BfR

Analysis of the information flow of pesticide related metabolism studies

Part
==Draft== proposals for improvement

Table of Content

1 Background

1.1 Context of this report

It should be noted 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
- Metabolites in the context of "Metabolomics".

The terms "*Metabolite*" and "*Metabolism study*" as used in this report are defined in Chapters 5.2.1.and 5.2.2.

1.2 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. Consequently, Regulation 1107/2009[1] defines the term "*Metabolite*" for the field of European plant protection (see chapter 5.2.1.3).

Since 2010, the US *EPA* had been advocating a standardized evaluation of metabolism studies with the goal of building a metabolite information system. The OECD MetaPath User Group (*OECD 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 US *EPA* initiated a study[2] 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 organize the results of metabolism studies into a systematic database to facilitate data comparisons and evaluations."[3]

This development work by the US EPA / 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.

That's why *OECD* Guideline 194 (2014)[4] has defined techniques / methods for data gap filling, an "analogue approach" and the "category approach" (see chapter 5.3.2). 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.

---

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009R1107

[2] https://archive.epa.gov/pesticides/ppdc/web/pdf/screening-casestudy.pdf

[3] https://doi.org/10.1016/j.yrtph.2012.02.013

[4] ENV/JM/MONO(2014)4

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 defined, 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* is 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.

1.3 The current EFSA process

The *EFSA* published a document "Reporting structured results of metabolism studies on rats, plants and livestock"[5] with a description of the underlined current European process steps.

The current "big picture" is a generic authority view of the current process with the following parts:

- input and output information according to a legal act,
- tools to process the data,
- information collections needed,
- looped process to improve information collections.

This current "big picture" is also useful for applicants to proactively adapt their strategies in data generation and handling.

---

[5] European Food Safety Authority. (2021, May 25). Process Steps for Metabolism Data. Zenodo. http://doi.org/10.5281/zenodo.4785179
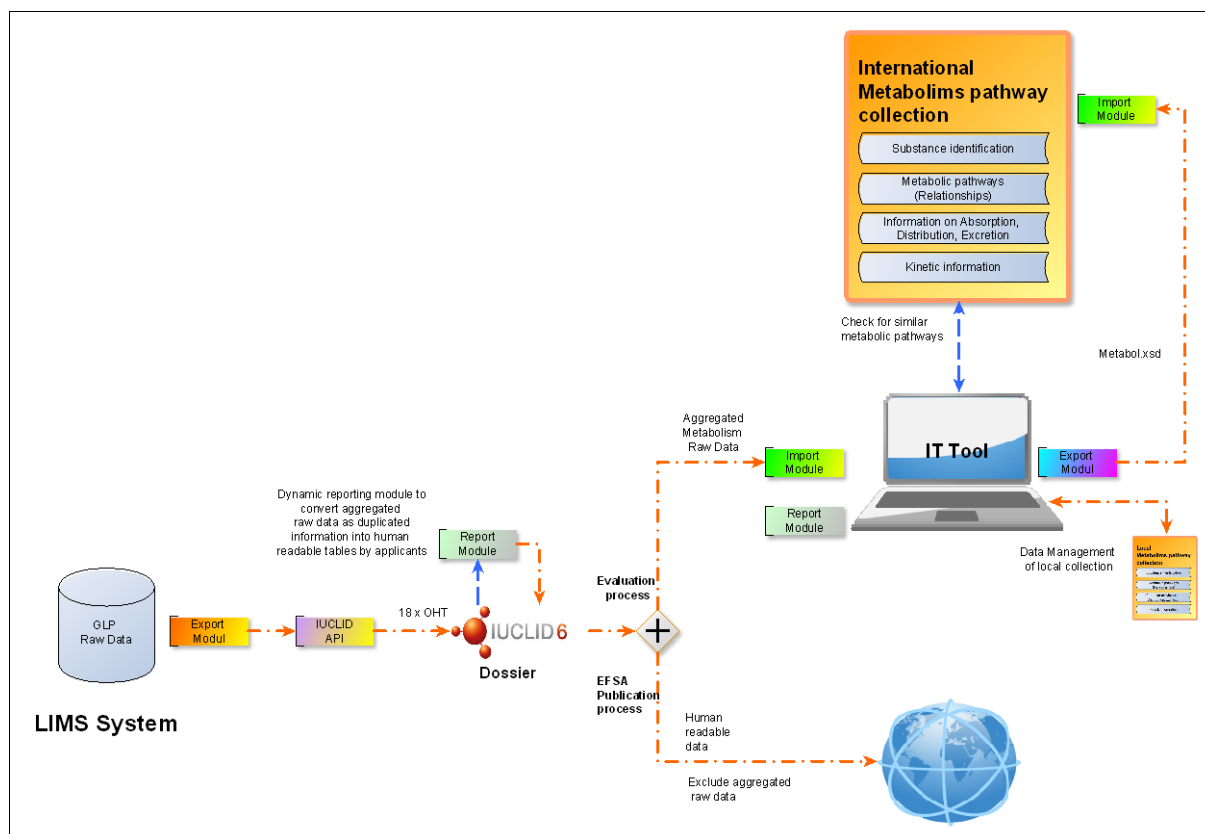
Figure 1: The current "Big picture" of handling metabolism study information in an authority

The following symbols were used:

⬤	Start event	⬤	Termination
📄	Physical data object	▭	Task
⟶	Connection between tasks	·· ▶	Connection between tasks, not analysed in detail
— · ▶	Task sending information to a data storage or makes a request to get information from a data storage	— · ▶	Task creates physical data object or reads physical data object
International Metabolims pathway collection	International collection of results of metabolism studies Local "*IT-Tool*" with a local collection of results of metabolism studies	Set of QSAR Tools with a set of QSAR Models	Different (Q)SAR tools with different models and different training sets
IT Tool	Generic *IT-Tool* to use / manage collections of results of metabolism studies		

The processes shown in Figure 1 assumes that it is essential to have an "*IT-Tool*" that supports the assessment of metabolism studies and the production of assessment reports.

The generic name "*IT-Tool*" used here indicates that the implementation of the required user function(s), which is more important than the concrete solution itself.

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

CR 1.3-1: A set of "*Aggregated Raw Data*" from metabolism studies is stored and managed in a local metabolism pathway collection.

CR 1.3-2: 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 1.3-3: *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 1.3-4: The current "*IT-Tool*" is named "*MetaPath*" and assists the user in process steps starting with the validation of incoming data sets, searching for similar metabolites / pathways.

2   Objectives for further development

The part of the report "Results of the international survey" 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* were summarised in chapter 7.4 of the framework of the Transparency Regulation implementation. Approximately two-thirds of the weaknesses identified showed a need for action to improve "*IT-Tools*".

From this, the *EFSA* formulated content-related objectives for improving the flow of information from metabolite studies and for their use in the assessment processes (Table 1).

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 big need not be considered. These objectives should only be scaled back 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"[6] or the model of an "innovation partnership solution"[7] should therefore be considered.

---

[6] https://www.eib.org/attachments/thematic/epec_flyer_en.pdf

[7] https://ec.europa.eu/growth/content/8699-innovation-partnerships-keep-public-services-date_en

Table 1: 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: OHT 58 BasicToxicokinetics OHT 85-2 MetabolismIn-Livestock OHT 85-3 MetabolismIn-Crops
				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	
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		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 must 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 have 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	
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	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	

3   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[8],
- 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
- To-do: stakeholder consultation

The *BfR* applied a holistic approach in the analysis of the information flow and for the development of the proposals for improvement. All steps of information flows were considered. This starts with the data generator (i.e. laboratory) and entails applicants and authorities that compile the different direct outputs: the assessment reports, the published meta data and the quality assured reference collection of metabolism studies. Furthermore, there are efforts by other data consumers to harvest data of this quality assured reference collection in their systems, e.g. the QSAR-Toolbox (chapter 6.5).
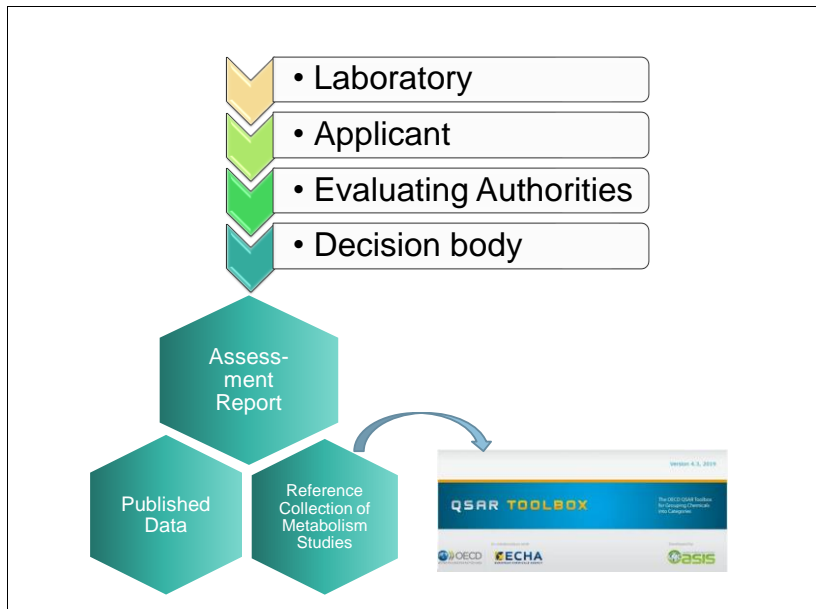
Figure 2:   Direct "end products" of the information flow and the QSAR-Toolbox as an example of a "harvesting" systems

[8] https://www.bfr.bund.de/en/analysis_of_the_information_flow_in_metabolism_studies_on_pesticides-272198.html

3.1 High-level Statements

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

3.1.1 High degree of overlap in stakeholder interests

O 3.1-1: 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 3.1-2: In the long-term, the development and the availability of such an information database for pesticide metabolism pathways will benefit both the applicants and the authorities. For this reason, both stakeholder groups should be equally interested in a generally improved IT support.

