAMS environment.
For statements you could use also the TEAMS chat. The chat will be recorded. **So no idea is lost.**

Should the migration start from the XML file?

GERMAN FEDERAL INSTITUTE FOR RISK ASSESSMENT

BfR
Bundesinstitut für Risikobewertung

Thank you for your attention

Stephan Worseck

German Federal Institute for Risk Assessment
Max-Dohrn-Straße 8-10 ● 10589 Berlin, GERMANY
Phone +49 30 - 184 12 - 0 ● Fax +49 30 - 184 12 – 99 0 99
bfr@bfr.bund.de ● www.bfr.bund.de/en