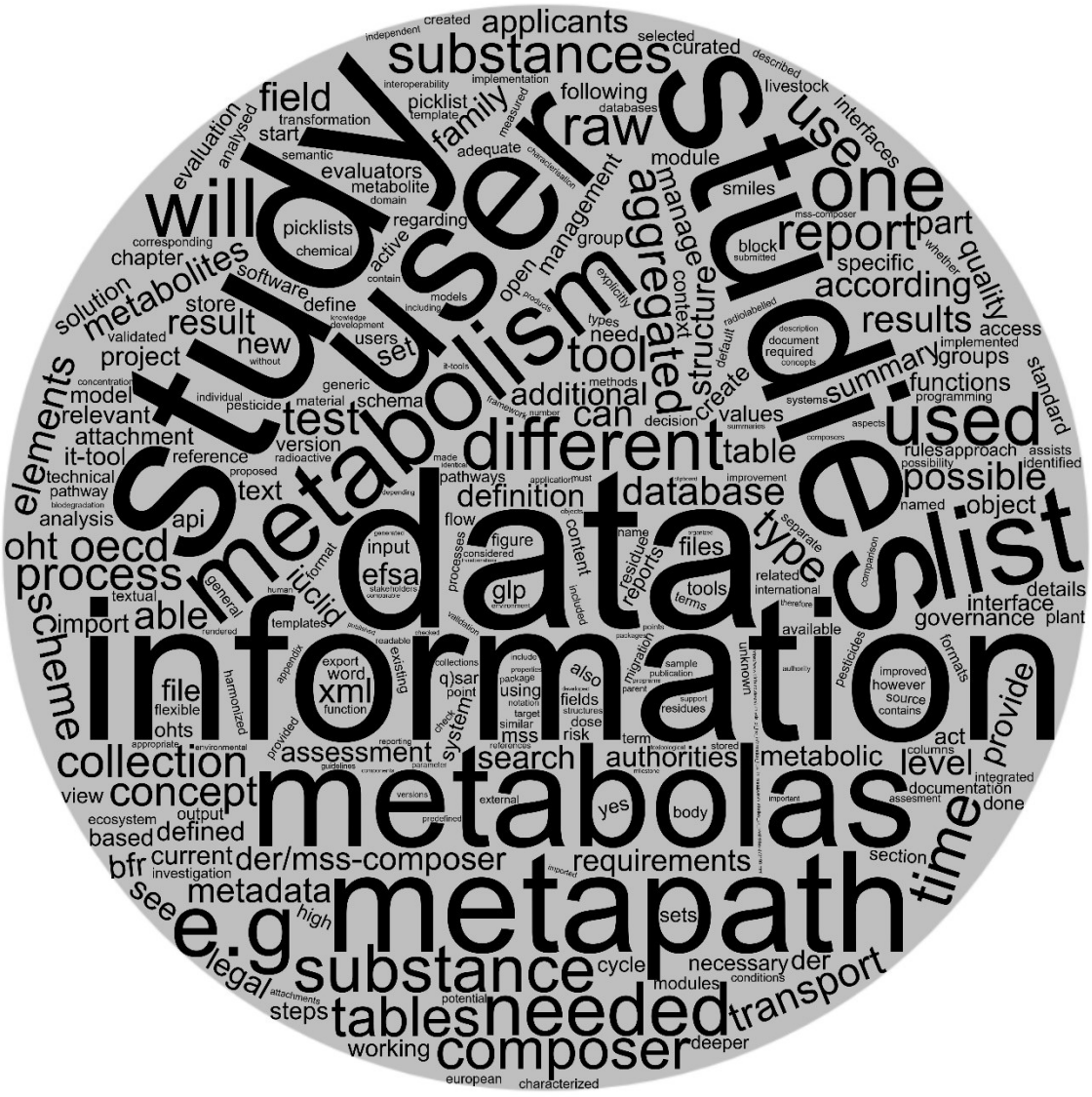
20.12.2021

Analysis of the information flow of pesticide related metabolism studies

Proposals for improvement

Stephan Worseck, Falko Frenzel, Tobias Opialla, Thomas Kuhl, Glenn Lurman, Alina Burchardt

German Federal Institute for Risk Assessment
Department Pesticides Safety



Generated by Wortwolken.com from the text of this report.

1 Summary

The *EFSA* had started a process for improving the information flow from pesticide related metabolism studies to build up a broader information database for metabolism pathways of pesticides in 2021.

The *BfR* has undertaken the following analyses/actions:

- status quo analysis utilizing a survey (*BfR 2021*)
- process analysis of the current processes within the European context
- analysis of the OHT58 and *DER* scheme descriptions
- analysis of the *MetaPath* user functions
- analysis of the current database implementation
- development of proposals for improvement
- stakeholder consultation

The *BfR* applied a holistic approach to the analysis of the information flow and for the development of the proposals for improvement. The aim is to consider all steps of the information flows. It starts with the data generator (e.g. laboratory) and entails applicants and authorities that compile the different direct outputs: the assessment reports, the published *Metadata* and the quality assured reference collection of metabolism studies that are the basis to create respective models (see Figure 1). Furthermore, there are efforts by other data consumers to harvest data of this quality assured reference collection in their systems, e.g. the *OECD* (Q)SAR-Toolbox (chapter 9.7).

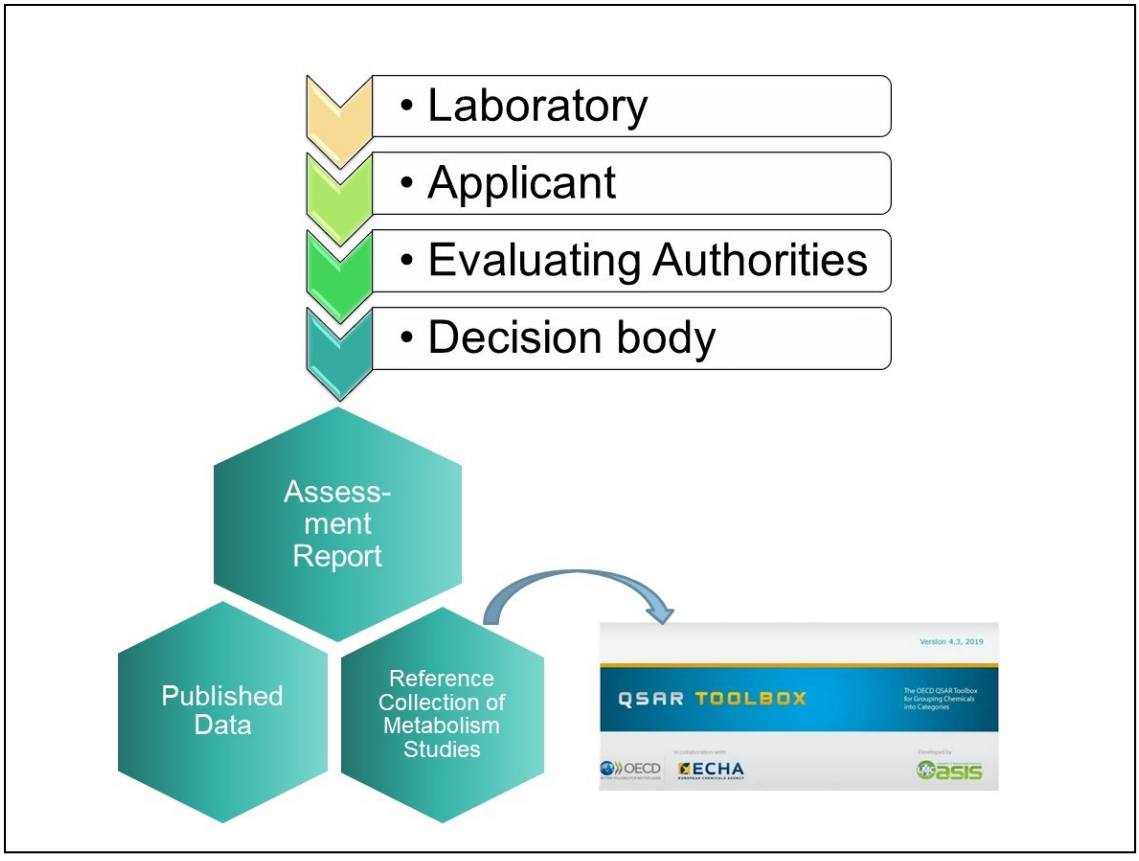


Figure 1: Direct “end products” of the information flow and the (Q)SAR-Toolbox as an example of a “harvesting” systems

1.1 Executive summary

BfR suggested to *EFSA* to initiate the decision making process for the improvement process. A set of high-level decisions are needed. All relevant stakeholders (e.g. applicants, authorites, *OECD*, (Q)SAR modeller) should be included in this process.

D 1.1-1: Guided by *EFSA* and in compliance with the Metapath User Group (*MUG*), decisions should be taken about the necessity, priority and time horizon to build up the *MetaPath II Ecosystem* with the needed components regarding the information flow of metabolism studies.
➔ Yes / No
➔ short-term / mid-term/ long-term

D 1.1-2: A bug review process prioritises the issues that need to be fixed in *MetaPath* and in the programs of the *DER/MSS-Composer family* to ensure the needed functionality in the interim period until the new *MetaPath II Tool* could go into production.
➔ Priority list

D 1.1-3: Within a first project planning phase decision about the necessity, priority and time horizon to build up an international curated reference collection of metabolism studies as part of the *MetaPath II Ecosystem* should be taken.
➔ Yes / No
➔ short-term / mid-term/ long-term

D 1.1-4: By the lead of *EFSA*, a decision on *OECD* level about the ***Governace body*** of the *MetaPath II Ecosystem* should be initiated.
➔ *OECD* as the ***Governace body*** / ***Governace body*** outside of the *OECD*

D 1.1-5: Decision whether IT components of the ecosystem could be part of an "*Open Source Project*" in which the interested parties contribute to the community.
➔ Yes / No

D 1.1-6: If *OECD* takes on the role of the ***Governace body***, *EFSA* role within this improvement process should be precisely defined. An improvement process on basis of the experiences and algorithms of *MetaPath* but on a new technological level could be initiated by *EFSA*. The *MetaPath II Ecosystem* should be accessible to the international community.
If *OECD* takes not the role of the ***Governace body***, *EFSA* would organise the project *MetaPath II Ecosystem* in an adequate manner alone.

Based on this report and the political decisions taken it is recommended to have a cost estimate carried out. This would be substantial to ensure an adequate funding in an adequate period.
➔ Required project funding is organised

Considering all this, an adequate project structure should be established. For this purpose, *BfR* has developed proposals for a possible project plan, which have been submitted seperately to *EFSA*.

1.2 High-level statements

The following high-level statements should be considered for further decisions.

1.2.1 Need for harmonisation

O 1.2-1: There is a high degree of overlap of the user requirements of applicants and authorities, as both stakeholder groups work according to the same regulatory framework. The data requirements and the corresponding assessment guidelines are the driver for the semantic content of the required information flow of pesticide related metabolism studies.

O 1.2-2: In the long-term, the development and the availability of an information database for pesticide metabolism pathways will be beneficial for both, the applicants and the authorities. For this reason, both stakeholder groups should be equally interested in a generally improved IT-support.

O 1.2-3: The high degree of overlap of said interests could be the basis to start a substantial improvement process regarding the flow of information on metabolism studies.

P 1.2-4: Further harmonisation is required. The *OECD* as a major institution for global harmonisation would be the first address to lead such a project. *EFSA*, as a project initiator, could submit an official request to the *OECD* secretariat to create such an *OECD* project.

P 1.2-5: According to a vote of the *MUG* the working title could be “MetaPath II” to indicate the continuity in the approaches on the one hand and on the other hand, the radical changes in the used techniques. According to the *MUG* vote, this project should be put on the same organisational level as the *OECD* (Q)SAR-Toolbox.

P 1.2-6: Following the proposed scenario, *OECD* would have the role of the *Governace body* in this project.

1.2.2 Need for a generalised concept of the term metabolism

O 1.2-7: There are 18 *OECD* Harmonised Templates where results with radioactive labelled test material could be summarised.

O 1.2-8: The analysis showed that the lack of an overall *OECD* term definition level for the technical guidance’s provides freedoms, but makes comparisons between similar *Study Types* thus difficult.

P 1.2-9: The *BfR* proposes a “generic” concept of the term metabolism, which is suitable to build up a generalised scheme to transport *Aggregated raw data* from metabolism studies, which covers all studies where radioactive labelled test material are used according the Test Guidelines. No distinction is made between biotic or abiotic processes causing these transformations.

1.2.3 Need for an ecosystem of components

P 1.2-10: The *BfR* proposes to build up an ecosystem of all needed components where each part of this ecosystem could be used by applicants and authorities because both stakeholders need tools to connect with the same interoperable functionality (*Governance concept*, user forum, picklists and picklist elements, scheme definition, IT-Tool, API, reference collection see chapter 9.2).

P 1.2-11: The *Governance concept* should be open for metabolism studies from all areas independent of the legal context (pesticides, biocides, pharmaceuticals, chemicals).

P 1.2-12: The *Governance concept* should contain rules on how to deal with competing interests.

1.2.4 Need for an appropriate transport concept of metabolism study metadata

O 1.2-13: *IUCLID* could be used in two different ways to transport the required *Metadata*. However, these transport concepts differ very significantly in the way they are implemented. These differences have consequences for the data interfaces, data collection, data presentation, data usage, supporting tools required, the publication process and ultimately in the resulting overall effort (see Table 14 for an overview).

P 1.2-14: A comparison of the efforts of the identified transport concepts had shown, that the transport concept of a new *OECD Attachment Type* for metabolism studies is the preferable solution. The *BfR* proposes expanding the *OECD* house architecture with the new category *OECD Attachment Type* (see chapter 9.8.3.2).

O 1.2-15: If the *Aggregated raw data* from metabolism studies are transported as an attachment there is the need for one generic scheme which covers all studies where radioactive labelled test material is used according to the Test Guidelines. There are no consequences in the user front end of *IUCLID*. It is possible to make a stepwise approach and to include one *Study Type* after the other. The XML-scheme should contain information parts, which are stable over time. The variability can be customized using picklists.

O 1.2-16: The complete human readable information of the metabolism study will be provided on the attachment level and on the study summary record level of the dossiers. According to the transparency regulation both should be published depending on the confidentiality rules. Because the *Aggregated raw data* contain the same semantic information there is no direct need to publish the machine-readable (non-human readable) data on application level. However, on the other side: Applicants have already the possibility to attach a full and a sanitized version. Therefore, it could not be a problem to solve confidentiality aspects.

O 1.2-17: If the *Aggregated raw data* from metabolism studies is transported and integrated in the *OHTs* there is the need for one generic *OECD* domain type which should cover all studies where radioactive labelled test material could be used according the Test Guidelines. 18 *OHTs* need to be updated in *IUCLID*.
All *Aggregated raw data* will be shown in the user front end.
From *BfR*s point of view this solution is feasible but has several disadvantages (see chapter 9.8.3.1).

P 1.2-18: *EFSA*, *ECHA* and *OECD* should examine the argumentation for the two transport concepts proposed and build an appropriate decision-making and organisational concept.

1.2.5 Need for a curated reference collection of metabolism study metadata

O 1.2-19: The *EFSA* decision to build up a curated reference collection of metabolism study *Metadata* and to update it after submitting new studies is supported by *BfR* (*BfR 2020*). This represents a significant step towards the goal of avoiding further tests on vertebrate animals as well as reducing uncertainty in human exposure assessments without lowering the level of protection.

O 1.2-20: The generic concept proposed by the *BfR* is intended to enable the curated reference collection of metabolism studies to be opened up for all types of metabolism studies that have not been considered so far (compare P 1.2-9). This approach is more open to the scientific community, increases transparency and could help reduce uncertainties in environmental risk assessment.

O 1.2-21: The current process organisation and the IT-support of the information flow from pesticide related metabolism studies is not optimal. However, the basic idea and the basic structure of this information flow must be retained.

P 1.2-22: The *BfR* proposes making a clear cut between the transport of the metabolism study *Metadata* and building up and maintaining a curated reference collection of metabolism study *Metadata*.

P 1.2-23: The *BfR* proposes embedding the required curated reference collection of metabolism study *Metadata* in an ecosystem (target system) with all necessary tools, definitions, master data and an adequate *Governance concept* (see chapter 8). These components could be used by applicants and authorities because both require the same functionality. One element of this ecosystem is an IT-Tool with the working title *MetaPath II Tool*.

P 1.2-24: The curated reference collection of metabolism study *Metadata* should only contain data that have their origin in the metabolism study itself. However, the *MetaPath II Tool* should be able to assist the assessment process by managing secondary *Metadata* from other sources. A decision is needed, whether these data should or should not be transferred into the curated reference collection of metabolism study *Metadata*.

R 1.2-25: Clear rules need to be defined under which conditions applicants should extract copies of the *Aggregated raw data* from the curated reference collection for repeated submission to the authorities. When organising such an information loop between applicant, authority and curated reference collection, there is an acute risk of losing data.

O 1.2-26: From *BfR* point of view, *IUCLID* was designed as a dossier transport system for applicants. *IUCLID* is not suitable to be the database management system for the curated reference collection of metabolism study *Metadata*.

R 1.2-27: The current IT-Tools (*MetaPath* and *DER/MSS-Composer family*) should be used until the new target system and an adequate migration tool for the current collections of metabolism study *Metadata* is available.
The needed migration is independent of choosing the *Transport concepts for aggregated raw data of metabolism studies*.

1.2.6 Need for an improved data management and data handling procecure

O 1.2-28: The content related concept of *MetaPath* is up-to-date and useful for the evaluation steps.

P 1.2-29: It is estimated that 1/7th of the start-up effort will be required to maintain this software. *MetaPath*'s underlying database technology and the front end are outdated, the amount of separate programs of the *DER/MSS-Composer family* need to be adapted, the number of different "custom versions" of *MetaPath* and the *DER/MSS-Composer family* and also the number of open user requests in relation to the user functions that do not require any change should be arguments for a radical change.

O 1.2-30: From *BfR*s point of view, the time has come to move the valuable concept and information contained in *MetaPath* to a new technological level. Otherwise, there is a risk of permanent dissatisfaction of the users.

O 1.2-31: The approach of P 1.2-29 is a chance to move away from the current strategy of "individual MSS-Composer" programs for each metabolism *Study Types* to a single, harmonised approach.

O 1.2-32: The current split of data input, data management and data use between the *DER/MSS-Composer family* and *MetaPath* should be discontinued. A continuation of this artificial separation, that a data input would now only make sense via the *OECD* template is not useful.

1.2.7 Need for improved reports

O 1.2-33: Each of the "individual "MSS-Composer" programs has its own report module (RENDER module). The list of bugs and proposals for improvement is long. From *BfR*s point of view, the time has come to check to current reporting concept used in *MetaPath*.

P 1.2-34: The best practice/algorithm of the current report should be implemented in a new reporting module of the new *MetaPath II Tool*. The quality and the reusability of formatted text blocks will become the indicator of user acceptance.

1.2.8 Need to organise the improvement process in an interim period

W 1.2-35: The missing project structure for the current *MetaPath* makes it difficult for stakeholders to organise the interim period well. There are no processes defined on how to decide on necessary improvements, that will result in a change request to *LMC* but which will have an impact for all users of *MetaPath*.

P 1.2-36: It is proposed that the *MUG* should be the forum for the interim period that provides substantial technical support for the funding stakeholders before they are commissioning change requests.

1.2.9 Need for an improved authority process

O 1.2-37: The current *EFSA* process (see chapter 5.3) described in the document "Reporting structured results of metabolism studies on rats, plants and livestock" (*EFSA 2021*) needs improvements, too. In particular, the *EFSA* processes to manage the historical XML files and to manage different databases (MTB) are work-intensive and carry a high risk of data inconsistencies between different versions.

P 1.2-38: The role of the XML files in the process is pending revision. In future, the XML files are to be only considered as temporary transport containers. The *Metadata* of the XML files need to be 100% importable into a *MetaPath II Collection* and also exportable from there.

P 1.2-39: International *Authorities MetaPath II collections* are required. The optimal solution would be to build up only **one** worldwide *Authorities MetaPath II collection*but other scenarios are possible.

P 1.2-40: The improved authority process should guarantee the principle:
“**One substance – one assessment**”

2 Table of content

1 Summary 2
1.1 Executive summary 3
1.2 High-level statements 4
1.2.1 Need for harmonisation 4
1.2.2 Need for a generalised concept of the term metabolism 4
1.2.3 Need for an ecosystem of components 5
1.2.4 Need for an appropriate transport concept of metabolism study metadata 5
1.2.5 Need for a curated reference collection of metabolism study metadata 6
1.2.6 Need for an improved data management and data handling procecure 7
1.2.7 Need for improved reports 7
1.2.8 Need to organise the improvement process in an interim period 7
1.2.9 Need for an improved authority process 7
2 Table of content 9
3 Abbreviations 12
4 Document structure, used nomenclature and methodology 14
5 Background 16
5.1 Context of this report 16
5.2 Previous efforts 16
5.3 The current EFSA process 17
6 Objectives for further development 18
7 Terms, user requirements and concepts 22
7.1 IT related terms 22
7.1.1 Chemical structure notation 22
7.1.2 Interoperability 22
7.1.3 Information package 23
7.1.4 Metadata 23
7.1.5 Picklist 23
7.1.6 Object type 23
7.1.7 Validation 23
7.2 Study related terms 24
7.2.1 Substance 24
7.2.2 Metabolism study 26
7.2.3 Object of investigation 27
7.2.4 Balance room 27
7.2.5 Balance of activity 27
7.2.6 Application 27
7.2.7 Dosing scheme 28
7.2.8 Transformation process 28
7.2.9 Test substance 28
7.2.10 List of metabolites 28
7.2.11 Metabolic pathway 28
7.2.12 Other metabolism related terms 29
7.3 Assessment related terms 30
7.3.1 Framework conditions 30
7.3.2 Data gap filling 34
7.3.3 Consider metabolites in the dietary exposure 35
7.3.4 Metabolites considered in toxicology 36
7.3.5 Consider metabolites in the residue definition 38

7.4	Applicants´ information packages	38
7.4.1	GLP study raw data	39
7.4.2	GLP study report	39
7.4.3	Aggregated raw data	43
7.4.4	Applicants study summary	44
7.4.5	Study summary metadata	45
7.4.6	Predefined study summary tables	46
7.4.7	Endpoint summaries	46
7.4.8	Dossier	46
7.5	Authorities' information packages	46
7.5.1	Authority study summary	47
7.5.2	Assessment report	47
8	**Detailed analyses**	**49**
8.1	OECD harmonised templates for metabolism studies	49
8.1.1	Owner	49
8.1.2	Information	49
8.1.3	Functionalities	50
8.1.4	Life cycle and process	50
8.1.5	Publication and documentation	52
8.1.6	Interoperability / output	52
8.2	Comparison of OHT58 „Basic toxicokinetics“ and DER Composer	52
8.2.1	Semantic aspects	53
8.2.2	Weaknesses of the two templates	53
8.2.3	Aspects of format and supporting tools	53
8.3	MetaPath	56
8.3.1	Owner	56
8.3.2	Contained Information	56
8.3.3	Functionalities	58
8.3.4	Database implementation	59
8.3.5	Life cycle and processes	63
8.3.6	Publication and documentation	63
8.3.7	Interoperability / output	64
8.4	DER/MSS-Composer family	65
8.4.1	Owner	65
8.4.2	Contained information	65
8.4.3	Functionalities	65
8.4.4	Life cycle and process	68
8.4.5	Publication and documentation	69
8.4.6	Interoperability / output	69
9	**Solution approaches**	**70**
9.1	Disclaimer	70
9.2	MetaPath II Ecosystem	70
9.2.1	Governance concept	70
9.2.2	User forum	71
9.2.3	Picklists and picklist elements	71
9.2.4	Scheme definition	72
9.2.5	MetaPath II Tool	73
9.2.6	MetaPath II Tool API	76
9.2.7	Authorities MetaPath II collection	77
9.3	Core structure for a relational database model	78
9.3.1	Disclaimer	78
9.3.2	Proposed core structure	78

9.4 User requirements for the MetaPath II Tool ... 85
9.4.1 User communication ... 87
9.4.2 General requirements ... 87
9.4.3 Create new study ... 92
9.4.4 Import ... 92
9.4.5 Search modules ... 93
9.4.6 Result list of substances ... 94
9.4.7 Substance details ... 95
9.4.8 Substance edit ... 96
9.4.9 Interoperability to other systems in the user front end ... 96
9.4.10 Visualisation ... 97
9.4.11 Study details ... 98
9.4.12 Study edit ... 102
9.4.13 Compare ... 103
9.4.14 Management of attachments ... 103
9.4.15 User set management module ... 103
9.4.16 Report ... 107
9.4.17 Documentation ... 108
9.5 System requirements for the MetaPath II Tool ... 109
9.5.1 Management ... 109
9.5.2 Import / export / validation ... 110
9.5.3 Assist the transport step via IUCLID ... 110
9.6 Internal stakeholder MetaPath II-instances ... 111
9.7 Usage of information of metabolism studies in (Q)SAR ... 111
9.8 Transport concepts for aggregated raw data of metabolism studies ... 111
9.8.1 Analogy of transport concepts ... 111
9.8.2 Requirements to transport aggregated raw datafor metabolism studies ... 113
9.8.3 Identified concepts to transport aggregated raw data ... 117
10 Discussions in the MetaPath User Group ... 121
10.1 Agreement to manage the transition period ... 121
11 Appendix ... 122
11.1 Bibliography ... 122
11.2 Attachments on file level ... 124
11.3 List of Harmonised Templates where radioactive labelled test material could be used ... 125
11.4 List of weak points identified in the survey ... 128
11.5 Comparison of possible solutions by the provided functions ... 129
11.6 The compromise between distinct data fields and free text fields ... 133
11.7 Standard tables for the presentation of metabolism studies ... 134

3 Abbreviations

Short	**Meaning**
(Q)SAR	(**Q**uantitative) **S**tructure **A**ctivity **R**elationship
AD	**A**dministered **D**ose
ADME	**A**bsorption, **D**istribution, **M**etabolism and **E**xcretion
ANSES	The French Agency for Food, Environmental and Occupational Health & Safety
API	**A**pplication **P**rogrammable **I**nterface
BfR	**B**undesinstitut **f**ür **R**isikobewertung
CR	**C**urrent high level user **R**equirement
CXSMILES	**C**hemAxon e**x**tended **SMILES**
DAR	**D**raft **A**ssessment **R**eport
DER	Metabolism Study Summary according to the **D**ata **E**valuation **R**ecord Templates used in USA - Canada.
DER-Composer	Software to store Metabolism Study Summaries in a defined XML schema; Copyright by OASIS LMC
ECHA	**E**uropean **C**hemicals **A**gency
EFSA	**E**uropean **F**ood **S**afety **A**uthority
GLP	**G**ood **L**aboratory **P**ractice
InChI	**In**ternational **Ch**emical **I**dentifier
IUCLID	**I**nternational **U**niform **C**hemica**L** **I**nformation **D**atabase
IUPAC	**I**nternational **U**nion of **P**ure and **A**pplied **C**hemistry
JDBC	**J**ava **D**ata**B**ase **C**onnectivity
LMC	**L**aboratory of **M**athematical **C**hemistry
MetaPath	Software and knowledge base for the purpose of archiving, sharing and analysing experimental data on metabolism and metabolic pathways; Copyright by OASIS LMC (Metapath doesn't contain
MSS	**M**etabolism **S**tudy **S**ummary
MSS-Composer	Software family to store Metabolism Study Summaries in a defined XML schema; Copyright by OASIS LMC
MUG	**M**etaPath **U**ser **G**roup
OECD	**O**rganisation for **E**conomic **C**o-operation and **D**evelopment
OECD MUG	**M**etaPath **U**ser **G**roup of the OECD
OHT	**O**ECD **H**armonised **T**emplates
QA	**Q**uality **A**ssurance

Short	**Meaning**
R	User **R**equirement
Rich-Text	Text according the Rich Text Format (RTF)
RMS / EMS	**R**apporteur **M**ember **S**tate / **E**valuating **M**ember **S**tates
Ruedis	**Rü**ckstands**I**nformations**S**ystem
SMARTS	**SM**ILES **AR**bitrary **T**arget **S**pecification
SMILES	**S**implified **M**olecular **I**nput **L**ine **E**ntry **S**ystem
TRR	**T**otal **R**adioactive **R**esidue
USEPA	United States **E**nvironmental **P**rotection **A**gency

4 Document structure, used nomenclature and methodology

The task of the current report is to give proposals for short and long-term improvements of the information flow to reduce the identified weak points. The information basis to compile this report are:

- the *BfR* 's assessment activities within the processes of consumer health protection of plant protection products and biocides including the experience in the use of (Q)SAR methods,
- the experiences around the *EFSA MetaPath* project,
- the experiences from the engagement in EU and *OECD* working groups,
- the results from the stakeholder survey and the intensive feedback and discussions with stakeholders in the *MetaPath* User Group and
- the high similarity with the *Ruedis* project.

However, *EFSA* had made available all that could be made available but some information are not publicly available and remain less accessible. Therefore, the *BfR* got no detailed information regarding data interfaces between

- *DER/MSS-Composer family* and *MetaPath*,
- *MetaPath* and *OECD* (Q)SAR Toolbox,
- *IUCLID* and *OECD* (Q)SAR Toolbox.

Furthermore, this report does not represent a literature review.

Chapter 4 contains a description of the relevant terms and concepts regarding the information flow of metabolism studies.

Italicized terms in quotation are cross-references to the respective terms explanation inside this report. An example: The terms were defined for *GLP* conditions. That is why the term “Study Report” was used as *GLP study report* in this report. The reader can follow this cross reference.

At the same time as terms and concepts were developed, statements with different objectives were formulated. The following statement types are used in this report and are organised as one sequence per chapter number of the 1st level with a starting letter added with the chapter number of the 2nd level e.g. R 9.2-76 (see Table 1).

The user requirements were formulated without a concrete technical solution. The listed user requirements describe needed functionalities to assist the process steps that can be used to compile the needed *Information packages*. The user requirements were written to get a level of interoperability of the systems, which ensures that data once entered in IT-systems does not need be re-entered manually again.

At the level of user requirements, an attempt was made to formulate them without preference for specific technical implementations. If such technical solutions were mentioned, this was only to make them easier to understand and clarify.

According to the list of user requirements, different solution approaches are possible. Chapter 9 contains the solution approaches, which are in line with the defined terms in chapter 7. It should be mentioned that the suggestions have been given based on the previously elaborated weak point analysis and the claim of generalising the information flow of pesticide related metabolism studies. Therefore, these proposals have not been justified a second time in this report.

Table 1: Statement types and their meaning

Type	Meaning	Objective	Count
A	Deeper **A**nalysis	A deeper analysis is needed for this topic to create realistic user requirements	26
B	**B**enefit	BfR proposed work packages. EFSA could achieve these benefits at the specified milestone.	8
CR	**C**urrent high-level user **R**equirement	Part of the implementation actions following the entry into force of the Transparency Regulation requirements.	4
D	Needed **D**ecision	The decision process should be organised by EFSA	19
M	**M**ilestone	Milestone in the proposed work packages by BfR.	20
O	**O**pinion	Assessment of the current state by BfR	25
P	**P**roposal	BfR recommendations for further development	84
R	user **R**equirement	User requirement collected by BfR assisted by stakeholders. The more requirements from this list are provided in the future IT-Tool, the more it will satisfy the users. These requirements are independent of the chosen transport concept or the transport will be realized as an IUCLID attachment (Chapter 9.8.3.1 or 9.8.3.2)	339
S	**W**eak Point	Weak Point identified in the Survey	22
W	**W**eak Point	Additional analysed weak points which are not part of the “List of weak points identified in the survey” (Chapter 11.4) This list should be minimized with a new IT-Tool.	116

A list of all statements of the current report is attached as a spreadsheet (see chapter 11.2: List_of_statement.xlsx).

5 Background

5.1 Context of this report

Please note that similar terms are used in different contexts. The following scopes of aspects are not the subject of this report:

- Studies on general metabolism in organisms,
- Metabolites in the context of the use of microorganisms as pesticides and
- Metabolites in the context of “Metabolomics”.

The terms *Metabolite* and *Metabolism study* as used in this report are defined in Chapters 7.2.1.and 7.2.2.

5.2 Previous efforts

The fact that not only the actual pesticides but also their *Metabolites* can have effects on human health or the environment is well known, and their potential qualitative and quantitative impact on different species is an integral part of the assessment process. Consequently, *EU 1107/2009* defines the term *Metabolite* for the field of European plant protection (see chapter 7.2.1.3).

Since 2010, the *USEPA* had been advocating a standardised evaluation of metabolism studies with the goal of building a metabolite information system. The *OECD* MetaPath User Group (*MUG*) was formed and the necessary concepts were developed within this international scientific community. *MetaPath* and several *MSS-Composers* for data ingestion of relevant metabolite study *Metadata* have been developed.

In 2011, the *USEPA* initiated a study (*Manibusan 2011*) to demonstrate the applicability and usability of *MetaPath* as a predictive model in regulatory practice, so that it “enables efficient and systematic metabolite comparisons across chemicals, species, and environmental media of potential risk concerns” with all types of metabolism studies. “The ‘*MetaPath*’ system grew out of the need to compile and organise the results of metabolism studies into a systematic database to facilitate data comparisons and evaluations” (*Kolanczyk 2012*).

In 2018 the former *OECD* project *MetaPath* was only mentioned in a footnote in the document ENV/JM/PEST/RD(2018)1 and at the 33rd Meeting of the Working Group on Pesticide the deprioritization probably took place. From this time on, the group no longer operated under the umbrella of *OECD*.

This development work by the *USEPA* /*MUG* was necessary and is not in any way discredited with the current analysis of the status and weaknesses.

One problem in the risk assessment of pesticide *Metabolites* is that a pesticides *Active Ingredient* can break down into a large number of *Metabolites*, depending on the conditions, and there is usually little or no knowledge about the properties of these degradation products.

OECD Guidance 194 (*OECD 2014*) has defined techniques / methods for data gap filling, an “analogue approach” and the “category approach” (see chapter 7.3.2). Both approaches starting with a step 0: “Check whether the chemical is a member of an existing category.” Adequate information sources for existing categories are needed. For the “analogue approach”, the first step is named “Identification of potential analogues” where methods are used to look for structural similarities. This step should also identify analogues according to the potential mechanism or mode of action of the test substance.

Future evaluations of pesticide *Metabolites* should preferentially use non-animal test methods wherever possible. The question then arises: Which methods are available to evaluators to support this goal and reduce the need for vertebrate studies? *MetaPath* can be used as an

information database to identify similar *Metabolites* or substructures from different compounds, as well as overlapping *Metabolic pathways* within and between different taxa or regulatory definitions, i.e. cumulative assessment groups. This is a prerequisite for a read-across assessment.

That means that as the number of metabolism studies deposited in the information database increases and efficient strategies become available, the chance of circumventing vertebrate studies will increase. The *EFSA* has recognized this problem and has initiated various projects to improve the information database. For example, the *BfR* and *ANSES* are processing 1200 studies on metabolic behaviour, which will be integrated into *MetaPath* as such an information database.

With the implementation of *IUCLID* as the sole delivery format for pesticide dossiers in the European Union as of 2021, there is an opportunity to reorganise the information flow of pesticide related metabolism studies. The *EFSA*'s objective is to ensure that the new metabolism studies provided in the application procedures are immediately incorporated into the information database for the modelling of the metabolic behaviour of pesticide active substances.

According to the specific agreement under the framework partnership agreement No GP/EFSA/AMU/2020/02, proposals for the improvement of the current information flow of metabolism studies should be developed.

The authors would like to thank the members of the MetaPath Users Group for the engaging technical discussions (compare chapter 9.8.3.3).

5.3 The current EFSA process

The *EFSA* published a documentation "Reporting structured results of metabolism studies on rats, plants and livestock" (*EFSA 2021*) with a description of the current European process steps.

There are the following current high-level user requirements (CR) relevant for applicants as well as for authorities:

CR 5.3-1: A set of *Aggregated raw data* from metabolism studies is stored and managed in a local metabolism pathway collection.

CR 5.3-2: A data interface exists for a data exchange of *Aggregated raw data* from metabolism studies between different metabolism pathway collections. This data interface can import *Aggregated raw data* submitted with a study in context of a legal act.

CR 5.3-3: *QA* checked *Aggregated raw data* of metabolism studies are collected in an international *Metabolic pathway* collection. A *Quality control body* uses a *Set of quality standard rules* prior to including the data sets into this collection.

CR 5.3-4: The IT-Tool specified for doing so is called *MetaPath* and assists the user in process steps starting with the validation of incoming data sets, searching for similar metabolites / pathways.

It is known that this process organisation, which have been set for a short-term horizon, have various weaknesses in itself. Therefore, a parallel improvement process has been initiated by *EFSA*.

The weakness W 5.3-5 was accepted by *EFSA*, because at this stages it was considered not beneficial to ask stakeholders to modify maps which have just been quality checked by *BfR*/ *ANSES* and The *EFSA*.

W 5.3-5: *EFSA* has not included the possibility of correction of XML-files in the current process, which are available in 1) Regulatory Legacy collection of maps *or* in 2) *EFSA* public collection of maps, which have been selected according a Standard Operation Procedure developed during project OC/EFSA/PRES/2019/01 *OECD* Metapath-Incorporation of Pesticide Residue Data (*BfR 2020*).

6 Objectives for further development

The part of the report “Results of the international survey” (*BfR 2021*) had shown weaknesses in the current information flow for metabolism studies and the available IT-Tools. The identified weaknesses using *MetaPath*, as it is currently required by the *EFSA* by the framework of the Transparency Regulation implementation, were summarised in chapter 11.4. Approximately two-thirds of the weaknesses identified showed a need for action to improve IT-Tools.

Starting from this point, the *EFSA* formulated content-related objectives for improving the flow of information from metabolite studies and for their use in the assessment processes (Table 2).

It should be noted that this evaluation matrix only considers content-related aspects and may simply be wishful thinking in some cases. At this point, concerns that the project might be too ambitious need not be considered. These objectives should only be scaled down if the decision-makers are not able to organise a project plan with individual project stages that can be financed within a manageable timeframe. The possibility of a "public-private partnership" (*EPEC 2015*) or the model of an "innovation partnership solution" (*BIG 2021*) should therefore be considered.

As the project processed, it became clear, that one key framework question was not included inTable 2, the *EFSA* objectives, how to handle weakness W 5.3-5.

Table 2: EFSA objectives for the further development of the information flow of pesticide related metabolism studies

No	Group	Objective	Justification	Priority (3 high, 2 medium, 1 low)	Notice
1.1	Generic approach	The provided solution should be usable to subsume all types of studies in which at least knowledge of the "identity of transformation products" is obtained.	All study types, where radiolabelled test substances could be used, should be a potential data source. It does not matter whether these transformations are triggered by biotic or abiotic processes.	1-2	Not for short term
1.2	Generic approach	The provided solution should be applicable in the harmonised OECD templates where the use of radiolabelled test substances is possible.		3	Phase 1: OHT58 BasicToxicokinetics OHT85-2 MetabolismInLivestock OHT85-3 MetabolismInCrops
				1-2	Phase 2: Other OHTs
2.1	Architecture	A new generic approach should be able to cover all types of metabolism studies with the same IT components.	It is impossible to finance and manage a life cycle for a set of high-differentiated MSS Composer for each metabolism study type.	1-2	Not for short term
2.2	Architecture	The number of needed data interfaces and export / import modules should be minimized. With a focus on the reuse of existing APIs (https://iuclid6.echa.europa.eu/public-api) and analysis of the need for additional APIs. Interfaces already developed by LMC under OECD and other projects should be analysed.	Each additional interface generates additional costs	3	
2.3	Architecture	It should be possible to start the data flow of meta data as an output of the GLP systems of the laboratories (LIMS).	Aggregated raw data of metabolism studies could be compiled at time point of "GLP Study Report".	1	
2.4	Architecture	The format for downloading the metadata from the curated repository should be the same as used for submitting a new metabolism study in a dossier.	Applicants have to be able to upgrade / correct the data in the same format as it was downloaded and be able to feed it back it in the processes.	1	

No	Group	Objective	Justification	Priority (3 high, 2 medium, 1 low)	Notice
3.1	Substance model	The provided solution could handle a set of “unknown” metabolites inside of one study	It is necessary to transport meta data for distinct but not yet identified substances.	2	
3.2	Substance model	The provided solution can manage a retrospective matching of identical “unknown” substances of different studies.	It is a normal case that metabolites are “unknown” in the earliest metabolism studies and named only by a code. However, this “unknown” metabolite could be identified later. Therefore, a flexible matching of substances between older and recent studies is necessary.	2	
4.1	Evaluation	Evaluators on the applicants and authorities side should use the same set of meta data for risk assessment.	Having the same starting point will minimise missunderstandings between APPL and RMS/EMS.	3	
4.2	Evaluation	The provided solution should make use of Metapath as is – but areas for improvement should be identified.	The MetaPath functions to manage metabolic trees, visualize metabolic trees, search for similar substances, compare metabolic trees are the most important essential functions.	3	
4.3	Evaluation	The provided solution should identify manual data transformations steps inside of the evaluation process, for prediction of metabolism pathways, for grouping of metabolites and prediction of toxicological parameters should be minimized (Q)SAR – and indicate which steps could be automated in a later phase.	The evaluators have to be able to check and evaluate the multitude of individual results against the legal requirements with scientific accuracy within a certain time frame.	3	
4.4	Evaluation	User should be able to create an overview (report) of relevant metabolism studies of a specific test substances inside of a local collection of metabolism studies which could be incorporated in an IUCLID flexible summary.	Evaluators should be able to summarize a set of studies.	3	
4.5	Evaluation	Users should be able to modify standard reporting table templates. The provided solution includes additional user functions for interactive grouping and reporting of results.	Static reports could assist only standard cases.	2	
4.6	Evaluation	All known weak points should be improved			
5.1	Reference collection	It should be possible to build up an international reference collection of metabolism studies under the Metapath project and user group. A publicly accessible interface should be defined.		3	

No	Group	Objective	Justification	Priority (3 high, 2 medium, 1 low)	Notice
5.2	Reference collection	Only QA checked metabolism studies should be part of a reference collection of metabolism studies.	Only QA checked data should be included in (Q)SAR models and will then be referenced in the QSAR Model Reporting Format (QMRF).	2	
6.1	Publication	The provided solution should be compatible with the needed publication process of EFSA.	Aggregated raw data of metabolism are not subject of publication because these data are part of Rich-Text fields in the study summaries.	2	
7.1	(Q)SAR model	It should be easy to include needed meta data of the QA checked metabolism studies into (Q)SAR models itself to improve the training data set.	(Q)SAR models should be improved for agrochemicals.	1	

7 Terms, user requirements and concepts

An attempt has been made in this report to enforce a uniform use of terms. This was to ensure that the user requirements in this report could be interpreted identically by all readers.