O 3.1-3: The high degree of overlap of said interests should be the basis to start a substantial improvement process regarding the flow of information on metabolism studies.

3.1.2 Need for a generalised concept of the term metabolism

O 3.1-4: There are 18 OECD Harmonized Templates where results with radioactive labelled test material could be summarised.

P 3.1-5: The *BfR* proposes a "generic" concept of the term metabolism, which is suitable to build up a generalised scheme to transport aggregated metadata 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.

3.1.3 Need for an ecosystem of needed components

O 3.1-6: The *BfR* proposes 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 the same interoperable functionality (Governance concept, user forum, picklists and picklist elements, scheme definition, IT-Tool, API, reference collection see chapter 6.2).

O 3.1-7: The Governance concept should open for metabolism studies independent of the legal context (pesticides, biocides, chemicals).

O 3.1-8: The Governance concept should contain rules on how to deal with competing interests.

3.1.4 Need for a curated reference collection of metabolism study metadata

O 3.1-9: The *EFSA*'s decision to build up a curated reference collection of metabolism study metadata and to update it after submitting new studies is supported by *BfR*. 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 3.1-10: The generic concept proposed by the BfR is intended to enable the curated reference collection of metabolism study metadata to be opened up for all types of metabolism studies that have not been considered so far (compare P 3.1-5). This approach is more open to the scientific community, increases transparency and could help reduce uncertainties in environmental risk assessment.

O 3.1-11: 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.

O 3.1-12: 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 will be published depending on the confidentiality rules. The "*Aggregated Raw Data*" contain the same semantic information and there is no further need to publish the machine readable (non-human readable) data.

P 3.1-13: 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.

O 3.1-14: IUCLID was designed as a dossier transport system for applicants. IUCLID is not currently suitable to be the database management system for the curated reference collection of metabolism study metadata.

P 3.1-15: 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 (compare chapter 6). 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 "*MetabolAS Tool*".

O 3.1-16: It is proposed that the OECD plays the role of the governance body, but this could also be organized by ambitious stakeholders.

R 3.1-17: 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.

3.1.5   Need for an appropriate transport concept of metabolism study metadata

O 3.1-18: IUCLID could be used in three 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 collection, data presentation, data usage, supporting tools required, the publication process and ultimately in the resulting overall effort.

P 3.1-19: The *BfR* proposes expanding the OECD house architecture with the new category "OECD Attachment Type". The *BfR* prefers this technical transport mechanism (chapter 6.6.2) of the "*Aggregated Raw Data*" from Metabolism studies because of many obvious advantages of this solution.

O 3.1-20: If the "*Aggregated Raw Data*" from metabolism studies are transported as an attachment there is the need for <u>one</u> 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. This is the *BfR´s* preferred solution.

O 3.1-21: If the "*Aggregated Raw Data*" from metabolism studies is transported and integrated in the OHTs there is the need for <u>one</u> generic OECD domain type which should cover all studies where radioactive labelled test material could be used according the Test Guidelines. 18 OHTs should 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 (chapter 6.6.3).

3.1.6 Necessary improvement of data management and data handling

O 3.1-22: The content related concept of "*MetaPath*" is up-to-date and useful for the evaluation steps.

P 3.1-23: It is estimated that $1/7^{th}$ of the start-up effort will be required to maintain this software.
The age of the "*MetaPath*" technology used for the database and the front end, the number of DER/MSS composers to be adapted and the number of open user requests in relation to the user functions that do not require any change are arguments for a radical change. From the *BfR*'s point of view, the time has come to move the valuable concept and information contained in "*MetaPath*" to a new technological level.

O 3.1-24: The approach of P 3.1-23 is a chance to move away from the current strategy of "individual "MSS-Composer" programs for each metabolism "Study type" to a single, harmonised approach.

3.2 Executive Summary

<mark>To-do: Until 12/2021</mark>

4 Document structure and used nomenclature

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.

Chapter 5 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 types of statements are used in this report and are organised as one sequence per chapter number of the 1nd level with a starting letter added with the chapter number of the 2nd level e.g. R 7.1-1.

The following starting letters were put in front of the sequence:

Starting with	Meaning	Objective
CR	**C**urrent high-level user **R**equirement	Part of EFSA's transparency offensive
O	**O**pinion	Assessment of the current state by BfR
R	user **R**equirement	Collected by BfR assisted by stakeholders
Q	open **Q**uestion	To answer by stakeholders
P	**P**roposal	BfR recommendations for further development

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 have to 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.

According to the list of user requirements, different solution approaches are possible. Chapter 6 contains the solution approaches, which are in line with the defined terms in chapter 5. 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.

You're reading through draft version right now, which is for public consultation. This draft version was circulated in September 2021 to get feedback from the stakeholders.

If you want to give us a feedback, please use the predefined commenting table which could be downloaded from the BfR website: https://www.bfr.bund.de/en/analysis_of_the_information_flow_in_metabolism_studies_on_pesticides-272198.html

The response to this public consultation will be summarized and presented in a workshop organized by members of the "*MetaPath*" User Group (*OECD MUG*). If you are not member of the *OECD MUG* and you are interested in this workshop please contact Richard Kolanczyk (Kolanczyk.Rick@epa.gov).

5 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.

5.1 IT related terms

5.1.1 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 have to be changed.

R 5.1-1: 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.

5.1.2 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 5.1-2: 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[9].

R 5.1-3: Systems using Chemical Structure Notation should be downward compatible.

There are the following open questions:

Q 5.1-4: Are the problems of the Markush/generic structures solved by the InChI notation?

Q 5.1-5: Are there any other proposals to solve the problems of the Markush/generic structures?

5.1.3 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*".

5.1.4 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 agreed upon standards. These standards should consider the needs of all data producers and users. Due to standardization, 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.

---

[9] https://jcheminf.biomedcentral.com/articles/10.1186/s13321-021-00517-z

5.1.5   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.

5.2   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.

5.2.1   Substance

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

R 5.2-6:   A "*Test Substance*" could be transformed in the "*Object of Investigation*" by "*Transformation Processes*" into Metabolites.

Note 1: The term "*Substance*" is used differently in the IUCLID concept. The equivalent in the IUCLID concept would be the term "Reference Substance".

Note 2: Currently, however, different definitions with different objectives are used internationally for "*Metabolite*", which are not consistent. To give an overview, on the following, 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 5.2-6).

5.2.1.1   FAO

The guideline Codex Alimentarius, FAO, 2017 (REP17/PR, Appendix XIII Definition Annex)[10] 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.)"

---

[10] http://www.fao.org/fao-who-codexalimentarius/sh-proxy/en/?lnk=1&url=https%253A%252F%252Fworkspace.fao.org%252Fsites%252Fcodex%252FMeetings%252FCX-718-49%252FREPORT%252FREP17_PRe.pdf

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

5.2.1.2 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*" (compare chapter 7.3, column "Test Guideline") like: Bioaccumulation, Bioconcentration, Biodegradation, Biomagnification, Hydrolysis, Metabolism, Mineralization, Transformation.

5.2.1.3 EU COM

The Regulation (EC) 396/2005 (Article 3 2c) [11] 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 Regulation (EC) 1107/2009 (Article 3, No. 32) [12] 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 <u>relevant</u> for the overall approval decision or for the definition of risk mitigation measures."

The document Sanco/221/2000-rev10 (25.02.2003, Chapter 3. Definitions)[13] 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,

---

[11] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02005R0396-20160513

[12] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009R1107

[13] https://ec.europa.eu/food/system/files/2016-10/pesticides_ppp_app-proc_guide_fate_metabolites-groundwtr.pdf

other taxa) or abiotic processes (hydrolysis, photolysis). The terms 'metabolite', 'breakdown product' and 'degradation product' are used interchangeably throughout this document."

5.2.1.4 EFSA

The *EFSA* used a slightly more restricted definition of metabolism with the aim of establishing a residue definition (EFSA Guidance on the establishment of the residue definition for dietary risk assessment, Chapter 1. Introduction)[14]:

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.