7.1 IT related terms

7.1.1 Chemical structure notation

There is a variety of notation forms available for chemical structure coding. It should be taken into account that the Chemical Structure Notation, like any natural language, is also subject to evolution. The current *MetaPath* tool set is using the *SMILES* concept. The conducted survey on the flow of information on metabolism studies, has emphasised, that the *SMILES* concept has limitations.

R 7.1-1: The information flow should be based on the more reliable chemical structure notation standard called *InChI* (International Chemical Identifier) developed and maintained by the *IUPAC* (*Goodman 2021*).

R 7.1-2: Systems using Chemical Structure Notation should be downward compatible.

R 7.1-3: It should be possible to choose a representative structure (Markush/generic structures) in the implemented Chemical Structure Notation.

R 7.1-4: The support of Markush/generic structures is a showstopper for the further improvement process.

A 7.1-5: A deeper analysis is needed to check whether the problems of the Markush/generic structures are solved by the *LMC* extension, which was programed for BASF in 2021.

7.1.2 Interoperability

Interoperability should be understood as "*the ability of different systems, devices, applications or products to connect and communicate in a coordinated way, without effort from the end user.*" (*TECHTARGET*).

There are different levels of interoperability on:

- syntactic,
- semantic and
- cross-domain (or organisational)

level.

This report uses this term only on the semantic level. "This is the ability of systems to exchange and accurately interpret information automatically. Semantic interoperability is achieved when the structure and codification of data is uniform among all systems involved" (*TECHTARGET*).

This report called a system "interoperable" if it is able

- to use provided services of other systems. This means, the system under review should be able to request other systems according the provided API service description of the data provider, get the response and to process the response according a defined procedure;

- to play the role of a data provider. To do this, the provided API and the format of the data interface should be defined and published.

7.1.3 Information package

The information flow of pesticide related metabolism studies is considered to be transported as *Information packages*, which are compiled according guidelines and transformed by adequate data interfaces. Therefore, its compilation should be flexibly defined according to agree upon standards. These standards should consider the needs of all data producers and users. Due to standardisation, the data can still be exchanged within the IT-Tool framework.

The term *Information package* for metabolism studies should be understood as real packages of objects, which contain the information on a specific level of aggregation according to the related format definitions. In essence, the information aggregation is highly depending on the stakeholders' point of view.

7.1.4 Metadata

The term *Metadata* should be used as an abstract term. *Metadata* provide additional information about data or, in other words, they are data about data.

One can find any number of compilations of *Metadata* for one object. It is therefore important to define the purpose of these descriptive *Metadata*. That means that the viewpoint of the potential data consumer should be the basis of the definition of a set of *Metadata* of an object. This perspective is the key to define a set of generally accepted *Metadata* for one object.

As soon as a new purpose is to be fulfilled with the descriptive *Metadata* of an object, the set of *Metadata* and possibly their formats need to change accordingly.

R 7.1-6: Based on this understanding, it is particularly important to describe the requirements and intended use of the *Metadata* as precisely as possible during the analysis phase.

7.1.5 Picklist

A *Picklist* is a list of the most frequently used terms that can be selected by the user in a specific field. The possible range of values for classifiable *Metadata* is controlled by a *Picklist*.

7.1.6 Object type

Pay attention. This term is used in two different meanings:

- the *Object Type*of a study specifies the Study Type at the highest level e.g. soil, sediment, crop, rotational crop, water, food, animal
- the object types in an IT-Tool differentiate different types of the *Shopping basket*.

7.1.7 Validation

The term "validation" is used in different contexts with different meanings. A validation process requires rules in each case, which are tested during the process of checking the validity. A 'formal' validation check is to check the validity of an exported or submitted XML file against an XML schema description of the data interface.

Another 'content based' validation review targets and checks the submitted study against the data requirements and test guidelines used.

7.2 Study related terms

In the following sections, an attempt was made to define a set of terms from the conceptual world of metabolism studies in such a way that they will be usable for a generic metabolism trial type. However, no term should be considered in isolation from this set of definitions, as each is incomplete on its own.

7.2.1 Substance

In the present report, the term *Substance* includes the *Test substance* and its *Metabolites*.

R 7.2-1: A *Test substance* could be transformed in the *Object of investigation* by *Transformation processes* into Metabolites.

Note: Currently, however, different definitions with different objectives are used internationally for *Metabolite*, which are not consistent. To give an overview, on the following pages, the definitions from FAO, *OECD*, EU COM, *EFSA* are presented. Nevertheless, all of these definitions should be covered by the used concept of this report (R 7.2-1).

7.2.1.1 FAO

The guideline Codex Alimentarius, (*FAO/WHO 2017* Appendix XIII Definition Annex) on performance criteria for methods of analysis for the determination of pesticide residues in food and feed defines different terms for the biotic und abiotic transformation as:

Metabolite: “Component of a pesticide residue occurring in a commodity as a result of biotic transformation (metabolism) of a pesticide in a biological system (e.g. plant, animal).”

Degradate “(degradant, degradation product): Component of a pesticide residue occurring in a commodity as a result of abiotic transformation of the pesticide (e.g. heat, light, moisture, pH, etc.)”

Here, biotic substance modification resulting in metabolites is considered separately from abiotic modifications, resulting in degradates.

7.2.1.2 OECD

No overall glossary was published which could be used for a consistent terminology for the *OECD*. This makes it harder to see similarities between the guidelines.

Many terms have been used that refer to similar or related transformation processes in the *OECD* (see chapter 11.3, column “Test Guideline”) like: Bioaccumulation, Bioconcentration, Biodegradation, Biomagnification, Hydrolysis, Metabolism, Mineralization, Transformation.

7.2.1.3 EU COM

The *EU 396/2005* Article 3 2c) on maximum residue levels of pesticides in or on food and feed of plant and animal origin and amending Council Directive 91/414/EEC is using a very generic definition:

Pesticide residues

"means residues, including active substances, metabolites and/or breakdown or reaction products of active substances currently or formerly used in plant protection products as defined in Article 2, point 1 of Directive 91/414/EEC, which are present in or on the products covered by Annex I to this Regulation, including in particular those which may arise as a result of use in plant protection, in veterinary medicine and as a biocide;"

The *EU 1107/2009* (Article 3, No. 32) used the following definition and created the term "relevant metabolite":

Metabolite "means any metabolite or a degradation product of an active substance, safener or synergist, formed either in organisms or in the environment.

A metabolite is deemed relevant if there is a reason to assume that it has intrinsic properties comparable to the parent substance in terms of its biological target activity, or that it poses a higher or comparable risk to organisms than the parent substance or that it has certain toxicological properties that are considered unacceptable. Such a metabolite is relevant for the overall approval decision or for the definition of risk mitigation measures."

The document *EU Sanco/221/2000*-rev11 (21.10.2021, Chapter 3. Definitions) on the assessment of the relevance of metabolites in groundwater is guidance for notifier and Member States in the context of the review of active substances and defined the term metabolite as:

Metabolite "for the purpose of this document, the term is used for all reaction or breakdown products of an active substance of a plant protection product, which are formed in the environment after the application, be it by biotic (microbials, other taxa) or abiotic processes (hydrolysis, photolysis). The terms 'metabolite', 'breakdown product' and 'degradation product' are used interchangeably throughout this document."

7.2.1.4 EFSA

The *EFSA* used a slightly more restricted definition of metabolism with the aim of establishing a residue definition (*EFSA 2016*, Chapter 1. Introduction):

Metabolite "The fate of pesticides after application on crops or soil may be affected by numerous biophysicochemical degradation processes resulting in a change of the chemical entity of the pesticide and occurrence of a mixture of compounds in harvestable commodities and the environment – the active substance (commonly called 'parent compound'), metabolites and degradates (in the following also termed 'metabolites'). "

This term refers to metabolism in plants, animals and in processing.

7.2.2 Metabolism study

In this report, the term “metabolism studies” is understood as a study type in which:

A test substance is investigated in an *Object of investigation*, and the absorption, distribution, metabolism and/or excretion kinetics are recorded under defined conditions.

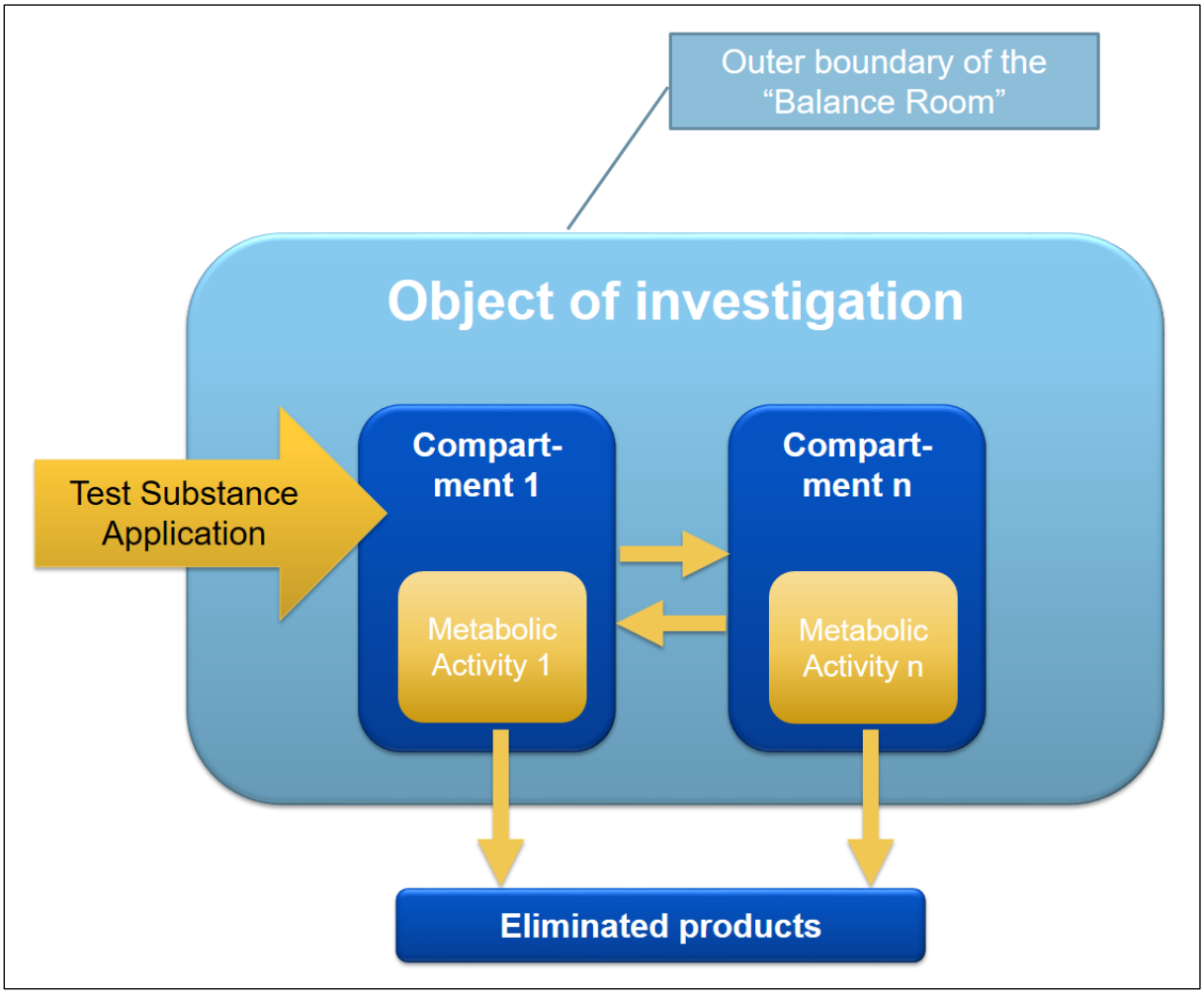


Figure 2: Model of the generalized term of metabolism study

R 7.2-2: The term *Metabolism study* should cover all types of studies in which at least the knowledge of the “identity of transformation products” is obtained. It does not matter whether these transformations are triggered by biotic or abiotic processes.

R 7.2-3: The term *Metabolism study* should cover the qualitative and the time dependent quantitative aspects of distribution and transformation.

R 7.2-4: In order to establish a *Balance of activity*, the use of radioactive *Test substances* is mandatory in the *OECD*. However, the proposed term *Metabolism study* could cover not only studies with radiolabelled substances. The labelling with non-radioactive isotopes in combination with modern analytical methods would allow additional experimental designs.

Experimental approaches that meet the above definition but do not specify the “identity of transformation products” allow only summary statements regarding the distribution. However,

the above term restriction does not mean that IT-Tools for processing “metabolism studies” should not be suitable to include these distribution-only trials.

Chapter 7.4.2 provides some information regarding *Metadata* of a *Metabolism study*.

7.2.3 Object of investigation

The term *Object of investigation*, used for a *Metabolism study*, should be understood as a generic representation for a test system (e.g. rat, mouse, plant, soil) where a *Test substance* is applied and being investigated (see Figure 2). Depending on the type of experiment, not all of the process steps of Figure 2 can be observed in the *Object of investigation*.

If several individual test systems are used in a study, they can be grouped together. All such groups are called *List of Study Object Groups*. As an example: For a rotational crop study different crops are used.

7.2.4 Balance room

The *Object of investigation* has an outer boundary, which encompasses the “Balance room”. This is the prerequisite to calculate the *Balance of activity*.

An *Object of investigation* can consist of individual parts (Compartments) which are separated from each other. Distribution processes between the *Compartments* are possible. Each compartment can have different enzymatic activities for the *Transformation processes* e.g. straw and grain; liver and kidney etc.

7.2.5 Balance of activity

Accounting the activity of the applied *Test substance* when leaving the *Balance room*. The “% of Administered Dose (AD)” is the most common form of specification of the *Balance of activity*.

If radiolabelled substances were used, results may be calculated as percentage of the applied used activity of the substance. These values could be used for balance results as well as for the remaining activity at the end of the experiment in different *Compartments*.

R 7.2-5: The sum of the *% of Administered Dose (AD)* of all *Compartments* as well as the eliminated products should be comparable to the initially used activity.

Pay attention:

- The term *Balance of activity* would not be correct when labelling with non-radioactive isotopes in combination.
- Some technical guidance’s use incorrect terms but meaning the same e.g. “Mass Balance” in TG 417.

7.2.6 Application

A *Test substance* is applied into / on an *Object of investigation* once or several times according a *Dosing scheme*. The mode of *Application* of the *Test substance* needs to be documented in detail e.g. i.v., i.p., oral.

Synonyms for *Application* are used in specific metabolism *Study Types* e.g. “Dosing” or “Feeding”.

7.2.7 Dosing scheme

The *Dosing scheme* describes the number and the frequency of *Applications* of an amount of the *Test substance*.

7.2.8 Transformation process

A chemical modification of a substance in a series of transformations processes (see also *Metabolic pathway*).

7.2.9 Test substance

A well-defined *Test substance* will be applied to the *Object of investigation*. The term *Test substance* could also be understood as a synonym for the term Test material that was used in different *OECD* guidance documents.

R 7.2-6: In most cases, the *Test substance* is also the Active Ingredient. However, there are also cases where synthesised *Metabolites*, are to be used as the *Test substance*. This case should also be covered by the data model.

7.2.10 List of metabolites

The *List of metabolites* is one of the main results of a *Metabolism study*. The *List of metabolites* is a flat list of *Metabolites* without any information about

- the sequence of the creation of the transformation products,
- the kinetics and
- the pathway as result of the *Transformation processes*.

A *Metabolite* could be “known” or “unknown” at the time point of writing the *GLP study report*.

A “**known metabolite**” should be characterised by at least one identifier of the molecule (compare chapter 7.1.1). The identification could be done by 2D structure information or, in some cases, stereo chemical information are needed.

The status “**unknown**” could only be correct at the time point of writing the *GLP study report*. An “unknown metabolite” could be identified time delayed in other *GLP study reports*.

7.2.11 Metabolic pathway

Please note that one *Metabolic pathway* should be seen only as an interpretation of the results of one metabolism study.

A *Metabolic pathway* involves the step-by-step *Transformation processes* of the initial *Test substance* to form transformation products in a specific *Object of investigation*. The *Metabolic pathway* describes the hierarchy of the transformations products. The result is: one *Metabolic pathway* for each test system in an *Object of investigation*.

The “Metabolic tree” should be understood as the visualisation of one *Metabolic pathway* information in a schematic diagram (see Figure 3). “Metabolic map” is a synonym for *Metabolic pathway*.

Within the same *Object of investigation* different aspects of the same *Metabolic pathway*, such as absorption kinetics or bile excretion, can be investigated.

Different *Metabolic pathways* are possible if several individual test systems are used in a study (e.g. rotational crops; different application regimes).

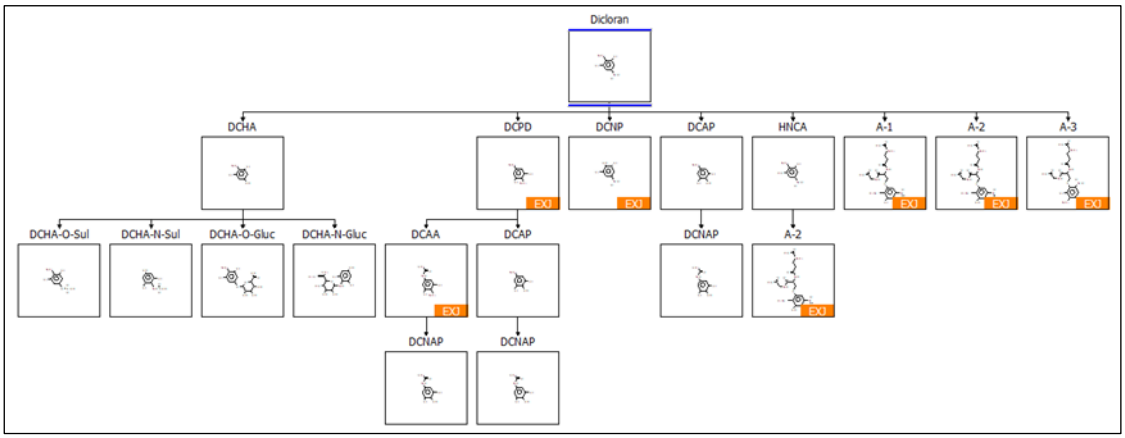


Figure 3: Example of a *Metabolic pathway* generated by *MetaPath*

7.2.12 Other metabolism related terms

There are more advanced terms in the context of metabolism studies, but historically they have only referred to certain types of experiments.

Examples of the related terms are listed in the following table.

Table 3: Other metabolism related terms

Term	**Meaning**	**Remark**
Absorption	Process(es) of uptake of substances into or across tissues.	
Accumulation	Increase of the amount of a substance over time after repeated exposure if the input rate is greater than the elimination rate.	It is essential to specify the basis for such values. Does one refer to the applied substance or to the sum of applied substance and metabolites?
ADME	Acronym for “Absorption, Distribution, Metabolism, and Excretion”;	Term is used for metabolism studies on animals and livestock
Bioaccumulation	Accumulation of a substance in biotic systems	
Bioavailability	The substance is available to biological processes and not bound in any inaccessible form.	
Distribution	Dispersal of a substance and its metabolites throughout the compartment(s) of the Object of Investigation	
Excretion	Process(es) by which an injected substance and/or its metabolites are removed from the Object of Investigation	
Route administration	Synonym for route of application (see 7.2.6)	
Extractable Portion	Samples are extracted with a series of solvents and/or solvent systems (including aqueous) with various polarities and other characteristics depending on the nature of the expected residues. These initially obtained residues are defined as extractable residues.	

However, these terms have only a limited scope in the generic approach and therefore their usage was avoided in this framework.

7.3 Assessment related terms

This chapter is not a description of any hazard and/or risk assessment procedures.

This chapter is a description of the process steps, techniques, approach tools and necessary information within these steps. The user requirements of this report derived from the hazard and risk assessment procedures are so universally valid that they will endure even if concrete procedures are revised. There is a need of integration of qualitative and quantitative data for the risk assessment process (see Figure 4).

R 7.3-1: *MetaPath* should contain crossreferences to adequate exposure data.

Because the *BfR* only has expertise in the field of the assessment processes for human health, the statements should be verified for other endpoints e.g. ecotox.

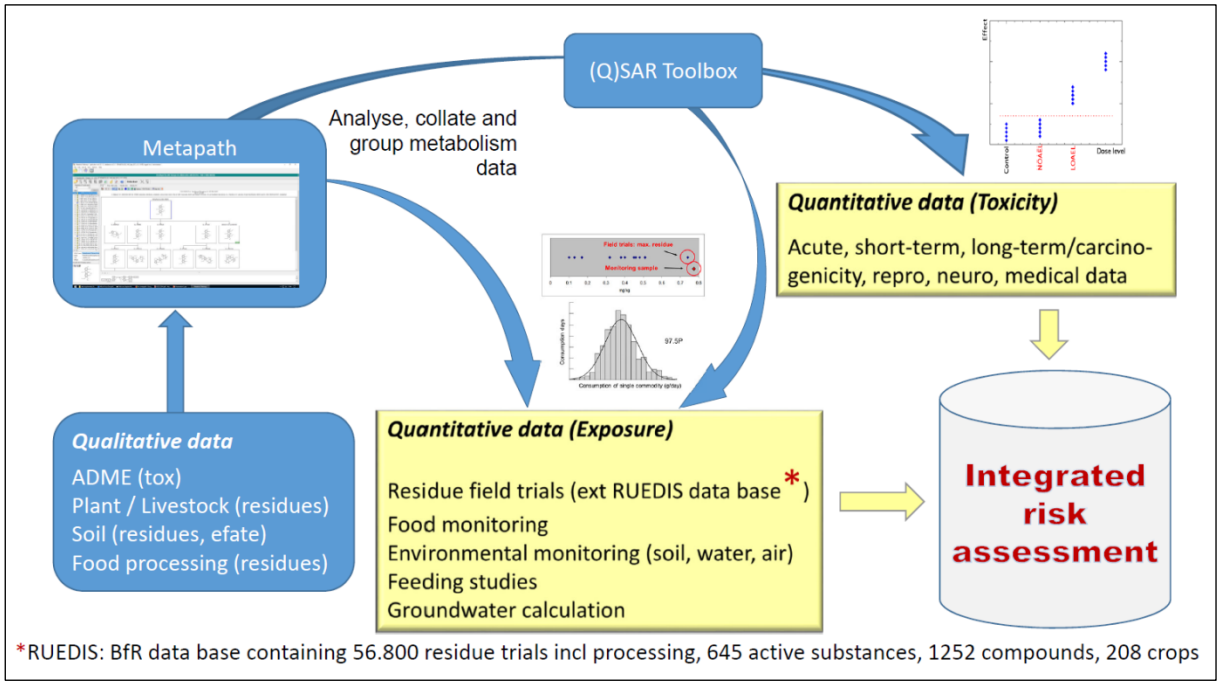


Figure 4: The need of integration of qualitative and quantitative data for the risk assessment process

The aim of this chapter is to formulate high-level user requirements for the future IT-support for these process steps.

7.3.1 Framework conditions

The driving force of the information flow are the data requirements for the evaluation of the substances. Without these data requirements, this information flow would not exist.

The test methods, guidance documents and models, which are to be used to address the data requirements of COM e.g. *EU 283/2013*, are listed in *EU 2013/C 95/01*.

This document refers to the *OECD* Guidelines according to which the tests are conducted. Comparable data requirements exist in other regulated areas.

R 7.3-2: The data requirements and the corresponding assessment guidelines thus determine the semantic content of the necessary *Information packages* for metabolites, and the specifications in the individual procedures determine the interfaces and IT-Tools to be used.

The user requirements are derived from these framework conditions.

There are no differences in user requirements between applicants and authorities, as both stakeholder groups work within the same regulatory framework.

For this reason, the term “user” can be understood as a representative *Evaluator* of the applicants or the authorities.

The differing requirements are described in separate chapters, *Applicants´ information packages* (7.4) and *Authorities’ information packages* (7.5) below.

The overall objective is to make best use of the available metabolism information for the risk characterisation and risk assessment of pesticide active substances (see Figure 5).

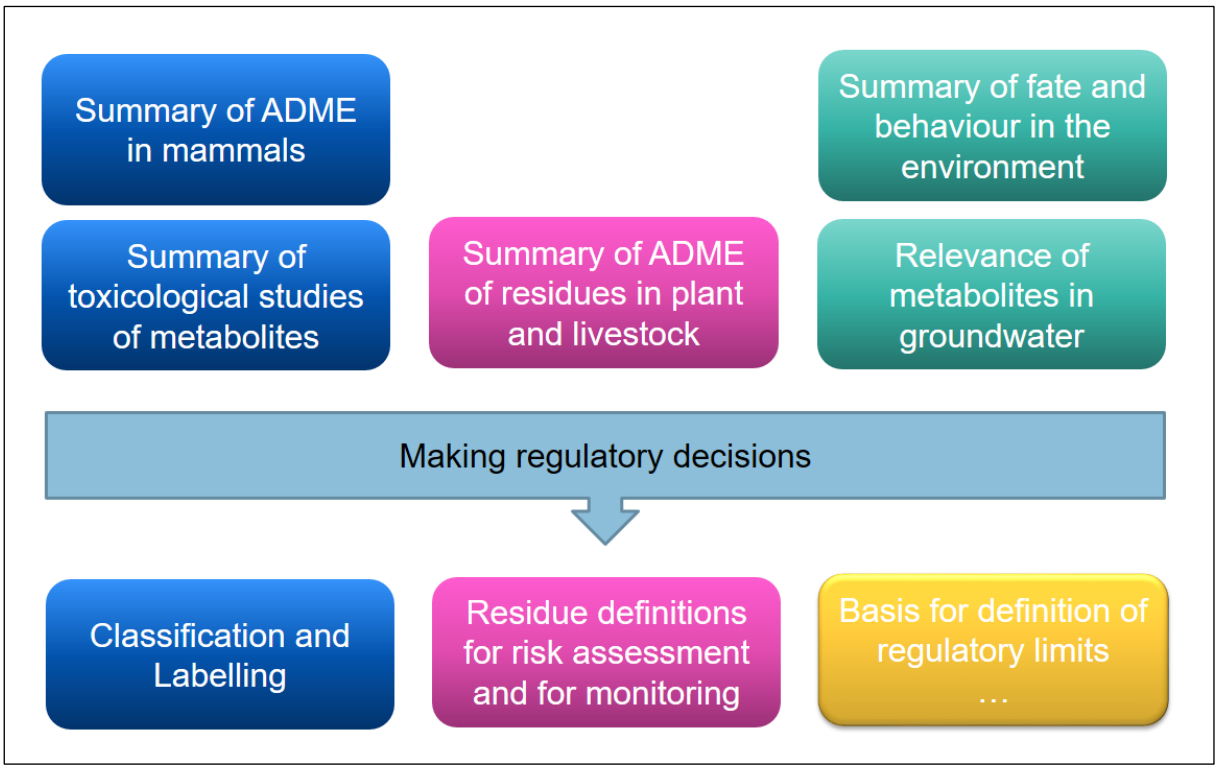


Figure 5: Information on metabolism influence regulatory decisions

The following Guidance documents are important in European context:

- *OECD* Guidance residue definition: ENV/JM/MONO(2009)30 (*OECD 2009*) revision ongoing (expected 2022)
- *OECD* Guidance on Grouping of Chemicals: ENV/JM/MONO(2014)4 (*OECD 2014*) revision started
- *EFSA* Guidance residue definition: EFSA Journal 2016;14(12):4549 [1] (*EFSA 2016*)
- SANCO Guidance document on the assessment of the relevance of metabolites in groundwater: SANCO/221/2000 (*EU Sanco/221/2000)*

[1] The methods for metabolite assessment in this guidance document represent the current standard of metabolite evaluation. This document is referred in current EFSA instruction but the guidance was not officially noted in EU. Regarding the decision criteria for the relevance of metabolites, reference was made to the OECD.

The user should be able to summarise the results of all submitted metabolism studies under consideration, supplemented by results from other *Active Ingredients*, the known toxicological properties of the active substance and metabolites supplemented by predicted toxicological properties of further metabolites.

The principle for creating a residue definition for the dietary risk assessment defines the two most important work tasks (ENV/JM/MONO(2009)30) (*OECD 2009*):

> *"The Metabolites, degradates, or other transformation products (hereafter collectively referred to as "metabolite/degradate") that significantly contribute to the dietary risk should be included in the exposure assessment. For each metabolite/degradate to be considered to contribute significantly to the risk, two factors must be addressed:*
>
> *1) the **potential for exposure** to the metabolite/degradate in the human diet; and*
>
> *2) the **relative toxicity** of the metabolite/degradate to the parent. Metabolites/degradates with higher potential exposures and toxicities are more likely to be included in the dietary assessment."*

There are additional data requirements, which may influence the human health risk assessment:

- *OECD* Test Guideline 307 (Aerobic and anaerobic transformation in soil)
- Scenarios and assessment models for residues in soil and groundwater (PEARL, PELMO, PERSAM, ESCAPE)

The results from the studies according to *OECD* Test Guideline 307 and related guidance documents, together with subsequent model results, determine whether environmental metabolites are to be considered for human health risk assessment.

From the survey on the flow of information on metabolism studies, it is known that the *Evaluator* is confronted with a flood of information that can be best managed with the help of an adequate IT-support.

The content of the information could be summarized as:

- Information of all metabolism studies according the harmonised templates submitted in the dossier according the data requirements of the legal act.
- Information on the metabolic pathway of additional substances, which are similar to the active ingredient or observed metabolites.
- Collected (Q)SAR based predictions related to observed metabolites and additional similar substances

The following user requirements are based upon the flood of information and the evaluation criteria.

R 7.3-3: *Evaluators* should be able to manage the huge amount of metabolism relevant information with the help of an adequate IT-support.

R 7.3-4: An IT-Tools is needed to store *Aggregated raw data* from metabolism studies.

The detailed requirements of such an IT-Tool are described in subsequent chapters.

The following high-level process steps are necessary for risk and hazard assessment (see Figure 6):

Figure 6: High-level process steps for a risk and hazard assessment of metabolites

The central processing steps in risk and hazard assessment are endpoint-independent and the process steps are always run in a loop over all known and unknown metabolites. The decision about the relevance of this metabolite is evaluated according to the relevant guidelines.

Some examples of the rules are:

According ENV/JM/MONO(2009)30 the major metabolites in the context of residues are (*OECD 2009*):

> *„For the purposes of discussion, major metabolites are considered to be those which at any point in time contribute to 10% or more of the total radioactive residue (TRR) in metabolism studies in plants, livestock, or rotational crops. Similarly, major environmental degradates are those which represent 10% or more of the applied dose in environmental fate studies at any point in time.“*
>
> The minor metabolites, which represent less than 10% of the TRR, should also be considered in the following situations:
>
> - *„Minor metabolites are known, or suspected, to be considerably more toxic than the parent compound.*
> - *The analytical method for data collection is a common moiety method and includes several metabolites, including minor ones.*
> - *Very few or no major residues are observed and numerous minor metabolites of toxicological significance collectively comprise a substantial portion of the TRR.“*

For residues, not only the relative content but also the concentration is relevant. Please have a look to the “Table 1” in TG 501/502/503, which clearly defines under which circumstances metabolites need to be characterised and identified.

In SANCO/221/2000 – rev.11 the “relevance” of groundwater metabolites are defined (*EU Sanco/221/2000*):

- *“This document describes a stepwise scheme, of increasing complexity, to identify “relevant metabolites” for which the above provision of Annex VI and thus the limit value of the Drinking Water directive should apply. The document further describes a scheme for the assessment of those metabolites, which are not identified as relevant, but which have to be evaluated previous to a decision on the inclusion of an active substance in Annex I to Directive 91/414/EEC.”*
- *“Consequently, this document describes a scheme to determine whether a metabolite*

is relevant (and thus subject to the 0.1 µg/L limit) or not relevant using criteria of biological activity, genotoxicity and toxicological hazard but also other, pragmatic administrative criteria to allow efficient and transparent regulatory decision-making."

- *"A metabolite is considered "relevant" if its toxicological properties lead to a classification as toxic or very toxic (T or T+)" according to Directive 67/548/EEC."*

7.3.2 Data gap filling

Regarding the risk assessment of metabolites data gap filling could be used for predicting:

- the *Metabolic pathway* and
- for toxicological endpoints.

7.3.2.1 Read-across and (Q)SAR

Read-across is regarded as a technique for extrapolating or interpolating endpoint information for one substance (target substance), by using data for the same endpoint from (an)other substance(s), (source substance(s)) (*TOXIT*)

The *OECD* Guideline 194 (*OECD 2014*) has defined two approaches for Read-across data gap filling, the "analogue approach" and the "category approach". Both approaches starting with a step 0: "Check whether the chemical is a member of an existing category." Adequate information sources on existing categories are needed (see chapter 7.3.2.2).

For the "analogue approach", the first step is named "Identification of potential analogues" where common analogue identification methods look for structural similarities. This step should also identify analogues according to the potential mechanism or mode of action of the test substance.

A (Q)SAR model is a predictive (quantitative) relationship between structure, i.e. one or more molecular descriptors and the biological activity (i.e. toxicity). (Q)SAR models are build using large sets of data derived from multiple substances. Based on those models the intention is to find a trend, which can then be applied to the target substance including a certain statistical error.

The (Q)SAR technique is a field of the computational toxicology using mathematical methods to calculate similarities, trends and probabilities.

The *OECD* has agreed the following principles (*OECD 2007*):

"To facilitate the evaluation of a (Q)SAR model for regulatory purposes, the following information must be supplied:

1. a defined endpoint;
2. an unambiguous algorithm;
3. a defined domain of applicability;
4. appropriate measures of goodness-of-fit, robustness and predictivity;
5. a mechanistic interpretation, if possible."

Adequate training sets are necessary and the regulatory inventories should be updated regularly.

7.3.2.2 Information base

All information used to predict properties, the *Metadata* of the training sets and used models should be subsumed by the term information database.

"Periodic review and update of category assessments provides a means of incorporating new information, re-affirming or strengthening the scientific basis of the original hypothesis for the

category, and ensuring that the methodology associated with category assessments is continually improved" (*OECD 2014*).

It should be noted that *EFSA* has already done a lot of preparatory work to improve the risk assessment of metabolites. The following user requirements describe overall aspects of an optimal improvement process on basis of the *Aggregated raw data*.

The raw *GLP study raw data* is aggregated over several stages and IT-Tools until they finally find their way into the (Q)SAR models (see Figure 7).

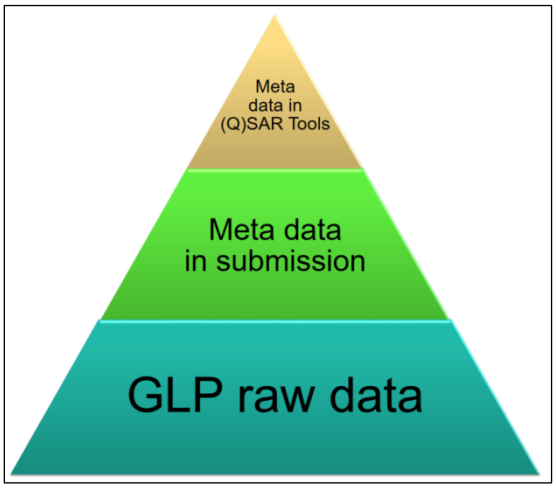


Figure 7: Aggregation level of metadata of metabolism studies

An advantage but at the same time a disadvantage is the multitude of available (Q)SAR tools and the (Q)SAR models / training sets as they require redundant maintenance and sometimes rely on the same standard definitions. This requires an ever-increasing high level of maintenance and will likely lead to inconsistencies between the tools. As such, the same data source has to fit multiple targets. If there is only a single overall schema of requirements, maintenance and interoperability is much more likely.

BfR does not develop (Q)SAR models, however *BfR* is fully convinced that international developers of (Q)SAR models will make use of published *Aggregated raw data* of the validated results of metabolism studies. International developers of (Q)SAR models should provide input for further requirements.

R 7.3-5: Sanitisation and confidentiality aspects should be clarified by authorities prior publication of the *Aggregated raw data* of the validated results of metabolism studies.

R 7.3-6: Authorities need to organise the publication process of the *Aggregated raw data* of the validated and *QA* checked results of metabolism studies, which should be an output of the evaluation process starting from the *Metadata* submitted.

R 7.3-7: (Q)SAR model creators, which provide their algorithms in a commercial manner, should have access to the published results of the validated and *QA* checked metabolism studies.

R 7.3-8: *OECD* would organise the improvement process of the *OECD* (Q)SAR-Toolbox models by including validated and QA checked results of metabolism studies.

7.3.3 Consider metabolites in the dietary exposure

The acute and chronic dietary risk assessment for pesticides is based on the exposure to all quantitatively relevant compounds in food and/or feed and by the toxicological characterisation of their effects.

The underlying dietary exposure assessment combines existing food and feed consumption data and residue occurrence data, provided that these residues are considered as toxicological relevant and included in the residue definition for risk assessment. Apart of treatment related metabolites, the occurrence of similar metabolites resulting from uses of other pesticide active substances or from uses in other regulated areas (e.g. biocides, fertilisers and veterinary drugs) may need to be identified and considered by experts.

The result of the dietary exposure assessment is the calculated chronic and/or acute intake of toxicological relevant residues (active substance and its metabolites, if relevant).

7.3.4 Metabolites considered in toxicology

Toxicological expertise for relevant endpoints is required for all active substance related compounds to which humans may be exposed.

The toxicological expertise required for two assessment aspects. One is the characterisation of the *ADME* properties of the parent substance incl. the toxicological characterisation of its metabolites and the other is the characterisation of the genotoxic potential of relevant metabolites.

7.3.4.1 Characterisation of the ADME properties

It is not possible to describe the scientific content for the characterisation of the *ADME* properties according this *Study Type* in this report. Here, the intention is to describe,

- which functions of an IT-Tool could help Evaluators in the assessment steps in a concrete legal act and
- which validated aggregated data could be useful for the improvement of (Q)SAR models

R 7.3-9: It should be possible to transport and import all needed Aggregated raw data of *ADME* studies into the IT-Tool.

R 7.3-10: *Evaluators* should be able to visualize the *Metabolic pathway* and the concentration time curves of different compartments (see chapter 9.4.10.2) with the help of the IT-Tool.

R 7.3-11: *Evaluators* should be able to use a flexible reporting module where the *Aggregated raw data* could be flexibly grouped (see chapter 9.4.16) with the help of the IT-Tool.

R 7.3-12: If calculations should be done, then there would be a need to include own scripts e.g. from R or python the IT-Tool. It is an open point how to document the algorithms used and should be discussed in a later project stage.

R 7.3-13: *Evaluators* should be able to calculate / check needed parameters (see chapter 7.4.2.4) with the help of internal functions of the IT-Tool.

R 7.3-14: *Evaluators* should be able to calculate concentration factors of measured values in a matrix in relation to another e.g. organ concentrations in relation to plasma concentrations with the help of the IT-Tool.

R 7.3-15: If calculated values should be stored, these values should be marked transparently.

R 7.3-16: The IT-Tool should manage all *Aggregated raw data* and *Aggregated result data* which are needed for an improvement of (Q)SAR models.

7.3.4.2 Check the metabolites toxicity

The toxic moiety may be unaffected, modified, or totally removed from the molecule in the process of metabolism/degradation. Alternatively, a new toxic moiety might be created. Toxicologists could be involved in the toxicological characterisation of relevant ground water metabolites or residue relevant metabolites.

An appropriate toxicological characterisation should be provided by toxicologists for each quantitatively relevant element of a *Set of substances*.

Within the assessment of ground water metabolites, identical properties are assumed for the metabolite, if the *Active Ingredient* (parent) has a relevant classification regarding the

- acute toxicity,
- repeated exposure toxicity,
- repro-/ developmental toxicity,
- carcinogenic toxicity

until evidence indicates otherwise.

If no identical properties could be assumed, there are two constellations for the applicants:

- Depending on threshold values, the data requirements demand to synthesise the metabolite and submit results of *in vitro* tests or
- To provide *in silico* data to characterise the expected toxicity.

If the calculated or measured concentration will be > 0.1 µg/L of the metabolites, a screening of the genotoxic potential of these metabolites is needed.

This will be done by evaluating the submitted *in vitro* studies or, if necessary, evaluating the *in vivo* studies according the list of required or recommended test guidelines and the *EFSA* scientific opinions.

The considered IT-Tool should support the following work steps:

R 7.3-17: *Evaluators* should be able to group the metabolites of the study according the *OECD* Guideline 194 (2014) by using (Q)SAR models. A group is characterised by a user defined name.

As a long term vision the (Q)SAR Tools should be usable as services. If such an interoperability is organised, it makes sense to consider the following user requirements for a possible IT-Tool:

R 7.3-18: *Evaluators* should be assisted to loop over a *Set of substances* and to start a (Q)SAR analysis in different (Q)SAR Tools as external services with different models based on different data sets and parameters.

R 7.3-19: The response results list of the (Q)SAR Tool contains a list of similar substances. It would be helpful to manage a user storable lists “List of similar substances” by selecting individual relevant substances from each of the (Q)SAR Tool response.

A 7.3-20: It would be helpful to manage (Q)SAR results of each substance from different (Q)SAR Tools according the *ECHA* guide (*ECHA 2016*) in the requesting IT-Tool (see Figure 8). A deeper analysis is needed to implement this function.

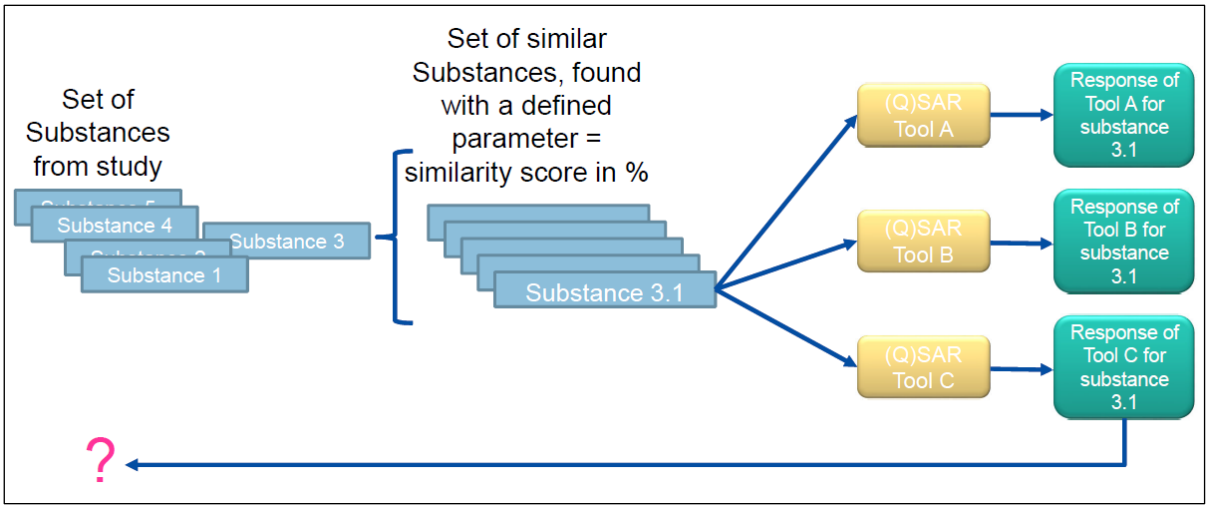


Figure 8: The usage of (Q)SAR Tools create information which should be managed

A 7.3-21: Would it be helpful to manage toxicity data for the metabolites (read across / predicted) in the requesting IT-Tool? A deeper analysis is needed for this function.

Moreover, if the calculated or measured concentration by lysimeter will be > 0,75 µg/L of the metabolites a refined and a cumulative risk assessment are needed.

7.3.5 Consider metabolites in the residue definition

Residue definitions are required for monitoring as well as for risk assessment.

A relevance assessment is performed for all metabolites detected in metabolism studies, and only those metabolites, which are quantitatively (exposure) and qualitatively (toxicity) relevant for humans, will be considered in the residue definitions for risk assessment. While one (or more) indicator compounds are sufficient for monitoring, the residue definition for risk assessment considers all compounds, which contribute to dietary exposure.

Necessary steps are the identification of treatment related metabolites, the evaluation of their quantitative relevance and the potential impact of similar metabolites from other pesticides (biocides etc.). Therefore, the IT-Tool should provide the needed functionalities in the user interface.

Since further "cold" studies may be additionally used to establish the residue definitions, jumps to external residue databases, such as *Ruedis*, should be possible as well (see chapter 9.4.9).

7.4 Applicants´ information packages

The following chapters contain high-level user requirements and some descriptions of main *Information packages*. Specific user requirements for data handling are described in chapter 8.

The submission of *Aggregated raw data* of metabolism studies became mandatory in the European context with the introduction of transparency regulations in the EU in April 2021.

7.4.1 GLP study raw data

The raw data of the metabolism studies are the data collected under *GLP* conditions in the laboratories (*OECD 1998*):

> *"**Raw data** means all original test facility records and documentation, or verified copies thereof, which are the result of the original observations and activities in a study. Raw data also may include, for example, photographs, microfilm or microfiche copies, computer readable media, dictated observations, recorded data from automated instruments, or any other data storage medium that has been recognised as capable of providing secure storage of information for a time period …"*

R 7.4-1: The *GLP study raw data* are subjected to *GLP* rules, but do not usually leave the laboratories. **These GLP data are not part of the needed information flow from applicants to authorities.**

R 7.4-2: User functions are needed to aggregate the *GLP study raw data* according the guidelines to write the *GLP study report*.

R 7.4-3: The *GLP* IT-systems of the laboratories should be able to
- assist the process step of writing the *GLP study report* and / or
- export the needed data into a data interface to write the *GLP study report* externally.

R 7.4-4: If an adequate external reporting/editing IT-System is necessary, a data interface should exist to import the aggregated information from the *GLP* IT-System.

R 7.4-5: If there is no adequate direct data interface to the *GLP* IT-System possible, an additional customisable data interface of the additional reporting/editing IT-System is needed to import CSV or spreadsheets at least.

R 7.4-6: The minimal request for the additional reporting/editing IT-System is, that an appropriate *User Interface* exits to record the needed data manually.

7.4.2 GLP study report

The *OECD* has described the principles of "Reporting of Study Results" under *GLP* conditions. The term "Final Report" is a synonym for *GLP study report*. A *GLP study report* is written by co-workers of the "Test Facility" and signed and dated by the Study Director.

The content of the *GLP study reports* is mainly subjected to *Evaluators* in the commissioning companies and *Evaluators* in the authorities. This information container is used to transport the achieved results unchanged from the test facility via the applicant to the authority.

The content and structure of the *GLP study report* is usually determined by the used test guideline. It is written by the "Test Facility" and contains the information in form of free text, tables and images. The *GLP* regulations define a basic structure of the *GLP study report*. The used Test Guideline contains the necessary information for the presentation of the data and its reporting.

The *OECD* has defined the principles of the life cycle of a *GLP study report* (*OECD 1998*) as follows:

> *"Corrections and additions to a final report should be in the form of amendments. Amendments should clearly specify the reason for the corrections or additions and should be signed and dated by the Study Director."*

At the same time, however, the *OECD* defined that a “reformatting” of the *GLP study report* does not constitute a correction, addition or amendment to the final report.

The *GLP* Test Facility” and the applicant (Study sponsor) are responsible to organise the process of the document life cycle.

The traditional users of the *GLP study report* consumed the content of the *GLP study report* by reading like a book.

In the following section, an attempt is made to outline which study report data usually arises in metabolism studies and are to be included in the *GLP study report* (see Figure 9).

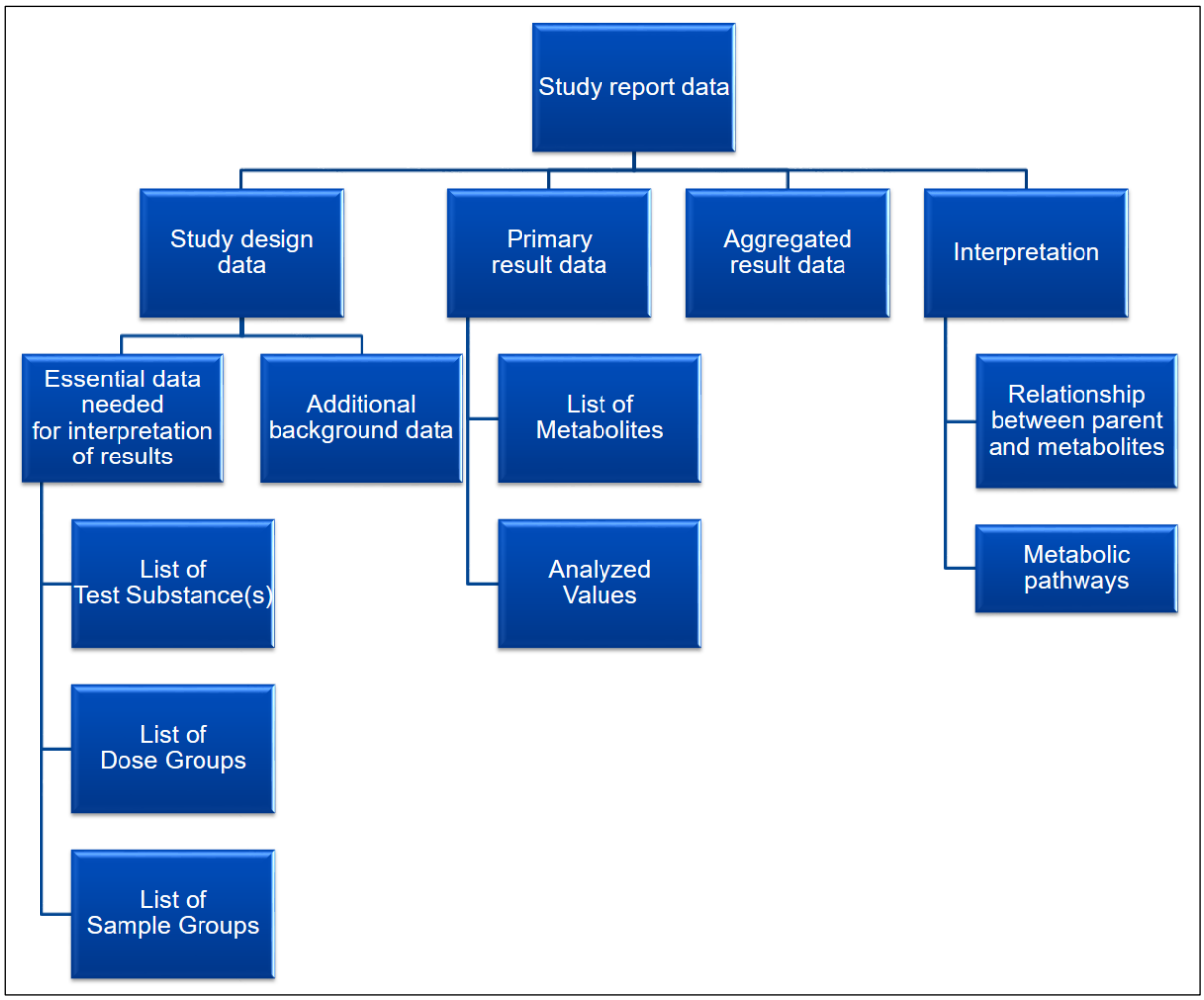


Figure 9: Hierarchy of study report data of metabolism studies

7.4.2.1 Study design data

The “Study design data” contains two groups.

R 7.4-7: The group of “essential study design data” is needed for grouping of the result data according the used *Test substance*, dose groups or sample groups.

R 7.4-8: “Additional background data” which are needed to understand the context of the study. This textual information cannot be applied for grouping of result data.

The “Essential study design data” are:

- **“List of Test Substances”** with all substance *Metadata*
  (e.g. Several variants of the radioactive labels of a substance can be used in one experiment; radiochemical purity and specific activity)
- “List of Dose Groups”
  - differences in dosing parameters (control, dose, dose replication, dose regime, dose interval, route of *Application*)
- “**List of Object Groups**” Normally the studies are investigating different groups of the *Object of investigation*. The reason for differences could be found in
  - the characteristics of the individual parameter of the *Object of investigation* (e.g. sex, age, strain, food but also crop, soil type)
- “**List of Sample Groups**”. Details on the sampling regime are important for the interpretation of the results (matrix, timing, sample interval, used methods).

R 7.4-9: If necessary, the *List of Dose Groups* could be modelled as a collection of individual *Object of investigation*.

All “Study design data” that do not belong in the group of “essential study design data” but are required by the technical guidance’s can be grouped together in the group of *Additional background data*. These *Additional background data* could not be used for grouping of the *Primary result data*. Some examples:

- Characterisation of the *Object of investigation*
  (e.g. biological, chemical, physical test conditions, origin, location, arrangement, size)
- Characterisation of the outside environment around the *Object of investigation*
  (e.g. environmental conditions)
- Characterisation of the storage stability
- Characterisation of the used analytical methods
  (e.g. capability of used analytical methods, extractability, fractionation, precision, sensitivity, limit of detection, recovery, characterization or identification of degradation products)

7.4.2.2 Primary result data

The following primary result data can be obtained from the experiment:

- *List of metabolites* of known and unknown identity (distinct peaks, not assigned to specific molecular entity)
- The “List of analysed Values” contains all analysed values with references to the corresponding elements of the
  List of Study Object Groups
  List of Dose Groups
  List of Sample Groups
  List of Substances
- Summarised observation of substances via the excretion pathways from the object under investigation (*Balance room*).
- Concentration-over-time pairs for substances in selected *Compartments* of the *Object of investigation*

R 7.4-10: The *List of Test Substances* and the *List of metabolites* should be merged to the “**List of Substances**”. The elements of this union list will be a grouping parameter for the result tables.

R 7.4-11: The *List of Dose Groups*, the *List of Sample Groups*, the *List of Substances* and the “List of analysed Values” are the source data for filtered data and for presenting the results.

7.4.2.3 Presentation of results tables

The compressed presentation of the analysed individual values in dependence of the

- *List of Study Object Groups*
- *List of Dose Groups*
- *List of Sample Groups*
- *List of Substances*

is a very complex task and quite challenging due to the immense amount of detailed information.

The *MetaPath* and the *DER/MSS-Composer family* are storing analysed values in a cells of a complex table structure. There are no functions in *MetaPath* to get an additional benefit of these stored analysis values than to read these analysis values as part of a static text table. It is impossible to display the analysis values in other groupings.

R 7.4-12: A new approach is needed to create flexible pivot tables by the author of the *GLP study report* to connect residue data to the chemical structures.

7.4.2.4 Presentation of metabolic pathways

R 7.4-13: A common type of the visualization of the results of a *Metabolism study* are figures of the *Metabolic pathway*.

7.4.2.5 Aggregated result data

Some aggregated result data could be calculated from the primary information e.g.:

- Maximum (peak) concentration
- Area under the curve (AUC)
- Order of the kinetic / transport process
- Half-life if the kinetic is of 1st order
- Clearance
- …

R 7.4-14: The proposed IT-Tool should provide basic functions of calculation of the most important aggregated result data.

7.4.2.6 Interpretation of the results

All detail results should be summarised, discussed and interpreted in context of the knowledge from other studies. These summaries are always textual interpretations, including text-tables. There are a lot of different aspects for textual interpretations.

R 7.4-15: It is necessary to manage textual summaries of the interpretation of the results for each aspect type.

7.4.3 Aggregated raw data

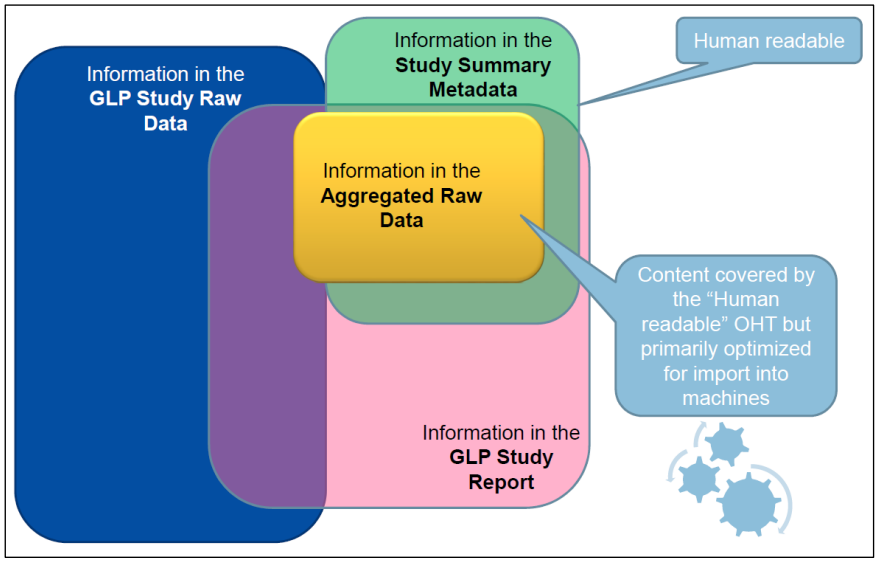


Figure 10: Relation of the information content of different objects

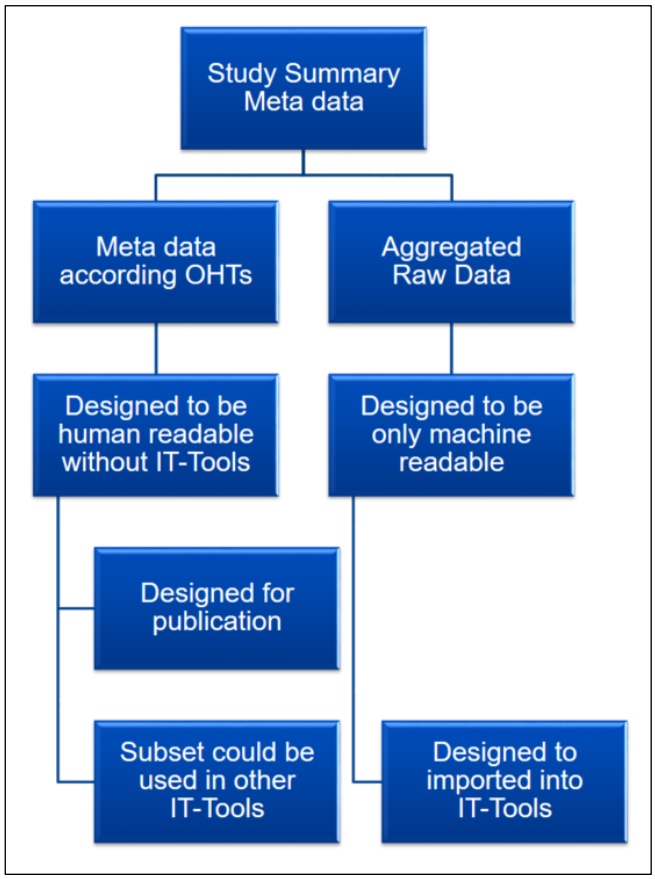


Figure 11: *Aggregated raw data* and *Metadata* according the *OHTs*

Please have a look at the relation (overlapping) of the information contained in *GLP study raw data*, *GLP study report*, *Study summary metadata* (*OHT*) and *Aggregated raw data* (see Figure 10 and Figure 11).

Aggregated raw data contain:

- NO information of the current application / legal act e.g. the reliability, data waiving, justification and the applicant's summary and conclusion,
- ONLY semantic duplicates of the data summarized in the human readable compilations of the *Study summary metadata* BUT NOT masked in word processing tables, but submitted as separate field values

The *Aggregated raw data* are not foreseen for publication, because

- they are not “human readable” without an adequate IT-Tool,
- they are not yet part of a quality assured curated reference collection of metabolism study *Metadata*.

R 7.4-16: The *Aggregated raw data* should be validated after the creation by the builder program.

R 7.4-17: The used builder program and its version, as well as the used schema definition versions, should be logged into the *Aggregated raw data* set.

R 7.4-18: The *Aggregated raw data* of metabolism studies should be extractable for import into an adequate IT-Tool.

R 7.4-19: Only validated *Aggregated raw data* should be imported into other IT-Tools.

7.4.4 Applicants study summary

A “Study summary” is a textual information container that provides the main information of a “GLP study report” in a human readable form.

There are different “Creator-Roles” in different steps of the legal process for study summaries. The first study summary will be written by the applicants when a legal act will be prepared and the *Metabolism study* should be part of the *Information package*, which will be submitted.

The *OECD* Harmonised Templates are standard data formats for reporting such information. An “Applicants study summary” has a life cycle and should also be revised if an amendment or corrigendum of a “GLP study report” is necessary. Furthermore, it should make a clear reference to the corresponding version of the “GLP study report”.

The “Applicants study summary” could be separated into two parts:

- The “**Pure Study Summary**” which contains all information about the used material, methods and the results. This “Pure Study Summary” should not contain conclusions referencing specific legal processes.
- The “Additional information in context of the legal act”.
  These are administrative data like:
  - the element “Adequacy of study” to indicate the adequacy of a (robust) study summary in terms of usefulness for hazard/risk assessment purposes depending on the relevant legislation
  - the flag “Robust study summary”
  - the flag “Used for classification”
  - the flag “Used for Safety Data Sheet (SDS)”

- the element “Reliability”
- the element “Rationale for reliability incl. deficiencies”
- the elements “Data waiving”, “Justification for data waiving”, “Justification for type of information”
- the block “Attached justification”
- the elements “Data access”, “Data protection claimed”
- the block “Applicant’s summary and conclusion

R 7.4-20: The content of the “Applicants study summary” of a study summary could only be the actual viewpoint of the applicant at the point in time of preparing the “Applicants study summary” for the current legal act according to actual data requirements. It is impossible to write this section at the point in time of writing the “GLP study report”.

7.4.5 Study summary metadata

If somebody should make a statement “*What is part of the Metadata of a Study summary and in which format?”,* then the answer depends on the purposes the user wants to consume this data (see also chapter 7.1.4).

- If the user “only” wants to store the data and publish them (depending on confidentiality) then almost any format is acceptable, since the publisher is not interested in the content of the information. Only the *Metadata* for the main search / access routes need to be defined.
- If authorities would like to build up other data collections for other user purposes, then additional *Metadata* are needed for other / or complex search / access routes.
- If authorities want to validate calculations of the applicant, *Evaluators* should be able to use these *Metadata* without complex transformations as input values for calculations.

Therefore, the term *Study summary metadata* should cover the user requirements of all process steps, which are needed in a legal act. The *OECD* Harmonised Templates so far cover a large part of the user needs for *Metadata* on the study summaries. In cases where new user requirements have been signalled, attempts were made to adapt the *OHTs* accordingly.

R 7.4-21: The provided *Study summary metadata* should be suitable if authorities would like to validate calculations of the applicant. *Evaluators* should be able to use these *Metadata* without complex transformations as input values for calculations.

R 7.4-22: Authorities should be able to create alternative tabular summaries from the reported results with the help of the *Study summary metadata*.

R 7.4-23: The *Study summary metadata* for a *Metabolism study* should cover the requirements defined in chapter 7.4.2.

7.4.6 Predefined study summary tables

Applicants obliged to fulfil different requirements for the presentation of aggregated data depending on the endpoint. The *OECD* provides multiple *Predefined study summary tables* per *OHT*.

R 7.4-24: It would be helpful to have internationally recognised table formats for summarising results of metabolism studies implemented on the *OECD* level as *Predefined study summary tables*.

R 7.4-25: While writing the *GLP study report*, the IT-Tool should be able to generate all other requested summary tables from the *Aggregated raw data*.

7.4.7 Endpoint summaries

Some authorities have created duplicate requirements for the presentation of the summary results: as endpoint summaries and as an attachment.

The *EFSA* has defined such a specific presentation format of the results of metabolism studies, the Appendix G “Template for presenting metabolism residues trials”. These spreadsheets are helpful in the period of the expert discussions because they present all the important information in condensed form.

R 7.4-26: An IT-Tools should be able to provide reports on a set of studies for different stakeholders in different formats. The *EFSA’s* Appendix G is only one report template.

7.4.8 Dossier

The *Dossier* is the compilation of the needed information of different studies and endpoint summaries for a concrete legal act. The data requirements are describing the content of the needed information and the published administrative guidance defines additional format requirements on the submission of the dossiers. The *Dossier* also has a life cycle.

The dossier container is the physical representation of a submission.

7.5 Authorities’ information packages

In principle, the outcome of a scientific assessment by an authority should not depend on the dossier format used at the time of submission, but only on the content of the documents submitted.

As the authorities parallelise the necessary evaluation processes in order to be able to prepare the opinions within the legal deadlines, standardised formats for applicant *Dossiers* and *Metadata* to be attached are a crucial prerequisite for timely processing. Therefore, authorities published administrative guidance documents to define format requirements on the submission of the *Dossiers*. These format specifications are meant to ensure that authorities are able to compose the needed evaluation reports according the guidance documents.

The completeness check process is work-intensive. Although *IUCLID* offers a comment function to parallelise the processes, this is not being used productively in any check procedure until today. However, the comments collected are only a temporary information package with regard to follow-up requests to the applicants and are therefore not considered any further within this report.

7.5.1 Authority study summary

Summarising a *GLP study report* is done by different additional actors in different legal processes with various templates suitable for multiple addresses (“Decision makers”). That means there are different study summaries derived from one single *GLP study report*. If one wants to refer to a specific study summary, one has always to refer to the legal process and to the creator of such summary e.g.

- *Applicants study summary*
- *RMS / EMS* study summary
- *EFSA* study summary

The *OECD* harmonised templates are an attempt to synchronise the different templates used worldwide on a semantic level. In most cases, a large part of the study’s descriptions will be identical. However, this high degree of similarity will pave the way for accusations of plagiarism, that authorities are only copying content of the applicants.

The processes suffer from the fact that the source of the text/the authors contribution is not verifiable at every level or that those could have been adopted intentionally after examination. Specific commenting boxes for the authorities indicate who had written which part but it will be difficult to read such assessment texts as the reading flow will be compromised. Alternatively, the authorities should have text processing functions at their disposal to clearly mark quoted text sections of the applicant’s text. It must be possible to edit flat texts, tables and graphics equally well via these copy/mark functions.

However, these IT functions could not be part of the considered IT-Tool, as those are user requirements for a text-processing tool independent of MS-Word.

The Study Summaries of the *RMS / EMS* will be part of an *Assessment report*.

R 7.5-1: *Evaluators* need the possibility to validate / recalculate results on study level of the submitted *Aggregated raw data*.

7.5.2 Assessment report

An *Assessment report*, which is written by a Rapporteur Member State (*RMS / EMS*) within the European pesticide evaluation procedures is a compilation of different report levels. The “B” chapters of volume 3 contain for each study the *Authority study summary* with the authority’s statement according to the acceptability / reliability and the applicability in the further procedure.

The following table shows the main metabolism related chapters of the European *DAR* templates.

Table 4: Main chapters of the European *DAR* templates (*EU DAR/CLH*) where results of the metabolism studies using radiolabelled test substances were presented and discussed including the mandatory Excel attachment „Appendix G” (*EFSA 2019*)

Number	**Chapter**
Vol 3 B 6.1.	Absorption, distribution, metabolism and excretion in mammals
Vol 3 B 6.8.1.	Toxicity studies on metabolites and relevant impurities
Vol 3 B 6.9.5.	Diagnosis of poisoning (determination of active substance, metabolites), specific signs of poisoning, clinical test
Vol 3 B 7 Appendix G	Residue data Template for presenting metabolism residues trials
Vol 3 B 7.2.	Metabolism, distribution and expression of residues

Number	Chapter
Vol 3 B 7.5.1.	Nature of the residue
Vol 3 B 7.6.1.	Metabolism in rotational crops
Vol 3 B 8.1.	Fate and behaviour in soil
Vol 3 B 8.2.	Fate and behaviour in water and sediment
Vol 3 B 8.3.	Fate and behaviour in air

8 Detailed analyses

8.1 OECD harmonised templates for metabolism studies

8.1.1 Owner

The *OECD* is the owner of the Harmonised Templates.

8.1.2 Information

The *OECD* Harmonised Templates (*OHT*s) are standard data formats for reporting information on studies regarding chemicals. Chapter 11.3 summarises 18 *OHT*s, which could found the pool of information regarding the transformation of chemicals because radio labelled substances could be used. However, only 13 of these templates can report details about the "identity of transformation products".

W 8.1-1: The possibility of transmitting detailed *Metadata* via the metabolism studies *OHT*s varies. This ranges from "not possible" to summary information on "degradation" to detailed data.

W 8.1-2: The *OECD* has not used a uniform principle for metabolism studies when defining *OHT*s. Only two of the metabolism studies *OHT*s are suitable for transmitting data on different radiolabelled substances.

W 8.1-3: There are no adequate Test Guidelines for four *OHT*s reported on the knowledge on the "Identity of transformation products". Without such Test Guidelines, the results of these studies on these knowledge areas are not internationally comparable.

In 2004, the *OECD* had formulated a guideline regarding the transmission of distinct fields and free text fields in the Harmonised Templates as follows (*OECD 2004*):

> *„The type of field included (fixed-field versus free text), should:*
> *i. be based on the needs of the reviewer and not the electronic technology requirements;*
> *ii. consider how often that field will be searched and by whom (i.e., searching is easier with fixed-fields than free-text fields);*
> *iii. consider the need for future manipulation of both text and numeric data in specific fields, e.g., extracting text blocks and/or numeric data into evaluation reports, performing statistical analyses, data mining, or other mathematical operations. For these tasks, fixedfields generally provide a greater ability than free-text fields;*
> *iv. Consider whether (and the degree to which) old, unstructured free-text data will be migrated into a field (migration to free-text fields is easier than to fixed-fields).“*

Please compare the corresponding figure in chapter 11.6.

It seems that the first condition, ***“be based on the needs of the reviewer and not the electronic technology requirements”*** has been increasingly forgotten over the years, and more and more subject-based *Metadata* has been incorporated into *OHTs*.

The third condition, the need to manipulate numeric values and to calculate with them may be valid for *Primary result data* but not for *Aggregated raw data*. It would not make sense to

use the submission transport IT-system to create pivot tables in the phase of writing the *GLP study report*. During the dossier submission process the applicant provides the compiled and condensed data which emphasises the applicants point of view based on their raw data interpretation. Nevertheless, *RMS / EMS* should have access to the raw data in order to analyse and in some cases recalculate the applicants raw data interpretation. Also, the *RMS / EMS* should be able to create their own grouping. Therefore, the study raw data should be included as an attachment and not embedded in *IUCLID* front end, in order to not compromise the dossiers readability.

BfR is fully aware that the *OECD*, also on the initiative of *BfR*, has extended some *OHT*s to include aggregated raw data. It was not analysed which *OHT*s this concerns in addition to the OHT85-5.

O 8.1-4: The fact, that the transfer of *Aggregated raw data* within the *OHT*s is generally possible does not mean that this implementation is a) actually used in the procedures b) that users consider that this implementation is helpful in the *OHT*s and c) that corresponding data users also "consume" these aggregated raw data. Only after a representative survey on the degree of use and the user satisfaction an assessment can be made regarding the success or failure of this implementation.
The Publication of a standard itself is not enough to be declared as a success.

8.1.3 Functionalities

The benefit of the *OHT*s should be for developers and maintainers of databases on chemicals the usage of the collected data in other IT-Tools.

W 8.1-5: The lack of an approach to standardise information on the metabolic behaviour leads to a high diversity of the corresponding *OHT*s. This can be a disadvantage for using the *Metadata* of metabolism studies for modelling.

8.1.4 Life cycle and process

A continuous improvement process has been implemented for the *OHT*s. The process is described on the *OECD* website[2]. As soon as a need for revision has been identified, this *OHT* could be revised. However, the annual capacity to revise the templates is limited.

W 8.1-6: The *BfR* had not found an answer to the question: “Is it possible and is it allowed, that study summaries which are created with different versions of *OHT*s would be compiled into one dossier?”

W 8.1-7: There is no version number of the *OHT* schema definition in the XML data file, which was used at creation time for this XML data file.

W 8.1-8: No statement was found on the backward compatibility of XML files for import into *IUCLID* if they have not been created according to the current schema definition outside *IUCLID*.

Example OHT85-5

[2] https://www.oecd.org/ehs/templates/overview-previous-templates.htm

The OHT85-5 has no relevance for metabolism studies. However, this *OHT* was the first, which should be improved to transport *Aggregated Raw Data*. It is a very helpful example to analyse the way of "The wish to receive residue data from applicants".

A first XML-interface XRUEDIS was developed in 2003 in Germany to reduce the time-consuming manual data input into *Ruedis*, the German database of residue results from controlled residue trials. However, no German XML-interface could be established under consideration of the development of *OHT*. So the idea of XRUEDIS was not supported by the *OECD* member states (*RUEDIS 2006*).

15 years later, a status was reached that the OHT85-5 would - in principle – be able to transmit individual residue values. But

- these *OHT*s could only be completed by applicants, who themselves have residue databases by adequate reporting tools,
- the *IUCLID* user interface is destroyed by so much textual references between the block "Materials and methods" and the block "Results and discussion" as well as the count of repeatable blocks,
- only reports could interpret this bulk of raw data into a human readable format and no realistic scenario could be seen to get such a *IUCLID* report,
- the internal references of residue data of products with a 2nd active ingredients are implemented inadequate and
- a publication of such data with the existing dissemination procedures would make no sense for the public.

These disadvantages are substantial, although perhaps in the future a data flow can be established through this interface.

The current approach of the *OHT*s makes a clear distinction between the blocks "Materials and Methods" and "Results and Discussion". This makes sense if the reported results refer only in a semantic textual way to specific used methods, trials, samples etc.

W 8.1-9: The current approach of the *OHT*s is only useful to a limited extent for transporting aggregated raw data. In the case described below, the restrictive definitions for the *OHT*s are even reasons for non-optimal data structures, which result in an overloaded user front end of *IUCLID*.

Example: The analytical methods are defined in OHT85-5 in the block Analytical Methods. Each method should have a MethodID[3]. This MethodID will be re-entered in the "Results and Discussion" block on the level of an analysed value[4] otherwise a proper contextual linkage would not be possible.

P 8.1-10: The current schema definitions of the *OHT*s should be improved if textual semantic references of fields could be replaced by schema internal key references. The goal should be to identify internal data inconsistencies by validating the data against the *OHT* schema.

P 8.1-11: The *IUCLID* user front end should support this relation by editing the field value only once a time. The depending field should provide only the already defined field values e.g. in a list of values.

[3] XPath in XML: ResiduesInRotationalCrops/MaterialsAndMethods/AnalyticalMethods/AnalyticalMethod/entry/MethodID

[4] XPath in XML: ResiduesInRotationalCrops/ResultsAndDiscussion/SummaryOfRadioactiveResiduesInCrops/SamplingAndResidues/entry/ResidueLevels/entry/MethodID

W 8.1-12: After more than 18 years of efforts, the result regarding OHT85-5 is not satisfactory. In retrospective, sending a separate XML with the aggregated raw data would have been the more efficient approach for the submission of residue results from controlled residue trials.

Until today, there is no experience with the data flow of *Aggregated Raw Data* in the residue area using the *OHT* 85-5!

8.1.5 Publication and documentation

On the *OECD* website, for the *OHT*s, *OECD* has organised:

- a publication system for the
  - finalised *OHT*s (Word table),
  - predefined tables & executive summaries
  - schema definition files
- a textual description of the history of the revised *OHT*s

The following statement on the *OECD* website (*OECD 2021*) should be subjected to critical review:

> "*The* OECD *Templates are also not prescriptive as to the order of appearance of any data entry fields or how the fields are technically implemented, as long as this does not affect the harmonised and agreed upon data exchange forma*t."

because the sequence is defined in a static way by the schema definition file (XSD) published by the *OECD* itself.

P 8.1-13: The *OECD* should discuss about the relevance of the upper statement regarding the sequence of the *Metadata* in the *OHT*s. This statement should either be expanded and justified or deleted if necessary.

W 8.1-14: The *OECD* does not provide an archive documenting no longer valid/outdated *OHT* scheme definitions.

8.1.6 Interoperability / output

Each *OHT* should be provided with textual description and a corresponding schema definition file (XSD). According the schema definition file, a study endpoint record could be written in XML syntax.

W 8.1-15: There is no common validator tool available, which could be used to validate a study endpoint record XML file against the corresponding XSD schema definition file.

W 8.1-16 The existence of several concurrent reporting formats, provided in the *DER/MSS-Composer family*, in *MetaPath* and in the *OHT*s are a barrier for harmonisation.

8.2 Comparison of OHT58 „Basic toxicokinetics" and DER Composer

A separate report was written regarding the comparison of OHT58 „Basic toxicokinetics" and the DER Composer". The full report was published on the *BfR* website (*BfR 2021*).

The following conclusions were made:

8.2.1 Semantic aspects

There is a large semantic overlap regarding the possibility to transport study summary information of metabolism studies via the DER composer schema and via the OHT58, created by *IUCLID*. For details please have a look into the full report.

Most of the additional elements are from type “Legal act”. From these elements, there is no threat of loss of information regarding the transmission of the study summaries.

The reason for these differences lies in the historical development of both templates. The template for DER-composer should be comparable to the *OECD* template OHT58 in terms of content, but at the same time should correspond to the specifics of *USEPA* and PMRA.

However, if one analyses the significance of the additional schema elements of the OHT58, there are also fields of the generic approach of the chemical legislation.

Based on this analysis, it can be concluded that both templates can be considered as semantic equivalent for study summaries for metabolism studies in terms of the content submitted.

8.2.2 Weaknesses of the two templates

No systematic analysis was performed by *BfR* to identify weaknesses in both templates. The following weakness was identified, when redesigning the database with respect to the element "Method of analysis".

The user can describe the analytical method in the free field "Details on dosing and sampling" in OHT58 according the free text template description:

> *Complete description including: limit of detection and quantification, variability and recovery efficiency, matrix used for standard preparations, internal standard*

The user of the DER Composer schema should be accompanied with a description for the method in the free text field “B. Study design and methods / 2. Dosing and sample collection / Sample Handling and Preparation”.

W 8.2-1: No values for "Limit of Detection (LOD)" and "Limit of quantitation (LOQ)" could be stored for the individual methods in the different matrices.

A 8.2-2: A deeper analysis is needed to describe the analytical methods in a way that avoids misunderstandings for an interpretation of the measured values (*FISK 2021*).

8.2.3 Aspects of format and supporting tools

The DER composer was used as the comparative model. Only those aspects were summarised in Table 5, which format usage or implemented supporting tools were reasons for a different quality of user functions.

The Table 5 contains a semi-quantitative rating of the aspects.

Based on this analysis, the DER composer concept has considerably more advantages than disadvantages compared to the OHT58 in *IUCLID* in terms of implemented formats and supporting tools.

Table 5: Semi-quantitative comparison of format and supporting tools aspects of OHT58 (IUCLID) and DER (Composer)

DER Chapter	Aspect	Comment	Score*	
			OHT58	DER
I. General Info	Test Material Purity	The DER composer has no validation of the numeric value implemented.		-1
II. Material and Methods/ A Materials	Test Compound	IUCLID uses a high sophisticated substance model with the levels: Test material ➔ substance ➔ Reference substance Metabolism studies were carried out with substances, not with products. Only a simple substance model is needed.	-1	1
II. Material and Methods / A Materials Test Compound/ Radiolabelled test material	Radio – labelled purity Specific activity and unit	The DER composer has no implementation of numeric value validation. No units picklist is implemented.		-1
II. Material and Methods / A Materials Test Compound/ Radiolabelled test material	Structure	It should be possible to create a structure for each different radiolabelled position. IUCLID has only one reference to one test material. The OHT58 has no repeatable block with a reference to a radiolabelled "Reference substance" characterised by its SMILES notation code. Regarding "Radiolabelled", IUCLID could only store values like: „Yes, No, other". This element is not helpful. The structure characterisation of the radiolabelled test material and the 2D structure editor are the most important advantages of the DER composer.	-1	3
II. Material and Methods / A Materials Test Compound/	Physicochemical Properties	IUCLID would be able to refer specific sections for the phys-chem properties. The DER data model could produce inconsistent data values for one compound via data input of different studies. It is not clear if evaluators need this information for the interpretation of metabolism studies.		-1
II. Material and Methods / B Study design and methods	Table 1a – Group arrangement	The detailed summary of the treatment groups is essential for all interpretations. IUCLID is very open for a textual description of the treatment groups. The DER uses an input template, which is suitable to describe the dose groups and the study design regarding other parameters. The dose route should be converted into a picklist value.		2
II. Material and Methods / B Study design and methods	Table 2a – Sample collection	The detailed summary of the sample collection is essential for all interpretations. IUCLID is very open for a textual description of the sample collection. The DER uses an input template, which is suitable to describe the sample collection for each defined matrix. The metadata for the sample collection are free text. There are dependencies between "Appendix 1a" and "Table 1a - Group Classification" that prevent a sample collection of a specific matrix to be used as a time series.		-1
III. Results / A Pharmacokinetic studies	Absorption	The Total Radioactive Residues (TRRs) are basic values, which could be used in toxico kinetic models.		1

DER Chapter	Aspect	Comment	Score*	
			OHT58	DER
		IUCLID uses a predefined table in a Rich-Text field, which could be modified flexibly. The DER offers to create flexible tables. The structure description is similar to HTML and would thus be very transparently usable for potential interfaces to models.		
III. Results / A Pharmacokinetic studies	Excretion	Statements on the elimination balance are essential. IUCLID uses a predefined table in a Rich-Text field, which could be modified flexibly. The DER offers to create flexible tables. The structure description is similar to HTML and would thus be very transparently usable for potential interfaces to models.		1
III. Results / B Metabolite characterization studies	Distribution of parent and metabolites in matrices	These values are in addition to the absorption TRRs. IUCLID uses a predefined table in a Rich-Text field, which could be modified flexibly. The DER offers to create flexible tables. The structure description is similar to HTML and would thus be very transparently usable for potential interfaces to models.		1
V. Appendix	Appendix 1a	The description of the “Dose groups” is needed for all calculations and reports. It is good that the DER composer defines such an important table. There is an algorithm available to subsume test-numbers and create corresponding rows for table II. Material and Methods / B study design and methods.		1
V. Appendix	Appendix 2	The list of metabolites detected, their structure, if applicable, and the presumed relationships to the applied labelled substances should be the highlight of a metabolism study summary. IUCLID has only a text field implemented for this list. The DER composer provides the 2D structure editor for the parent and for the metabolites.	-1	3
Data Container	Metabolism study export file	IUCLID exports the study as an i6z container with XML files and attachment files for all object types. The DER XML file contains all information in one file. This file is “self-contained”. An external editor could modify this XML file.		2
Overall sum			-3	11

*

Textual interpretation	Score
A deficit in functionality	-1
Default	
A benefit in functionality	1
A significant advantage in functionality	2
An important advantage in functionality	3

8.3 MetaPath

The *MetaPath* objective was described in 2012 as follows: *"The MetaPath knowledge base was developed for the purpose of archiving, sharing and analysing experimental data on metabolism, Metabolic pathways and crucial supporting Metadata. The MetaPath system grew out of the need to compile and organize the results of metabolism studies into a systematic database to facilitate data comparisons and evaluations*." (*Kolanczyk 2012*).