5.2.2 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.
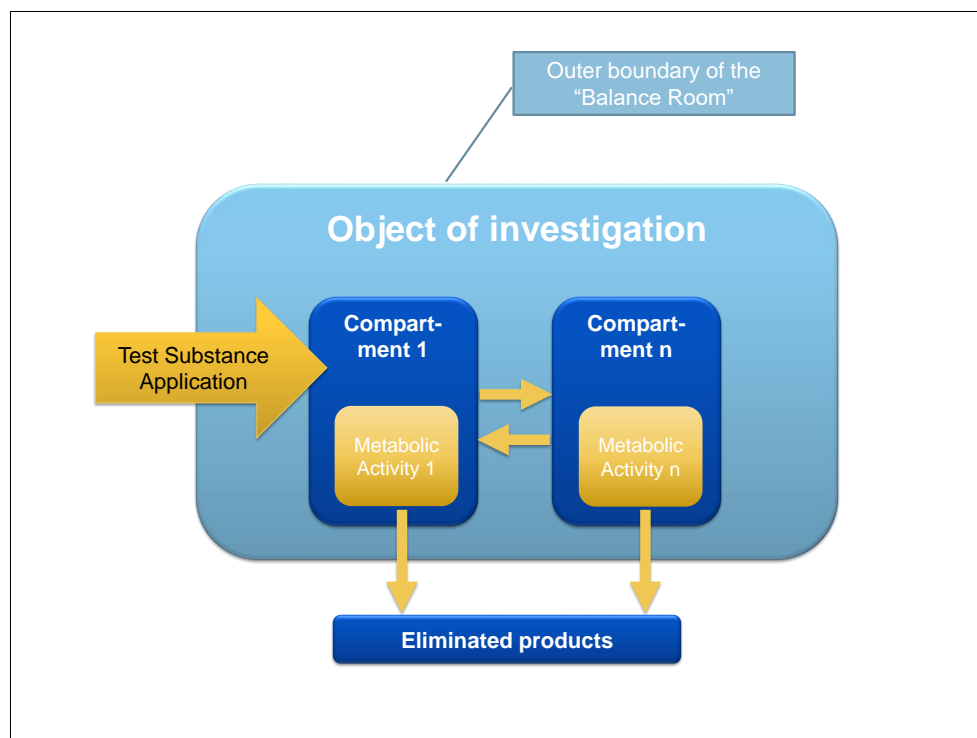
Figure 3: Model of the generalized term of metabolism study

---

[14] https://doi.org/10.2903/j.efsa.2016.4549

R 5.2-7:   The term "*Metabolism study*" should cover all types of studies (Study type) 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 5.2-8:   The term "*Metabolism study*" should cover the qualitative and the time dependent quantitative aspects of distribution and transformation.

R 5.2-9:   In order to establish a "*Material Balance*", 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.

Chapter 5.4.2 provides some information regarding metadata of a "*Metabolism study*".

5.2.3   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 (compare Figure 3). Depending on the type of experiment, not all of the process steps of Figure 3 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.

5.2.4   Balance Room

The "*Object of Investigation*" has an outer boundary, which encompasses the "Balance Room". This is the prerequisite to calculate the "*Material Balance*".

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.

5.2.5   Material Balance

Accounting of "*Test Substance*" entering and leaving the "*Balance Room*". The "% of Administered Dose (AD)" is the most common form of specification of the "*Material Balance*".

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 5.2-10:   The sum of the *AD* of all "*Compartments*" as well as the eliminated products should be comparable to the activity used.

5.2.6 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*" must 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".

5.2.7 Dosing Scheme

The "*Dosing Scheme*" describes the number and the frequency of "*Applications*" of an amount of the "*Test Substance*".

5.2.8 Transformation Process

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

5.2.9 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.

5.2.10 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 characterized by at least one identifier of the molecule (most preferred InChI; compare chapter 5.1.2). 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*".

5.2.11 Metabolic Pathway

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. "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).
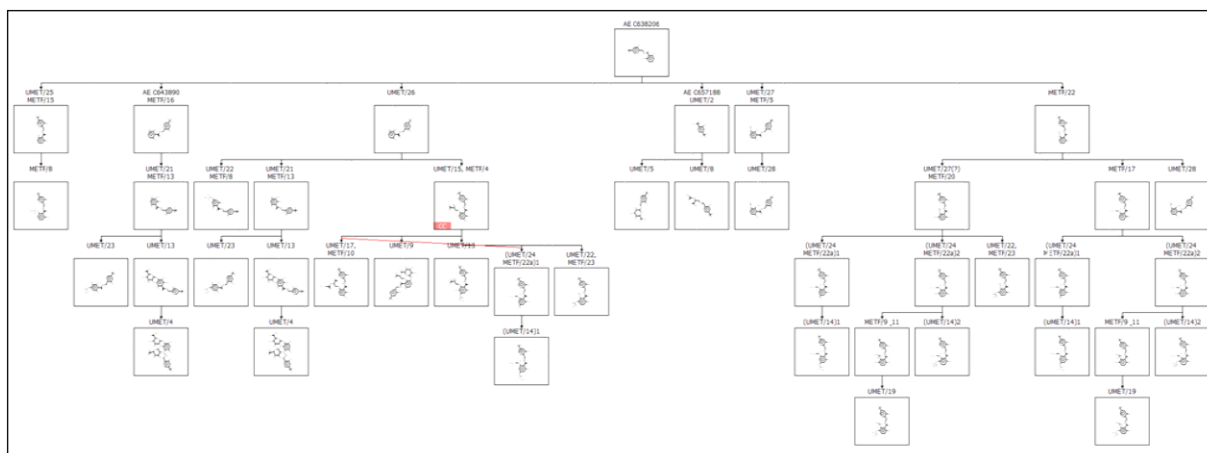
Figure 4: Example of a "*Metabolic Pathway*" generated by "*MetaPath*"

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

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

Table 2: 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 5.2.6)	
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.	

5.3 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, approaches 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.

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.

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

5.3.1   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[15], are listed in EU COM 2013/C 95/01[16].

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

R 5.3-11:   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*" (5.4) and "*Authorities' information packages*" (5.5) below.

The overall objective is to make best use of th eavailable metabolism information for the risk characterisation and risk assessment of pesticide active substances.

---

[15] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013R0283

[16] https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2013:095:0001:0020:EN:PDF
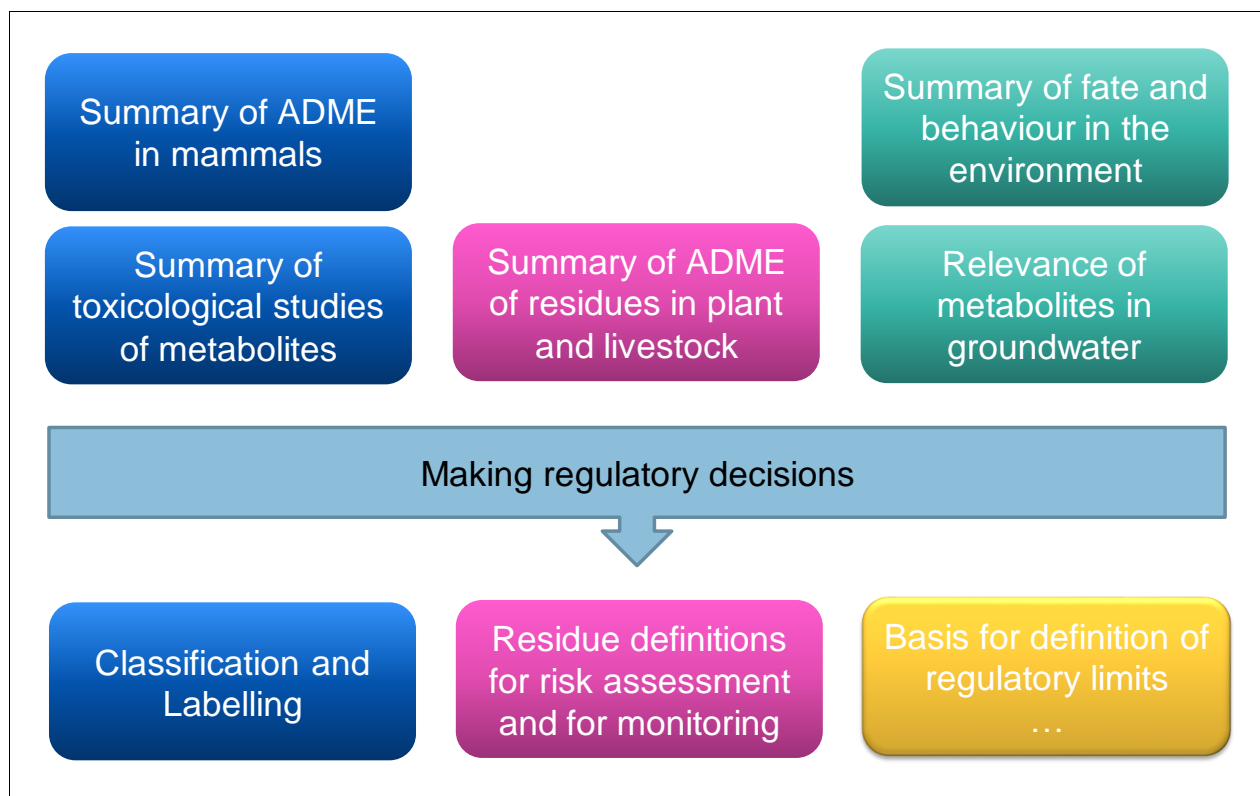
Figure 5:   Information on metabolism influence regulatory decisions

The following Guidance documents are important in European context:

- OECD Guidance residue definition: ENV/JM/MONO(2009)30
  revision ongoing (expected 2022)

- OECD Guidance on Grouping of Chemicals: ENV/JM/MONO(2014)4
  revision started

- EFSA Guidance residue definition: EFSA Journal 2016;14(12):4549 [17]

- SANCO Guidance document on the assessment of the relevance of metabolites in groundwater: SANCO/221/2000 – rev.10 – final (2003)

The user has 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) :

"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."*

---

[17] 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.