LMC described 2021 objectives for the improvement of *MetaPath* as follows: *"Currently, efforts are underway through an Organization for Economic Co-operation and Development (OECD) work group to extend the use of DER Composers as harmonized templates for rat metabolism, livestock residue, plant residue and environmental degradation studies. These efforts are being coordinated by the MetaPath Users Group (MUG), a subgroup of the OECD Working Group on Pesticides. The main collective goal of the MUG is to harmonize the QA protocols and checklists developed for efficient standardized and accurate data entry on metabolism or catabolism."(LMC Metapath*)

The *EFSA* published a document "Reporting structured results of metabolism studies on rats, plants and livestock" (*EFSA 2021*) with a description of the current European process steps.

Please note that due to this publication of the *EFSA*, *MetaPath* plays an important role in this information flow, although several weak points were already identified in 2020.

The importance of *MetaPath* will increase by using this archive as a data source for the *OECD* (Q)SAR-Toolbox.

The analysis of weaknesses contained in the following chapter is intended to identify potential for improvement. This analysis was made on basis of *MetaPath* version 5.1.0.39 (beta). *BfR* did not directly contact the software producer *LMC*.

8.3.1 Owner

The software producer *LMC* has marked *MetaPath* with a copyright information. However, there exists no public available governance model.

W 8.3-1: It is not transparent, which rights the commissioning authorities have, which had sponsored the development of *MetaPath*.

W 8.3-2: The rights of the data donors of a public version of *MetaPath* are not described.

W 8.3-3: The obligations of someone who intends to use information of *MetaPath* for own commercial interests were not described.

8.3.2 Contained Information

Although data entry of study data is possible in *MetaPath*, this function was outsourced to specialised input programmes like the *DER/MSS-Composer family*. However, the contained information of *MetaPath* is only a subset of information collected by programmes of the *DER/MSS-Composer family*.

A logical data model of *MetaPath* is not publicly available. Therefore, a detailed analysis of the current database implementation was made by *BfR* (see chapter 8.3.4).

W 8.3-4: *MetaPath* does not store *Aggregated raw data*. The analysed values are stored in cells of a complex but flexible table structure, organised similar to HTML tables (see Figure 12). This is the origin of various weak points.

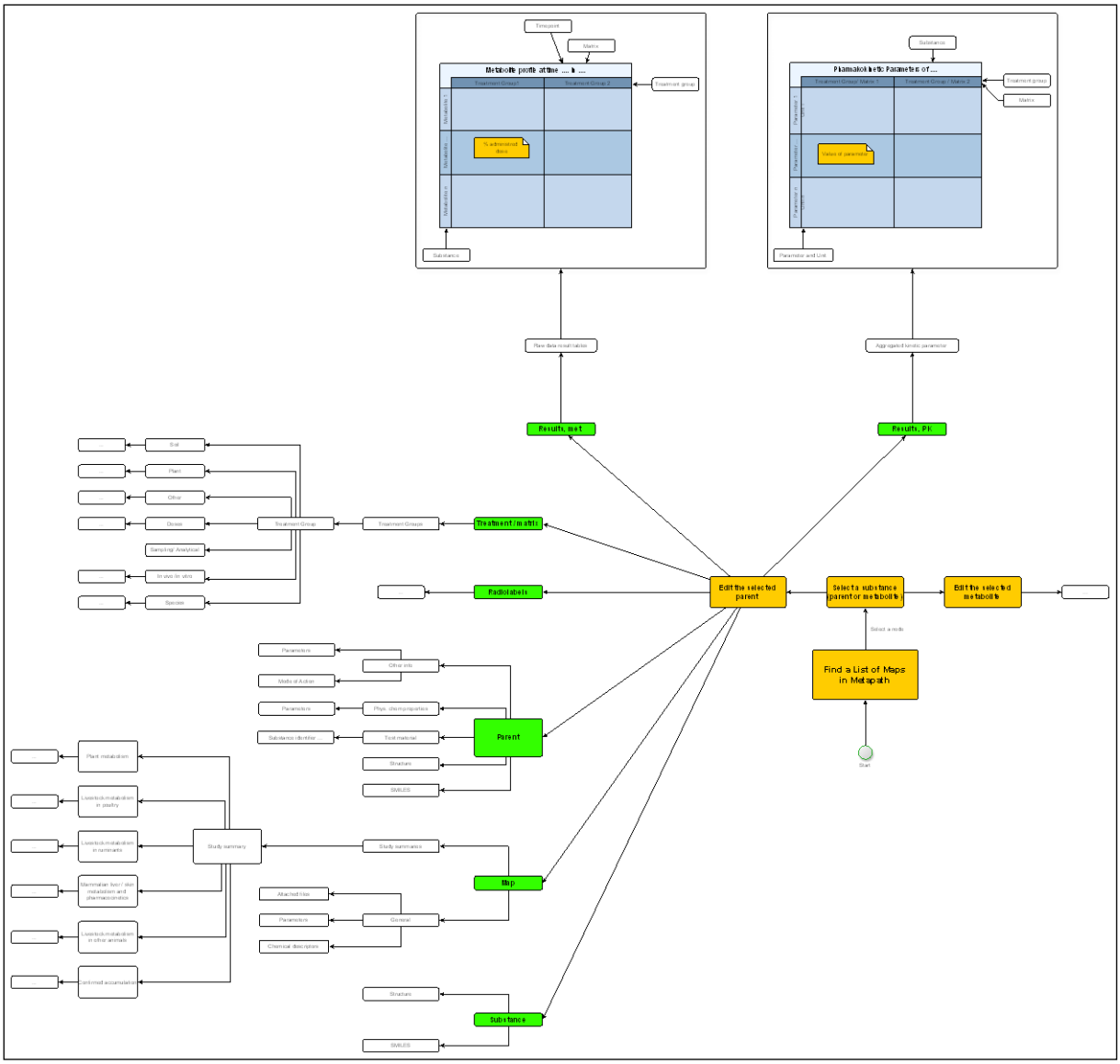


Figure 12: Example of the program modules hierarchy of *MetaPath* (see also 11.2 MetaPath_Modules.svg)

Another point is a missing picklist concept to code the objects itself and the matrix under investigation. These *Metadata* are stored in the labels of the rows / columns of the above-mentioned complex data tables.

LMC provides different databases for *MetaPath* metabolism trees on their web site. In addition, there are further collections from other authorities such as *EFSA*.

W 8.3-5: There is no concept to compile all metabolism information into one information backbone. The current *MetaPath* tool would not be able to manage such a compiled information volume with a satisfied question-answer time response for the user.

W 8.3-6: It is not documented, which part of the collected information of the *DER/MSS-Composer family* will not be imported into *MetaPath*.

W 8.3-7: The *MetaPath* approach of the flexible data tables is a hurdle for using the measured data values for calculations or for reports in other context.

W 8.3-8: The *MetaPath* approach of species / matrix *Metadata* storing in the data tables is an obstacle for searching data efficiently.

8.3.3 Functionalities

MetaPath is an information archive (database and an IT-Tool) to store background information and results from metabolism studies. *MetaPath* is one of the most prominent public data collection which can be used to search for metabolism pathways of pesticides in different matrices using substance *Metadata*. The most imported function is the visualisation of these metabolism pathways and to read in the background information of those depicted metabolism studies.

It is not the purpose to generate additional knowledge from this archive but it is a very helpful tool to search for details and to compare results between studies.

However, it is also possible to use *MetaPath* to add / modify / delete information on parent substances, metabolites, and studies which were imported via XML files.

W 8.3-9: No documentation is available regarding a role concept of *MetaPath*.
The normal login will be done as “administrator” with the world wide identical password “ccr” (see Figure 13).

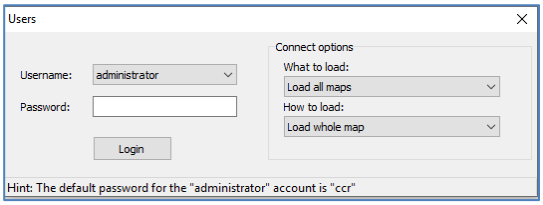


Figure 13: The *MetaPath* default administrator password

W 8.3-10: It is not easy to understand the hierarchy of the *MetaPath* functions because there are a lot of different ways to activate the functions e.g. the main menu, different additional tab menus, right click context menus and buttons.

W 8.3-11: The initialisation phases to open *MetaPath* is too slow. It needs 90 seconds for 715 maps.

W 8.3-12: The visualization of *Metabolic pathways* might be misleading by showing the same metabolite several times under each possible parent.

W 8.3-13: *MetaPath* has input functions, which are already implemented in the *DER/MSS-Composer family*. No documentation was available which additional data are entered by the *DER/MSS-Composer family*.

W 8.3-14: No documentation was available whether the *MetaPath* input modules assists the same functions compared with the *DER/MSS-Composer family* e.g. whether the same rules for validation are integrated. There is a high risk to get the same level of internal QA checks in the different user front ends.

W 8.3-15: This technological approach is a real risk in terms of data integrity. No information, added / modified in *MetaPath* will be reembedded to the original XML file.

W 8.3-16: *MetaPath* does not facilitate common keyboard shortcuts like <Ctrl>+<v> for paste the content of the clipboard into the *MetaPath* field. Manual typing of each values in tables is very error-prone.

W 8.3-17: *MetaPath* does not posess a function to copy messages into the clipboard e.g. the textual representation of a query like “Q1: Chemical name containing ‘Meco’”.

It is possible to generate a Quality Assurance Report for metabolism maps. Status information for the types “Completed” and for “’Reviewed” were shown for treatment groups and for the structure information. A summary will be generated with the following meaning:

- Not completed
- Completed on Level 1 (Basic)
- Completed on Level 2 (Basic + *in vivo/in vitro*)
- Completed on Level 3 (Complete)

MetaPath offers the following search options:

- Chemicals
- Reactions
- Similarity
- Tables
- Transformation

It is possible to store search queries.

W 8.3-18: The frame to compose a search query is cumbersomely designed. The message “The search clause is not correctly complete!” does not provide enough feedback for improving the query.

W 8.3-19: The export / import of complex search queries is not possible.

W 8.3-20: The initialisation phases to open a search for transformations, chemicals, similarity or tables is too slow and needs 30 to 60 seconds for 715 maps.

Applicants could use *MetaPath* in the phase of the development of new *Active Ingredients* for their own documentation.

W 8.3-21: The idea of *MetaPath* and *MetaPath* as an IT Tool are not self-explanatory. The learing curve to master the tool is steep.

W 8.3-22: There is no function to export all needed information into an alternative format like the Appendix G (Excel) for the European assessment processes.

8.3.4 Database implementation

8.3.4.1 Database management system

MetaPath works with Firebird as the underlying database management system. According the DB-Engines website (Knowledge Base of Relational and NoSQL Database Management Systems *DB-ENGINES*) Firebird is on place 17th of the relational database management systems.

W 8.3-23: As long as *MetaPath* is bound to Firebird, it is not realistic to look for software providers other than *LMC*. This is a very high risk for a project, which is supposed to be future-oriented.

W 8.3-24: The Firebird technology is a challange for an eventual interoperability with other systems.

In *BfR* point of view, as long as *MetaPath* is bound to Firebird, it is not likely that the current system can be further improved. In OECD QSAR Toolbox project, *LMC* has shown, that a conversion to a newer database management system (such as PostgreSQL) is possible.

W 8.3-25: It seems that *MetaPath* supports only the Windows platform.

W 8.3-26: The *BfR* is not informed of any server implementation of *MetaPath* in a multi-user environment.

8.3.4.2 Implemented database model

A very open substance concept is used by *MetaPath* to manage information about metabolites, although the identity of these metabolites may not be clarified until later in the course of a number of subsequent studies.

An attempt was made by *BfR* to understand and to document the database model of *MetaPath*.

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

You will find this assumed database model of *MetaPath* in Figure 14. Pay attention that some contexts could be misinterpreted in detail. However, this compilation is a necessary step in order to design a migration of the data for any follow-up project.

Table 6: Used symbols in Figure 14

Symbol	Meaning	Symbol	Meaning
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 interpreted on basis of an existing index		
	Relation based on foreign keys between used tables		Relation based on foreign keys but between a used and an empty table
	Relation interpreted by substrings in the fieldname e.g. ATTACHEDFILES.FILEID= MAPFILES.FILEID		Relation interpreted by substrings in the fieldname but between a used and an empty table e.g. TRANS.TRID= TRANSNAMESLIST.TRID

Figure 14 is provided as a vector graphic alongside this report in order to zoom into the details without sacrificing readability (see also 11.2: MetaPath_ER_Schema.svg).

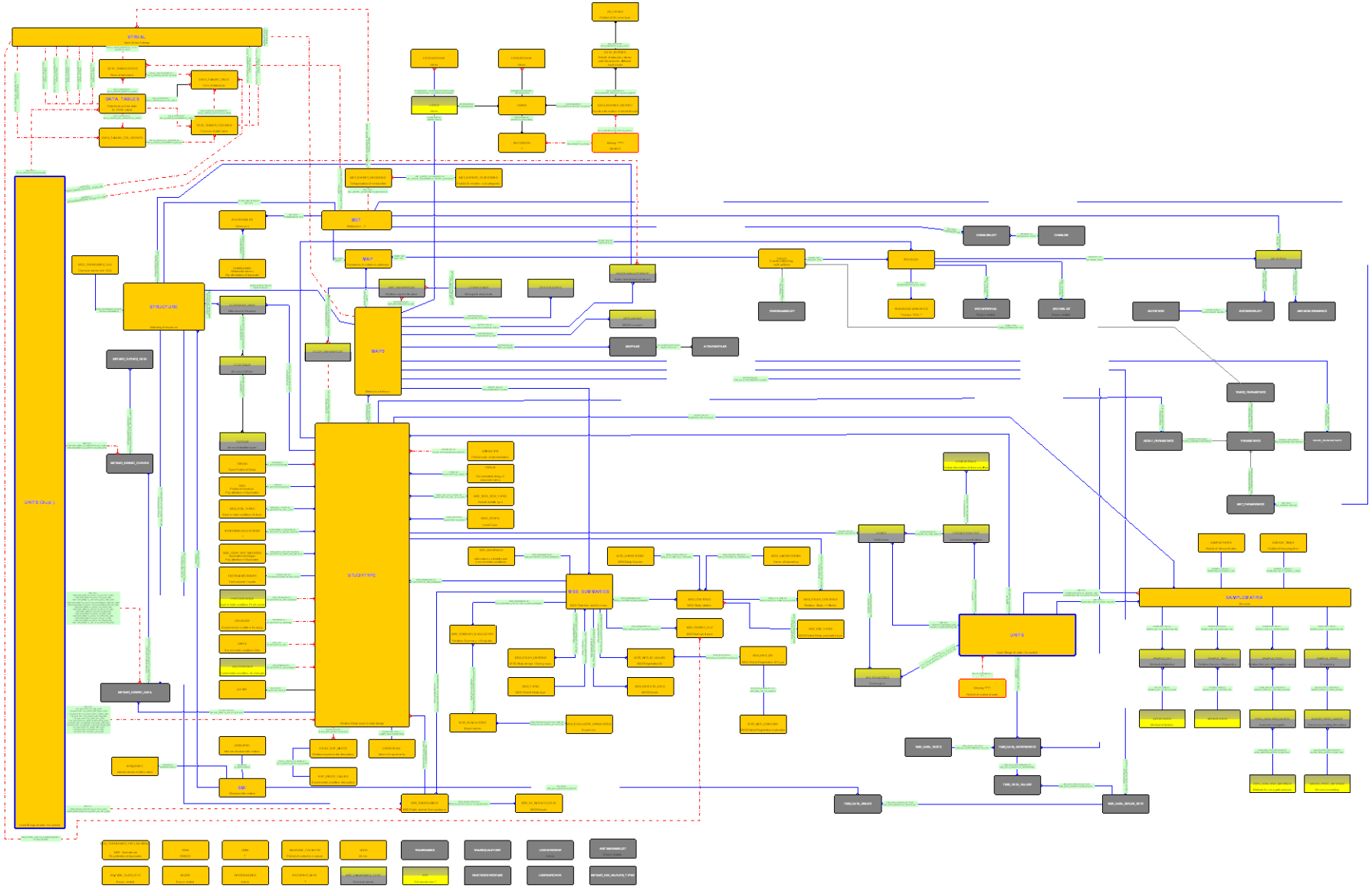


Figure 14: Database model of *MetaPath*

The current data model of *MetaPath* organises open picklists for each field separately. A generic administration of picklists was not foreseen for *MetaPath*. The *DER/MSS-Composer family* as the input programmes, do not make use of the *MetaPath* picklists, they use own module-intern programmed picklists. With such an inadequate coding via open picklists, the quality of the captured *Metadata* is insufficient for data analysis across the whole database or faceting of the result lists after a full-text search.

Findings based on the data model:

W 8.3-27: The database implementation indicates that several developers with different preferences were involved in this project over time. There are several parallel naming concepts for field names.

W 8.3-28: The *MetaPath* database uses duplicate data structures according to the DER Composer and the MSS Composer. No attempt was made to implement a uniform, expandable concept. Many database tables are supported only by the DER-Composer (e.g. DETECTIONS).

W 8.3-29: Many database tables are empty. Therefore, a lot of functions / program lines are never used.

P 8.3-30: An improvement of the *MetaPath* database should be combined with
- a generic redesign of the used database model,
- a purging of unused database tables and
- using of clear naming conventions, used consistent all over the project.

8.3.4.3 Data analysis example

The data model for the use of units seems to be sophisticated in *MetaPath*. The table UNITS contains fields such as:

- internal ID
- decoded value UNIT and
- FID_CONTEXT as a grouping attribute.

A further investigation of the UNITS table regarding the used units for the age (FID_CONTEXT=2), reveals what BfR considers a deficit in the administration of the used units.

```
SELECT FID_AGE_DIM, UNIT, count(*)
  FROM STUDYTYPE, UNITS
 WHERE STUDYTYPE.FID_AGE_DIM = UNITS.ID
   AND STUDYTYPE.FID_AGE_DIM is not null
   AND upper(UNIT) like '%Y%'
 GROUP BY 1, 2
```

Table 7: Units referencing the semantic content “years” for a field “Age” in the collection “Public MetaPath Db_EFSA project.MTB”

FID_AGE_DIM	UNIT	COUNT
76	years	408
273	Years	32
287	yrs	75
291	year	102
349	year old	7
486	yr	5

FID_AGE_DIM	UNIT	COUNT
722	2 years old	10
761	(2-3 years old)	7
792	year,11months and 1year,7months	7

Findings regarding the picklist of units for the field “Age”:

W 8.3-31: The *MetaPath* concept for the units of a field “Age” is inconsistent.

W 8.3-32: There are fields available in the database, which are not supported by the actual composers as an input field (e.g. STUDYTYPE.AGE_MAX)”.

W 8.3-33: The DER Composer supports separate fields for the value and the unit of age of the test animal. The MSS-Composer for livestock supports only one field for the age value and the unit.

W 8.3-34: The concept of unit picklists actually implemented in *MetaPath* fails. No administration and quality check of coding the units is conducted.

W 8.3-35: A subsequent implementation of a central maintenance of picklists of the current project would be an unrealistic illusion, since five input programmes would need to be adapted and a cyclic distribution of the picklists to all installed *MetaPath* and *DER/MSS-Composer family* instances would need to be set up.

8.3.5 Life cycle and processes

MetaPath has an intensive life cycle. The version number of *MetaPath* could be used as an indicator for that fact. There are some events which are triggering the life cycle:

- *MetaPath* depends on all modifications in the export data interfaces of the *DER/MSS-Composer family*. The *MetaPath* data model and the import functions should follow these modifications.
- Modifications or new user requirement regarding *MetaPath* are an additional source for further software versions.
- It is possible, that the usage of *MetaPath* sub-information in the *OECD* (Q)SAR-Toolbox and the needed data interface would have an influence on *MetaPath* itself.

W 8.3-36: There is no governance model for the further development of *MetaPath*. Who is responsible to coordinate the different user requirements and who is allowed to decide about priorities?

W 8.3-37: Furthermore, *LMC* supports “custom versions” of *MetaPath*. Therefore, there is no single version of the tool *MetaPath*. The different “custom versions” of *MetaPath* require additional resources for each new version.

LMC had announced in the *MUG* meeting in 12/2021, to end the support of “custom versions” in 2022.

8.3.6 Publication and documentation

W 8.3-38: *LMC* offers a web platform for publication of the software. No documentation and training material is available.

W 8.3-39: There is no user documentation integrated into the *MetaPath* tool.

W 8.3-40: There is no version history of *MetaPath* published.

W 8.3-41: No communication was organised to inform users regarding updates of *MetaPath*.

W 8.3-42: No system documentation (e.g. the used data model) of *MetaPath* is publicly available.

8.3.7 Interoperability / output

It is possible to import

- DER XML-files,
- MSS XML files and
- parent compounds from an Excel file into *MetaPath*.

Some simple export functions are implemented to create text files for different components of *MetaPath*. It should be possible to export:

- parent data to tab delimited or SMI-files,
- metabolite data to tab delimited or SMI-files
- treatment groups data,
- time dependent data to a tab delimited file.

It is possible to create subset *MetaPath* databases from a selection of metabolism maps. However, it is not possible to merge different *MetaPath* databases into one database.

Via non-publicly documented functions, a data transfer to the *OECD* (Q)SAR Toolbox is possible.

W 8.3-43: No documentation regarding the implemented import functions is available. Without documentation, it is not possible to understand messages of the importer if errors were found. So it is not transparent, which modifications are necessary to import the data successfully.

W 8.3-44: After import, the user only gets a summary, prompting
“xx XML files were successfully imported” and
“yy XML files failed to import.”
The user gets no information, which files were successfully imported and no information about the reasons of failed imports.

W 8.3-45: New versions of *MetaPath* have problems to read older composer XML files.

W 8.3-46: No separate validation function is implemented to check an XML file according the corresponding DER- or MSS-Composers schema prior of the import.

W 8.3-47: No interaction with other IT-Tools is foreseen.

W 8.3-48: There are no report functions for users implemented.

W 8.3-49: There is no documentation available, which *MetaPath* data will be included into the *OECD* (Q)SAR Toolbox.

8.4 DER/MSS-Composer family

These tools are part of the *MetaPath* software package. The *DER/MSS-Composer family* consist of four different programmes specialised for different objects:

Table 8: Programs of the *DER/MSS-Composer family*

Composer Programs	Object of investigation	Scheme
DER Composer	Rat, Goat, Hen, Cow,	DER[99].xsd
MSS Livestock Composer	Poultry, Lactating Ruminants, other Animals	MSSLivestock.xsd + MSSCommonTypes.xsd
MSS Plants Composer	Crops	MSSPlants.xsd + MSSCommonTypes.xsd
MSS Crops Composer	Rotational Crops	

8.4.1 Owner

USEPA and PMRA prepared the scientific background. The “Laboratory of Mathematical Chemistry” (*LMC*) made the IT-Tools. The software contains a legal copyright for *LMC*.

W 8.4-1: There is no clear license rule available for the use of this software.

8.4.2 Contained information

The information that was collected from one metabolism study could be stored in an XML file by these programmes. The MSS Composers are working without any database.

The contained information should be comparable with the corresponding *OHTs*. An analysis was made only for the DER Composer (compare 8.2.1).

The contained information is more than the information, that is imported into *MetaPath*. The data structure of the *DER/MSS-Composer family* and *MetaPath* are aligned.

8.4.3 Functionalities

The “Data Evaluation Record (DER) Composer" and the “Metabolism Study Summary (MSS) Composers” are IT-Tools to help to summarise the needed study information developed by the *USEPA* and PMRA. These tools are called the *DER/MSS-Composer family*.

A major advantage of the composers is the flexibility of the label captions, which could be used for table labels, columns headers and rows.

W 8.4-2: The *DER/MSS-Composer family* programmes are working without any user authorisation and therefore without any role concept. If the user has access to the XML file, the user could delete all information on file level.

W 8.4-3: The information contained in the *DER/MSS-Composer family* XML file is comparable but follow a very different XML-scheme compared with the *OHT* XML-scheme.

W 8.4-4: The *DER/MSS-Composer family* schemes are based on many free text fields. Without picklists regarding animals, plants, analysed materials etc. it would be very difficult to consume these information from the XML file by other tools.

W 8.4-5: The predefined tables of the *DER/MSS-Composer family* are not suitable for “freestyle” studies not conducted following the newest *OECD* TG. However, the amount of such studies is relevant (approximately > 1000 studies in the *BfR*) for building a data foundation.

W 8.4-6: Additional non-guideline experiments are hardly possible to code in the *DER/MSS-Composer family* (e.g. stem injection, cell cultures) but can principally be coded by free-text fields (with: character restrictions!).

W 8.4-7: Tentative results are difficult to handle in the *DER/MSS-Composer family*.

User front end:

W 8.4-8: Is it not possible to use the smaller "<"-character in tables of the *DER/MSS-Composer family*.

W 8.4-9: There is no field for the Lot/Batch # of the radiolabel Section IIA of the *DER/MSS-Composer family*.

W 8.4-10: Section IIIA of the *DER/MSS-Composer family*: In the Total Radioactive Residues section, it is not possible to add extra lines to the extraction efficiency table, however studies often analyse more than one matrix/matrix type and therefore the ability to add extra lines would be welcome.

W 8.4-11: The limitation of the free text fields of the *DER/MSS-Composer family* does not allow the same text quality as in *OECD* summaries, e.g. III. D. free text field.

W 8.4-12: Bug in Section V, Appendix 3 of the *DER/MSS-Composer family*: Scrolling through this table changes the entries, and it is easy to do this accidentally if the cursor goes across the table. Given the importance of this table, it would be better to demand a click into the cell to change the “linked” entry.

W 8.4-13: Section IIIA of the MSS Composer Livestock, Table B.7.2.1-5: It is not very meaningful to include % TRR values in this table (value is 100% TRR for each tissue), but it would perhaps be useful to include % of dose values here for each tissue and also the total amount recovered value (as % of dose).

W 8.4-14: Bug in Section II. Materials and Methods_B. Study Design: The text in Experimental Conditions is not stored. The software deletes it each time after the Word document creation.
Bug in Section IIB: Text entered in the “Experimental Conditions” text box is deleted from the XML file when the MSS Composer software is closed, and is also omitted from the MS Word rendered version.

W 8.4-15: Bug in MSS Livestock Composer: Lactating Ruminants - III. Results and Discussion – A. Total Radioactive Residues: Text field for Quantitation missing.

W 8.4-16: Bug in MSS Livestock Composer: Lactating Ruminants Appendix 1_Dose (measured): the units mg/kg bw/day are missing in the Word document.

W 8.4-17: Pasting text from an MS Word document into the Composer sections does not auto adjust the font size to a uniform size and there is no way to do so within the Composer itself.

W 8.4-18: There is no “undo” button in the *DER/MSS-Composer family*.

W 8.4-19: The tab button does not work to move through the input fields in the *DER/MSS-Composer family*.

W 8.4-20: The predefined tables of the *DER/MSS-Composer family* are limited to seven columns per radiolabel. This is not sufficient and the origin for improvisations.

W 8.4-21: The *DER/MSS-Composer family* is not optimised for a manual data input of tables of raw data. *DER/MSS-Composer family* is not able to analyse pasted tables from clipboard as a whole (e.g., from Excel or Word) to save time and to avoid errors. The current way to enter data is error-prone.

Structure drawing:

W 8.4-22: The structure drawing program of the *DER/MSS-Composer family* is poor.

W 8.4-23: The *SMILES* code generated in the *DER/MSS-Composer family* seems to be Composer specific and cannot be read by other Chemistry Drawing software - also vice versa: *SMILES* pasted from other software into Composers will be translated into "composer-dialect" (this is very error-prone).

W 8.4-24: The limitations regarding the Markush/generic structures are relevant. Often only generic structures are given in the reports (very often e.g. OH-position phenyl-ring, conjugate-position).

Compatibility:

W 8.4-25: Users of the *DER/MSS-Composer family* are able to create XML files which could not be imported into *MetaPath*. The used restrictions are only available in the MSS Composer Manual.

W 8.4-26: *DER/MSS-Composer family* provide no *API* to allow direct data import.

Reporting:

The *DER/MSS-Composer family* programmes provide functionalities to render the information into MS Word reports. Each program generates a report according to its own template.

W 8.4-27: *DER/MSS-Composer family* does not redact the authors names on reports that will end up in a public domain.

W 8.4-28: Section IIA of the *DER/MSS-Composer family*: The structures for the radiolabels are difficult to read as displayed on the screen or in the MS Word rendered version.

W 8.4-29: MSS Plant Composer: Pasting text from a Word document into the Composer sections does not auto adjust the font size to a uniform size and there is no way to adjust it within the Composer itself.

W 8.4-30: Bug in MSS Plant Composer: Section III. Results and Discussion_A. Total Radioactive Residues: once the word file is rendered, the report changes from plant to livestock metabolism.
Section IIIA: In the Total Radioactive Residues section, the MSS Composer screen shows “Extraction efficiency of radioactive residues from plant metabolism study using residue enforcement method”, but the MS Word rendered version changes this to “Extraction efficiency of radioactive residues from livestock metabolism study using residue enforcement method”.

W 8.4-31: Bug in MSS Composer (Plant and Crop): Section II. Materials and Methods_A. Materials_3.Soil Type and Environmental conditions can be entered. When the World doc is rendered only Soil Type appears, but not environmental conditions.

W 8.4-32: Bug in MSS Composer: Section V. Appendix_Appendix 2: the expertise comment is not transferred to the Word document.

W 8.4-33: Bug in MSS Composer: Section III. Results and Discussion, General Health of Animals: Text did not transfer to MS Word document.

W 8.4-34: Bug in MSS Composer: Section II Materials and Methods_A. Materials- lot number has to be entered under other synonyms, as it is missing. In the MS Word document, it is however present but then blank.

W 8.4-35: Bug in MSS Composer: Section IIA: There is no field to enter the Lot/Batch # for the non-radiolabelled test material on the screen, but a row for this information is included in the Word rendered version.

W 8.4-36: Bug in MSS Composer (Plant and Crop): Section IIIE: The 2D structures are not included in the list below the *Metabolic pathways* in the MS Word rendered version.

W 8.4-37: Bug in MSS Composer (Plant and Crop): Section V, Appendix 1: The MSS Composer screen shows “Application Rate” in the header to column 4 of the table, but the MS Word rendered version changes this to “Application route”

W 8.4-38: Bug in MSS Composer Livestock: Section IIIB and Section IIID: The row for “Total bound residues (PES)” is omitted from each table in the MS Word rendered version.

W 8.4-39: Bug in MSS Livestock Composer: Lactating Ruminants Section V, Appendix 1: The MSS Composer screen shows e.g. “1.2 mg/kg bw/day” in column 5 of the table (Dose (nominal), but the MS Word rendered version changes this to e.g. “1.2 1.2”

W 8.4-40: MSS Composers: Section V Appendix: *SMILES* codes are shown in Appendix 2. Once the MS Word document is rendered, the *SMILES* codes are not shown under 2D structures.

8.4.4 Life cycle and process

There is a real life cycle of these IT-tools. Today the composers for plant /crop /livestock have the version number v 1.9 but there is no information available, which were the triggers to produce new composer versions. It seems that no improvement process was implemented for the composers.

W 8.4-41: It seems that no process for the life cycle of these IT tools was implemented.

W 8.4-42: There is no governance model for the further development of the composer. Who is responsible to coordinate the different user requirements and who is allowed to decide?

W 8.4-43: The used different technological approach to have different specialised input programmes drastically increases the software maintenance effort.

W 8.4-44: Furthermore, *LMC* supports “custom versions” of the *DER/MSS-Composer family*. Moreover, there is not one version of the *DER/MSS-Composer family*. The different “custom versions” of *DER/MSS-Composer family* bind additional resources for each new version.

8.4.5 Publication and documentation

The Composer tools are available in a bundle with *MetaPath*.

W 8.4-45: *LMC* offers a web platform for publication of the software. No documentation and training material is available.

W 8.4-46: There is no user documentation integrated in the *DER/MSS-Composer family* tools.

W 8.4-47: There is no version history published for the composers.

8.4.6 Interoperability / output

The most important benefit is, that the output of the *DER/MSS-Composer family* tools can be used to feed the database *MetaPath*.

The output is written in an XML syntax. Therefore, generally spoken: the output is interoperable.

W 8.4-48: The *EFSA* process to provide XML files of already evaluated studies is difficult to handle for the applicants.

W 8.4-49: The actual XML schema definition files of the *DER/MSS-Composer family* and the outdated versions are not publically available.

W 8.4-50: New versions of Composers have problems to read previously XML files.

W 8.4-51: Each MSS-Composer supports a report generator to render MS Word files. The rendered files are sometimes not readable.

W 8.4-52: There is no common validator tool available, which could be used to validate the composer XML file against the corresponding schema definition file XSD.

W 8.4-53: Transformation tools between the “Data Evaluation Record Templates” of *USEPA* and PMRA and the *OHTs* are not existing.

A 8.4-54: The fact that two different, comparable, but not identical approaches to summarising metabolic studies have been implemented in *OECD* member countries is an indication for a need for a deeper analysis of the differences and the need for further harmonisation.

9 Solution approaches

9.1 Disclaimer

Starting with this chapter, the improved IT-Tool should get the temporary name *MetaPath II Tool*. Additionally, the question whether *MetaPath II Tool* is

- a completely new development or
- an improved version of the existing system "*MetaPath*" or
- an improvement of *IUCLID* with the support of all additional user functions defined in this report

is left open for discussion.

9.2 MetaPath II Ecosystem

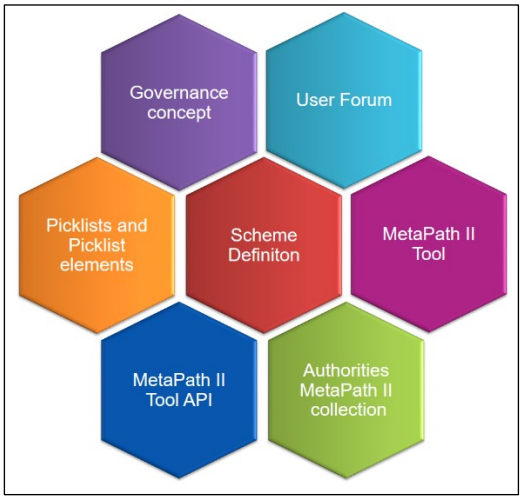


Figure 15: The MetaPath II Ecosystem

MetaPath II should be a synonym for the final improved system. It is proposed to build up an ecosystem of different components where each part of the MetaPath II Ecosystem could be used by applicants and authorities because both stakeholders need the same interoperable functionality. The components of the *MetaPath II Ecosystem* (see Figure 15) are described in the next chapters.

What the *MetaPath II Ecosystem* should not be:

The MetaPath II should not contain

- information and methods to predict exposure and
- methods to predict toxicological properties.

However, it has become clear that the technical solution to the transport issue of the *Aggregated raw data* on metabolism studies will only meet a small portion of the user requirements. The greatest benefit is expected in the reconceptualisation and extension of the *MetaPath* idea and the deployment of an improved IT-Tool for collecting, processing and the visualisation of the results from metabolism studies.

9.2.1 Governance concept

R 9.2-1: A *Governance concept* is needed for this ecosystem.

R 9.2-2: The Governace body has the responsibility for the *Scheme definition*; the schema description to transport raw data specific for a *Metabolism study*.

R 9.2-3: The *Governace body* additionally has the responsibility for the needed *Picklists and picklist elements*.

R 9.2-4: The other IT components of the ecosystem could be part of an *Open Source Project* in which the interested parties contribute to the community.

9.2.1.1 The OECD as the governance body

The preferred solution would be, that the *OECD*

- plays the role of the new *Governace body*,
- will improve its own transport mechanisms for study summaries by the new *OECD Attachment Type* and
- has also the responsibility of the needed *Picklists and picklist elements*.

9.2.1.2 A governance body outside of the OECD

If *OECD* is not willing to hold the role of the new *Governace body*, this could also be organised by ambitious stakeholders. This body may perform either all or some of the tasks mentioned above. Via this route, an in-official quasi-standard could be agreed on outside the *OECD*. The approach would then be an improved level of the current approach via the *MSS-Composer* family and *MetaPath*. For this, a well defined foundation for constitutiing this *Governace body* would be needed.

9.2.2 User forum

R 9.2-5: This user forum should be used to inform the *Governace body* about new identified weaknesses, errors and requirements regarding the *MetaPath II Ecosystem* and to assist the *Governace body* for an adequate priorisation.

R 9.2-6: The members of the user forum should be invited to take part to test new beta versions of the *MetaPath II Ecosystem*.

R 9.2-7: This user forum could be used to inform users regarding changes in the *MetaPath II Ecosystem*.

9.2.3 Picklists and picklist elements

R 9.2-8: If the *OECD* will adopt the role of the *Governace body*, the *Picklists and picklist elements* should be defined by *IUCLID* mechanisms according to the adequate *OHT*. This way, it would be possible to reduce the list of values of sensible picklist elements for a specific metabolism *Study Type*. Otherwise, the *Governace body* should organise adequate mechanisms.

R 9.2-9: If the *OECD* will adopt the role of the *Governace body*, the life cycle management of the *Picklists and picklist elements* should be included in the *OHT* life cycle. Otherwise, the *Governace body* should reorganise the life cycle management.

R 9.2-10: Each element of the *Scheme definition* that is coded by picklist mechanisms should have the attribute “catalogue” with a fixed string containing the picklist id. Today, this fixed string of the attribute “catalogue” contains a generic “draft” name, depending on the respective type e.g. “study_type_class”. Please have a look to the needed *Picklists* in chapter 9.2.3.

P 9.2-11: There are logical references between items of different *Picklists*. If a specific study_type_class was selected, only a sub group of picklist items of the object_class is useful.

R 9.2-12: There are logical references between data elements of the *Scheme definition* and the units in which the respective data are given. The unit_class should be divided into different *Picklists*.

9.2.4 Scheme definition

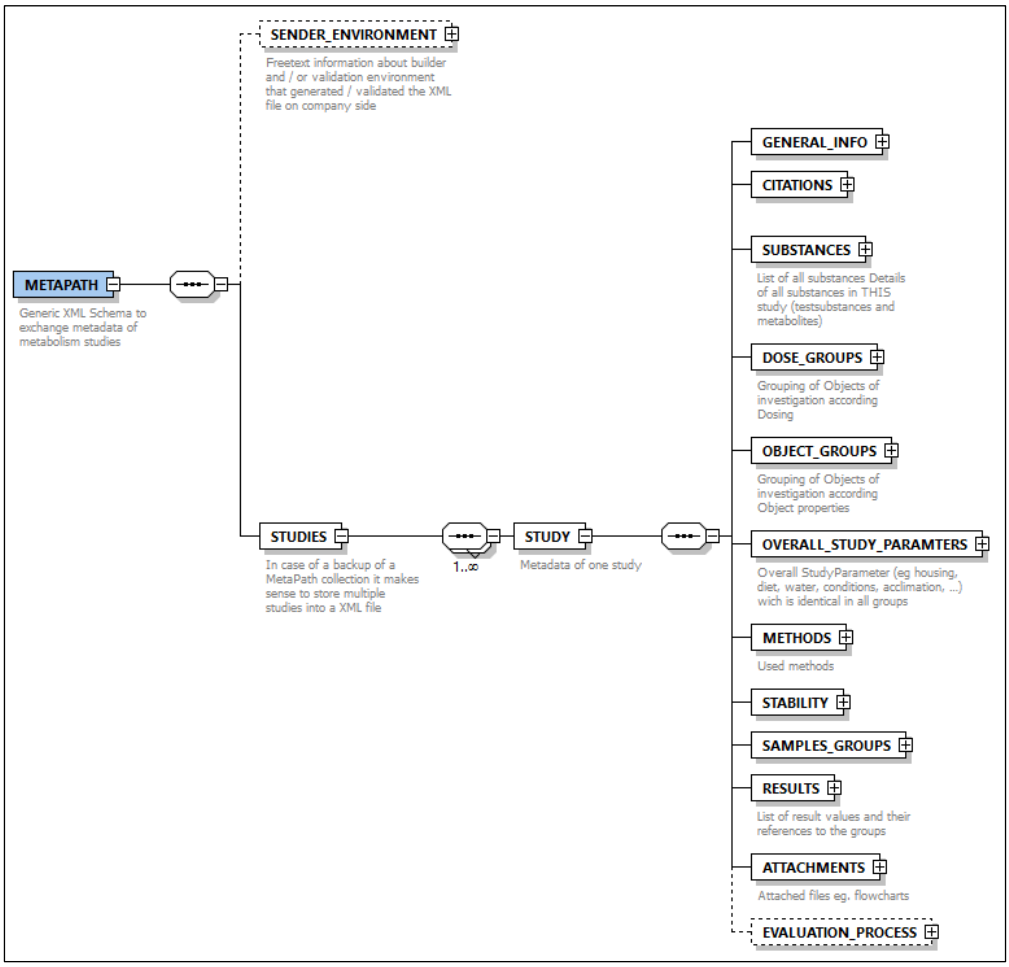


Figure 16: MetaPath II scheme definition (see also 11.2 MetaPath_II.xsd or MetaPath_II_xsd.png)

A schema definition is needed as a basic interface to solve interoperability issues. The XML file according the *Scheme definition* (see Figure 16) will be called *MetaPath II.XML* and accompanied with this report.

R 9.2-13: The MetaPath II *Scheme definition* (see chapter 11.2) should cover all *Study Types* where radioactive labelled test material is used according to the Test Guidelines. The variability can be customised using picklists. The customisation process should start with the *Study Types* assisted by the current *DER/MSS-Composer family*.

R 9.2-14: This *Scheme definition* contain internal reference techniques that e.g. only defined substances could be used for the analytical values.

R 9.2-15: This *Scheme definition* should be used as a data interface between different IT-Tools and should be applicable for all types of Metabolism studies.

R 9.2-16: This schema definition should re-use schema types derived from the *OHTs* to minimise the efforts for the life cycle management of the schema itself and to ensure an interoperability on the level of the XML files.