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

R 5.3-12: "*Evaluators*" should be able to manage the huge amount of metabolism relevant information with the help of an adequate IT support.

R 5.3-13: An "*IT-Tool*" is needed to store "*Aggregated Raw Data*" from metabolism studies.

The detail requirements of such an "*IT-Tool*" are described in subsequent chapters.

The following high-level process steps are necessary for risk and hazard assessment:

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:

> „*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 when metabolites have to be characterised and identified.

In SANCO/221/2000 – rev.10 – final (2003) the "relevance" of groundwater metabolites are defined:

> • *"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."*

5.3.2 Data gap filling

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

- the metabolic pathway and
- for toxicological endpoints.

5.3.2.1 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)).[18]

The *OECD* Guideline 194 (2014)[19] 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 5.3.2.2).

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.

---

[18] https://www.toxit.it/en/services/read-across

[19] ENV/JM/MONO(2014)4

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[20]:

"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.

5.3.2.2   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."[21]

The following user requirements describe all aspects of an optimal improvement process. 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.
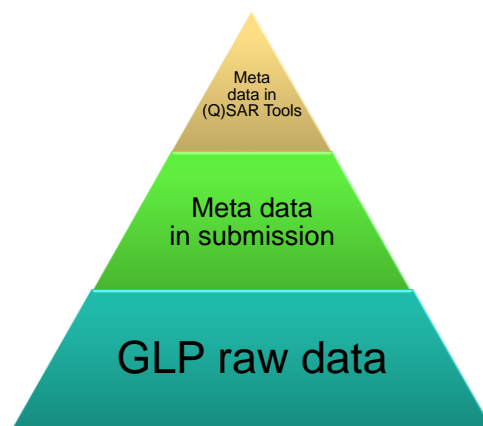
Figure 7:   Aggregation level of metadata of metabolism studies

R 5.3-14:   Authorities have to clarify the confidentially and sanitization aspects of a publication of the "*Aggregated Raw Data*" of the validated results of metabolism studies.

R 5.3-15:   Authorities have to organize 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 meta data submitted.

R 5.3-16:   Commercial (Q)SAR providers should have access to the published results of the validated and "*QA*" checked metabolism studies.

---

[20] ENV/JM/MONO(2007)2
https://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2
[21] ENV/JM/MONO(2014)4

R 5.3-17:   OECD has to organize the improvement process of the models of the OECD QSAR-Toolbox by including validated and "QA" checked results of metabolism studies.

5.3.3   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 feedand 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, 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).

5.3.4   Metabolites considered in Toxicology

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

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.

5.3.4.1   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 5.3-18:   It should be possible to transport and import all needed "Aggregated Raw Data" of ADME studies into the "*IT-Tool*".

R 5.3-19:   "*Evaluators*" should be able to visualize the metabolic pathway and the concentration time curves of different compartments (compare chapter 6.3.8) with the help of the "*IT-Tool*".

R 5.3-20:   "*Evaluators*" should be able to use a flexible reporting module where the "*Aggregated Raw Data*" could be flexibly grouped (compare chapter 6.3.12) with the help of the "*IT-Tool*".

R 5.3-21:   "*Evaluators*" should be able to calculate / check needed parameters (compare chapter 5.4.2.4) with the help of internal functions of the "*IT-Tool*".

R 5.3-22:   The "*IT-Tool*" should manage all "Aggregated Raw Data" and "*Aggregated result data*" which are needed for an improvement of (Q)SAR models.

There are the following open questions:

Q 5.3-23:   Is it helpful for the "*Evaluator*" 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*"?

Q 5.3-24:   Is there a need to include own scripts e.g. from R or python the "*IT-Tool*"?

5.3.4.2   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 characterization of relevant ground water metabolites or residue relevant metabolites.

Toxicologists have to provide an appropriate toxicological characterisation 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 synthesize the metabolite and to submit results of *in vitro* tests or
- To provide *in silico* data to characterize 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 the evaluation of the submitted *in vitro* studies or if necessary of 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 5.3-25:   "*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 characterized by a user defined name.

As a long term vision the (Q)SAR Tools should be usable as services. If such an interoperability is organized, the following user requirement make sense for the considered "*IT-Tool*":

R 5.3-26:   "*Evaluators*" should be assisted to loop over the "*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.

There are the following open questions:

Q 5.3-27:   Is it necessary for the "*Evaluator*" to manage user storable lists "List of similar substances" by selecting individual relevant substances from the response results list of the (Q)SAR Tool?

Q 5.3-28: Is it necessary to manage toxicity data for the metabolites (read across or predicted) in the requesting "*IT-Tool*"?

Q 5.3-29: Is it necessary to manage (Q)SAR results of each substance from different (Q)SAR Tools according the ECHA guide[22] in the requesting "*IT-Tool*"?

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.

There is the following open question:

Q 5.3-30: Are there any user requirements for the "*IT-Tool*" for the refined or for the cumulative risk assessment?

5.3.5 Consider Metabolites in the Residue Definition

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

A relevance assessment has to be 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, potential impact of similar metabolites from other pesticides (biocides etc). The "*IT-Tool*" therefore 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 (Please compare chapter 6.3.6).

5.4 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 6.

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 2021.

5.4.1 GLP Study Raw Data

The raw data of the metabolism studies are the data collected under "*GLP*" conditions in the laboratories[23]:

> "***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 …*"

---

[22] ECHA: Practical guide - How to use and report (Q)SARs, version 3.1 from 2016

[23] ENV/MC/CHEM(98)17

R 5.4-31: The "*GLP Study Raw Data*" are subject to "*GLP*" rules, but do not usually leave the laboratories. **These GLP data are not the content of the needed information flow from applicants to authorities.**

R 5.4-32: User functions are needed to aggregate the "*GLP Study Raw Data*" according the guidelines to write the "*GLP Study Report*".

R 5.4-33: 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 5.4-34: 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 5.4-35: 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 5.4-36: The minimal request for the additional reporting/editing IT-System is, that an appropriate user interface exits to record the needed data manually.

5.4.2   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 directed 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 reporting.

The OECD has defined the principles of the life cycle of a "*GLP Study Report*"[24].

> *"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 organize 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.

---

[24] ENV/MC/CHEM(98)17

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*".
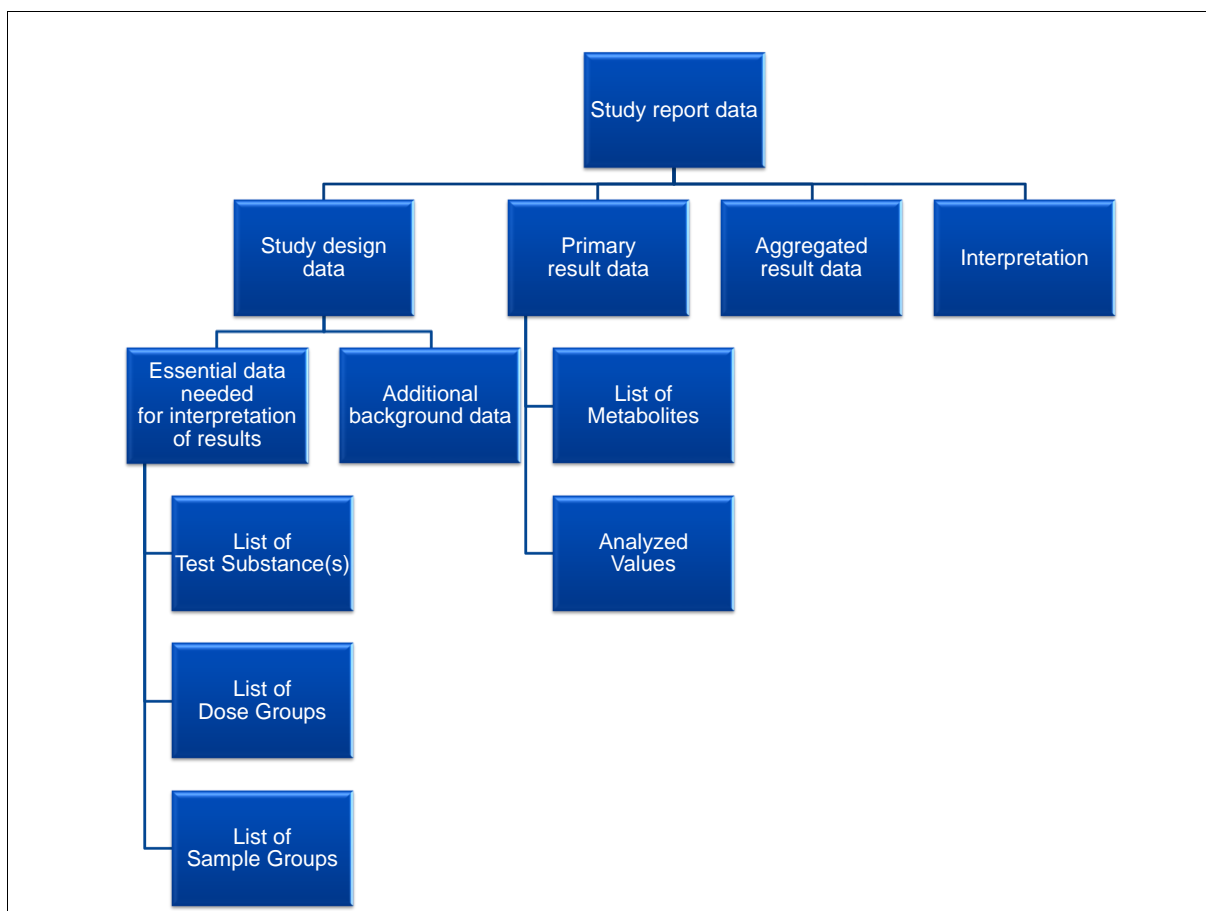
Figure 8: Hierarchy of study report data of metabolism studies

5.4.2.1 Study design data

The "Study design data" contains two groups.

R 5.4-37: 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 5.4-38: "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**". Normally the studies are investigating different groups of the "*Object of Investigation*". The reason for differences could be found in
   o differences in dosing parameters (control, dose, dose replication, dose regime, dose interval, route of "*Application*")
  o 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 5.4-39:   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 <u>not</u> 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)

5.4.2.2   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 5.4-40:   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 5.4-41:   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.

5.4.2.3   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 MSS-Composer Family and "*MetaPath*" 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 5.4-42:   A new approach is needed to create flexible pivot tables by the author of the "*GLP Study Report*"

5.4.2.4   Presentation of metabolic pathways

R 5.4-43:   A common type of the visualization of the results of a "*Metabolism study*" are figures of the "*Metabolic Pathway*".

5.4.2.5   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
- …
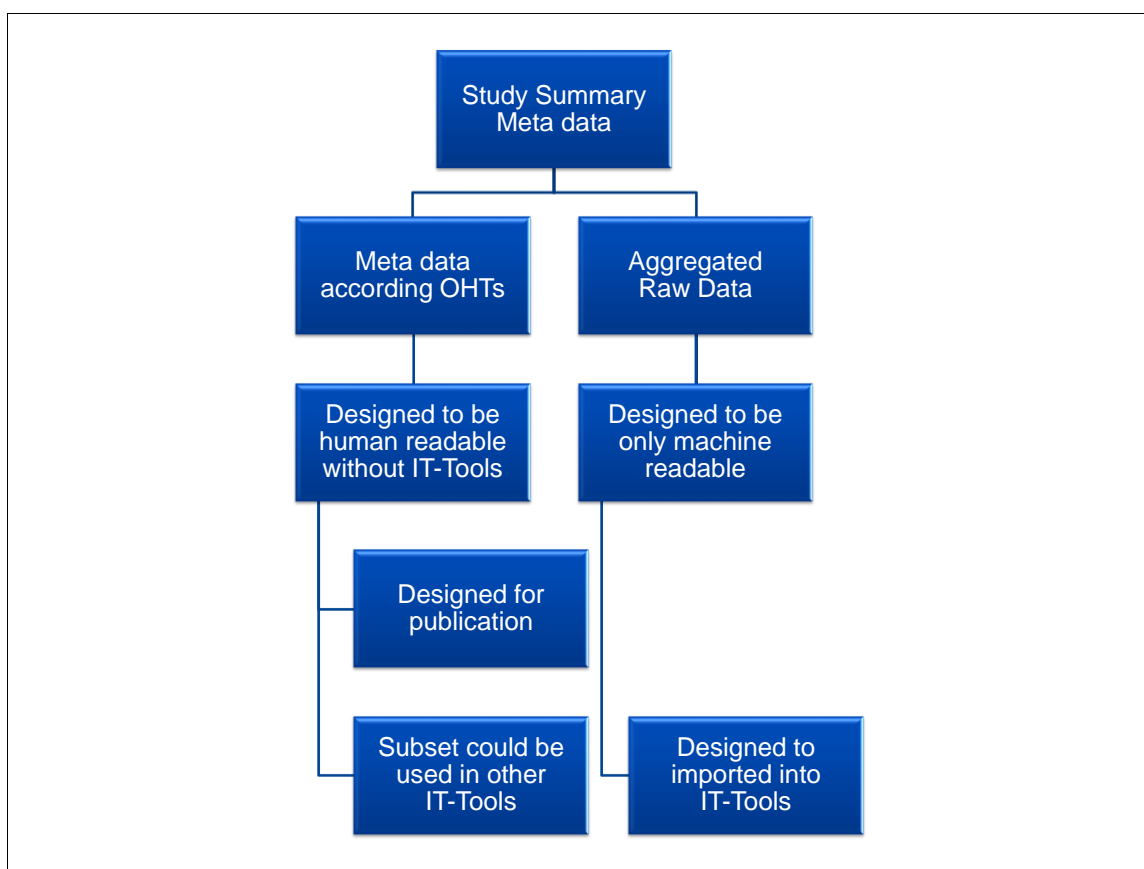
Figure 9: "*Aggregated Raw Data*" and meta data according the OHTs

There are the following open questions:

Q 5.4-44: Should the proposed "*IT-Tool*" really include basic functions of calculation of such aggregated result data?

Q 5.4-45: If the proposed "*IT-Tool*" should not provide basic functions of calculation of such aggregated result data, which data interfaces should be of assistance to transfer needed data into specialized tools?

5.4.2.6 Interpretation of the results

All detail results should be summarized, 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 5.4-46: It is necessary to manage textual summaries of the interpretation of the results for each aspect type.

5.4.3 Aggregated Raw Data

The term "*Aggregated Raw Data*" of a study report should be understood as the set of "*Metadata*" which were defined in chapter 5.4.2. (Figure 8).

The "*Aggregated Raw Data*" are not identical with the "*GLP Study Raw Data*" (Please compare chapter 5.4.1.) The "*Aggregated Raw Data*" should be provided to the authorities according to a "*Transport Concept*" yet to be created. The "*Aggregated Raw Data*" are not foreseen for publication, because

- they are not "human readable" without an adequate "*IT-Tool*" and
- semantic duplicates of the data are summarized in other human readable compilations.

Please have a look at Figure 8 for the relation of "*Aggregated Raw Data*", "*Study Summary Metadata*" and Meta data according the "*OHT*s".

R 5.4-47: The "*Aggregated Raw Data*" should be validated after the creation by the builder program.

R 5.4-48: 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 5.4-49: The "*Aggregated Raw Data*" of metabolism studies should be extractable for import into an adequate "*IT-Tool*".

R 5.4-50: Only validated "*Aggregated Raw Data*" should be imported into other "*IT-Tools*".

5.4.4 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. 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:
  - ₒ 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
  - ₒ the flag "Robust study summary"
  - ₒ the flag "Used for classification"
  - ₒ the flag "Used for Safety Data Sheet (SDS)"
  - ₒ the element "Reliability"
  - ₒ the element "Rationale for reliability incl. deficiencies"
  - ₒ the elements "Data waiving", "Justification for data waiving", "Justification for type of information"
  - ₒ the block "Attached justification"
  - ₒ the elements "Data access", "Data protection claimed"
  - ₒ the block "Applicant's summary and conclusion

R 5.4-51:   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".

5.4.5   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 5.1.1).

- 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 have to build up other data collections for other user purposes, then additional "Metadata" are needed for other / or complex search / access routes.

- If authorities have 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 Harmonized 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 "*OHT*s" accordingly.

R 5.4-52: The provided "*Study Summary Metadata*" should be suitable if authorities have to validate calculations of the applicant. "*Evaluators*" should be able to use these "*Metadata*" without complex transformations as input values for calculations.

R 5.4-53: Authorities should be able to create alternative tabular summaries from the reported results with the help of the "*Study Summary Metadata*".

R 5.4-54: The "*Study Summary Metadata*" for a "*Metabolism study*" should cover the requirements defined in chapter 5.4.2. (Figure 8).

5.4.6   Predefined Study Summary Tables

Applicants have 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 5.4-55: 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 5.4-56: At the time point of writing the "*GLP Study Report*" the "*IT-Tool*" should be able to generate all other requested summary tables from the "*Aggregated Raw Data*".

5.4.7   Endpoint Summaries

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

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 5.4-57: An "*IT-Tool*" should be able to provide reports on a set of studies for different stakeholders in different formats. The *EFSA* Appendix G is only one report template.

5.4.8   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*" has a life cycle.

The dossier container is the physical representation of a submission.

5.5 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 "Dossier*s*" 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 "Dossier*s*". These format specifications are meant to ensure that authorities are able to compose the needed evaluation reports according the guidance documents.

5.5.1 Authority Study Summary

Different additional actors have to summarize a "*GLP Study Report*" in different legal processes with different templates suitable for different addressees ("Decision makers"). That means there are different study summaries from one origin "*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 study summary
- EFSA study summary

The OECD harmonized templates are an attempt to harmonize 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[25].

The Authority Study Summaries will be part of an "*Assessment Report*".

R 5.5-58: "*Evaluators*" need the possibility to validate / recalculate results on study level from the submitted "*Aggregated Raw Data*".

5.5.2 Assessment Report

An "*Assessment Report*" which is written by a Rapporteur Member State (RMS) 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.