R 9.2-17: The needed flexibility of the *Scheme definition* should be possible by using different *Picklists and picklist elements* depending on the type of the metabolism study.

R 9.2-18: The *Scheme definition* describes only data from one specific study.

R 9.2-19: The *Scheme definition* should contain only the information parts, which are stable over time. That means, exclusion of submission depending *Metadata*. The schema contains all information on level of the *Aggregated raw data*.

R 9.2-20: As optional elements, information regarding actions already carried out by the authorities as part of the assessment process.

R 9.2-21: If the *OECD* will hold the role of the *Governace body*, the life cycle management of the *Scheme definition* should be embedded in the *OHT* life cycle. Otherwise, the *Governace body* would be obliged to organise the life cycle management.

Modelling the *Scheme definition* was not finished, only the basic principles and the main data organisation were modelled until now. Further specification can be made, when procedural questions are answered.

A 9.2-22: A deeper analysis is needed to match all needed information into the *Scheme definition*.
Therefore, all current information contained in all composers
+ additional elements of *MetaPath*
+ all additionally needed fields according the weak points list
should be addressed.

A 9.2-23: A deeper analysis would be needed to fulfil the idea to transport information about the evaluation process steps (e.g. authority, status, remarks) and results of the evaluation steps between different *MetaPath II Collections* regarding the metabolism studies.

A 9.2-24: A deeper analysis is needed to include relevant information of the status of the evaluation process steps (e.g. authority, status, remarks).

9.2.5 MetaPath II Tool

The *MetaPath II Tool* is the *User Interface* of a *MetaPath II Collection* realised with a database management system.

The *MetaPath II Tool* should cover already implemented functions of the *MSS-Composer* family and of *MetaPath* improved by needed additional functions. This report represents an attempt to list all necessary user requirements.

R 9.2-25: A user documentation, a version history and a technical documentation are needed. These documents should be updated with each new version.

R 9.2-26: The user documentation should be integrated into the tools.

R 9.2-27: The *MetaPath II Tool* should work with an open source database management system on different common used server operating systems.

R 9.2-28: The database management system of the *MetaPath II Tool* should support common used access methods (e.g. JDBC, ODBC) and programming languages (e.g. Java, PHP, Python).

R 9.2-29: A “Role Concept” is required for implementing the user access rights to the functions in the *MetaPath II Tool*.

R 9.2-30: The *MetaPath II Tool* provides all modules as a web *User Interface* programmed as an *Open Source Project*. Code Maintenance should be governed by the *Governace body*

R 9.2-31: The life cycle management of this *MetaPath II Tool* should be organised according the *Governance concept*.

R 9.2-32: The *MetaPath II Tool* is able to manage *Aggregated raw data* of Metabolism studies in a *MetaPath II Collection*. The tool should be able to manage different collections but not in parallel.

R 9.2-33: Each *MetaPath II Collection* should have its own database management system to get a user friendly response time.

R 9.2-34: The *MetaPath II Tool* manages all needed information on the study level (Study Data Set) for the substance identification, the relationships between these substances (*Metabolic pathways*), information about absorption, distribution and excretion and kinetic information.

R 9.2-35: The *Test substance*, used on study level, has a reference to a unique *Substance*.

R 9.2-36: If a *Substance* is radiolabelled at different points, these are different *Test substances* which are referencing to the same unique *Substance*.

R 9.2-37: The *MetaPath II Tool* should be able to manage unknown metabolites on study level under the same name e.g. “M1” in different studies.

The main entities of the *MetaPath II Tool* are similar to the structure of the *Scheme definition*.

The following statements describe the relationships of the main entities to each other. The description of the needed attributes should be part of a further technical concept.

R 9.2-38: The *MetaPath II Tool*stores information on the level of a *Study*.

R 9.2-39: The “*Study*” is characterised by at least one citation of a *GLP study report* and with general information. The bibliographic *Metadata* has individual fields at least for author, title, report number, report year, source which should be in line with the *IUCLID* literature reference entity.

R 9.2-40: A *GLP study report* could have amendments with its own bibliographic *Metadata*.

R 9.2-41: A "*Study*" could be divided into a field and an analytical part (*GLP study reports*) with its own bibliographic *Metadata*.

R 9.2-42: All documents, which could be covered under the umbrella "*Study*", should be stored as an attachment.

R 9.2-43: To be compatible with already entered studies, one study could have more than one citation to a *GLP study report*.

R 9.2-44: The value for the author of vertebrate studies will be sanitised automatically whilst compiling reports for the public.

R 9.2-45: The "*Study*" can contain textual descriptions (text blocks) of the study such as remarks, justifications, conclusions etc.

R 9.2-46: The "*Study*" will be evaluated in a legal act identified by specific ID formats according a "Legal Act Type".

R 9.2-47: The "*Study*" can contain information of more than one *Metabolic pathway*.

R 9.2-48: One *Study* contains information to more than one *Substance*.

R 9.2-49: A *Substance* can be a *Test substance*, a Radiolabelled Test Substance or a *Metabolite*.

R 9.2-50: A *Test substance* could be inactive or radiolabelled. All used radiolabelled *Test substances* and the inactive *Test substance* should be defined. The parent of all radiolabelled *Test substances* will be the inactive *Test substance*.

R 9.2-51: A radiolabelled *Test substances* is characterised by Lot & Batch numbers.

R 9.2-52: Each *Substance*is characterised by a set of predefined *Metadata*. Users should be able expand the substance *Metadata* by user defined elements e.g. an own substance identifier which should be used to jump into an own external substance database (see chapter 9.4.8).

R 9.2-53: A *Substance*could be characterised by values of different phys-chem. properties. However, evaluators need only the logP_{ow} value for interpretations. (Q)SAR Tools are able to calculate a logP_{ow} from the chemical structure. Therefore, it is not highly necessary to store phys-chem. properties.

R 9.2-54: A known *Metabolite*, used on study level, has a reference to a unique *Substance*.

R 9.2-55: A *Metabolite* can have different *Substance* parents.

R 9.2-56: If a *Metabolite* has the *Test substance* as the parent, the inactive *Test substance* will be used as parent. This will be done although the data have been derived from a defined *Radiolabelled Test Substance*.

R 9.2-57: A *Substance* can be transformed into more than one *Metabolite*.

R 9.2-58: The reference of a *Metabolite* to the parent *Substance* is stored on study level. It should be possible to store contradictory assumed parent relations in different studies.

R 9.2-59: Each "*Study* set is characterised by at least one status. This status should not represent the status of the scientific analysis, this should be a technical status e.g. "Migrated from Metapath", "Imported but outstanding QA", "Quality checked", …

R 9.2-60: Each “*Study* could be used to investigate the *Metabolic pathway* in more than one *Object of investigation* e.g. mice and rats.

R 9.2-61: The “*Study*” and the *Object of investigation* are characterised by more than one “Study Parameter” referencing to a “Parameter Type”.

R 9.2-62: The individuals of the *Object of investigation* could be grouped in the “*List of Study Object Groups*”. Each *Study Object Group* consists of a number of individual objects. All parameter should be identical for the individual objects of the *Study Object Group*.

R 9.2-63: A *Test substance* could be applied by different administration procedures.

R 9.2-64: The *List of Study Object Groups* are used to define *Dose Groups*. The *Dose Group* is characterised by one *Test substance* and the application parameter.

R 9.2-65: Samples were collected from the members of the *Study Object Group* at different time points / time intervals from different matrices. Samples should be grouped in *Sample Groups*.

R 9.2-66: It should be possible to store different analytical methods with different extraction flowcharts for different matrices.

R 9.2-67: Each sample could have more than one analytical result for different substances, analysed by different methods / fractions.

A 9.2-68: A deeper analysis is necessary regarding a correct model of the relation between samples, methods, analysed fractions and measured values.

A 9.2-69: It is necessary to be backward compatible. A deeper analysis is advised. It should be defined, for which formats / functions this compatibility is necessary. Algorithms should be developed on which specifications a mapping of the content should follow.

R 9.2-70: It should be possible to store output files of the reports with the results of the analysis as attachments.

R 9.2-71: It should be possible to store at least a reference to the legal act where the study was evaluated the first time.

A 9.2-72: Within a further analysis it should be identified whether there is a need to store further references to legal acts where this study was also part of the submission package.

Some other entities could be needed. The proposed *Scheme definition* gives an impression on how the definition could look like (see chapter 11.6).

9.2.6 MetaPath II Tool API

The *MetaPath II Tool* should provide an application programmable interface (*API*).

R 9.2-73: The *API* should provide functions for reading and writing data from / into a *MetaPath II Collection* on element and record level according the *Role Concept*.

R 9.2-74: The *API* should provide a data interface that can be used by (Q)SAR tool providers for harvesting validated data sets for their models.

R 9.2-75: It should be possible to access a specific data set of a *MetaPath II Collection* via REST *API* from external tools.

9.2.7 Authorities MetaPath II collection

A central element of the *MetaPath II Ecosystem* should be an international *Authorities MetaPath II collection*, which will be supported by a federation of international authorities. When organising an information loop between applicant, authority and curated reference collection, there is an acute risk of data loss.

R 9.2-76: The best organisational concept for an *Authorities MetaPath II collection* pesticides related data collection could be clarified in an *OECD* meeting of the Chemicals and Biotechnology Committee (CBC).

D 9.2-77: It should be decided how to handle metabolism studies that are outside the pesticides domain.

R 9.2-78: An expandable “Set of quality standard rules” is needed, which should be checked prior declassification of new data sets or modifications of the data sets. The highest priority should be an *Authorities MetaPath II collection*, that contains only validated *Aggregated raw data* of metabolism studies of pesticides.

A 9.2-79: The *Authorities MetaPath II collection* should only contain data that have their origin in the metabolism study themselves. This means, that users can collect secondary *Metadata* from other sources and store them within the *MetaPath II Tool* into their local *MetaPath II Collection*. A decision is needed, whether these data should or should not be transferred into the *Authorities MetaPath II collection*. A deeper analysis is required.

A 9.2-80: A deeper analysis is needed under which conditions applicants are obliged to extract copies of the *Aggregated raw data* from the curated reference collection for repeated submission to the authority.

O 9.2-81: The cycle described in A 9.2-80 makes only sense for a) submissions of revised *Aggregated raw data* by the owner of the study, which were originally entered by the authorities for the first time, and b) submissions of an addendum / corrigendum to the original study by the study owner, if data errors have been corrected or previously unknown metabolites have been identified.

O 9.2-82: Applicants who want to use a metabolism study already contained in the curated reference collection for a read-across for another legal act should not resubmit the *Aggregated raw data*. The applicant has to be the data owner or has the access rights from the data owner. In such cases, the applicant has to give only an updated endpoint summary on the *OHT* level with regard to the actual legal act.

A 9.2-83: A deeper analysis is needed to define “Set of quality standard rules”.

R 9.2-84: A “Quality control body” is required to ensure an appropriate data quality with the help of the *Set of quality standard rules*.

R 9.2-85: The time point to include the data set into the *Authorities MetaPath II collection* depends on the legal aspects of the different jurisdictions. This including process should be initialised by the responsible authority.

R 9.2-86: The *Authorities MetaPath II collection* is primarily to be optimised for a non-restricted access by the authorities. If only authorities have access to this collection, the inclusion process (R 9.2-85) has not the character of a publication.

R 9.2-87: The development of a confidentiality concept is required, if other stakeholders would be granted access to the *Authorities MetaPath II collection*.
One approach could be: All data fields containing confidential information must be listed in a “Field list of confidential information”. The corresponding front end, the *MetaPath II Tool*, would hide confidential data fields depending on the access rights to all rows.

R 9.2-88: The *Authorities MetaPath II collection* has to be organised as a browser based cloud service.

9.3 Core structure for a relational database model

9.3.1 Disclaimer

A core structure for a relational database model is proposed as a result of the deep analysis of the information flow and the user requirements. This database model should show that the project is ambitious but not utopian.

Moreover, this database model is not completely thought through and is not a separate part of the *MetaPath II Ecosystem*.

A 9.3-1: The proposed core structure contains no tables for the management of users, user roles, reports or supporting tables to speed up certain queries. A deeper analysis is needed.

A 9.3-2: If the *MetaPath II Tool* should be able to manage secondary *Metadata* from other sources than the metabolism study itself, a deeper analysis is needed.

The proposals for the flag “mandatory” and option of “Delete cascade” are not final and should be checked in the future. Sometimes it would be better to handle such validation or delete steps by the front end because the user could be warned in a more drastic manner.

This future database model is intended to show that far fewer tables will be needed and one will still be open to metabolism studies on other objects in contrast to the current complex database model of *MetaPath* (compare 8.3.4.2).

9.3.2 Proposed core structure

The proposed core structure should be in line with

- the textual logical model for the *MetaPath II Tool* (Chapter 9.2.5) and
- the scheme definition (chapter 9.2.4).

The prosed tablenames are bold in this chapter.

The proposed core structure (see Figure 17) was designed with the “Oracle SQL Developer Data Modeler” (*Data Modeler*). The internal used notation for the relations in the diagram could not be changed.

The following symbols are used:

Example	Cardinality	Action after delete
	0:N	No action
	0:N	Delete the referencing records
	1:N	Delete the referencing records

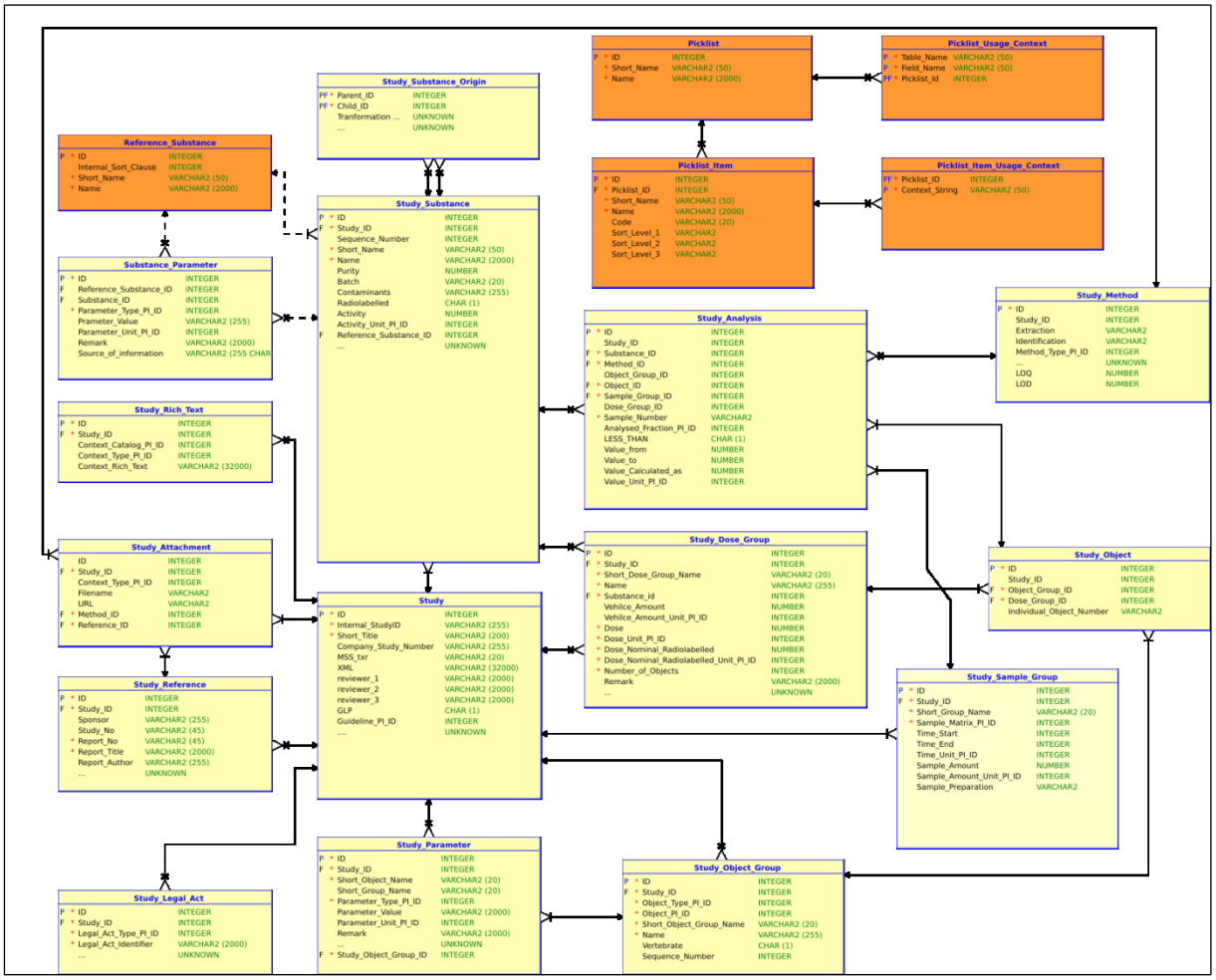


Figure 17: Proposed core structure for a relational database model (see also chapter 11.2 - MetaPath_II_DB.svg and MetaPath_II_DB.zip)

The following chapters describe the design principles, which are used. However, it is not specifically mentioned that each table has its own ID as primary key.

9.3.2.1 Picklist model

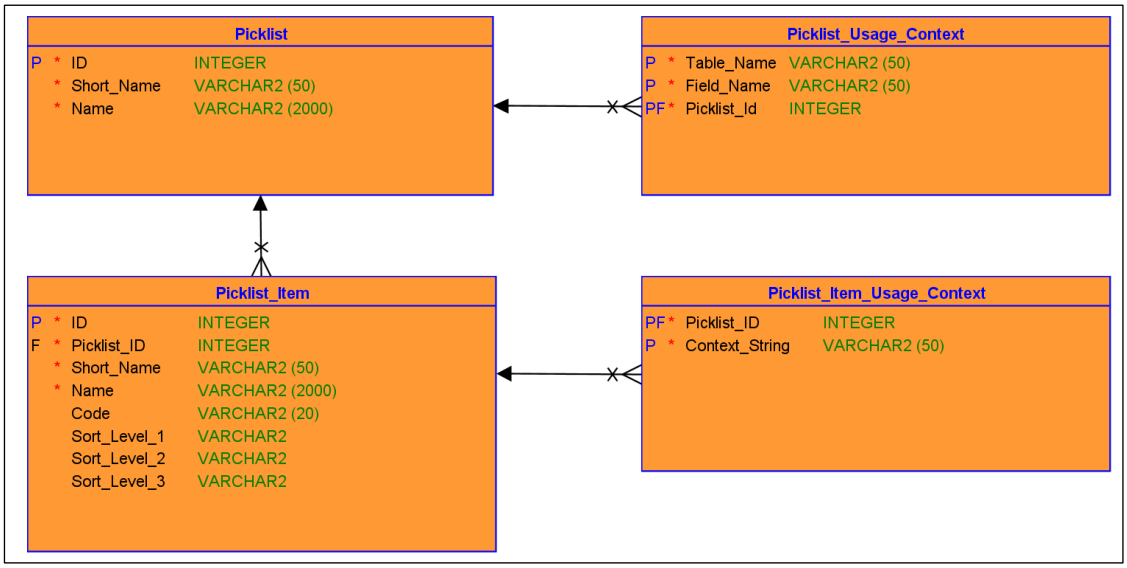


Figure 18: Details for the picklist model

Table 9: Current modelled code fields

Table name	Field name
Study	Guideline_PI_ID
Study_Analysis	Value_Unit_PI_ID
Study_Analysis	Analyzed_Fraction_PI_ID
Study_Dose_Group	Dose_Nominal_Radiolabelled_Unit_PI_ID
Study_Dose_Group	Dose_Unit_PI_ID
Study_Dose_Group	Vehilce_Amount_Unit_PI_ID
Study_Legal_Act	Legal_Act_Type_PI_ID
Study_Method	Method_Type_PI_ID
Study_Object_Group	Object_PI_ID
Study_Object_Group	Object_Type_PI_ID
Study_Parameter	Parameter_Type_PI_ID
Study_Parameter	Parameter_Unit_PI_ID
Study_Sample_Group	Sample_Amount_Unit_PI_ID
Study_Sample_Group	Sample_Matrix_PI_ID
Study_Sample_Group	Time_Unit_PI_ID
Study_Substance	Activity_Unit_PI_ID
Study_Text	Context_Catalog_PI_ID
Study_Text	Context_Type_PI_ID
Substance_Parameter	Parameter_Type_PI_ID
Substance_Parameter	Parameter_Unit_PI_ID

The picklist model (see Figure 18) of *MetaPath II Tool* is an advanced *IUCLID* picklist model with the following ideas:

- The table **Picklist** contains records of all picklists, needed in the *MetaPath II Tool*. At least a name and a short name are needed.
- For each picklist there are multiple records in the table **Picklist_Item**. At least a name and a short name are needed for each picklist item. If all picklists items are also in *IUCLID*, the Picklist_Item.ID could be identical to *IUCLID*. The code is optional.
- Each picklists should be used in at least one *MetaPath II Tool* table field, which is stored in the table **Picklist_Usage_Context**. One picklist could be used in different table fields (see Table 9).
- The fields Sort_Level_1 to Sort_Level_3 are optional fields to reuse the picklist management to arrange a chapter organisation for reports. However, a deeper analysis is needed for reports.
- A picklist could be segmented according a context string in sub-picklists. This could be managed by the table **Picklist_Item_Usage_Context**. E.g. picklists for units could contain all used units. The context_string could split the picklist items e.g. for Densitiy and Boiling Point etc.
  An alternative additional and equivalent approach is, to organise individual picklists for each context type.

9.3.2.2 Substance model

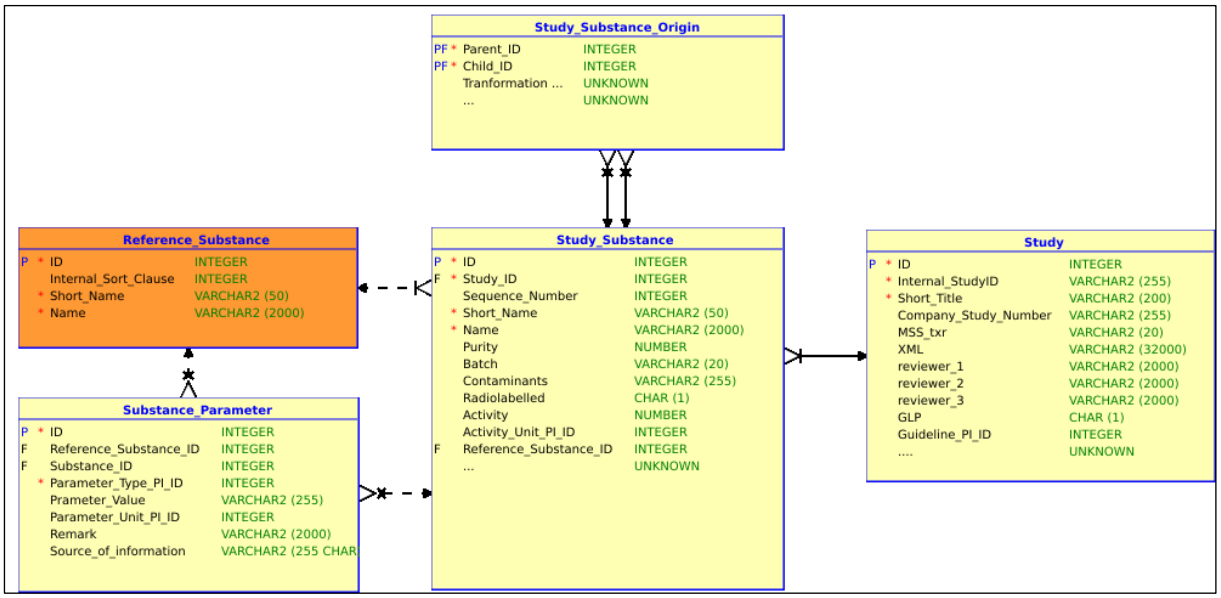


Figure 19: Details for the substance model

The substance model (see Figure 19) has the following ideas:

- All substances of the table **Study_Substance** are defined in context of the metabolism study:
  - Inactive parent substance,
  - Radiolabelled substance(s),
  - Metabolites

  by at least a name and a short name.
- All substances of a **Study** could be sorted by a study internal sequence number.

- The *Metabolic pathway* is stored in the table **Study_Substance_Origin**. Each record of this table has references between the parent compound and its child.
- A **Study_Substance** should have no or one reference to a **Reference_Substance** of an internal inventory. The **Reference_Substance** is always the non-radiolabelled substance. Unknown metabolites have no link to a **Reference_Substance.**
- The table **Reference_Substance** contains records from external inventories. This concept is identical to *IUCLID*. The table **Reference_Substance** should be administered separately. Trained staff should include newly identified substances continuously.
- Each of the **Study_Substance**s could have a set of **Substance_Parameter**. Each record of this set could contain a Parameter_Value of a defined Type, an optional Unit and an optional remark.
- The parameter types are defined by a picklist of the parameter_class. As parameter are foreseen:
  - CAS-Name,
  - *IUPAC*-Name,
  - Synonyms,
  - *SMILES,*
  - *CXSMILES,*
  - *InChI*,
  - CAS-No,
  - EINCS,
  - *IUCLID*-Reference Substance UUID,
  - log P_{ow}
  - …
- The source of all of these parameter values stored in this table is indirectly the metabolism study itself.
- We accept that inconsistent data for duplicates of substances could be stored. Therefore, these data fields are not the default searching fields for substances! These data are representing the knowledge / the content of the study. It is necessary to clarify how to deal with obvious errors.
- However, on the other side, this data model is open to manage different chemical notations per substance. In addition, this could be one possible way for handling generic structures.
- A “chemical target notation” should be defined for the *MetaPath II Tool* e.g. the parameter type <*InChI*>. The substance search should be designed in a way to search over all records of the table Substance_Parameter with the parameter type <*InChI*>. The result list of substances is the distinct list of found Substance_IDs.
- One possible solution for generic structures: If a generic structure information is stored in the table **Substance_Parameter** with a parameter type <generic>, a procedure should be activated after commit to resolve this generic structure into all possible structures and store them into the table **Substance_Parameter** with the target chemical notation e.g. *InChI*. So one generic structure could produce e.g. 20 possible structures.
  Modifying the generic structure should be made with caution: All inserted possible structure records should be deleted before creating the new records!
- Each of the **Reference_Substance**s is defined by a minimal set of **Substance_Parameter**:

- At least one structure information (*SMILES*, *CXSMILES*, *InChI*, …) should be provided.
- The source of information should have the sub structure of the LITERATURE object of *IUCLID*. However, how to proceed with snapshot information from international database records e.g. Pubmed or eChemPortal would be part of an extensive analysis.
- If the **Study_Substance.Reference_Substance_ID** is set, the **Substance_Parameter** should be managed only via the **Reference_Substance**. An adequate *Role Concept* is needed to differentiate editing of parameters of a **Reference_Substance** and parameters of a **Study_Substance**.

9.3.2.3 Study model

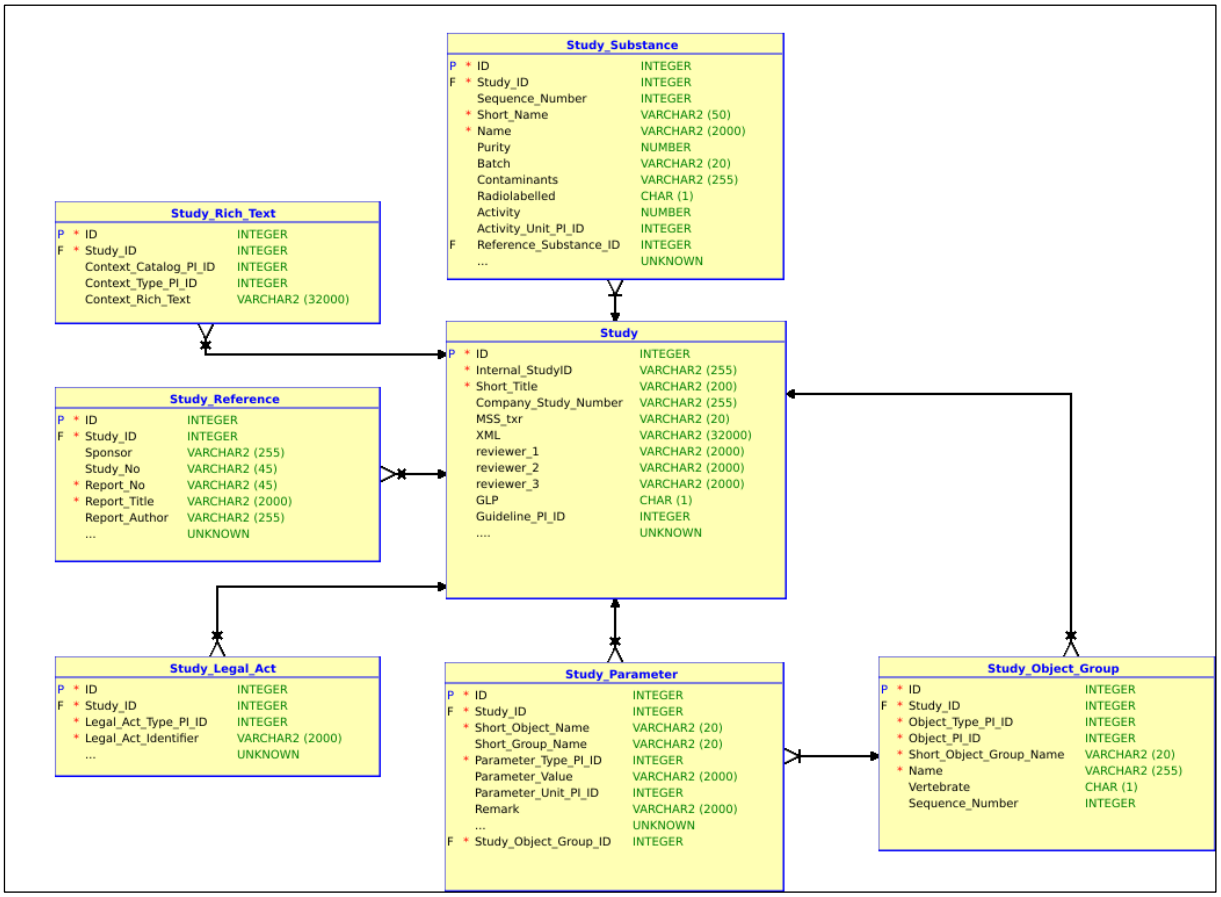


Figure 20: Details for the study model

The study model (see Figure 20) has the following ideas:

- Each study has one record in the table **Study.** It should be noted, that the field list of this table is not ready for the following reasons:
  - The reviewer fields should not be managed on study level but instead a Study_Evaluation_Process table is needed where the fulfilled actions are logged.
  - Some identifiers, which have been submitted with the *DER/MSS-Composer family* in the past, are not yet fully clearified. Identifiers related to a legal act should be managed in designated legal act fields (Study_Legal_Act) and it could be possible, that some other identifiers should be managed in the citations on study level of the Study_Reference, e.g. the GLP study and its addendum.

- A study must contain at least one record in the table **Study_Legal_Act**. The Legal_Act_type_PI_ID and the corresponding Legal_Act_identifier code for the legal act in which the metabolism study was evaluated the first time.
  It would also be possible to manage additional legal acts, where this study was also part of a submitted dossier. If this should be available, a 1:N relationship would be needed.
- A study must contain at least one record in the table **Study_Reference**. Each Study_Reference could have multiple references to the attached documents managed by the *MetaPath II Tool* or as an URL's to external documents.
- It is possible to store different context texts regarding context codes (Context_Catalog_PI_ID, Context_Type_PI_ID) in table **Study_Text**. It is easy to store many different context texts but the most important task is to report these text blocks in a defined order. This sorting mechanism could require additional fields.
- One study could referto many **Study_Substances**.
- A study can have different objects of investigation as its **Study_Object.** The study object is classified by a type (the Object_Type_PI_ID e.g. for rotational crops the 1st and the 2nd crop).
- Each study object could be describd by different parameters (**Study_Parameter**) according different Parameter_Tpye_PI_ID /Parameter_Values eg. for animals: strain, age, weigth, … for crops: seeding, BBCH, …
- A study could be characterised by different **Study_Parameter** according a Parameter_Type_PI_ID and a Parameter_Value.

9.3.2.4 Model of object, object groups, application and analysis

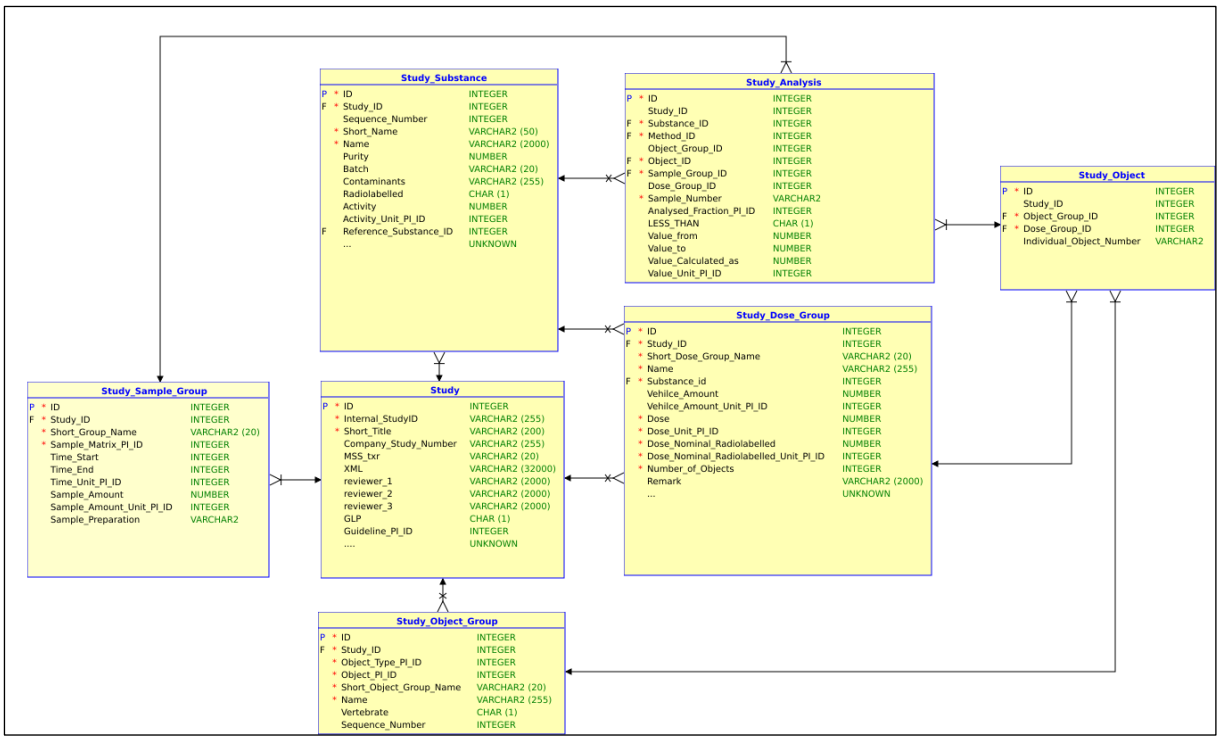


Figure 21: Detail for the object, object groups, application and analysis model

The object-, object groups-, application- and analysis model (see Figure 21) has the following ideas:

- The table **Study_Object** contains records for each individual object, which has been tested.
- The table **Study_Object_Group** is used to group the individual objects according an Object Type e.g. Soil, Sediment, Plant, Water, Food, Animal… and by the object itself e.g. Wheat, Carrots. It must be noted that this classification is made by using the pick-lists.
- The table **Study_Object_Parameter** contains additional parameter for the object group (e.g. male / female).
- The table **Study_Dose_Group** is used to define the different dose groups by using different test substances / different routes / different doses, which were applied. Each **Study_Dose_Group** can have different individual objects.
- The table **Study_Sample_Group** is used to group the individual analysed samples according identical sampling *Metadata*. The most important parameter is the Sample Matrix (Sample_Matrix_PI_ID).
- The individual analysed values are stored in the table **Study_Analysis**. A sample for one individual object, identified by the sample_number could have different analysed values, e.g. for each analysed study substance.

9.3.2.5 Other tables and views

R 9.3-3: The table **Study_Method** contains information about the used analysis methods including the management of flow charts.

A 9.3-4: A deeper analysis is required regarding the data model of the method description (extraction efficiency table).

R 9.3-5: The core structure, which is filled with internal references via integer ID values, should be more understandable with the help of defined views. The internal table structure could be encapsulated.
According to the *Role Concept*, the views should be provided for reading.

9.4 User requirements for the MetaPath II Tool

All essential user functions should be implemented in the “User Interface”. The *MetaPath II Tool* consists of different modules (see Figure 22), which could be called by parameters (see Figure 23). The user requirements for these modules are described in the following chapters.

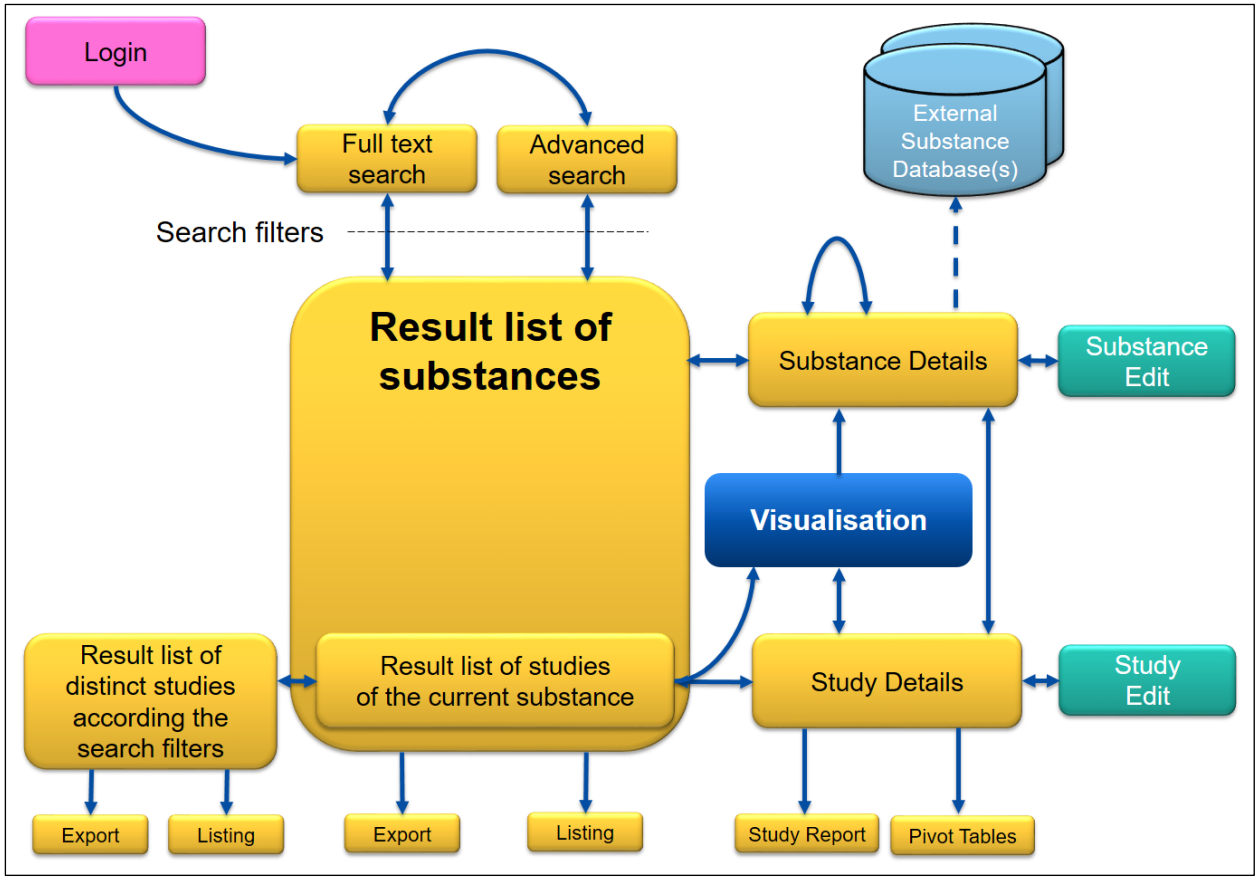


Figure 22: Main modules of the MetaPath II Tool

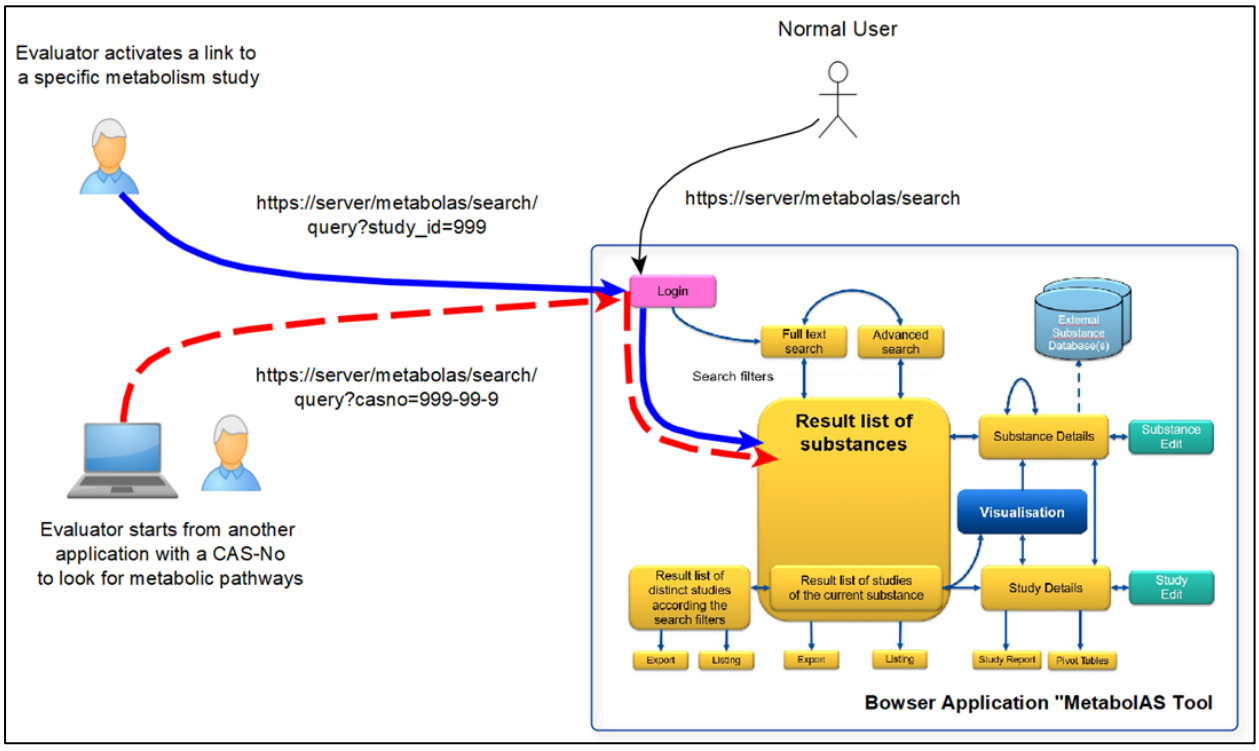


Figure 23: Examples for possible entry points of the *MetaPath II Tool*

There are some general user requirements for the *User Interface* of the *MetaPath II Tool*.