---

[25] A new approach, independent of MS-Word would be necessary. In Jupyter/RMarkdown there is this functionality that creates structured text. With Git, one can also find the history of individual characters in a document, or individual markings can be easily defined.

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

Table 3: Main chapters of the European DAR templates[26] where results of the metabolism studies using radiolabelled test substances were presented and discussed including the mandatory Excel attachment "Appendix G"[27]

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	Residue data
Appendix G	Template for presenting metabolism residues trials
Vol 3 B 7.2.	Metabolism, distribution and expression of residues
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

There is the following open question:

Q 5.5-59: Are there any requirements for the "*IT-Tool*" which could assist "*Evaluators*" to write the higher summary levels regarding the risk and hazard assessment of metabolites?

---

[26] https://ec.europa.eu/food/system/files/2019-04/pesticides_ppp_app-proc_guide_doss_12592-2012.zip

[27] doi: 10.2903/sp.efsa.2019.EN-1612

6   Solution approaches

This chapter will raise the impression that the proposed solutions have been formulated as final and irrevocable. It is true that the attempt has been made to describe a project that is consistent in itself.

However, the reader is advised to check the details in this project's stage very carefully. Some open questions are marked in this draft report to receive feedback during the commenting period and the planned feedback loops may add additional user requirements leading to changes.

In the onset to this report, there have already been controversial discussions from various sides about details, in particular about the different technical approaches. However, the technical solution that is finally chosen is actually of secondary importance. What is more important is the integrated presentation of user requirements.

6.1   Disclaimer

A generic NO-name "*IT-Tool*" was introduced in the big picture (Figure 1). Starting with this chapter this "*IT-Tool*" should get the temporary name "*MetabolAS Tool"* without claiming that this name must then be kept later. This new name was created to avoid confusion with existing IT systems and to prepare helpful feedback discussions.

Additionally, the question is left open whether "*MetabolAS 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!

6.2   MetabolAS ecosystem

Figure 10: The MetabolAS ecosystem

MetabolAS should be a synonym for **Metabol**ism **A**ssessment **S**ystem. It is proposed to build up an ecosystem of different components where each part of the MetabolAS ecosystem could be used by applicants and authorities because both stakeholders need the same interoperable functionality. The components (as in the figure) of the "*MetabolAS ecosystem*" are described in the next chapters.

What shall the "*MetabolAS ecosystem*" not be:

The MetabolAS 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 seen in the reconceptualization 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.

6.2.1   Governance Concept

R 6.2-1:   A "*Governance Concept*" is needed for this ecosystem.

R 6.2-2:   The Governance Body has the responsibility for the "*Scheme Definition*", the schema description to transport raw data specific for a "*Metabolism study*".

R 6.2-3:   The Governance body additionally has the responsibility for the needed "*Picklists and Picklist elements*".

R 6.2-4:   The other IT components of the ecosystem could be part of an open source project in which the interested parties contribute to the community.

6.2.1.1   The OECD as the Governance body

The preferred solution would be, that the OECD

- plays the role of the new "*Governance 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*".

6.2.1.2   A Governance body outside of the OECD

If OECD is not willing to hold the role of the new "*Governance Body*", this could also be organized 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, constitution of this "*Governance Body*" will have to be defined.

6.2.2   User Forum

It may be helpful to create a user forum.

6.2.3   Picklists and Picklist elements

R 6.2-5:   If the OECD will take on the role of the "*Governance Body*", the "*Picklists and Picklist elements*" should be defined by IUCLID mechanisms according to the adequate "*OHT*". This way, it is possible to reduce the list of values of sensible picklist elements for a specific metabolism "*Study type*". Otherwise, the "*Governance Body*" has to organise adequate mechanisms.

R 6.2-6:   If the OECD will take on the role of the "*Governance Body*", the life cycle management of the "*Picklists and Picklist elements*" should be included in the "*OHT*" life cycle. Otherwise, the "*Governance Body*" has to organise the life cycle management.

R 6.2-7:   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". At least the following "*Picklists*" are required.

Table 4: Needed Picklists

Picklist name	Remark
study_type_class	Type of harmonized template e.g. ""Basic Toxicokinetics""
regulatory_id_class	Type of Regulatory IDs format
guideline_class	Used guideline
name_class	Type of namespaces e.g. IUPAC, CAS
structure_info_class	Type of structure coding systems e.g. SMILES, CXSMILES, InChI, EINECS, ...
unit_class	List of units
object_class	Type of the object of investigation e.g. Soil, Sediment, Plant, Water, Food, …
matrix_class	Type of the sampled matrix e.g. Serum, Bile, Urine, Fruit, Egg, …
method_class	Type of analytical method e.g. HPLC, LC-MS, NMR, …
textual_result_class	Name for result blocks in the report e.g. tissue absorption, distribution excretion, ...
parameter_class	Type of parameters e.g. feeding, pH. Log KoW, body weight, melting point, ...

There are the following open questions:

Q 6.2-8:   Are there any additional elements, which should be coded by "*Picklists*"?

Q 6.2-9:   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 are useful? Should this relation be modelled?

Q 6.2-10:   There are logical references between data elements of the "*Scheme Definition*" and the units in which the respective data are given. Should the unit_class be divided into different "*Picklists*"?

6.2.4   Scheme Definition
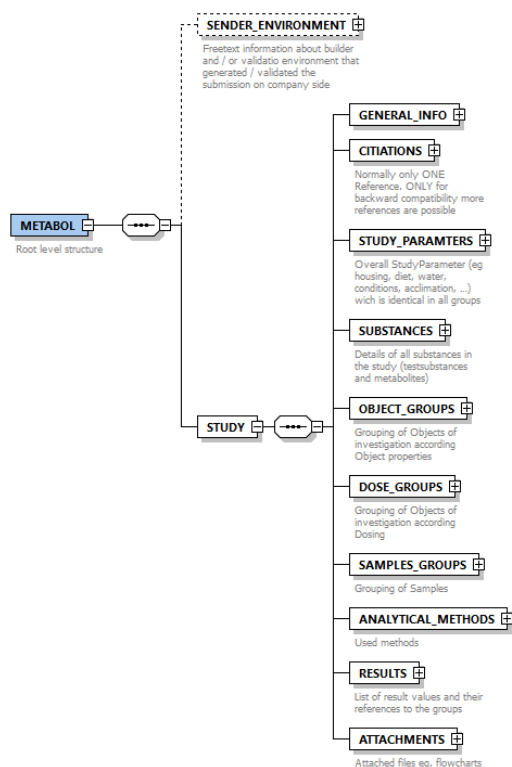
Figure 11:   MetabolAS scheme definition

A schema definition is needed as a basic interface to solve the interoperability issues. The XML file according the "*Scheme Definition*" will be called "*METABOL.XML*" in this report.

Optional To-do: Create proposal of schema definition Metabol.xsd

R 6.2-11:   The "*Scheme Definition*" (Metabol.xsd[28]) should be used as a data interface between different "*IT-Tools*" and which should be applicable for all types of Metabolism studies.

R 6.2-12:   This schema definition should re-use schema types from the "*OHT*s" to minimize the efforts for the life cycle management of the schema itself and to ensure an interoperability on the level of the XML files.

R 6.2-13:   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 6.2-14:   The "*Scheme Definition*" describes only data from a specific study.

R 6.2-15:   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*".

---

[28] Optional To-do: Attach draft Metabol.xsd schema.

R 6.2-16: If the OECD will hold the role of the "*Governance Body*", the life cycle management of the "*Scheme Definition*" should be included in the "*OHT*" life cycle. Otherwise, the "*Governance Body*" has to organise the life cycle management.

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

There are the following open questions:

Q 6.2-17: There are approximately ~2500 XML MSS Composer files in the world. A data mapping is necessary. What should happen with the data elements and values which were not mapped to the new schema? Recommendations are requested.

Q 6.2-18: Is there a need for elements to transport information about the evaluation process steps (e.g. authority, status, remarks) and results of the evaluations steps?

6.2.5 MetabolAS Tool

The "*MetabolAS Tool*" is the "*User interface*" of a "*MetabolAS collection*" realized with a database management system.

The "*MetabolAS 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 6.2-19: The "*MetabolAS Tool*" should work with an open source database management system.

R 6.2-20: A "Role Concept" is required for implementing the user access rights to the functions in the "*MetabolAS Tool*".

R 6.2-21: The "*MetabolAS Tool*" provides all modules as a web "*User interface*" programmed as an Open Source Project. Code Maintenance should be governed by the "*Governance Body*".

R 6.2-22: The life cycle management of this "*MetabolAS Tool*" should be organized according the "*Governance Concept*".

R 6.2-23: The "*MetabolAS Tool*" is able to manage "*Aggregated Raw Data*" of Metabolism studies in a "*MetabolAS collection*". The tool should be able to manage different collection but not in parallel.

R 6.2-24: Each "*MetabolAS collection*" has its own database management system.

R 6.2-25: The "*MetabolAS 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 6.2-26: The "*Test Substance*", used on study level, has a reference to an unique "*Substance*".

R 6.2-27: The "*MetabolAS 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 "*MetabolAS Tool*" are similar to the structure of the "*Scheme Definition*".

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

R 6.2-28:   The MetabolAS store information on the level of a "*Study*".

R 6.2-29:   The "*Study*" is characterized by 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.