R 9.4-1: The *MetaPath II Tool* should support single sign-on techniques.

R 9.4-2: The web user interfaces should start with minimal delay.
The loading time and the preparing time to show a search result should be acceptable when less than 2 seconds. Currently, *MetaPath* is too in-performant.

R 9.4-3: The web user interfaces should show the "*Result list of substances*" first where the "*Search - full text*" is integrated. However, the user can switch from here into the "*Advanced search*" back and forth.

R 9.4-4: The web based user interfaces should provide the possibility to search for text strings in the interface by using the browsers search functionality.

R 9.4-5: The web based user interfaces should provide the possibility to copy elements from input forms and report elements into the clipboard.

R 9.4-6: The layout frames should not be fixed. The users should have the possibility to resize or move frames to separate browser views according to the user's needs.

R 9.4-7: The web based user interfaces should provide the drag and drop functionality.

R 9.4-8: The proposed result lists should be an entry point for REST *API* calls.

R 9.4-9: The proposed result lists should be provided as configurable table including the possibility for including or excluding columns.

9.4.1 User communication

The organisation's goal should be to establish effective user communication on different layers.

R 9.4-10: The application should inform about planned actions at the application instance and the non-availability of the service.

R 9.4-11: A central user help desk is mandatory.

R 9.4-12: Users should be able to subscribe to a newsletter for this project.

R 9.4-13: A user documentation according chapter 9.4.17 is mandatory.

R 9.4-14: A release note management is recommended.

9.4.2 General requirements

R 9.4-15: The tool should be user friendly and usable in terms of daily practice.

R 9.4-16: The focus on the fields should be activated using the tab key to navigate to the next field according a sequence, which has been carefully designed.

R 9.4-17: *Rich-Text* fields should be unlimited in the text length. They should allow the same text quality as in the *OHTs*.

R 9.4-18: The *Rich-Text* fields should provide an adequate editor to format the text. The user should be able to reformat pasted text e.g. from an MS Word document.

It should be possible to start a context sensitive User Documentation from all modules (see R 9.4-173).

9.4.2.1 General layout

R 9.4-19: A recognisable logo should be created for the *MetaPath II Tool*.

R 9.4-20: A well-defined collection of icons without any conflicts regarding copyrights should be used in the *MetaPath II Tool*.

Table 10: Used icons in the *IUCLID* of this report

Icon	Meaning
	Edit record
	Save modifications
	Cancel action without saving
	Add record
	Delete record
	Import a list of items into the current list

R 9.4-21: The following objects should be highlighted with different colours: Mandatory fields, Fields with validation errors, the current row and the current cell of each tables.

R 9.4-22: The background colour should change when in editing mode. Doing so, will increase the users awareness when critical actions are performed.

R 9.4-23: The length and the width of the user front end should not be limited. Accorddingly, the user will be able to search for texts in the browser tab with the browserbased search functionality. In addition, complex tables can be displayed entirely.

9.4.2.2 Sessions and transactions

R 9.4-24: The user can start the *MetaPath II Tool* in multiple sessions in different browsers or browser tabs. The *MetaPath II Tool* manages these applications by multisessions.

R 9.4-25: Each write operation should end in a defined way with a commit or a rollback. Write modules need elements for “Save” or “Cancel” explicitly.

R 9.4-26: If it is possible to skip to the next or to the previsions record in an open writing process, a user message will be prompted: “You have unsaved changes which will be lost. Do you want to proceed? Yes/No/Cancel”

R 9.4-27: If the user initiates a writing process, the data object in concern should be locked in a pessimistic way.

R 9.4-28: By the time point an editing session is initiated, a timer is started, whose remaining time is displayed to the user. With each action event triggered by the user, this timer will be reset. If the timer ends without the “Save” action, a roll-back action is initiated and the object will be unlocked.

R 9.4-29: Field restrictions should be depicted transparently in the user front end or tagged with a hint after a mouse over event. After leaving an input field with field restrictions, it should be validated. In case of errors, a comprehensible user message has to be generated and the focus falls back to the incorrect input field. The user can modify the field value or cancel the changes.

R 9.4-30: If there are validation rules on transaction level, those will be checked before saving the data. In case of errors, a comprehensible user message must be generated and the focus falls back to the first incorrect input field. The user can modify the field values of the record or cancel the changes.

9.4.2.3 Main menu

R 9.4-31: The user should be able to navigate into the main modules from the main menu (see Figure 24).

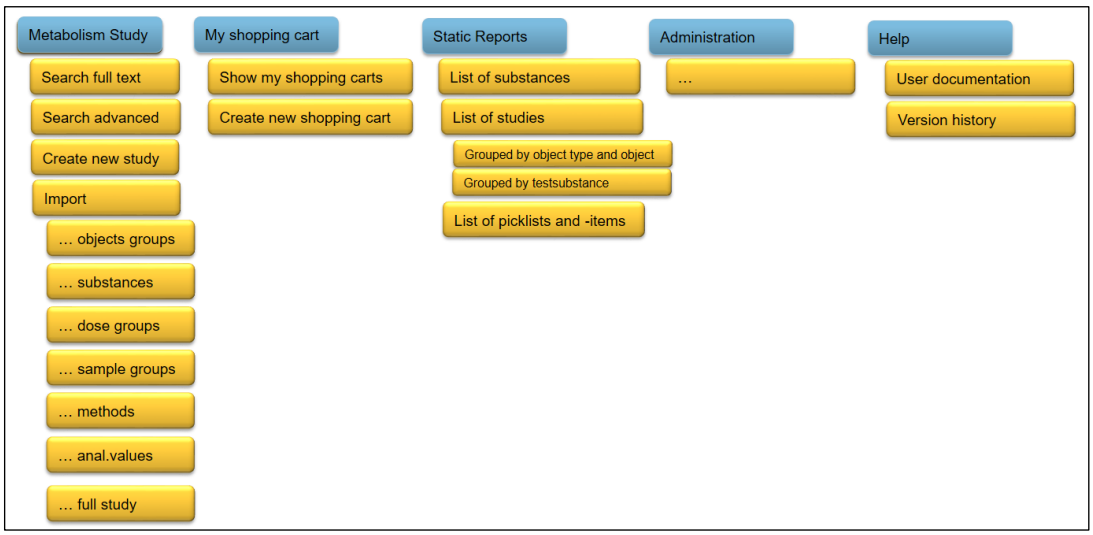


Figure 24: Mock-up of the main menue

R 9.4-32: If the user activates an action in the main menu, the current module will be closed. If there is an ongoing write operation, a user message will be displayed: “Do you want to save changed data: Yes or No?”

R 9.4-33: Additional menu modules could be needed for “user preferences” in addition to Figure 24.

9.4.2.4 Sortable list

R 9.4-34: Lists of objects could be used as headers of rows or columns. If sorting by a default sorting field string value does not apply, this sequence should be edited using a list consisting of movable elements, which can be rearranged in the list.

R 9.4-35: The manual edited sequence should be storeable. This sequence should be used as the default-sorting field for the current study.

Example (see Figure 25): The following list of sample groups should be defined in this example. The list of sample groups will be sorted according the sample group name as the default-sorting field.

Nonetheless, it would make more sense to move the urine sample groups next to the faces (excreta). So, both sample groups selected by the mouse can be moved between faces and fat. After that, a reordering of the plasma sample groups should be done.

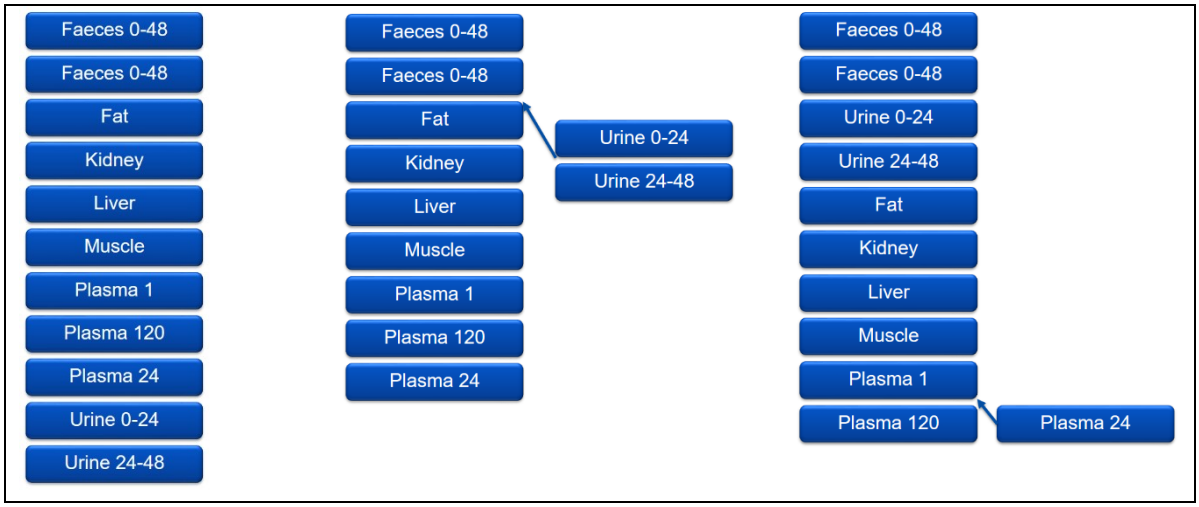


Figure 25: Example of sorting a list of sample groups with moveable rows

9.4.2.5 Flexible table input

R 9.4-36: Especially for manual data transfer, it is reasonable that the user is able to configure the input tables according to the template of the study report in a flexible way. Only by this approach the chance to carry out a quality control of the input values is given.

Solution approach[5] based on an example of analysed values:

R 9.4-37: Before entering this module, the user is asked to provide all needed groups. E.g., to enter analysed values, the user would be forced to define at least the following groups in advance (see Figure 26). Each group has a short name, which can be used for headers of table columns or table rows.

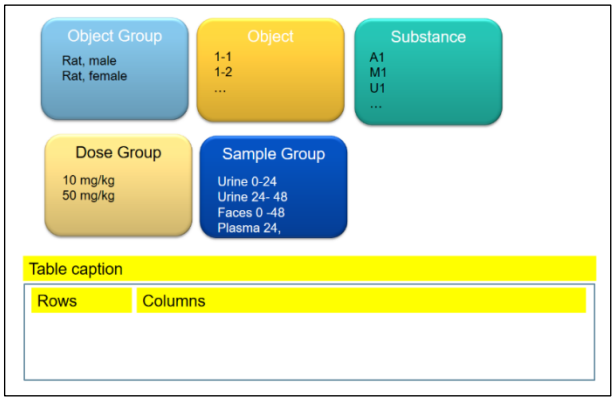


Figure 26: Start to configure a flexible table input

[5] It should be noted: There might be use cases where the user does need other grouping constellations.

One or more of the moveable objects can be chosen as the table caption. Move e.g. the “Sample Group” to the “Table Caption” region by using the mouse and select a value of the list of values (see Figure 27).

All selected values of the moveable objects in the table caption region will be used as fixed values.

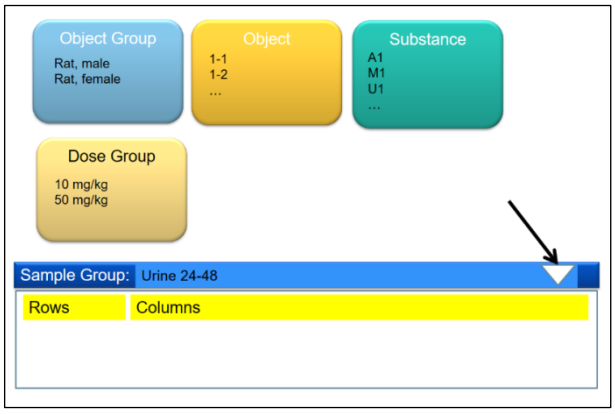


Figure 27: Sample group was moved into the table caption

Move e.g. the “Substance” to the “Row” region by using the mouse!

Move e.g. the “Dose Group” to the “Columns” region by using the mouse!

Move e.g. the “Object Group” to the “Columns” region by using the mouse!

Doing so, will result in the following table matrix (see Figure 28).

R 9.4-38: It should be possible to hide (or delete) row or columns from this *Flexible table input* where no data values are available.

R 9.4-39: It should be possible to make a manual data input in the flexible created table. The references to the groups should be stored adequately.

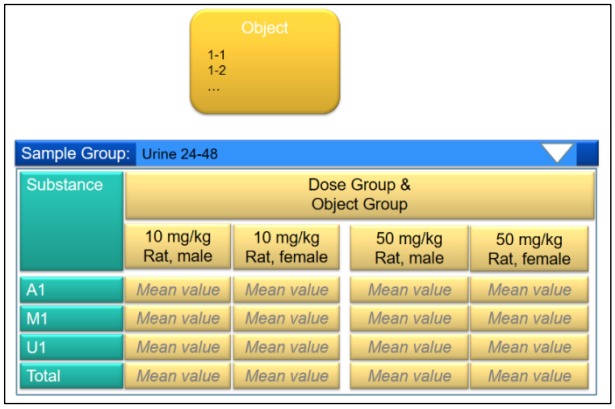


Figure 28: Configured flexible table for manual data input

R 9.4-40: Alternatively, it should be possible to move a file (Excel or CSV) with the same structure to the red marked region (see Figure 29). The module should extract the values. It is recommended that the spreadsheet file contains column headers, which are skipped by default during data transfer.

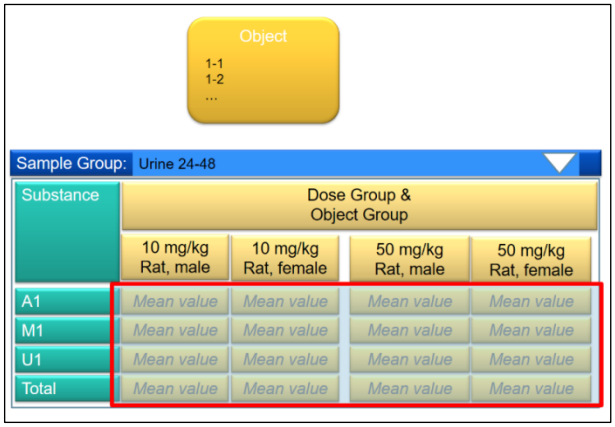


Figure 29: Configured flexible table for data import

R 9.4-41: The values of this table template could be stored by a save action. After that, the user is able to modify this table template to input other values of other group constellations.

R 9.4-42: Pay attention: In contrast to rearranging a *Sortable list*, the parameter for the arrangement of the table will not be saved. This is only done before of importing table values.

A 9.4-43: A deeper analysis is recommended, whether this function could also be used to interpret result table data from the clipboard (spreadsheet).

9.4.2.6 Rich-Text fields

A 9.4-44: A deeper analysis is needed, which functions should be supported by the built-in editor for *Rich-Text* fields.

9.4.3 Create new study

R 9.4-45: It should be possible to create a new empty study. After this action, the user can add data manually or import predefined lists.

9.4.4 Import

The preferred option is to import *Metadata* for a full study from an XML file. The adequate requirements are described in chapter 9.5.2 Import / export / validation.

This chapter has the focus on the incremental import of lists of items on basis of already partially digitalised studies.

Before an incremental import, the user is asked to select the study, where predefined lists should be imported. If one had created a new study in the action before, this study is preselected as the default study. The following import according the predefined lists is only possible into an empty study and in an adequate empty list of items and only in the needed sequence! Otherwise, error messages and the opportunity to delete the requested items should be available.

R 9.4-46: It should be possible to import a "*List of Study Object Groups*" including the description of the *Object of investigation* of this study all at once.

R 9.4-47: It should be possible to import a *Substances* list including the *Test substances* and *Metabolites* together with their relations to this study at one go.

R 9.4-48: It should be possible to import a *List of Dose Groups* for this study at one go.

R 9.4-49: It should be possible to import a *List of Sample Groups* for this study at one go.

R 9.4-50: It should be possible to import a list of used analytical methods for this study at one go.

R 9.4-51: It should be possible to import a *List of analysed Values* for this study at one go.

R 9.4-52: Imports of a full study should fail if the study is already in the collection. Adequate algorithms to check for duplicates should be implemented. If the system had found a duplicate, the user decides: “A duplicate was found … Do you want to delete this study before the import of the selected XML file?“

9.4.5 Search modules

9.4.5.1 Search - full text

R 9.4-53: The full text filter should be located at the top of the *Result list of substances*. It should be possible to search over the full data sets in the opened collection for text strings including auto completion and suggestion functions.

R 9.4-54: The result set of the full text search, the *Result list of substances*, should filterable by facet’s according different *Metadata* of different depending objects:
the study e.g. year of the GLP report, legal act type, used guideline,
the “*Object Type*” e.g. soil, sediment, crop, rotational crop, water, food, animal…
the object of investigation e.g. rat, cow, hen, ground water, wheat, carrots, …
the type of application (e.g. i.v.),
the *Sample Matrix* (e.g. full plant, fruit, serum, bile, liver, urine, sediment)
the substance e.g. name or CAS No.

A 9.4-55: A deeper analysis is needed, how to present the result list of the *Search - full text* and how to link into the other result lists. Example: If a user searches for ‘bile’ the *Result list of substances* including studies where bile was used as a *Sample Matrix*, will be prompted to the user.

9.4.5.2 Advanced search

This module should be the most important searching procedure in the *MetaPath II Tool* (see Figure 30*)*. The application should be optimised for this way.

R 9.4-56: The *Advanced search* searches for *Test substances* and *Metabolites*.

R 9.4-57: All *Metadata* of *Test substance*, “*Metabolite*, “*Study*”, *Object of investigation*, “*Metabolic pathway*, etc. could be usable as a filter. The search filters are specific according the data type of the entity attributes.

R 9.4-58: All used filter clauses are concatenated with a logical ‘AND’.

R 9.4-59: It should be possible to search for substances depending on the objects, which are defined in the users “Shopping baskets”. It means e.g. search for all substances, which are elements of the shopping basket defined for substances, or which have references to studies, which used at least those matrices listed in the “Matrix Shopping basket”.

R 9.4-60: It should be possible to use logical expressions inside of a specific search filter field e.g. in SQL syntax ((‘red’ OR ‘*blue’) AND NOT ‘dark blue’).

R 9.4-61: The initialisation time for some search options and for preparing the search results of the current *MetaPath* is not acceptable.

R 9.4-62: It should be possible to search for structure similarities of the substances to find comparable *Metabolic pathways* in pathway collections. Different types of search strings (*SMILES*, *CXSMILES*, *InChI*, *SMARTS*) could be used which are automatically transformed internally into the needed format. A graphical tool to draw a (sub)structure to be searched for should be included.

R 9.4-63: The size of the result list of the similarity search (R 9.4-62) depends on the chosen algorithm and the similarity factor. Users should be able to modify the default values (comparable with *MetaPath*: Chemical similarity options)

R 9.4-64: The similarity search (R 9.4-62) is only an additional filter clause which are concatenated with a logical ‘AND’. This way, it is possible to search for metabolism studies e.g. including a) *Substances* with a certain substructure b) in a given species (*Object of investigation*) c) with a specific treatment type (mode of *Application*)

R 9.4-65: If a user has asked for an unrealistic structure or realises a mistake in the provided structure, it should be possible to interrupt the searching process. Currently, *MetaPath* search process can not be stopped.

R 9.4-66: It should be possible to search for attributes of the *Transformation processes*.

R 9.4-67: It should be possible to search for substances by their names. Nonetheless, it should be emphasised that the *MetaPath II Tool* has no overall substance model for unknown substances (see chapter 9.3.2.2).

R 9.4-68: The users should be able to store the used search filter options locally and to load a stored request.

R 9.4-69: “Result list of substances” is the result set of the search module. It contains columns with short information (substance type, name, and identifiers) on a first level.

R 9.4-70: User can expand the “Result list of substances” to see a 2nd level with a table listing all studies which the current substance occurs in. This table is called Result list of studies. The fields visibilities can be configured by the user.

R 9.4-71: The proposed concept of the “Result list of substances” has the effect, that a specific study would be shown in all “*Result list of studies*” on the 2nd level where the *Substances* match the search filters options. Therefore, studies could appear as duplicates! There is a need to create an additional result list of the distinct “*Result list of studies*”.

9.4.6 Result list of substances

R 9.4-72: User can select substances and / or studies by checkboxes. By default, all rows are selected.

R 9.4-73: Selected rows could be used for reports and exported as XML-result-sets files. All substance *Metadata* should be exported.

R 9.4-74: Both lists could be sorted by the provided columns.

R 9.4-75: Users can activate a separate frame called *Substance details* by selecting one substance. By scrolling down in the *Result list of substances*, the individual *Substance details* frame will be refreshed.

R 9.4-76: Users can activate a separate frame called *Study details* by selecting one study. Scrolling down refreshes the individual *Study details* frame.

R 9.4-77: Users can choose two options for showing *Study details*: "*Metabolic tree"* or "Study information".

R 9.4-78: Users can select a "maximum rows shown" option with a maximum number of rows or all rows of the result set. From today's perspective, a scrolling mechanism between different pages is not needed.

Search for metabolism studies

Full text	**Open advanced search**	**Options**	**Show results**
Search string _____		Show also linked test substances _Yes_ Show also linked metabolites _Yes_ Show study details as: **Metabolic tree** Study information	Max: 999 / **all**

Facet 1	
Facet Item 1.1	999
Facet Item 1.2	999
Facet Item 1.3	999
Facet Item 1.4	999
Facet Item 1.5	999
Facet 2	
Facet Item 2.1	999
Facet Item 2.2	999
Facet Item 2.3	999

Result list of substances

Trivial Substance Name **IUPAC Name**	**CAS-No**	**PubChem**	**Type**
Imidacloprid (NE)-N-[1-[(6-chloropyridin-3-yl)methyl]imidazolidin-2-ylidene]nitramide	138261-41-3	86287518	Active Substance

Result list of studies

Meta-Path II-ID	Object	Type	Dose range	MetaPath II-ID
287	rat	Biokinetic	1 – 150 mg/kg bw	
288	rat			

6-Hydroxy-nicotinsäure 6-Hydroxypyridine-3-carboxylic acid	5006-66-6	329751860	Metabolite of Imidacloprid

Result list of studies

Meta-Path II-ID	Object	Type	Dose range	MetaPath II-ID
287	rat	Biokinetic	1 – 150 mg/kg bw	

…			

Figure 30: Mock-up of the module "Search for metabolism studies"

R 9.4-79: It is recommended to implement the possibility to switch between the *Result list of substances* and a tree structure (similar to the left panel side of *MetaPath*).

9.4.7 Substance details

R 9.4-80: The frame *Substance details* should contain all needed information of the selected *Substance* and an overview on the observed relationships of this substance ("is formed from" and "can be transformed into").

R 9.4-81: It should be possible to toggle between the *Substances* listed in “is formed from” and “can be transformed into”.

R 9.4-82: It should be possible to start the module *Substance edit* when having selected a *Substance*.

9.4.8 Substance edit

R 9.4-83: All needed *Metadata* should be considered to characterise the *Test substances*, *Metabolites* and their references.

R 9.4-84: It should be possible to use one or more different codes of chemical structure notations for the input. These codes should be transformed into one chemical notation internally which is used for similarity searches.

R 9.4-85: It should be possible to scroll from the current selected *Substance* to the next or to the previsions *Substance* in the *Result list of substances*.

9.4.9 Interoperability to other systems in the user front end

It is necessary to be able to start from the *MetaPath II Tool* into external information bases (knowledge sources) via API calls. “Extern” means from outside the *MetaPath II Tool* and therefore could be included within the user's network or also international information data bases.

9.4.9.1 Jump into external substance databases

To jump into external substance databases is reasonable in order to get more detailed toxicological data, to look for possible exposition scenarios or to find out about residue tests on this substance.

R 9.4-86: It should be possible to jump into different predefined external substance databases from a selected *Substance*.

R 9.4-87: The local application manager should be able to configure the list of the assisted preferred predefined external substance databases.

R 9.4-88: The local application manager should be able to define the substance identification field, which is necessary to be used as the dynamic parameter for the call to jump into external substance databases.

9.4.9.2 Jump into external process management systems

If *Metadata* of a study could be used to link into external process management systems, this should be assisted e.g. by using the field “Legal_Act_Identifier”. To jump into external process management systems is useful for authorities’ workflows.

R 9.4-89: It should be possible to jump into a specified different predefined external process management system from a selected “Legal_Act_Identifier”.

R 9.4-90: The local application manager should be able to configure the list of the supported predefined external process management systems.

9.4.9.3 Jump into external document archives

If *Metadata* of a study could be used to link into external document archives this should be supported.

R 9.4-91: If a study reference is refering to an external document, the *MetaPath II Tool* should provide a link to open this document in this external archive.

R 9.4-92: The local application manager should be able to configure the list of the supported predefined external document archives.

9.4.10 Visualisation

9.4.10.1 Visualisation of metabolic pathways

R 9.4-93: It should be possible to start into the visualisation of a specific *Metabolic pathway* from a current record of a "*Result list of studies*" or from the *Study details* of a selected study.

R 9.4-94: If there is the need to merge *Metabolic pathways* of different studies, these studies should be collected in a "Shopping basket" of studies (compare R 9.4-126). From the "Shopping basket" of studies, it should be possible to start into a merged visualisation, which is depicting all of these *Metabolic pathways*.

R 9.4-95: When starting the visualisation module from the selected *Shopping basket* of studies (R 9.4-94), a separate frame should be generated, where all "*Object Type* (species, plant, food, …), *Study Object Groups* (male / female, …), *Dose Groups*, "*Sample Matrix* (serum, urine, milk, eggs, whole plant, fruit, …) and analysed *Substances* are listed with checkboxes. The evaluator would then be able to show/hide entries by checking and unchecking elements of this list. Depending on the update time, the visualisation module could refresh after changing one checkbox or after activation of an additional button "Refresh".

R 9.4-96: If the evaluator uses the mouse, a Quick Substance Information should be prompted by a mouse-over event.

R 9.4-97: The content of the *Quick Substance Information* should be configurable by the user. Embedded in ther user profile, the user is able to select via checkboxes, which study substance *Metadata* will be displayed in which sequence e.g. name, structure, origin, percentage of the applied dose.

R 9.4-98: It should be possible that the user can jump from a selected substance into the visualisation module and/or external substance databases (R 9.4-86).

R 9.4-99: The *MetaPath II Tool* should inherit all *MetaPath* functions for visualising the *Metabolic pathway*.

R 9.4-100: There are many different possibilities pathways can be visualised. Specifying an interface like *Cytoscape* might be a desirable. *Cytoscape* is an open source tool with performant plugin-structure for network visualisation.

A 9.4-101: A deeper analysis is recomended for the visualisation algorithm to prevent misinterpretations of the *Metabolic pathway*.

9.4.10.2 Visualisation of concentration time curves

A 9.4-102: A deeper analysis is needed whether an internal visualisation of concentration-time-curves should be implemented or an adequate Excel-Output is preferred.

9.4.10.3 Visualisation of charts

A 9.4-103: A deeper analysis is needed whether an internal visualisation of charts should be implemented or an adequate Excel-Output is preferred.

9.4.11 Study details

R 9.4-104: On *Study details* level, the *Visualisation of metabolic pathways* as a *Metabolic tree* should be opened in parallel.

R 9.4-105: All study information objects are grouped by topics in a Study detail element hierarchy (see Figure 31). The navigation could be organised as a list with distinguishable style format. The lowest hierarchy has one simple style format.

R 9.4-106: It should be possible to set display options for building the *Study details* on the highest level. These display options are valid for the current session or could be saved as a preset in the user profile.

R 9.4-107: It should be possible to collapse / expand all nodes at once or separately by a click on the *Study detail element*. The content will be revealed after a click on the Study Detail Element in the next table row.

R 9.4-108: If the user has the adequate role, it should be possible to start into the edit-mode on each *Study detail element* by clicking a specific “Edit” icon.

R 9.4-109: It should be possible to add elements into the adequate *Shopping basket* with a specific icon.

R 9.4-110: A “Delete” icon should be displayed in the expanded mode if the user has the adequate role.

R 9.4-111: The content of a *Study detail element* is compiled from a list of element fields and different repeating blocks of child tables depending of the “*Object Type*”.

Study <name> Show only Study – Rich-Texts: ☐ Hide empty fields: ☐ …. ☐		
Study General Info's		
Study Substances		
Test substance		
<name>		
<name>		
Metabolites		
<name>		
Content of the 1st metabolite		
<name>		
Content of the 2nd metabolite		
…		
Study Objects Groups		
M		
Content of the 1st study object group "M"		
<name>		
…		
Study Dose Groups		
<name>		
…		
Study Sample Groups		
<name>		
<name>		
…		
Study Methods		
<name>		
<name>		
…		
Study Analysed Values		

Figure 31: Mock-up of the module "Study details"

9.4.11.1 Study detail element: Study general Info's

To do in the project planning phase.

9.4.11.2 Study detail element: Test substance

To do in the project planning phase.

9.4.11.3 Study detail element: Metabolite

To do in the project planning phase.

9.4.11.4 Study detail element: Study object group

Study Objects Groups

M

Object Group Name	Male Group
Vertebrate	True
Object Type	Animal
Object	Rat

Object Group Parameter *(specific for Object Type)*

Parameter Type	**Value**
	Remark (optional row)
Sex	Male
Strain	My strain
Age at study initiation	6 weeks
Weight at study initiation	250 – 350 g
Source	
Housing	
…	
Acclimatisation period	7 days

Object Group Remarks

Context_type	The selected context type
RTF-Text	

Elements of the Object Group

Object Number	**Dose Group**
M1	DG 50
M2	DG 50
M3	DG 100
M4	DG 100
M5	DG 300
M6	DG 300

Figure 32: Mock-up of the module: Study Detail Element: *Study Object Group*

The content block of a *Study Object Group* consists of isolated fields on the group level and accompanied with three depending tables (see Figure 32)

- List of individual object numbers which are elements of the *Study Object Group*
- List of classification parameter of the *Study Object Group*
- List of *Rich-Text* descriptions of the *Study Object Group*

9.4.11.5 Study detail element: Study object group

To be completed during the project planning phase.

9.4.11.6 Study detail element: Study object group

To be completed during the project planning phase.

9.4.11.7 Study detail element: Study method

To be completed during the project planning phase.

9.4.11.8 Study detail element: Study analysed values

R 9.4-112: The content block *Study analysed values* is the most important result data block for a study. It should be organised like a spreadsheet (see Figure 33).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Study ID	Object Group	Object ID	Dose Group	Testsubstance	Sample Group	Sample time start	Sample time end	Sample time unit	Sample No	Method ID	Analyzed fraction	Analyzed substance	LT_LOQ	value from	value to	Unit	calculated as
2	xyz	M	100	50	P1	0-24	0	24	h	U100-24	M185	Acetone	Not Identified		22		TRR	P1
3	xyz	M	100	50	P1	0-24	0	24	h	U100-24	M185	Hexene	Not Identified		10		TRR	P1
4	xyz	M	100	50	P1	0-24	0	24	h	U100-24	M185	0,1N KOH	Not Identified		3		TRR	P1
5	xyz	M	100	50	P1	0-24	0	24	h	U100-24	M185	1N KOH	Not Identified		5		TRR	P1
6	xyz	M	100	50	P1	0-24	0	24	h	U100-24	M185	Rest	Not Identified		3		TRR	P1
7	xyz	M	100	50	P1	0-24	0	24	h	U100-24	M185	All	P1		15		TRR	P1
8	xyz	M	100	50	P1	0-24	0	24	h	U100-24	M185	All	M1		25		TRR	P1
9	xyz	M	100	50	P1	0-24	0	24	h	U100-24	M185	All	M2		8		TRR	P1
10	xyz	M	100	50	P1	0-24	0	24	h	U100-24	M185	All	U1		0,02		TRR	P1
11	xyz	M	100	50	P1	0-24	0	24	h	U100-24	M185	All	U2		2		TRR	P1
12	xyz	M	100	50	P1	0-24	0	24	h	U100-24	M185	All	Not Identified		10		TRR	P1
13	xyz	M	100	50	P1	24-48	24	48	h	U100-48	M185	Acetone	Not Identified		10		TRR	P1
14	xyz	M	100	50	P1	24-48	24	48	h	U100-48	M185	Hexene	Not Identified		5		TRR	P1
15	xyz	M	100	50	P1	24-48	24	48	h	U100-48	M185	0,1N KOH	Not Identified		4		TRR	P1
16	xyz	M	100	50	P1	24-48	24	48	h	U100-48	M185	1N KOH	Not Identified		2,5		TRR	P1
17	xyz	M	100	50	P1	24-48	24	48	h	U100-48	M185	Rest	Not Identified		1		TRR	P1
18	xyz	M	100	50	P1	24-48	24	48	h	U100-48	M185	All	P1		7		TRR	P1
19	xyz	M	100	50	P1	24-48	24	48	h	U100-48	M185	All	M1		18,5		TRR	P1
20	xyz	M	100	50	P1	24-48	24	48	h	U100-48	M185	All	M2		6		TRR	P1
21	xyz	M	100	50	P1	24-48	24	48	h	U100-48	M185	All	U1	<	0,01		TRR	P1
22	xyz	M	100	50	P1	24-48	24	48	h	U100-48	M185	All	U2		1,5		TRR	P1
23	xyz	M	100	50	P1	24-48	24	48	h	U100-48	M185	All	Not Identified		13		TRR	P1

Figure 33: Mock-up of the module: Study detail element: Study analysed values

R 9.4-113: For each analysed value, one row exits in this table. Here, the design should be commited to the “tidy data” principle:
- each variable forms one column
- each observation forms one row
- each cell is a single measurement

R 9.4-114: It should be possible to change into edit mode. However, this should be used only for error correction of some values.

If manual data input is needed for the analysed values, the *Flexible table input* option should be used. The columns represent the *Metadata* of the analysed value. The user should be able to arrange an input table, similar to the structure of the table in the GLP study report.

R 9.4-115: The *Flexible table input* option should offer row filters by each column (comparable with Excel).

R 9.4-116: It should be possible to hide columns in the *Flexible table input*.

R 9.4-117: The content block *Study analysed values* is the module to initiate the creation of a pivot table.

9.4.12 Study edit

R 9.4-118: There are the following ways to store study data in the *MetaPath II Tool* a) manual data input b) import of a data set according the defined XML schema and c) via *API* from external modules.

R 9.4-119: It should be possible to create a *Study Data Set* manually in the *MetaPath II Tool* without using the import or *API* functions.

R 9.4-120: The *MetaPath II Tool* should be capable of partial import of tables for different entities (e.g. *List of Dose Groups*, "*List of Sample Groups*, "*List of Substances*) by using the flexible table input (compare 9.4.2.5) via an import icon.

R 9.4-121: It should be possible to modify an imported *Study Data Set* in general. However, the decision "Is it allowed to modify this current dataset – or not" depends on the status of the *Study Data Set*. Specific rules should be defined and enforced by the *Governace body*.

9.4.12.1 Study edit: Study object groups

R 9.4-122: Figure 34 should be an example for a module where a *Sortable list* of numbers (moveable items) is embedded. Such a list could be build up a) manually by the "Add number" button b) or by importing a spreadsheet.

To be completed during the project planning phase.

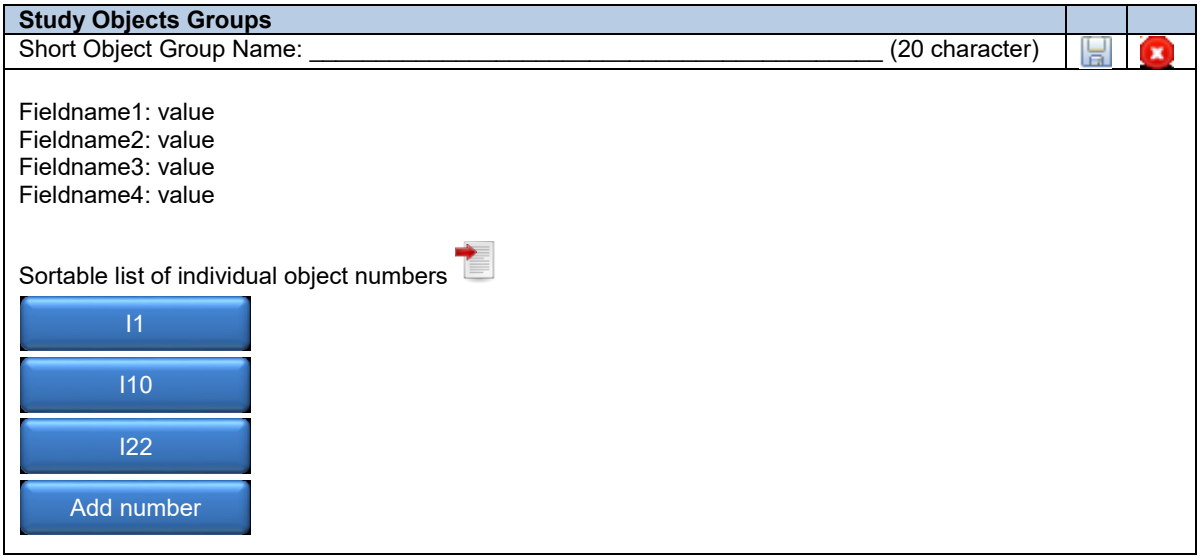


Figure 34: Mock-up of the module Study Edit: Study Object Groups

9.4.12.2 Study edit: Study analysed values

To be completed during the project planning phase.

R 9.4-123: It should be possible to use the smaller "<"-character for the result values.

9.4.12.3 All other Study edit modules

To be completed during the project planning phase.

9.4.13 Compare

R 9.4-124: The *MetaPath II Tool* should inherit all *MetaPath* functions for comparing *Metabolic trees* (e.g. data comparison within and across different taxa) and to visualise different and identical parts.

9.4.14 Management of attachments

R 9.4-125: The *MetaPath II Tool* should be able to manage attachments. At least attached study reports and flow charts should be stored.

9.4.15 User set management module

The following chapters describe helpful supporting user functions to handle sets like a *Shopping basket*. These functions are independent from the endpoint, which is in focus. However, it could be that the frequency of using the different functions will differ between the endpoint experts.

9.4.15.1 User “Shopping basket”

R 9.4-126: The user is able to create personal *Shopping baskets* for **all** different object types, which are managed by the *MetaPath II Tool* e.g. for studies, substances, object types, dose groups, sample groups, matrix, picklist-items.

R 9.4-127: The content of the users *Shopping basket* could be filled by checking the adequate objects in a result list. User can store these *Shopping baskets* with user defined names. The *Shopping basket* itself or some objects of the *Shopping basket* can be easily edited and/or deleted.

R 9.4-128: A *Shopping basket* could be understood as a task list. The user loads the objects, which should be modified or checked. Each object, which was modified or checked, will be deleted from the adequate user *Shopping basket* until it is empty.

R 9.4-129: A *Shopping basket* could also be used as a filter for the advanced search (see chapter 9.4.5.2).

R 9.4-130: A *Shopping basket* of studies could be used for a visualisation of the referenced *Metabolic pathways* as well (compare R 9.4-94).

9.4.15.2 The user working stack “List of relevant studies”

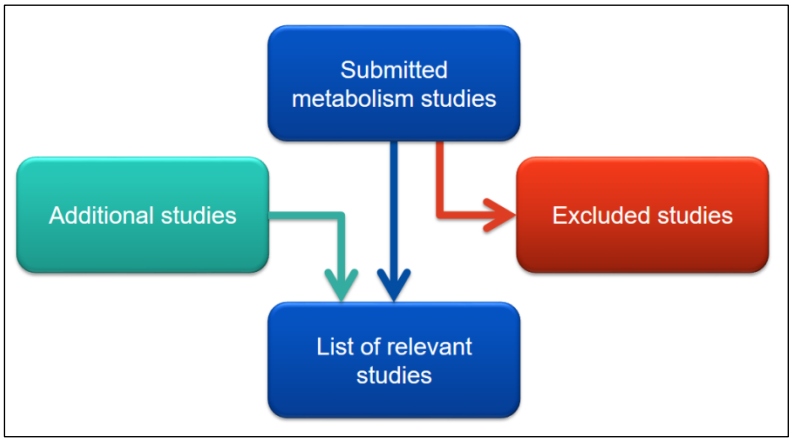


Figure 35: Define the “List of relevant studies”

The term *List of relevant studies* depends on the endpoint and the legal act addressed. The *List of relevant studies* is a specific user *Shopping basket* of studies (see Figure 35).

R 9.4-131: At the end of the assessment process, *Evaluators* should be able to define a user storable set of submitted and additional studies. *Evaluators* can specify a name for the *List of relevant studies*. This collection is not a separate collection; it is a user defined subset view of the whole study collection.

R 9.4-132: The *List of relevant studies* is specific for each section (Toxicology, Residues, and Environmental Fate).