R 6.2-30:   The value for the author will be sanitized automatically whilst compiling reports for the public based on the "*Study type*".

R 6.2-31:   The "*Study*" can contain textual descriptions (text blocks) of the study such as remarks, justifications, conclusions etc.

R 6.2-32:   The "*Study*" will be evaluated in a legal act identified by specific ID formats according a "Legal Act Type".

R 6.2-33:   The "*Study*" can contain information of more than one "*Metabolic Pathway*".

R 6.2-34:   One "*Study*" contains information to more than one "*Substance*".

R 6.2-35:   A "*Substance*" can be a "*Test Substance*" or a "*Metabolite*".

R 6.2-36:   Each "*Substance*" is characterized 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 6.3.5).

R 6.2-37:   A known "*Metabolite*", used on study level, has a reference to an unique "*Substance*".

R 6.2-38:   A "*Metabolite*" can have different "*Substance*" parents.

R 6.2-39:   A "*Substance*" can be transformed into more than one "*Metabolite*".

R 6.2-40:   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 6.2-41:   Each "*Study*" set is characterized by a status. There are rules for changing this status depending on the processing steps / user actions.

R 6.2-42:   Each "*Study*" could be used to investigate the metabolic pathway in more than one "*Object of Investigation*" e.g. mice and rats.

R 6.2-43:   The "*Study*" and the "*Object of Investigation*" are characterized by more than one "Study Parameter" referencing to a "Parameter Type".

R 6.2-44:   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 6.2-45:   A "*Test Substance*" could be applied by different administration procedures.

R 6.2-46:   The "*List of Study Object Groups*" are used to define "*Dose Groups*". The "*Dose Group*" is characterized by one "*Test Substance*" and the application parameter.

R 6.2-47: Samples were collected from the members of the "*Study Object Group*" at different time points / time intervals from different matrices. Samples could be grouped in "*Sample Groups*".

R 6.2-48: Each sample could have more than one analytical result for different substances, analysed by different methods.

Many other entities are needed. The proposed "*Scheme Definition*" gives an impression (compare chapter 7.6).

There are the following open questions:

Q 6.2-49: Is it necessary that the "*MetabolAS Tool*" contains references to all legal acts where this study was part of the submission packages or is it adequate to store only the legal act where the study was evaluated the first time?

Q 6.2-50: Is it necessary to be backward compatible? Today, one MSS composer XML file can have more than one citation to a "*GLP Study Report*"!

Q 6.2-51: Is it necessary to store phys-chem. properties of the substances in the "*MetabolAS Tool*"? This concept could lead to conflicting indication in the information system if they are stored on the level of a study.

Q 6.2-52: Is there a need to store results on the level of each individual object?

Q 6.2-53: Which additional entities are needed for the data management in the "*MetabolAS Tool*"?

6.2.6   MetabolAS Tool API

The "*MetabolAS Tool*" should provide an application programmable interface ("*API*").

R 6.2-54: The "*API*" should provide functions for reading and storing data from / into a "*MetabolAS collection*" on element and record level.

R 6.2-55: The "*API*" should provide a data interface to feed (Q)SAR models with validated data sets.

R 6.2-56: It should be possible to open a specific data set of a "*MetabolAS collection*" via REST "*API*" from external tools.

There are the following open questions:

Q 6.2-57: Which other "*API*" functions are needed?

Q 6.2-58: Definition of the permission / security level for the "*API*" functions.

6.2.7   Authorities MetabolAS collection

An optional element of the "*MetabolAS ecosystem*" could be an international "*Authorities MetabolAS collection*" which will be supported by a federation of international authorities.

R 6.2-59: The best organizational concept for an "*Authorities MetabolAS collection*" pesticides related data collection could be clarified by the joint meeting of pesticides.

R 6.2-60: There is a "Set of quality standard rules" which should be checked prior declassification of new data sets or modifications of the data sets. So the "*Authorities MetabolAS collection*" contains only validated "*Aggregated Raw Data*" of Metabolism studies of pesticides.

R 6.2-61: A "Quality control body" is needed to ensure an appropriate data quality with the help of the "*Set of quality standard rules*".

R 6.2-62: The time point to include the data set into the "*Authorities MetabolAS collection*" depends on the legal aspects of the different jurisdictions. The publication process has to be initialized by the responsible authority.

R 6.2-63: Because of transparency the public should have the access right for reading the data in the "*Authorities MetabolAS collection*".

6.3 User interface and essential functions of the MetabolAS Tool

All essential user functions should be implemented in the user interface. The "*MetabolAS Tool*" consists of different modules, which could be started by parameters.
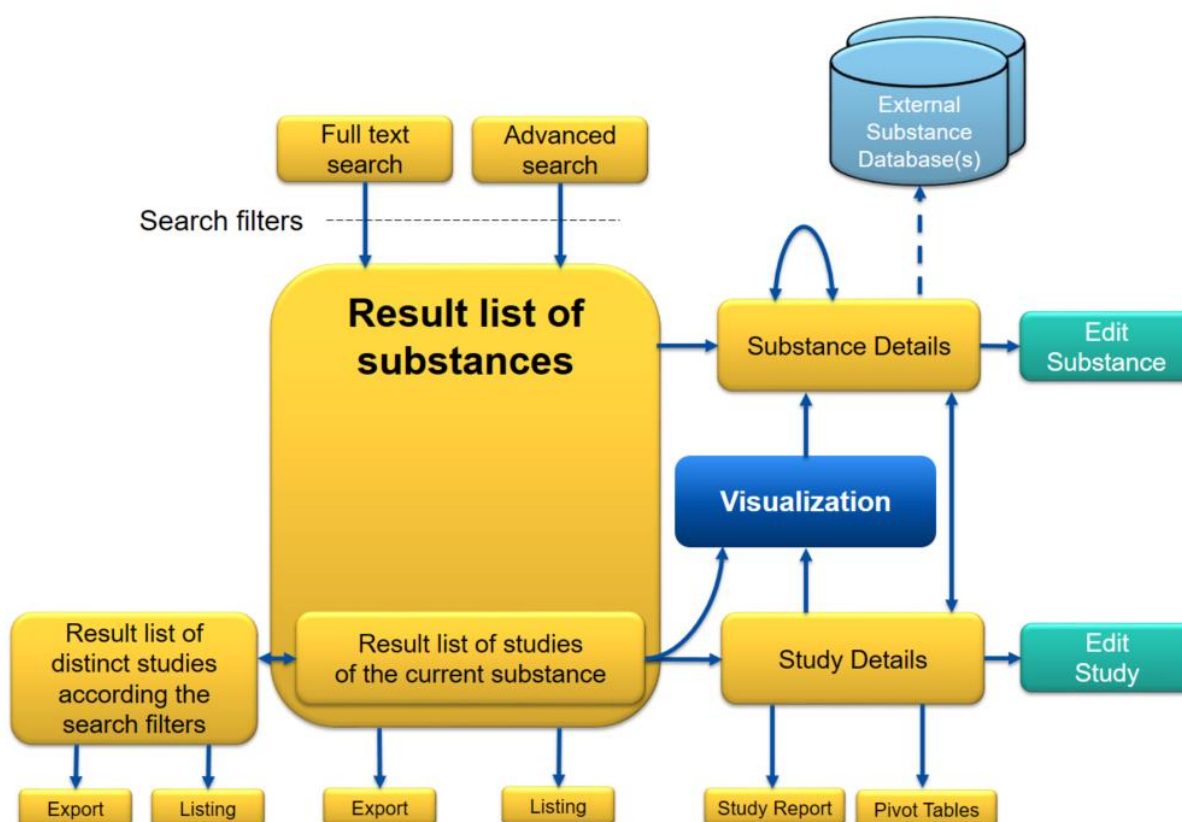

Figure 12: Main modules of the MetabolAS Tool

There are some general user requirements for the user interface of the "*MetabolAS Tool*".

R 6.3-64: The web user interfaces should start without minimal delay.
The loading time and the preparing time to show a search result should be acceptable. Currently, "*MetaPath*" is too in-performant.

R 6.3-65: The web user interfaces should start into the "Result list of substances" where the "*Full text search"* is integrated. However, the user can switch from here into the "*Advanced search*" back and forth.

R 6.3-66: The web user interfaces should provide the possibility to search for text strings in the interface by using the capabilities of the browser.

R 6.3-67: The web user interfaces should provide the possibility to copy elements from input forms and elements form reports into the clipboard.

R 6.3-68: The layout frames should not be fixed. The users should have the possibility to reduce, enlarge or move frames to separate browser views according to the user's needs.

R 6.3-69: The web user interfaces should provide the possibility of drag and drop.

R 6.3-70: The proposed result lists should be entry points for REST "*API*" calls.

R 6.3-71: The proposed result lists should be provided as configurable table. This is much more flexible e.g. for including or excluding of columns.

6.3.1   Full text search

R 6.3-72: 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 6.3-73: The result set of the full text search should filterable by facet's according the main metadata of the "*GLP Study Report"* e.g. year of the GLP report,
the "*Object of Investigation*" e.g. crop or animal,
the "*List of Dose Groups*" e.g. applied dose,
the "*List of Test Substances*" e.g. name or CAS No,
the "*List of Sample Groups*" e.g. the sample matrix.

There is the following open question:

Q 6.3-74: A deeper analysis is needed, how to present the result list of the "*Full text search*" 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.