R 9.4-133: *Evaluators* should be able to flag submitted studies with “not to consider” and to exclude them from the following consideration. The reason (justification) for excluding studies should be stored in the IT-Tool.

R 9.4-134: *Evaluators* should be able to screen for additional *Metabolic pathways* of the same *Active Ingredient* or comparable pathways from outside of the current legal act which are already stored in the reference collections.

R 9.4-135: It should be possible to complete the *List of relevant studies* with other studies where similar metabolites were found.

R 9.4-136: *Evaluators* are able to group the elements of the *List of relevant studies* into different groups. A group is characterised by a user defined name.

R 9.4-137: *Evaluators* can use the *List of relevant studies* for reports. The defined groups could be used for filtering or aggregation of results.

Because there are currently no uniform criteria for similarity searches in other reference collections, the *List of relevant studies* will vary between the *Evaluators*.

R 9.4-138: It should be possible to modify the *List of relevant studies* after the peer review process to include input from other member states.

R 9.4-139: *Evaluators* are able to include additional studies into the *List of relevant studies*, which are not currently in submitted dossiers but in the local collection of the *MetaPath II Tool*.

R 9.4-140: The *MetaPath II Tool* should allow archiving of the search criteria used for the screening step (R 9.4-134 e.g. name of the reference collection, timeliness, search criteria, ...)?

R 9.4-141: The archive of the used screening search strategies (R 9.4-140) should become a part of the *Aggregated raw data* package and thus also be delivered by the applicant to the agency?

9.4.15.3 The “View of substances to evaluate”

The term *View of substances to evaluate* (see Figure 36) should be understood as the distinct collection of all substances contained in the user-working stack *List of relevant studies*.

This list represents the maximum assessment framework for the substance level. Substances that are not included in this list cannot be considered further in the following process steps.

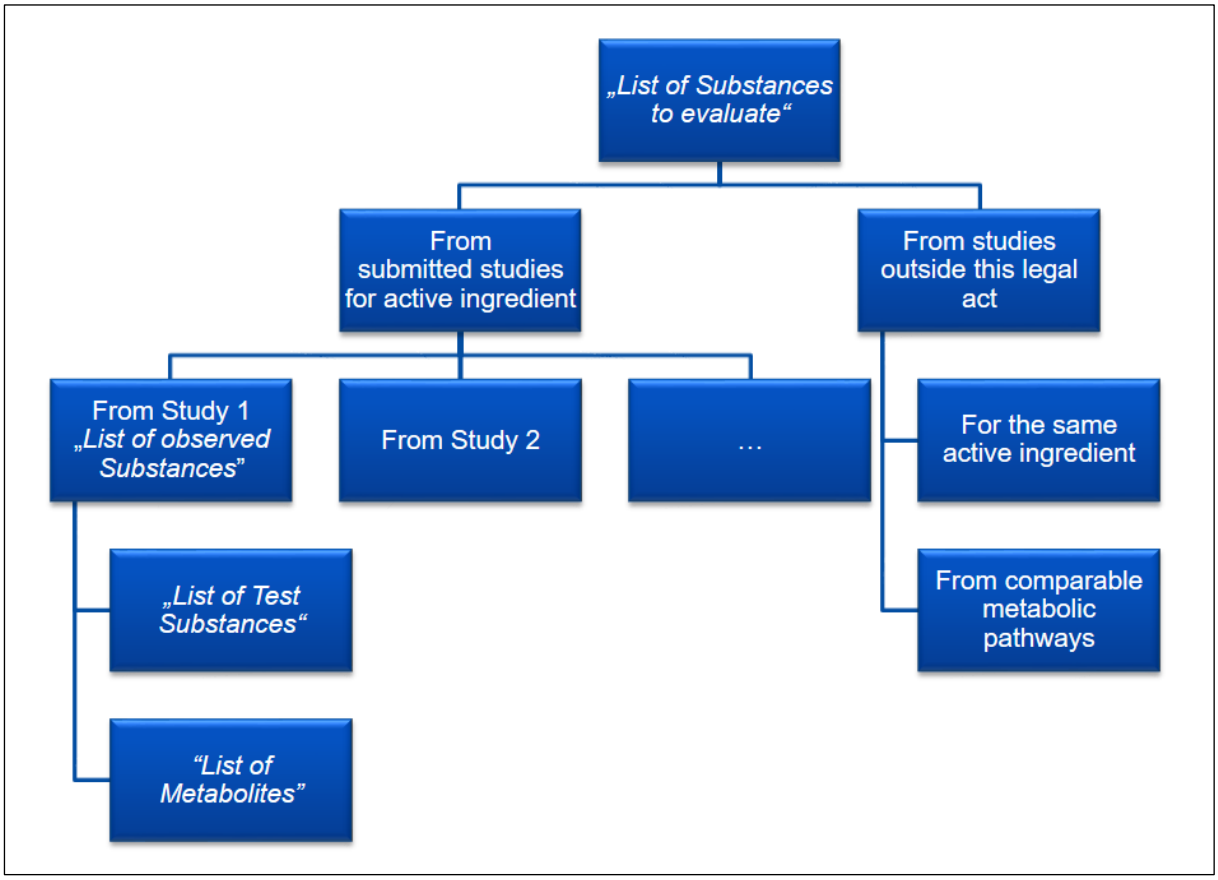


Figure 36: Sources of the “View of substances to evaluate”

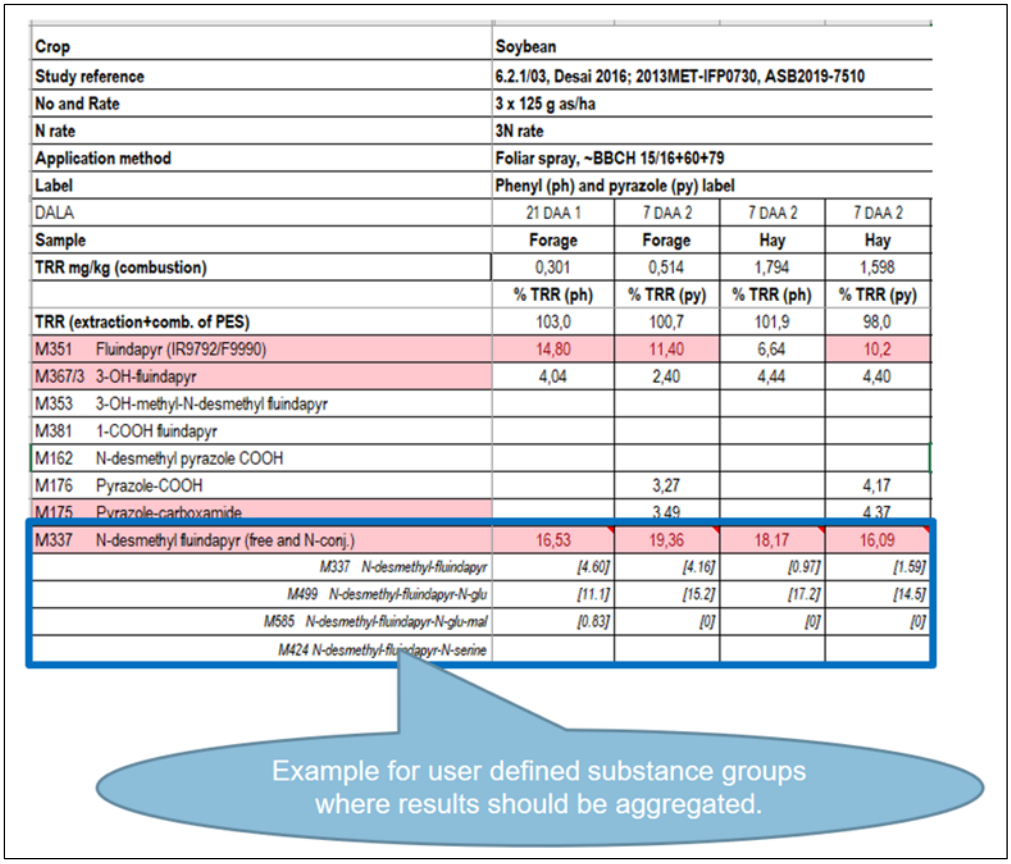
R 9.4-142: *Evaluators* should be able to get the overall view of all substances of the studies included in the *List of relevant studies*. This view is called *View of substances to evaluate*.

Because the *List of relevant studies* are different for the residue experts, toxicologists and ecotox experts, the *View of substances to evaluate* could also be different.

9.4.15.4 The users working stack “Set of substances”

R 9.4-143: *Evaluators* should be able to include all (default option) or only the relevant substances of the *View of substances to evaluate* in a user storable snapshot (user working stack of a *Set of substances* or user shopping cart). Evaluators can specify a name for the *Set of substances* specific for this legal act.

R 9.4-144: *Evaluators* can use the *Set of substances* for reports. The defined group names could be used for filtering or aggregation of results (see Figure 37).



Crop	**Soybean**			
Study reference	**6.2.1/03, Desai 2016; 2013MET-IFP0730, ASB2019-7510**			
No and Rate	**3 x 125 g as/ha**			
N rate	**3N rate**			
Application method	**Foliar spray, ~BBCH 15/16+60+79**			
Label	**Phenyl (ph) and pyrazole (py) label**			
DALA	21 DAA 1	7 DAA 2	7 DAA 2	7 DAA 2
Sample	**Forage**	**Forage**	**Hay**	**Hay**
TRR mg/kg (combustion)	0,301	0,514	1,794	1,598
	% TRR (ph)	**% TRR (py)**	**% TRR (ph)**	**% TRR (py)**
TRR (extraction+comb. of PES)	103,0	100,7	101,9	98,0
M351 Fluindapyr (IR9792/F9990)	14,80	11,40	6,64	10,2
M367/3 3-OH-fluindapyr	4,04	2,40	4,44	4,40
M353 3-OH-methyl-N-desmethyl fluindapyr				
M381 1-COOH fluindapyr				
M162 N-desmethyl pyrazole COOH				
M176 Pyrazole-COOH		3,27		4,17
M175 Pyrazole-carboxamide		3,49		4,37
M337 N-desmethyl fluindapyr (free and N-conj.)	16,53	19,36	18,17	16,09
M337 N-desmethyl-fluindapyr	*[4.60]*	*[4.16]*	*[0.97]*	*[1.59]*
M499 N-desmethyl-fluindapyr-N-glu	*[11.1]*	*[15.2]*	*[17.2]*	*[14.5]*
M585 N-desmethyl-fluindapyr-N-glu-mal	*[0.83]*	*[0]*	*[0]*	*[0]*
M424 N-desmethyl-fluindapyr-N-serine				

Figure 37: User defined aggregation groups on substance level

R 9.4-145: *Evaluators* are able to copy *Metadata* of the *Set of substances* into a clipboard. The definition of the required *Metadata* and the required formats should be done in a later project phase.

R 9.4-146: As studies from different applicants and / or different laboratories / different years are to be combined, different synonyms for one and the same metabolite may have been used in the *Aggregated raw data*. The *Evaluator* should be able to pool identical substances of different names across the studies.

R 9.4-147: For one or more selected elements of this list, *Evaluators* should be able to open a detailed view of all TRR results measured in the different studies. The users are able to select / deselect studies from the *List of relevant studies* for this detailed view. Reports should be generatable from the current view.

9.4.16 Report

R 9.4-148: The report generator assists different output formats, at least PDF, Word, Excel.

R 9.4-149: The report generator should support a full version and a sanitised version.

R 9.4-150: An important output format would be the Word format to use the result in other text processing contexts.

R 9.4-151: For the Word output format, paragraph and table templates provide a way to automatically adjust the used report content format to each target assessment report format template before inserting into an assessment report.

R 9.4-152: The rendered version of the *Metabolic pathways* images should have an adequate, publishable quality.

9.4.16.1 Listings

R 9.4-153: *Evaluators* should be able to create default scientific reports for each type of the result listincluding all or selected rows.

9.4.16.2 Default study reports

R 9.4-154: The *MetaPath II Tool* contains a *Report* which can be used to create a textual study summary over all study *Metadata* information of the selected study. This default study report should be uploaded into *IUCLID* (see also R 9.5-18).

R 9.4-155: The users should be able to recalculate results from one substance to another substance and need to be able to aggregate results according to their expert knowledge.

9.4.16.3 Pivot tables

The creation of pivot tables is an interactive process.

R 9.4-156: **The created pivot tables should not be stored in the *MetaPath II Tool*!** If the user want to have this created pivot table for a report, the table could be copied into the clipboard.

O 9.4-157: There is the risk of getting data inconsistencies between possibly updated *Aggregated raw data* and previously generated pivot tables, which now have been embedded in *Rich-Text* fields without storing the rearranged empty pivot table template (see R 9.4-156).

A 9.4-158: A deeper analysis is needed to define needed process rules to avoid follow-up data inconsistencies described in O 9.4-157.

A 9.4-159: A deeper analysis is recommended to define default pivot table templates.

R 9.4-160: *Evaluators* should be able to create default pivot tables from the *Aggregated raw data* of one selected study by choosing one of the *Predefined study summary tables*. Chapter 11.7 contains proposals for pivot tables.

R 9.4-161: Pivot tables should be free of maximum column limitation.

R 9.4-162: The users should be able to create flexible pivot tables from the *Aggregated raw data* of one selected study by using the defined groups in the study or those that have been defined by the *Evaluator* ("*List of Study Object Groups*", *List of Substances*, "*List of Dose Groups*, "*List of Sample Groups*).
The users should be able to store such "flexible pivot table" templates to reuse them in later similar problem settings.

R 9.4-163: The predefined substance groups ("known" and "unknown") could be used for the aggregation of the results.

R 9.4-164: *Evaluators* should be able to create additional substance groups by defined characteristics (e.g. according to functional groups, conjugates, …).

R 9.4-165: If mass balance data should be presented, the total values of the Total radioactive residue (TRR) should be calculated automatically in the pivot tables.

R 9.4-166: It should be possible to recalculate "Analysed Values" from one substance to another "calculated as substance".

R 9.4-167: It should be possible to calculate mean, standard deviation and the count for individual data and to add those in additional columns.

R 9.4-168: If concentration-over-time values are measured, corresponding summary tables of the results should be created automatically.

R 9.4-169: An adequate graphical output of measured concentration-over-time values of request R 9.4-168 is highly recommended (compare chapter 9.4.10.2).

9.4.16.4 Summary reports

R 9.4-170: *Evaluators* should be assisted to write the higher summary levels for active ingredients regarding the risk and hazard assessment of metabolites by specific reports. The created report should be in line in terms of format and content of the assessment reports (see chapter 7.5.2).

A 9.4-171: A harmonisation of the format of the assessment reports is needed on *OECD* level. Otherwise, many similar reports would have to be programmed, which would only increase the projects overall effort but would not contribute to an improved workflow. A deeper analysis is needed.

R 9.4-172: Appendix G, supposed to be generated by *EFSA* automatically.

9.4.17 Documentation

R 9.4-173: It should be possible to start a context sensitive User Documentation from all modules.

R 9.4-174: The User and the System Documentation should be part of the proposed *Open Source Project*.

9.5 System requirements for the MetaPath II Tool

9.5.1 Management

9.5.1.1 System management

R 9.5-1: The used database management system provides all the necessary system management functions for a secure and effective runtime of the applications (system updates, backup, analysis tools, …).

9.5.1.2 User management

R 9.5-2: The *MetaPath II Tool* supports a user management in combination with a *Role Concept*.

R 9.5-3: The *MetaPath II Tool* should be able to manage users in a local autonomous environment. However, it should be configurable in a way, that *MetaPath II Tool* is alternatively connected to an external user administration/authentication system (e.g. LDAP system).

R 9.5-4: A privileged user can manage other users/roles.

9.5.1.3 Substance management

R 9.5-5: A privileged user can manage the central list of *Substances*.

9.5.1.4 Picklist management module

R 9.5-6: It should be transparent for the users which picklist elements could be used in which editing module and which input field. This module creates a list of all *Picklists and picklist elements*, grouped by the *Picklists*.

R 9.5-7: Currently, there is a need for users to have a mechanism for electronically requesting new elements of the picklist from the ***Governace body*** and to be able to use them. The function could be very helpful for the general acceptance of the tool.

9.5.1.5 Specialised administrator module

All content related management functions should be summarised in a specialised administrator module

R 9.5-8: Merge two substances and their references, because these substances are duplicates.

R 9.5-9: Completion of pathways by related studies (same or other active substances).

R 9.5-10: Identification of structurally related compounds over all studies.

9.5.2 Import / export / validation

R 9.5-11: The *MetaPath II Tool* could import and export data sets of *Aggregated raw data* of Metabolism studies. The *Scheme definition* for the *MetaPath II.XML* file is the data interface description to transfer data sets of Metabolism studies between different local collections of the *MetaPath II Tool*.

R 9.5-12: Before of the import of the *Aggregated raw data* the data should be validated against the schema description. If errors occur, a meaningful user message has to be generated. The user would be able to cancel the import module.

R 9.5-13: The *MetaPath II.XML* file is one output which will be generated according the *Scheme definition* by the *MetaPath II Tool*.

R 9.5-14: The *MetaPath II Tool* or an external tool should be able to convert the existing MSS-Composer XML files into the new schema description according the *Scheme definition*.

R 9.5-15: The *MetaPath II Tool* assist the checking of validation rules on study level. This report should be generated with a detailed information about the emerged errors and warnings.

9.5.3 Assist the transport step via IUCLID

If the options *OECD Domain Type* or *As attachment* should be implemented there is no user requirement that *IUCLID* has to manage *Aggregated raw data* of metabolism studies. However, the transport of the *Aggregated raw data* as an attachment is sufficient.

R 9.5-16: According to the approach of the *MetaPath II Tool* the GLP report plus the adequate attached XML file should be submitted with the help of *IUCLID*. Both documents have the same document life cycle.

R 9.5-17: All data, which should be published, are content of the *OECD* harmonised templates in *IUCLID*. The Confidential Business Information (CBI) flagging and the needed publication rules are already implemented in *IUCLID*. Because the content in the *Aggregated raw data* is the same semantic information but only in a different format, a publication is not needed.

R 9.5-18: Applicants should be able to store the output of the *Report* of the *MetaPath II Tool* into the *OECD* harmonised template in the block “Applicant summary and conclusion”.

R 9.5-19: Applicants should be obliged to add information in the section “Administrative data” and “Applicant summary and conclusion”. There, relevant aspects of the study including the obtained conclusions in context of the regulatory context should be summarised.

The following function is not a user a requirement for the *MetaPath II Tool*. However, for *IUCLID* it would be worth striving for an import function that can read *Metadata* information of the attached “MetaPath II*.XML*” files, which are necessary for other *IUCLID* user functions e.g. to import *Metadata* of the *List of Substances* and the relationship between the substances (parent ➔ child).

Additional requirement if the option “*OECD Domain Type*” (see 9.8.3.1) should be implemented:

R 9.5-20: An additional import function is needed to extract the needed data from the *Study summary metadata* into the *MetaPath II Tool* if the option *OECD Domain Type* should be implemented. It has to be emphasised, that this import is a ONE WAY import. It would not be desirable to export these data back to a *Study summary metadata* schema.

O 9.5-21: If both, *IUCLID* and the *MetaPath II Tool* should be able to include the corresponding *Metadata*, the project effort increases significantly and additional project dependencies are created. These project dependencies require an update cycle synchronisation. The overall project cost for the creation and for the annual maintenance increases.

9.6 Internal stakeholder MetaPath II-instances

R 9.6-1: Stakeholders can create their own internal instances of the MetaPath II for specific questions, which are outside the *MetaPath II Ecosystem*.

9.7 Usage of information of metabolism studies in (Q)SAR

One benefit of the provided *Solution approaches* is that (Q)SAR Tools could use the published data for the development of (Q)SAR models.

R 9.7-1: The *Authorities MetaPath II collection* could be the official data source for the *OECD* (Q)SAR-Toolbox regarding the pesticide metabolism pathway and kinetics models.

R 9.7-2: The *Authorities MetaPath II collection* should support the interoperability with (Q)SAR Tools through an *API* (see section 9.2.6).

9.8 Transport concepts for aggregated raw data of metabolism studies

9.8.1 Analogy of transport concepts

The following chapters help to understand the characteristics of the *OECD* terminology of transport concepts by starting with an analogy.

9.8.1.1 Letter

Figure 38: The woman (sender) and the man (receiver) cannot speak directly. The information is transported in a container (“Letter”). It is not the primary role of the mailman to read the letter.

The message will be transferred without any interpreter.

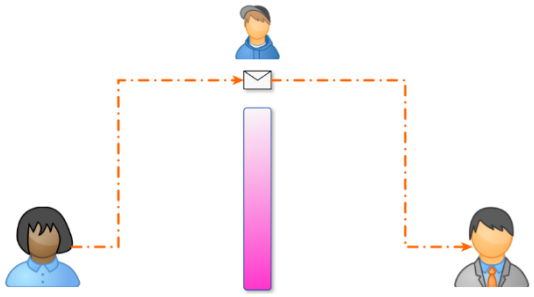


Figure 38: Analogy view for the term “attachment”

Table 11: Replacement of the analogy into the *OECD* terminology for the term “Attachment”

Analogy term	**Replacement into the OECD terminology**
Woman	Applicant
Man	Authority
Mailman	IUCLID
Envelope of the letter	IUCLID Dossier
Letter	Attachment

9.8.1.2 Call center

Figure 39: The woman (sender) and the man (receiver) cannot speak directly. In addition, it is forbidden to write a letter. Under this circumstances, the woman will give all information to the mailman orally, who is working in an call center. It is a clever mailman. The mailman heard and interpreted the oral message. Afterwards, he shareed this message with the receiver who listens and interpretes the oral message.

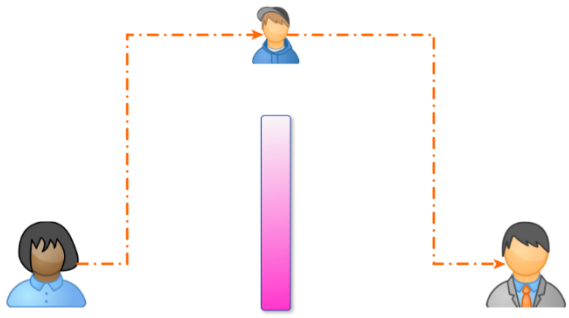


Figure 39: Analogy view of the term “OHT”

The content of the message will be transferred with two translation steps: speak ➔ hear ➔ interpret.

Table 12: Replacement of the analogy into the *OECD* terminology for the term *OHT*

Analogy term	**Replacement into the OECD terminology**
Woman	Applicant
Man	Authority
Mailman	IUCLID and IUCLID Dossier
Woman: Speak ➔ Hear	Export via IUCLID API
Mailman: Hear ➔ Interpret	IUCLID API to IUCLID Data model
Mailman: Speak ➔ Hear	IUCLID Export (or IUCLID Data model to IUCLID API)
Man: Hear ➔ Interpret	Import from IUCLID Export (or import from IUCLID API)

9.8.1.3 Combination of letter and mailman

Figure 40: The woman (sender) and the man (receiver) cannot speak directly. They can use a message transport container (“Letter”) to transport the information directly. However, the woman want to give the most important information also to her friend without retyping the content. Therefore, the woman ask the mailman in his office to read the letter and to tell the most important content to her friend.

The message will be transferred without any interpreter between the main actors.

The content of the message will be filtered with one translation step: read ➔ interpret

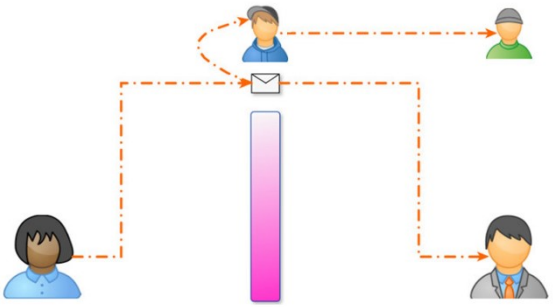


Figure 40: Analogy view of the term “Attachment in combination with an *IUCLID* Add-In”

Table 13: Replacement of the analogy into the *OECD* terminology for the combination of an “Attachment” and an *IUCLID* Add-In

Analogy term	**Replacement into the OECD terminology**
Woman	Applicant
Man	Authority
Mailman	IUCLID and IUCLID Dossier
Envelope of the letter	IUCLID Dossier
Letter	Attachment
Friend	OECD Toolbox
Read the most important content	New IUCLID Add-In extract the substance information and the relation between the substances

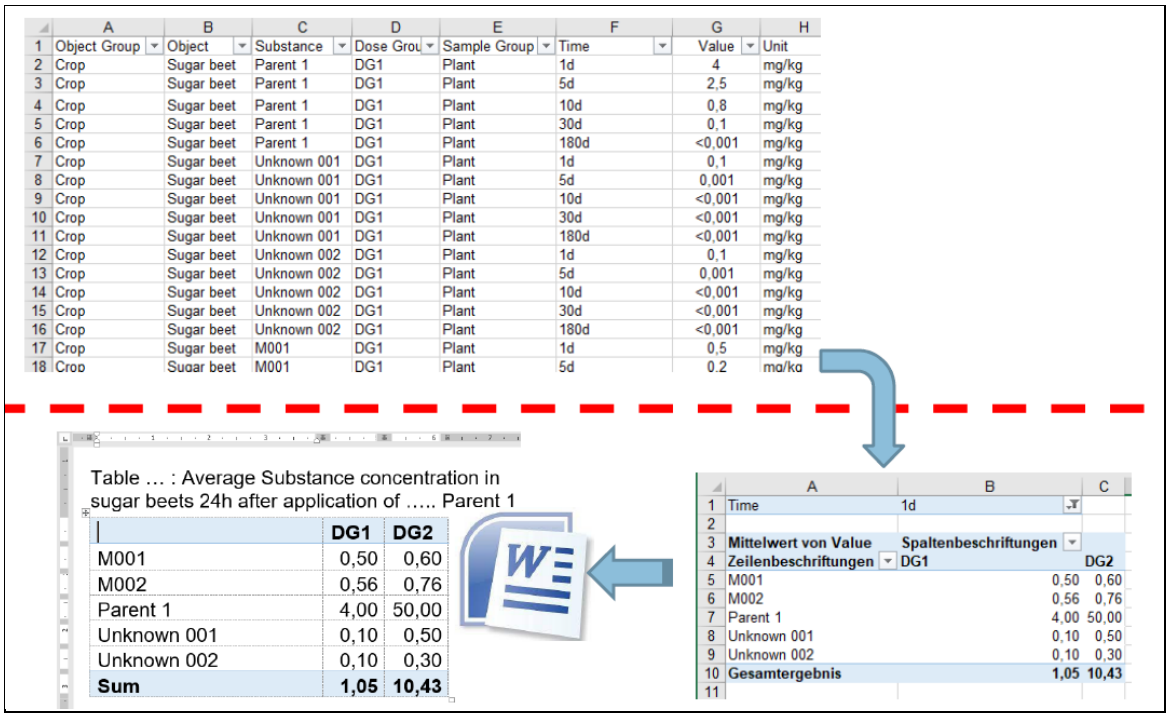
9.8.2 Requirements to transport aggregated raw datafor metabolism studies

9.8.2.1 Granularity of information

R 9.8-1: The information flow of *Aggregated raw data* should be organised according the specific needs on the level of *Aggregated raw data* in the needed granularity.

O 9.8-2: It is uncritical for the system if sematically identical information is transmitted in parallel in other compilations in a different format.

The following Figure 41 should illustrate the difference of *Aggregated raw data* and data in summarised tables in *Rich-Texts* using the example of the analysed values.



	A	B	C	D	E	F	G	H
1	Object Group	Object	Substance	Dose Grou	Sample Group	Time	Value	Unit
2	Crop	Sugar beet	Parent 1	DG1	Plant	1d	4	mg/kg
3	Crop	Sugar beet	Parent 1	DG1	Plant	5d	2,5	mg/kg
4	Crop	Sugar beet	Parent 1	DG1	Plant	10d	0,8	mg/kg
5	Crop	Sugar beet	Parent 1	DG1	Plant	30d	0,1	mg/kg
6	Crop	Sugar beet	Parent 1	DG1	Plant	180d	<0,001	mg/kg
7	Crop	Sugar beet	Unknown 001	DG1	Plant	1d	0,1	mg/kg
8	Crop	Sugar beet	Unknown 001	DG1	Plant	5d	0,001	mg/kg
9	Crop	Sugar beet	Unknown 001	DG1	Plant	10d	<0,001	mg/kg
10	Crop	Sugar beet	Unknown 001	DG1	Plant	30d	<0,001	mg/kg
11	Crop	Sugar beet	Unknown 001	DG1	Plant	180d	<0,001	mg/kg
12	Crop	Sugar beet	Unknown 002	DG1	Plant	1d	0,1	mg/kg
13	Crop	Sugar beet	Unknown 002	DG1	Plant	5d	0,001	mg/kg
14	Crop	Sugar beet	Unknown 002	DG1	Plant	10d	<0,001	mg/kg
15	Crop	Sugar beet	Unknown 002	DG1	Plant	30d	<0,001	mg/kg
16	Crop	Sugar beet	Unknown 002	DG1	Plant	180d	<0,001	mg/kg
17	Crop	Sugar beet	M001	DG1	Plant	1d	0,5	mg/kg
18	Crop	Sugar beet	M001	DG1	Plant	5d	0,2	mg/kg

	DG1	DG2
M001	0,50	0,60
M002	0,56	0,76
Parent 1	4,00	50,00
Unknown 001	0,10	0,50
Unknown 002	0,10	0,30
Sum	**1,05**	**10,43**

	A	B	C
1	Time	1d	
2			
3	**Mittelwert von Value**	**Spaltenbeschriftungen**	
4	**Zeilenbeschriftungen**	**DG1**	**DG2**
5	M001	0,50	0,60
6	M002	0,56	0,76
7	Parent 1	4,00	50,00
8	Unknown 001	0,10	0,50
9	Unknown 002	0,10	0,30
10	**Gesamtergebnis**	**1,05**	**10,43**
11			

Figure 41: The difference between *Aggregated raw data* (above the red line) and textual summaries embedded in text blocks (bottom left). The created word table contain a subset of analysed values filtered from the *Aggregated raw data* in a flexible pivot outlook design.

9.8.2.2 The direction of the informations flow

To design a transport concept and consider the pros and cons, it is important to know the types of the “Interested Parties”[6] (*ISO 9001*), who is the data consumer of the transported data.

Many *MetaPath II Collections* could be build globally in parallel according to chapter 9.2 in the long-term. In this context, communication partners should be able to exchange *Aggregated raw data* on demand (see Figure 42) without any restrictions (nondirected independent data exchange).

R 9.8-3: The most important type of data consumer tool will be another instance of the *MetaPath II Tool*. It should be possible to exchange *Aggregated raw data* of metabolism studies including the *Metabolic pathways* between different *MetaPath II Collections* directly without any conversion steps. No additional modules, other than the integral export and import module are needed.

[6] Compare ISO 9001

Quality management systems - Requirements (ISO 9001:2015); German and English version EN ISO 9001:2015

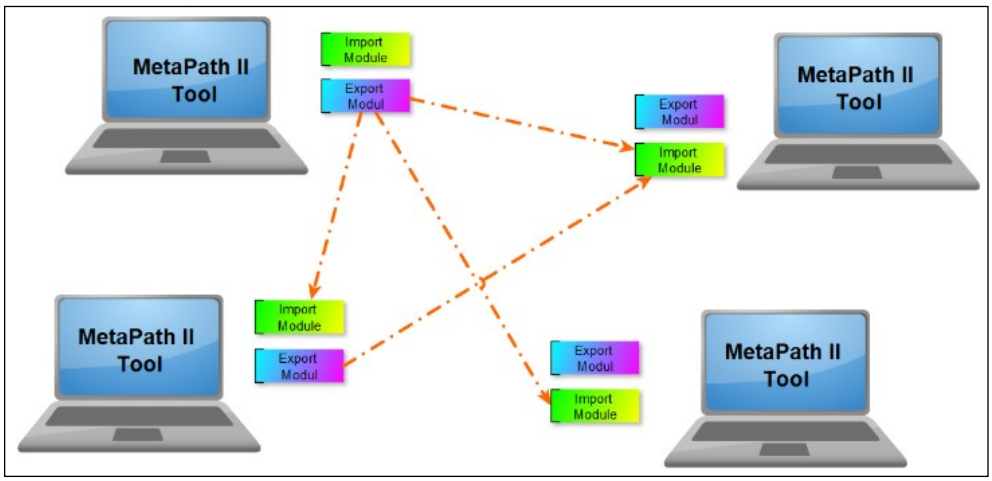


Figure 42: Nondirected independent data exchange beween different *MetaPath II Tool*.

R 9.8-4: The information flow according R 9.8-3 should work also without an instance of *IUCLID*.

R 9.8-5: If another generic data consumer tool would be considered, the supplied XML files are available and the required data can be extracted by the generic data consumer tool with an individual data extractor module (see Figure 43).

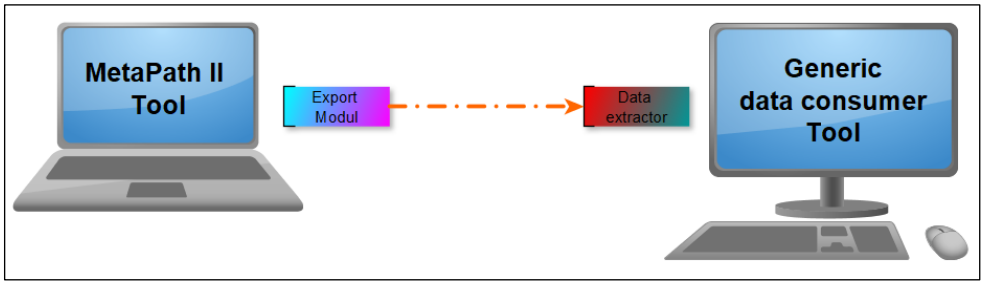


Figure 43: Generic data consumer of *Aggregated raw data* of metabolism studies

An example tool, which is facilitating data of the *Aggregated raw data*, could be *IUCLID*.

R 9.8-6: It would avoid duplicate work, if *IUCLID* could extract needed *Metadata* from the *Aggregated raw data* of metabolism studies e.g. all substance information and the relation between the substances. This action could be activated after attaching the XML of *Aggregated raw data* of metabolism studie in *IUCLID*.

R 9.8-7: If it is enforced by the authority, that *Aggregated raw data* from metabolism studies are asked to be submitted in context of the legal act, an information from the applicant is sent to the authority via an unidirectional transport system. No backward data flow should be considered.

R 9.8-8: If an international curated reference collection of metabolism study *Metadata* is foreseen, an additional transport should take place between authority and reference collection. In case of detected errors, the data flow should be bidirectional.

According to the objectives of the improvement process (Table 2) an information flow should be taken into account from the curated reference collection to the applicants. This case becomes more important as the agencies have already collected a considerable amount of metabolism studies through the *DER/MSS-Composer family* and thus this information will be available in the curated reference collection.

R 9.8-9: The format for downloading the *Metadata* from the curated repository should be the same as used for submitting a new metabolism study in a dossier.
The possibility to upgrade / correct the data in the same format as it was downloaded must be available to the applicants and therefore, they need to be able to feed it back into the processes (see Figure 44).

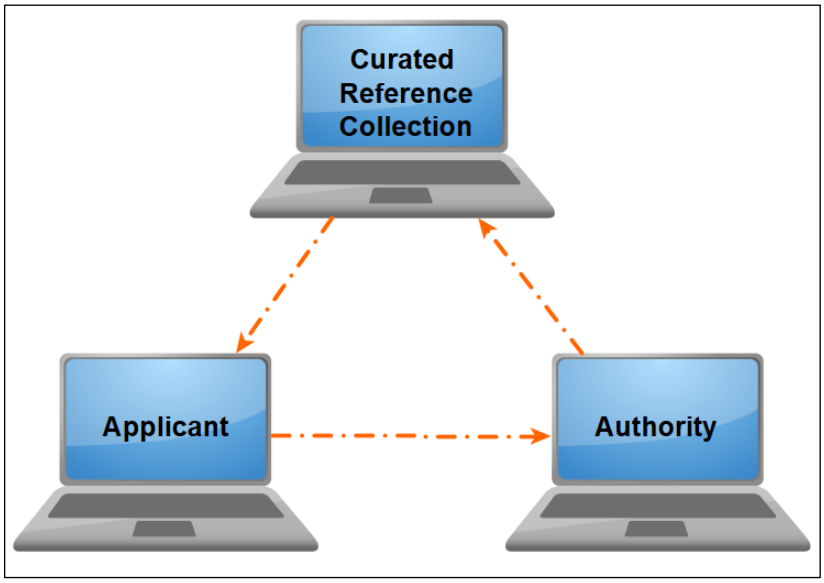


Figure 44: Needed “Information loop” of metabolism studies

9.8.2.3 Improve the curated reference information

R 9.8-10: The generally accepted goal to avoid further tests on vertebrate animals as well as reducing uncertainty in human exposure assessments without lowering the level of protection is the driving force for building up this information loop (see Figure 44).

R 9.8-11: If the information flow is organised in a loop, the used data interface format should always be the same. Therefore, all tool instances in the circle are using the same export and import modules.

R 9.8-12: Results of new metabolism studies will be integrated in this loop the first time initiated by applicants.

R 9.8-13: Authorities could add results of already submitted and assessed metabolism studies.

R 9.8-14: The authorities use the information to write the Draft Assessment Report (DAR) and to create a set of validated and *QA* checked results of metabolism studies (see R 7.3-6).

R 9.8-15: After the decision, the information will be published in a curated reference collection.

R 9.8-16: Applicants can reuse the published information.

Figure 45 was included in the report in such a design, which was used in the 3rd *MUG* meeting discussion.

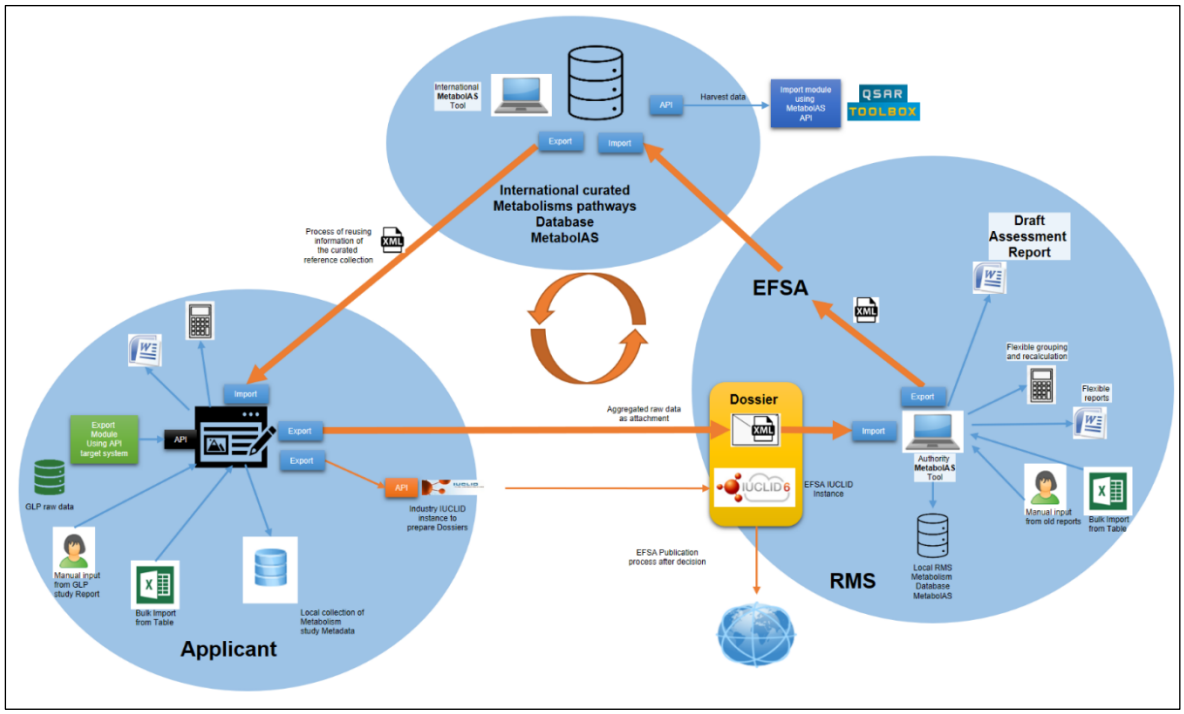


Figure 45: High-level information flow of metabolism studies

9.8.3 Identified concepts to transport aggregated raw data

Based on the illustration in Figure 45, the following two transport concepts were identified in the 3rd *MUG* meeting (see Figure 46).

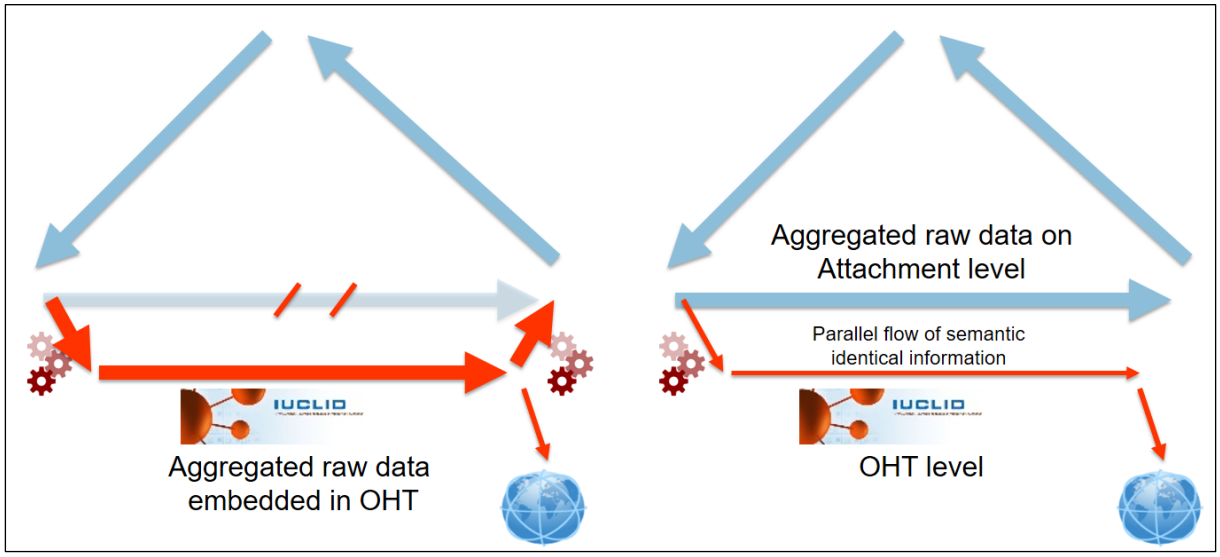


Figure 46: Two proposed transport concepts:
a) *Aggregated raw data* embedded in *OHT*
b) *Aggregated raw data* as attachment

9.8.3.1 Embedded in *OHT*

Because of other framework conditions outside of this project, it could be one option to build the information loop (Figure 45) with two different transport concepts.

To implement this identified transport option, a definition of a new *OECD Domain Type* for "Metabolism raw data" is needed, because many *OECD* Harmonised Templates" should be improved to transport the *Aggregated raw data* of metabolism studies (see Figure 47).

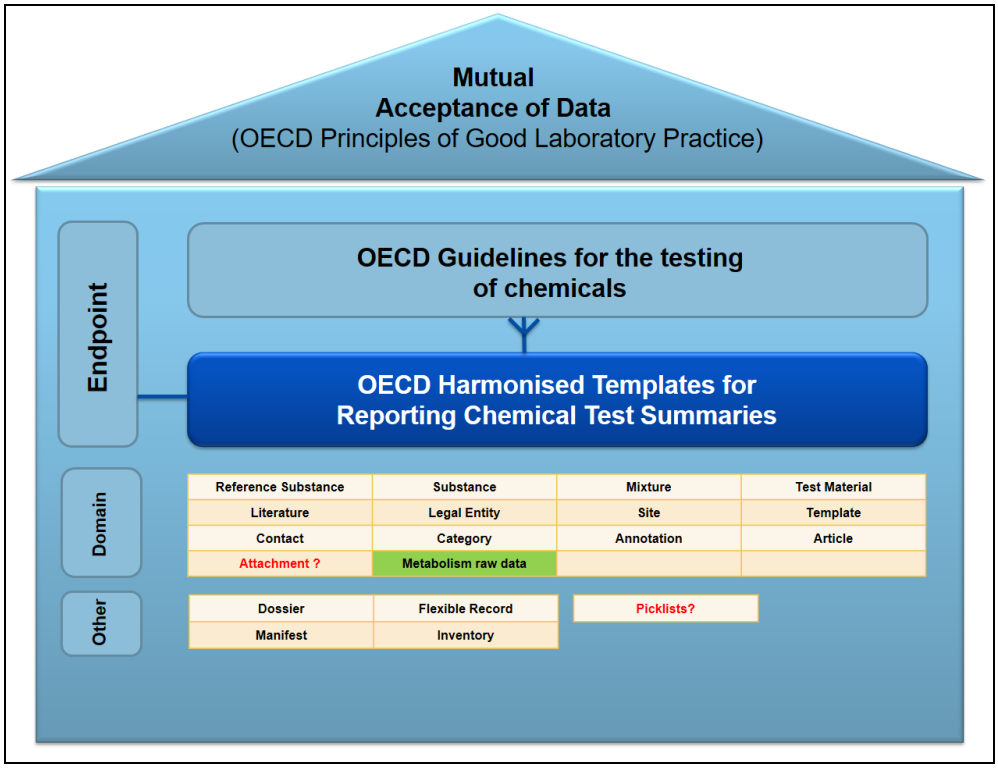


Figure 47: Expanding the *OECD* data architecture with a new domain type "Metabolism raw data"

The main principles for this solution are:

1. The initial transport step should be based on the *OHTs*. The *Aggregated raw data* of metabolism studies are embedded in the *OHTs* with the help of a new *OECD Domain Type* for metabolism studies.
2. This is a generic approach. All user interfaces for metabolism studies could refer to the new *OECD Domain Type* if *Aggregated raw data* would be submitted.
3. The authorities need an importer for the *Aggregated raw data*, embedded in the *OHTs*, to transfer them into the *MetaPath II Tool*.

9.8.3.2 As attachment

This concept is based on the current process, where the XML files, created by the *DER/MSS-Composer family* should be attached to the *IUCLID* dossiers. The *MUG* group has preferred only solutions under the governance of the *OECD*.

Because of this reason, a definition of a new category *OECD* Attachment Type is needed with its first representative called "Metabolism raw data".

Other attachment types for transporting other *Aggregated raw data* would be useful, e.g. for data on genetic toxicity (see Figure 48).

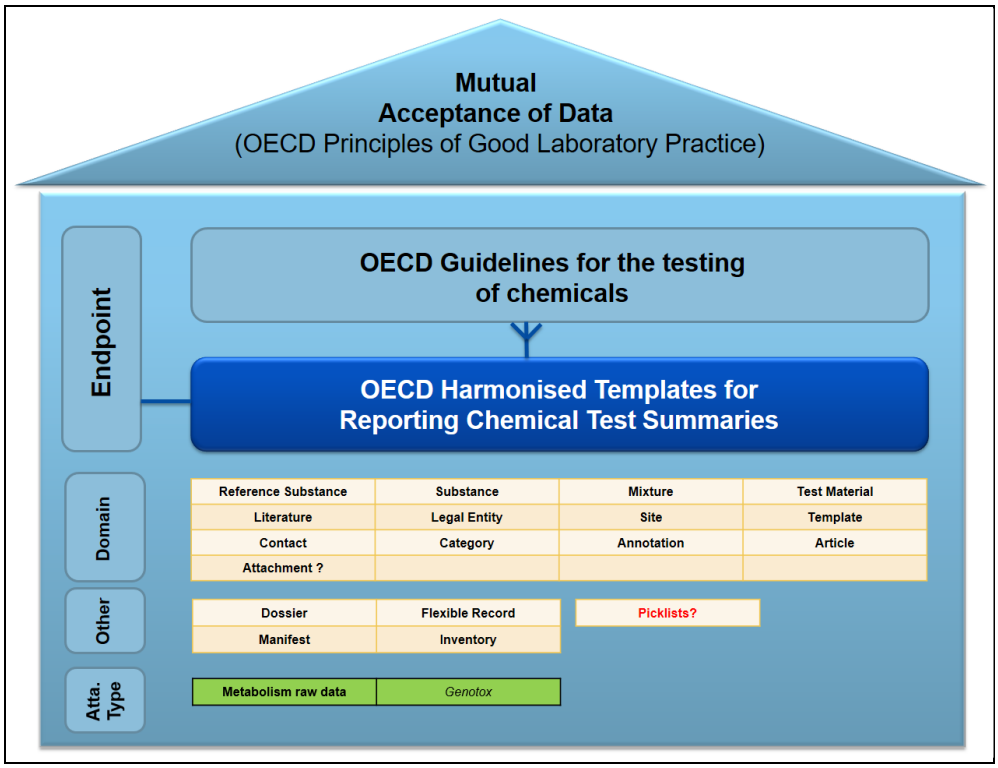


Figure 48: Expanding the *OECD* data architecture with the new category "OECD Attachment Type"

The main principles for this solution are:

1. All transport step should assit the dataflow of *Aggregated raw data* of metabolism studies with the help of a new *OECD Attachment Type* for metabolism studies.
2. No *OHT* user interface for metabolism studies should be modified, because no *Aggregated raw data* are embedded into the *OHTs*.

9.8.3.3 Semi-quantitative comparsion of the efforts of the identified transport concepts

The following semi quantitative comparison contain only efforts to implement these identified concepts.

The following efforts will not be included:

- to define / redefine the needed standards
- efforts for components which are present in both cases.

The semi-quantitative comparison of the efforts of the identified transport concepts (see Table 14) have shown, that the transport concept of a new *OECD Attachment Type* for metabolism studies is the prefered solution.

Table 14: Semi-quantitative comparison of the identified efforts by implementing the different transport concepts

Affected components		**In Case of embedded in "OHT"**		**In Case of attachment**	
Group	Component	Consequence	Effort*	Consequence	Effort*
Applicants system	Local IUCLID Instance	Manage a local instance of IUCLID.	2	No specific requirement of metabolism studies. May be that such local instance is needed because of other additional reasons.	0
Applicants system	Local MetaPath II Instance	Is needed at least for visualising the metabolism pathways.	2	Is needed at least for visualising the metabolism pathways.	2
Applicants system	Reuse of data from the reference collection	Not possible. Additional converter from the data structure of the reference collection into the assisted OHTs are needed. Only 4 OHTs have been included in the comparison. In total,18 converters would be needed.	4* 2	MetaPath II supports the reuse of data derived from the reference collection.	0
Applicants system	Manual data entry in the user front end	Stakeholder will avoid the effort and will implement an alternative solution.	5	MetaPath II is optimised for raw data input – but the effort to do manual data entry is high.	4
Applicants system	Report of a metabolism study. The quality of the needed reports should be comparable with the M3 Documents.	In total: 18 different reports for 18 OHTs should be supported. Only 4 Reports have been included in the comparison.	4 * 3	One report generator should support all study types. Therefore, it will be a very difficult report template. Calculated with factor 2.	2 *4
Applicants system	Include LIMS data	High effort to include LIMS data directly.	4	High effort to include LIMS data directly.	4
Authority	Import module OHT into MetaPath II	In total: 18 different import module for 18 OHTs should be supported. Only 4 importer have been imported in the comparison. All of these importers have the same core import modul.	4 * 3	No importer would be needed.	0
Overall sum			46		18

*

Interpretation	**Effort class**	**Example**
Not acceptable high	5	Stakeholder will avoid the effort and will implement an alternative
High	4	Complex data entry program / report / import / export is needed
Medium a	3	Data entry program / report / import / export is needed
Medium b	2	Converter program is needed / Manage a local instance
Low	1	Configuration effort
No	0	Function not needed or already included

10 Discussions in the MetaPath User Group

Because the *BfR* was interested to get feedback from stakeholders who are involved in the flow of information of metabolism studies, a non-formal commenting round based on a draft report version was organised in September 2021. This draft report was forwarded to the following groups:

- Pesticide Steering Network (PSN) *IUCLID* group
- EU Member States (MRL review) group
- EU Member States (peer review) group
- MetaPath user group (*MUG*)
- Pesticide companies
- Independent laboratories
- *EFSA*, *ECHA*, *OECD*

BfR and the *USEPA* had organised three *MUG* Meetings in November / December 2021 to discuss open questions

- regarding user requirements,
- to compare technical solutions and
- to highlight organisational questions.

BfR will publish the results of the *MUG* Meetings individually in 2022.

10.1 Agreement to manage the transition period

One of the most important results of the *MUG* Meetings in November / December 2021 is, that the MetaPath User Group (*MUG*) should be the forum where decisions on the priority of change requests should be discussed and decided.

R 10.1-1: *EFSA* should build up a repository with all change requests to the current tool set of *MetaPath* and the *DER/MSS-Composer family*. This repository should be open and transparent for all *MUG* participants for a prioritisation and discussion.

P 10.1-2: It is proposed to use management tools for R 10.1-1, which are already implemented in Open Source Projects.

D 10.1-3: *LMC* should classify the change requests according the effort into a) low effort b) mid effort c) high effort d) not achievable in the intermediate time.

D 10.1-4: *LMC* should give a statement to which extent the necessary resources can be provided in the required timeframe.

D 10.1-5: All stakeholders should decide, who would contract which change request in which period.

11 Appendix

11.1 Bibliography

Short	Meaning
BfR 2020	MetaPath - Incorporation of Pesticide Residue Data (MetaPath) https://www.bfr.bund.de/en/metapath___incorporation_of_pesticide_residue_data__metapath_-252118.html
BfR 2021	BfR: Analysis of the information flow in metabolism studies on pesticides. Berlin, Germany: BfR; 2021 https://www.bfr.bund.de/en/analysis_of_the_information_flow_in_metabolism_studies_on_pesticides-272198.html
BIG 2021	Bundesimmobiliengesellschaft mbH (BIG), Austria: Developing ICT solutions for smart and efficient building management, ICLEI European Secretariat, January 2021 https://procure2innovate.eu/fileadmin/user_upload/Case_studies/2020Procura_InnovationICTA-ward_BIG.pdf
Cytoscape	Cytoscape Consortium: Cytoscape Project. U.S. National Institute of General Medical Sciences of the National Institutes of Health https://cytoscape.org/
Data Modeler	Oracle: Oracle SQL Developer Data Modeler https://www.oracle.com/database/technologies/appdev/datamodeler.html
DB-ENGINES	DB-Engines: Knowledge Base of Relational and NoSQL Database Management Systems, solid IT gmbh https://db-engines.com/en/system/Firebird%3BMySQL%3BPostgreSQL
ECHA 2016	ECHA: Practical guide How to use and report (Q)SARs. ABC. 2016;3.1:37. https://doi.org/10.2823/81818
EFSA 2016	EFSA. Guidance on the establishment of the residue definition for dietary risk assessment. EFSA Journal. 2016;14(12). https://doi.org/10.2903/j.efsa.2016.4549
EFSA 2019	EFSA: Administrative guidance on submission of dossiers and assessment reports for the peer-review of pesticide active substances in: EFSA Supporting publications 2019;16(4). https://doi.org/10.2903/sp.efsa.2019.EN-1612
EFSA 2021	EFSA: Process Steps for Metabolism Data. Reporting structured results of metabolism studies on rats, plants and livestock. Description of process steps of the information flow in EU Zenodo: 2021 2021-05-25. http://doi.org/10.5281/zenodo.4785179
EPEC 2015	EPEC, European Investment Bank. The European Public Private Partnerships (PPP) Expertise Centre at a glance. 2015. https://www.eib.org/attachments/epec_flyer_en.pdf
EU 1107/2009	EC: REGULATION (EC) No 1107/2009 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. Official Journal of the European Union. 2009; L 309/1:50. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009R1107&from=DE
EU 2013/C 95/01	EC: NOTICES FROM EUROPEAN UNION INSTITUTIONS, BODIES, OFFICES AND AGENCIES. Commission Communication in the framework of the implementation of Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. Official Journal of the European Union. 2013;C 95/1. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2013:095:0001:0020:EN:PDF
EU 283/2013	EC: COMMISSION REGULATION (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market. Official Journal of the European Union 2013;L93/1.

Short	Meaning
	https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013R0283
EU 396/2005	EC: REGULATION (EC) No 396/2005 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on maximum residue levels of pesticides in or on food and feed of plant and animal origin and amending Council Directive 91/414/EEC. 2005 2640. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02005R0396-20160513
EU DAR/CLH	EC: Current Combined Templates for Draft (Renewal) Assessment Reports and CLH Reports 2019 https://ec.europa.eu/food/system/files/2019-04/pesticides_ppp_app-proc_guide_doss_12592-2012.zip
EU Sanco/221/2000	EC: GUIDANCE DOCUMENT ON THE ASSESSMENT OF THE RELEVANCE OFMETABOLITES IN GROUNDWATER OF SUBSTANCES REGULATED UNDER REGULATION (EC) No 1107/2009 Official Journal of the European Union 2003. Sanco/221/2000 – rev.11 21. October 2021 https://ec.europa.eu/food/system/files/2021-10/pesticides_ppp_app-proc_guide_fate_metabolites-groundwtr-rev11.pdf
FAO/WHO 2017	FAO, WHO: REPORT OF THE 49th SESSION OF THE CODEX COMMITTEE ON PESTICIDE RESIDUES. JOINT FAO/WHO FOOD STANDARDS PROGRAMMECODEX ALIMENTARIUS COMMISSION 2017 24 - 29 April 2017; Beijing, P.R. China. REP17/PR https://www.fao.org/fao-who-codexalimentarius/sh-proxy/jp/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252FMeetings%252FCX-718-49%252FREPORT%252FREP17_PRe.pdf
FISK 2021	Peter Fisk Associates Limited: Critical literature review of analytical methods applicable to environmental fate studies; Reference: PFA.750.011.001; 02/2021 https://echa.europa.eu/documents/10162/17228/pfab_750_06_wp4_echa_final_report_en.pdf
Goodman 2021	Goodman JM, Pletnev I, Thiessen P, Bolton E, Heller SR. InChI version 1.06: now more than 99.99% reliable. Journal of Cheminformatics. 2021;13:40. https://doi.org/10.1186/s13321-021-00517-z
ISO 9001	DIN: Qualitätsmanagementsysteme - Anforderungen (ISO 9001:2015); Deutsche und Englische Fassung EN ISO 9001:2015-11
Kolanczyk 2012	Kolanczyk RC, Schmieder P, Jones WJ, Mekenyan OG, Chapkanov A, Temelkov S, et al. MetaPath: An electronic knowledge base for collating, exchanging and analyzing case studies of xenobiotic metabolism. Regulatory Toxicology and Pharmacology. 2012;63(1):84-96. doi: 10.1016/j.yrtph.2012.02.013
LMC Metapath	OASIS Laboratory of Mathematical Chemistry of the University "Prof. Assen Zlatarov": Software METAPATH, Bourgas, Bulgaria. http://oasis-lmc.org/products/software/metapath.aspx
Manibusan 2011	Manibusan, Mary: Using Searchable Databases and Predictive Systems: MetaPath. US EPA Archive Document 2011. https://archive.epa.gov/pesticides/ppdc/web/pdf/screening-casestudy.pdf
OECD 1998	OECD: OECD Principles on Good Laboratory Practice (as revised in 1997) in: Series on Principles of Good Laboratory Practice and Compliance Monitoring No. 1, ENV/MC/CHEM(98)17 https://one.oecd.org/document/ENV/MC/CHEM(98)17/en/pdf
OECD 2004	OECD: FINAL Report of the Expert Group Meeting to Explore Harmonising Templates 14-16 June, 2004 Paris 7 October, ENV/JM/RD(2004)9; not published
OECD 2007	OECD: GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE)STRUCTURE-ACTIVITY RELATIONSHIPS [(Q)SAR] MODELS. OLIS. 2007;2007(2). ENV/JM/MONO(2007)2. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2
OECD 2009	OECD: GUIDANCE DOCUMENT ON THE DEFINITION OF RESIDUE (AS REVISED IN 2009) JT03268292; OECD Environment, Health and Safety Publications; Series on Testing and Assessment No. 63 and Series on Pesticides No. 31 ENV/JM/MONO(2009)30 http://www.oecd.org/officialdocuments/displaydocument/?cote=ENV/JM/MONO(2009)30& doclanguage=en
OECD 2014	OECD: GUIDANCE ON GROUPING OF CHEMICALS, SECOND EDITION. Series on Testing & Assessment No. 194. ENV/JM/MONO(2014)4, JT03356214. OLIS 2014;2014(4):141. https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2014)4&

Short	Meaning
	doclanguage=en
OECD 2021	Website of OECD “OECD Harmonised Templates - Introduction” https://www.oecd.org/ehs/templates/introduction.htm
OECD Test Guidelines	https://www.oecd.org/chemicalsafety/testing/oecdguidelinesforthetestingofchemicals.htm
OECD Harmonised Templates	https://www.oecd.org/ehs/templates/
RUEDIS 2006	Presentation “RUEDIS –German experience with electronic residue trial data” FAO/OECD workshop “Electronic field trial data for pesticides” Rom 24.02.2006; not published
TECHTARGET	TechTarget: Application Architecture Definitions https://searchapparchitecture.techtarget.com/definition/interoperability
TOXIT	Website of “S-IN Soluzioni Informatiche S.r.l.”: TOXIT - Read-across. https://www.toxit.it/en/services/read-across

11.2 Attachments on file level

Filename	Content
List_of_statement.xlsx	List of statements of the current report
MetaPath_II.xsd	Draft schema definition file of MetaPath II
MetaPath_II_xsd.png	Figure of the draft schema definition figure of MetaPath II
MetaPath_II_DB.zip	Draft database model ((Oracle SQL Developer Data Modeler)
MetaPath_II_DB.svg	Figure of the draft database model
MetaPath_Modules.svg	Figure of the current MetaPath modules
MetaPath_ER_Schema.svg	Figure of the current MetaPath database model

11.3 List of Harmonised Templates where radioactive labelled test material could be used

OHT Group	OHT No	ENDPOINT_STUDY _RECORD name	R 1)	L-TP 2)	RD 1)	Test Guide-line	Test Guideline Name	Definition terms regard-ing metabolism	Labelled test sub-stances
Environmental fate & behaviour	OHT24	PhototransformationInAir	-	+	D	none	-	-	-
Environmental fate & behaviour	OHT25	Hydrolysis	-	+	-	TG111	Hydrolysis as a Function of pH	Transformation products; Hydrolysis products	Yes
Environmental fate & behaviour	OHT26	Phototransformation	-	+	D	TG316	Phototransformation of Chemicals in Water – Direct Photolysis	Transformation (biodegradation, mineralization) and parameter of the transformation process	Yes
Environmental fate & behaviour	OHT27	PhotoTransformationInSoil	-	+	D	none	-	-	-
Environmental fate & behaviour	OHT28	BiodegradationInWaterScreen-ingTests	-	-	D	TG301	Ready Biodegradability	No separate definition block but bi-odegradation is explicitly named	Yes
						TG302A	Inherent Biodegradability: Modified SCAS Test	No separate definition block but bi-odegradation is explicitly named	Yes
						TG302B	Inherent Biodegradability: Zahn-Wel-lens/ EVPA Test	No separate definition block but bi-odegradation is explicitly named	No
						TG302C,	Inherent Biodegradability: Modified MITI Test (II)	No separate definition block but bi-odegradation is explicitly named	No
						TG306	Biodegradability in Seawater	No separate definition block but bi-odegradation is explicitly named	Yes
						TG310	Ready Biodegradability - CO2 in sealed vessels (Headspace Test)	Transformation (biodegradation, mineralization) and parameter of the transformation process	No
						TG311	Anaerobic Biodegradability of Organic Compounds in Digested Sludge: by Measurement of Gas Production	No separate definition block but bi-odegradation is explicitly named	No
Environmental fate & behaviour	OHT29	BiodegradationInWaterAndSed-imentSimulationTests	-	+	D	TG303A TG303B	Simulation Test - Aerobic Sewage Treatment -- A: Activated Sludge Units; B: Biofilms	No separate definition block but bi-odegradation is explicitly named	Yes
						TG308	Aerobic and Anaerobic Transformation	Transformation products and pa-	Yes

OHT Group	OHT No	ENDPOINT_STUDY _RECORD name	R 1)	L-TP 2)	RD 1)	Test Guide-line	Test Guideline Name	Definition terms regard-ing metabolism	Labelled test sub-stances
							in Aquatic Sediment Systems	rameter of the transformation pro-cess	
						TG309	Aerobic Mineralisation in Surface Water – Simulation Biodegradation Test	Transformation (biodegradation, mineralization) and parameter of the transformation process	Yes
						TG314A TG314B TG314C TG314D TG314E	Simulation Tests to Assess the Biodeg-radability of Chemicals Discharged in Wastewater	No separate definition block but degradation products are explicitly named	Yes
Environmental fate & behaviour	OHT30	BiodegradationInSoil	-	+	D	TG304A	Inherent Biodegradability in Soil	No separate definition block but degradation products are explicitly named	Yes
						TG307	Aerobic and Anaerobic Transformation in Soil	Transformation products and pa-rameter of the transformation pro-cess	Yes
Environmental fate & behaviour	OHT32	BioaccumulationAquaticSedi-ment	-	-	-	TG305,	Bioaccumulation in Fish: Aqueous and Dietary Exposure	Bioaccumulation, Bioconcentra-tion, Biomagnification	Yes
						TG315	Bioaccumulation in Sediment-dwelling Benthic Oligochaetes	Bioaccumulation, Bioconcentra-tion, Biomagnification	Yes
Environmental fate & behaviour	OHT33	BioaccumulationTerrestrial	-	-	-	TG317	Bioaccumulation in Terrestrial Oligo-chaetes	Bioaccumulation, Bioconcentra-tion, Biomagnification	Yes
Environmental fate & behaviour	OHT34	AdsorptionDesorption	-	+	-	TG106,	Adsorption - Desorption Using a Batch Equilibrium Method	Looks primarily at physical phe-nomena, but points to a possible transformation	Yes
						TG121	Estimation of the Adsorption Coefficient (Koc) on Soil and on Sewage Sludge using High Performance Liquid Chro-matography (HPLC)	Looks only at physical phenomena	Yes
Effects on biotic systems	OHT56	BiotransformationAndKinetics	-	+	-	none	-	-	-
Health effects	OHT58	BasicToxicokinetics	-	+	-	TG417	Toxicokinetics	Biotransformation, Metabolism	Yes

OHT Group	OHT No	ENDPOINT_STUDY _RECORD name	R 1)	L-TP 2)	RD 1)	Test Guide-line	Test Guideline Name	Definition terms regard-ing metabolism	Labelled test sub-stances
						TG319A, TG319B	Determination of in vitro intrinsic clear-ance …	No separate definition block but bi-otransformation and bioaccumula-tion is explicitly named	No
Health effects	OHT59	DermalAbsorption	-	-	-	TG427	Skin Absorption: In Vivo Method	No separate definition block but metabolism is explicitly named	Yes
						TG428	Skin Absorption: In Vitro Method	No separate definition block but metabolism is explicitly named	Yes
Pesticide residue chemistry	OHT85-1	MigrationOfResidues	-	+	-	none	-	-	-
Pesticide residue chemistry	OHT85-2	MetabolismInLivestock	+	+	+	TG503	Metabolism in Livestock	No separate definition block but metabolism is explicitly named	Yes
Pesticide residue chemistry	OHT85-3	MetabolismInCrops	+	+	+	TG501	Metabolism in Crops	No separate definition block but metabolism is explicitly named	Yes
						TG502	Metabolism in Rotational Crops	No separate definition block but metabolism is explicitly named	Yes
Pesticide residue chemistry	OHT85-8	NatureResiduesInProcessed-Commod	-	+	+	TG507	Nature of the Pesticide Residues in Pro-cessed Commodities - High Tempera-ture Hydrolysis	No separate definition block but metabolism is explicitly named	Yes
Pesticide residue chemistry	OHT85-10	StabilityOfResiduesInStored-Commod	-	(+)	+	TG506	Stability of Pesticide Residues in Stored Commodities	No separate definition block but metabolism is explicitly named	Yes

1) R: Different radiolabelled test substances are foreseen (+) or are not foreseen (-)

2) L-TP: List of transformation products could be reported (+) or could not be reported (-)

3) RD: Raw data could be reported (+) or could not be reported (-) or only results of degradation (D)

11.4 List of weak points identified in the survey

The internal numbers shown here were created in the report “Analysis of the information flow of pesticide related metabolism studies – Part Results of the international survey” (*BfR 2021*). Each internal number of the weak point is starting with the letter “**S**(urvey)”.

Internal number	Weak point	Improvement through a better
S 3.3.1-1	The identified knowledge gap that laboratories and applicants use appropriate but rather unknown IT-Tools to the authorities is an indication of a lack of exchanges of tools and practices between the different actors in this knowledge area.	Communication
S 3.3.2-1	A harmonised definition of the term Metabolism study is needed.	Concept
S 3.3.2-2	Dissatisfaction with current tools for storing, handling and disseminating metabolic data is an indication that improvements are needed.	IT-Tool
S 3.3.2-3	It seems that the current information flow is accompanied by a high level of duplication of effort.	Process Organisation, Harmonisation, IT-Tool
S 3.3.2-4	The stakeholders have a different understanding of the term “raw data”.	Concept
S 3.3.3-1	EFSA's March 2021 changes to the submission formats for metabolism studies do not appear to have been prepared with all stakeholders to the necessary extent.	Communication
S 3.3.3-2	The use of MSS Composer is necessary in the new information flow. Inadequate knowledge of how to use this IT-Tool poses a high risk for the implementation of this intermediary information flow.	Communication, Process Organisation
S 3.3.3-3	The current governance model of the MSS Composer could be a risk for the implementation of the MSS Composer in the European workflow.	IT-Tool
S 3.3.3-4	The MSS composers do not yet fully support the format of the Volume 3 of DAR/RAR.	IT-Tool
S 3.3.4-1	The use of Metapath is necessary in the new information flow. Inadequate knowledge of how to use this IT-Tool poses a high risk for the implementation of this intermediary information flow.	Communication, Process Organisation
S 3.3.4-2	There exists a need of more interoperability of Metapath with other IT-Tools.	IT-Tool
S 3.3.4-3	The current governance model of Metapath could be a risk for the implementation of the Metapath in the European workflow.	IT-Tool
S 3.3.5-1	Both IUCLID and Metapath (compare with 3.3.3.7) do not currently yet support the necessary reporting formats.	IT-Tool
S 3.3.6-1	The rejection of the statement “The pesticide-related (Q)SAR models are of sufficient quality for predicting metabolism pathways” suggests that the OECD (Q)SAR-Toolbox has weaknesses in this area.	IT-Tool Communication
S 3.4.1-1	There are elementary difficulties in encoding of structures (generic structures; stereochemistry). As long as these difficulties exist, IT-Tools for storing results from metabolism studies, searching for structure-like and predicting metabolic pathways will be imperfect.	IT-Tool
S 3.4.1-2	As long as there are elementary difficulties in encoding of structures, the IT-Tools provided will also only be of limited use.	IT-Tool
S 3.4.2-1	There seems to be a discrepancy between the wealth of information required for a risk assessment of metabolites and the suitability of the IT-Tools provided.	IT-Tool

Internal number	Weak point	Improvement through a better
S 3.4.2-2	An insufficient degree of harmonisation in the templates to be completed, the variety of IT-Tools to be used and the lack of data interfaces are the cause of duplication.	Process Organisation, Harmonisation, IT-Tool
S 3.4.2-3	The orientation towards EU specific requirements / formats complicates the efforts for a global harmonisation.	Harmonisation
S 3.4.3-1	Due to the modern analytical methods, the data basis to be provided for metabolism studies is growing to a level that risk assessors cannot cope without IT-support. Technical limitations of the IT-Tools, difficulties in data exchange between systems and in the visualisation of the results can lead to an excessive demand on the risk assessors.	IT-Tool
S 3.4.3-2	The QSAR tools currently available on the basis of the existing models and the existing database can only be used to a limited extent in the field of metabolic pathway prediction.	Concept IT-Tool
S 3.4.3-3	The OECD QSAR-Toolbox is limited in the prediction of the kinetics in different “objects of investigation” (species, crops, and environment) of a certain metabolite at different time points.	Concept IT-Tool

11.5 Comparison of possible solutions by the provided functions

		A	B	C
Function Group	**User function, independent of the transport step**	**Current MetPath Environment**	**Improved MetPath Environment**	**New MetaPath II Tool**
Generic approach	Cover all metabolism study types at once	0	0	1
Transparency	Fullfill the requirements	1	1	1
Transparency	Rights of the data donors of a public collection are clear	0	0	1
Chemical structure notation	Downward compatible	0	?	1
Chemical structure notation	Support different notations	0	?	1
Chemical structure notation	Modul to draw structure satisfies the users needs	0	?	1
Chemical structure notation	Markush/generic structures	0	?	?
Chemical structure notation	Search for structure similarities	1	1	1
Chemical structure notation	Search for similarities independent of the used chemical structure notation	?	?	1
Chemical structure notation	The similarity search filter could be combined with additional filter clauses	0	0	1
Substance model	Manage reference substances	0	?	1

		A	B	C
Function Group	**User function, independent of the transport step**	**Current MetPath Environment**	**Improved MetPath Environment**	**New Meta-Path II Tool**
Substance model	Assists references from substances to metabolites	1	1	1
Substance model	Option to merge substance duplicates and their references	?	?	1
Assesment process	Usable for "non-guideline experiments "	0	0	1
Assesment process	Manage "tentative results"	0	0	1
Assesment process	Usable for “freestyle” studies	0	0	?
Assesment process	Manage of textual summaries of the interpretation of the results	1	1	1
Assesment process	Flexible reporting by flexible groups (Pivot tables)	0	0	1
Assesment process	Limitation of 7 columns per table was removed	0	?	1
Assesment process	Recalculations of values from one to another substance	0	0	1
Assesment process	Calculation of concentration factors in relation to other matrix	0	0	1
Assesment process	Grouping of metabolites according the OECD Guideline	0	0	1
Assesment process	Manage Q(SAR) responses in a user storable “List of similar substances”	0	0	1
Assesment process	Integrated start into Q(SAR) Tools with SMILES as the parameter	0	0	O
Assesment process	Manage response from the Q(SAR) tools according ECHA guide	0	0	O
Assesment process	Integrated start into predefined external substance databases	0	0	O
Assesment process	Prediction of metabolic pathways	0	?	?
Assesment process	Pooling of identical substances of different names across the studies	?	?	1
Assesment process	Mange substance groups by defined characteristics (e.g. according to functional groups, conjugates, …)	0	0	O
Curated collection	A curated international collection of quality assured metabolism studies data could be created for a local usage	1	1	1
Curated collection	Assists a central curated international collection of quality assured metabolism studies data	0	0	1
Data organisation	Customisable data interface is needed to import CSV or spreadsheets	0	?	1
Data organisation	Rich text fields are "unlimited" in the text length	0	?	1
Data organisation	Assist a "List of analysed Values”	0	0	1
Data organisation	Maintain picklists	0	0	1
Data organisation	Maintain references between picklists (logical references between items of different “Picklists”)	0	0	O
Data organisation	Use of picklists for elements that are frequently used for search queries	0	0	1
Additional metadata	Manage phys. chemical properties	1	1	O

		A	B	C
Function Group	**User function, independent of the transport step**	**Current MetPath Environment**	**Improved MetPath Environment**	**New MetaPath II Tool**
Additional metadata	Manage toxicological properties	1	1	O
Interoperability	Export / import of data is possible	1	1	1
Interoperability	Only one generic XML Schema definition is needed for export / import	0	0	1
Interoperability	Import a list of mol files or multiple Smiles codes at one time	0	0	1
Interoperability	OECD Q(SAR) Toolbox can harvest required quality assured data by API	1	1	1
Interoperability	Other Q(SAR) modeller can harvest required quality assured data by API	0	0	1
Interoperability	Open a specific data set in the user interface via REST “API	0	0	1
Interoperability	API to read / write data	0	0	1
Interoperability	Validation of XML data files for data exchange after export / before of import	0	0	1
Interoperability	Logging of used IT-systems to build the XML data files for data exchange	0	0	1
Interoperability	Partial import of tables for different entities e.g. “List of Dose Groups”, “List of Sample Groups” by using the clipboard.	0	0	O
Visualization	Metabolic pathway	1	1	1
Visualization	Overlay (merge) of different Metabolic pathways	1	1	1
Visualization	Compare Metabolic pathways	1	1	1
Visualization	Concentration time curves	0	0	1
Project framework	Service could be organised in the web cloud	0	0	1
Project framework	The DBMS provides all the necessary system management functions for a secure and effective running of the applications	0	0	1
Project framework	A governance concept exits	0	0	1
Project framework	Testing of beta versions of modules is organised	0	?	1
Project framework	Lifecycle of modules is organised	0	?	1
Project framework	Open program sources	0	0	1
Project framework	Each write transaction will be finished with a save or cancel action	0	0	1
Project framework	Satisfying response time	0	?	1
Project framework	User / Role / Security concept exits	0	?	1
Project framework	User-friendly search module to create complex queries	0	?	1
Project framework	Multiuser environment to work in parallel	?	?	1

		A	B	C
Function Group	**User function, independent of the transport step**	**Current MetPath Environment**	**Improved MetPath Environment**	**New MetaPath II Tool**
Project framework	Export/ import of complex search queries	1	1	O
Project framework	Create complex search queries with logical expressions	0	0	O
Project framework	User-initiated server actions (e.g. reports) should beinterruptable by the user	0	0	1
Project framework	Support common used access methods and programming languages	0	0	1
User interface	Program modules are web programs	0	0	1
User interface	Mandatory fields are marked in the user interface	0	0	1
User interface	Restrictions on field level could be seen in the user interface	0	0	1
User interface	Copy & Paste of data values and system messages are possible	0	?	1
User interface	Tab button move focus through the input fields	0	?	1
User interface	Context sensitive user documentation integrated in the programs	0	?	1
User interface	Search for a text in the IT module used in headings, label or a data value	0	?	1
User interface	The users should be able to store the used search filter options locally and to load a stored request.	1	1	1
User interface	Mange user storable sets of submitted and additional studies	0	0	1
User interface	Mange user storable sets substances	0	0	1
User interface	Result lists could be sorted by all provided columns	?	?	1
Report	Rendering of one metabolism study according a default template	1	1	1
Report	The default report outputs is in line in format and content of the assessment reports	1	1	1
Report	The default report outputs could be used for the IUCLID section: “Applicant’’ summary and conclusion”	1	1	1
Report	Report generator assists a full version and a sanitized version	?	?	1
Report	Rendering of chemical structures in a sufficient quality	0	?	1
Report	Reports on a set of studies for different stakeholders to create an equivalent to EFSA Appendix G	0	?	1

Used symbols:

- 0 Not such function
- ? No information
- **1** Provided function
- **O** Optional function

11.6 The compromise between distinct data fields and free text fields

The following figure was found in an internal *OECD* document (*OECD 2004*):

FIELD TYPES

FIXED TEXT

Field No.	Field Name	Subfield
Animal		
5	Species	Picklist ▼
6	Strain	Picklist ▼
7	Sex	Picklist ▼

FREE TEXT (w/ subfields)

Field No.	Field Name	*Subfield*
. . .		
Animal		
5	Subheading: • Species *Enter Text* • Strain *Enter Text* • Sex *Enter Text*	

FREE TEXT (w/ guidance)

Field No.	Field Name	Subfield
. . .		
Animal		
5	Guidance: (Species, Strain, Sex) *Enter Text*	

High Searchability

Easy to manipulate or further process data

Low Searchability

More difficult to manipulate or further process data

Difficult to migrate old data into

Easy to migrate old data into

11.7 Standard tables for the presentation of metabolism studies

Table 15: <Generic title>.

Column Group name **Row group name**	**Column group**	**Column group**	**…**	**Optional Mean***	**Optional SD***
Row group 1					
Row group 2					
Row group 3					
Sum of all rows*					

* calculated values

Table 16: Characterisation and identification of extractable radioactive residues collected in excretion products of <*Object Group*> when dosed with <*Test substance* >

Dose group & Sample group **Substances**	**DG1 Urine 0-24**	**DG1 Urine 24-28**	**DG2 Urine 0-24**	**DG2 Urine 24-28**	**…**
Parent	%TRR / ppm of substance in matrix				
Metabolite 1					
Metabolite 2					
Metabolite 3					
Metabolite 4					
Unknown 1					
Unknown 2					
Unknown 3					
Sum of unknown TRR					
Sum					

Table 17: Concentrations of radioactive residues in matrices of <*Object Group*> when dosed with <*Test substance* > at <*time*>

Dose group & Sample group **Substances**	**DG1 Muscle**	**DG1 Fat**	**DG1 Liver**	**DG1 Kidney**	**…**
Parent	Concentration of radioactive residues in ppm				
Metabolite 1					
Metabolite 2					
Metabolite 3					
Metabolite 4					

Dose group & Sample group	**DG1 Muscle**	**DG1 Fat**	**DG1 Liver**	**DG1 Kidney**	…
Substances					
Unknown 1					
Unknown 2					
Unknown 3					

Table 18: Concentration factors of radioactive residues between matrices and plasma concentration in <*Object Group*> when dosed with <*Test substance* > at <*time*>

Dose group & Sample group	**DG1 Muscle**	**DG1 Fat**	**DG1 Liver**	**DG1 Kidney**	…
Substances					
Parent	Factor in relation to serum concentration				
Metabolite 1					
Metabolite 2					
Metabolite 3					
Metabolite 4					
Unknown 1					
Unknown 2					
Unknown 3					

Table 19: %TRR in matrices of <*Object Group*> when dosed with <*Test substance* > at <*time*>

Dose group & Sample group	**DG1 Muscle**	**DG1 Fat**	**DG1 Liver**	**DG1 Kidney**	…
Substances					
Parent	%TRR (extrapolated from the sample to the whole compartment)				
Metabolite 1					
Metabolite 2					
Metabolite 3					
Metabolite 4					
Unknown 1					
Unknown 2					
Unknown 3					

Table 20: Radioactive residues in different analysed fractions of different matrices of <*Object Group*> when dosed with <*Test substance*>

Sample group **Analysed Fraction**	**Matrix 1** **%TRR**	**Matrix 1** **ppm**	**Matrix 2** **%TRR**	**Matrix 2** **ppm**	**…**
TRR	%TRR / ppm of Analysed Fraction in matrix				
Fraction 1					
Fraction 2					
Fraction 3					
Fraction …					
Sum					

Table 21: Radioactive residues identified in <*Sample Matrix*> as function of time when dosed with <*Test substance*>.

Substances **Sample Interval**	**Parent** **%TRR**	**Metabolite 1** **%TRR**	**Metabolite 2** **%TRR**	**Metabolite 3** **%TRR**	**…**
Sample Interval 1	Identified %TRR in samples				
Sample Interval 2					
Sample Interval 3					
Sample Interval 4					
Sample Interval …					
Sum					

Table 22: Radioactive residues in excreta as function of time when dosed with <*Test substance*>.

Dose group **& Sample group** **Sample Interval**	**DG1** **Urine** **%TRR**	**DG1** **Faeces** **%TRR**	**DG1** **Milk** **%TRR**	**DG1** **Eggs** **%TRR**	**…**
Sample Interval 1	Identified %TRR in samples				
Sample Interval 2					
Sample Interval 3					
Sample Interval 4					
Sample Interval …					
Sum					

Table 23: Plasma concentration of radioactive residues as a function of time when dosed with <*Test substance*>.

Dose group / **Sample Time**	**DG1 ppm**	**DG2 ppm**	**DG3 ppm**	**DG4 ppm**	**…**
Sample Time 1	ppm				
Sample Time 2					
Sample Time 3					
Sample Time 4					
Sample Time …					