6.3.2   Advanced search

This module should be the most important searching procedure in the "*MetabolAS Tool*". The application should be optimized for this way.

R 6.3-75: The "*Advanced search*" searches for "*Test Substances*" and "*Metabolites*".

R 6.3-76: All metadata of "*Test Substance*", "*Metabolite*", "*Study*", "*Object of Investigation*"" "*Metabolic Pathway*", etc. could be used as a filter. The search filters are specific according the data type of the entity attributes.

R 6.3-77: All used filter clauses are concatenated with a logical 'AND'.

R 6.3-78: 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 6.3-79: The initialization time for some search options and for preparing the search results of the current "*MetaPath*" is not acceptable.

R 6.3-80: 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, extended SMILES (CXSMILES), InChI, SMARTS) could be used which are transformed internally into the needed format automatically. A graphical tool to draw a (sub)structure to be searched for should be included.

R 6.3-81: The size of the result list of the similarity search (R 6.3-80) 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 6.3-82: The similarity search (R 6.3-80) is only an additional filter clause which are concatenated with a logical 'AND'. So it is possible to search for metabolism studies e.g. including a) "*Substances*" with a specific substructure b) in a specific species ("*Object of Investigation*") c) with a specific treatment type (mode of "*Application*")

R 6.3-83: If a user has asked for an unrealistic structure or realises a mistake in the entered structure, it should be possible to interrupt the searching process. "*MetaPath*" search process could not be stopped.

R 6.3-84: It should be possible to search for attributes of the "*Transformation Processes*".

R 6.3-85: It should be possible to search for substances by their names. However, pay attention that the "*MetabolAS Tool*" has NO overall substance model for unknown substances (compare R 6.2-27).

R 6.3-86: The users should be able to store the used search filter options locally and to load a stored request.

R 6.3-87: "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 6.3-88: User can expand the "Result list of substances" to see a 2nd level with a table where all studies are listed in which the current substance occur. This table is called "*Result list of studies*". The fields to be visible can be configured by the user.

There are the following open questions:

Q 6.3-89: "*MetaPath*" provides five different types of searches for chemicals / reactions / similarity / Tables / Transformation. All search filters should be included in the web interface according to R 6.3-76.
It should be evaluated if there is a need for all of these different types of queries e.g. the search for similar maps?

Q 6.3-90: The proposed concept of the "Result list of substances" has the consequence, that a specific study would be shown in all "*Result list of studies*" on the 2nd level of the "*Substances*" matches the search filters options. Therefore, studies could be listed as duplicates! Is there a need to create an additional result list of the distinct "*Result list of studies*"?

Q 6.3-91: Which other independent search modules for the "*MetabolAS Tool*" should be created to show different result lists, **other** than the proposed "Result list of substances" e.g. a "Result list of test objects"?

Q 6.3-92: Is there a need to concatenate used filter clauses by OR?

Q 6.3-93: Which other structure language codes should be assisted additionally to SMILES and InChI?

6.3.3   Result list of substances

R 6.3-94: User can select substances and / or studies by checkboxes. By default, all rows are selected.

R 6.3-95: Selected rows could be used for reports and exported as XML-result-sets files.

R 6.3-96: Both lists could be sorted by the provided columns.

R 6.3-97: Users can activate a separate frame called "*Substance Details*" by selecting one substance. By scrolling down in the "*Result list of substances*", the separate substance details frame will be refreshed.

R 6.3-98: Users can activate a separate frame called Study Details by selecting one study. By scrolling down, the separate study details frame will be refreshed.

R 6.3-99: Users can choose two options for showing "*Study Details*": "*Metabolic tree*" or "Study information".

R 6.3-100: 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_		Max: 999 / **all**
		Show also linked metabolites _Yes_		
		Show study details as:	**Metabolic tree**	
			Study information	

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

Metabo-IAS-ID	Object	Type	Dose range	MetabolAS-ID
287	rat	Biokinetic	1 – 150 mg/kg bw	
288	rat			

Trivial Substance Name IUPAC Name	CAS-No	PubChem	Type
6-Hydroxy-nicotinsäure 6-Hydroxypyridine-3-carboxylic acid	5006-66-6	329751860	Metabolite of Imidacloprid

Result list of studies

Metabo-IAS-ID	Object	Type	Dose range	MetabolAS-ID
287	rat	Biokinetic	1 – 150 mg/kg bw	

…

Figure 13: Draft layout of the module "Search for metabolism studies"

There are the following open questions:

Q 6.3-101: "*MetaPath*" assists a tree structure (left side). Is there a need for changing between a result list table and a result list tree?

Q 6.3-102: If no additional result list tree is needed: Which additional essential user functions of this left side should be included in the proposed stacked (expandable) result list table?

6.3.4   Substance Details

R 6.3-103: 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 6.3-104: It should be possible to toggle between the "*Substances*" listed in "is formed from" and "can be transformed into".

6.3.5   Substance Edit

To-do

6.3.6   Jump into external substance databases

It is necessary to be able to start with a substance identification as a parameter in external substance databases, e.g. to view more detailed toxicological data or to find out about residue tests on this substance.

R 6.3-105: It should be possible to jump into different predefined external substance databases.

R 6.3-106: The users should be able to configure the list of the preferred predefined external substance databases.

R 6.3-107: The users should be able to expand the list of predefined external substance databases

R 6.3-108: The users should be able to define substance identification field, which should be used as the dynamic parameter.

6.3.7   Study Details

To-do

6.3.8   Visualization

R 6.3-109: 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 6.3-110: The "*MetabolAS Tool*" should provide all "*MetaPath*" functions for visualising the metabolic pathway.

R 6.3-111: There are many different ways pathways can be shown. Specifying an interface with cytoscape[29] might be a good idea. This is an open source tool with performant plugin-structure

6.3.9 Study Edit

R 6.3-112: It should be possible to create a "*Study Data Set*" in the "*MetabolAS Tool*" manually without the import function.

R 6.3-113: The "*MetabolAS 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*" which else?) by using the clipboard.

R 6.3-114: The tool should analyse the table, provide importing suggestions to the user and import the rows according a defined column structure.

R 6.3-115: 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*".

There is the following open question:

Q 6.3-116: Specific rules should be formulated and enforced by the "*Governance Body*".

6.3.10 Compare

R 6.3-117: The "*MetabolAS Tool*" should provide all "*MetaPath*" functions for comparing "*Metabolic trees*" (e.g. data comparison within and across different taxa) and to visualise different and identical parts.

6.3.11 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.

---

[29] https://cytoscape.org/

6.3.11.1 The user working stack "List of relevant studies"
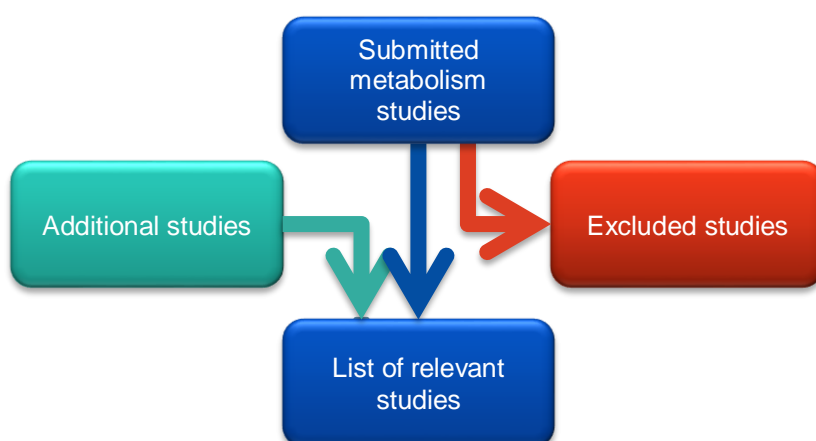
Figure 14: Define the "List of relevant studies"

The term "*List of relevant studies*" depends on the endpoint and the legal act addressed.

R 6.3-118: At the end of assessment, "*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 6.3-119: The "*List of relevant studies*" is specific for each section (Toxicology, Residues, Environmental Fate)´.

R 6.3-120: "*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 into the IT-Tool.

R 6.3-121: "*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 6.3-122: It should be possible to complete the "*List of relevant studies*" with other studies where similar metabolites are found.

R 6.3-123: "*Evaluators*" are able to group the elements of the "*List of relevant studies*" into different groups. A group is characterized by a user defined name.

R 6.3-124: "*Evaluators*" can use the "*List of relevant studies*" for reports. The defined groups could be used for filtering or aggregations 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 6.3-125: It should be possible to modify the "*List of relevant studies*" after the peer review process to include input from other member states.

R 6.3-126: *Evaluators*" are able to include additional studies into the "*List of relevant studies*" which are not currently in submitted dossier but in the local collection of the "*MetabolAS Tool*".

R 6.3-127: The "*MetabolAS Tool*" should allow archiving of the search criteria used for the screening in step (R 6.3-121 e.g. name of the reference collection, timeliness, search criteria, ...)?

R 6.3-128: The archive of the used screening search strategies (R 6.3-127) should become a part of the "*Aggregated Raw Data*" package and thus also be delivered by the applicant to the agency?

6.3.11.2 The "View of Substances to evaluate"

The term "*View of Substances to evaluate*" 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.
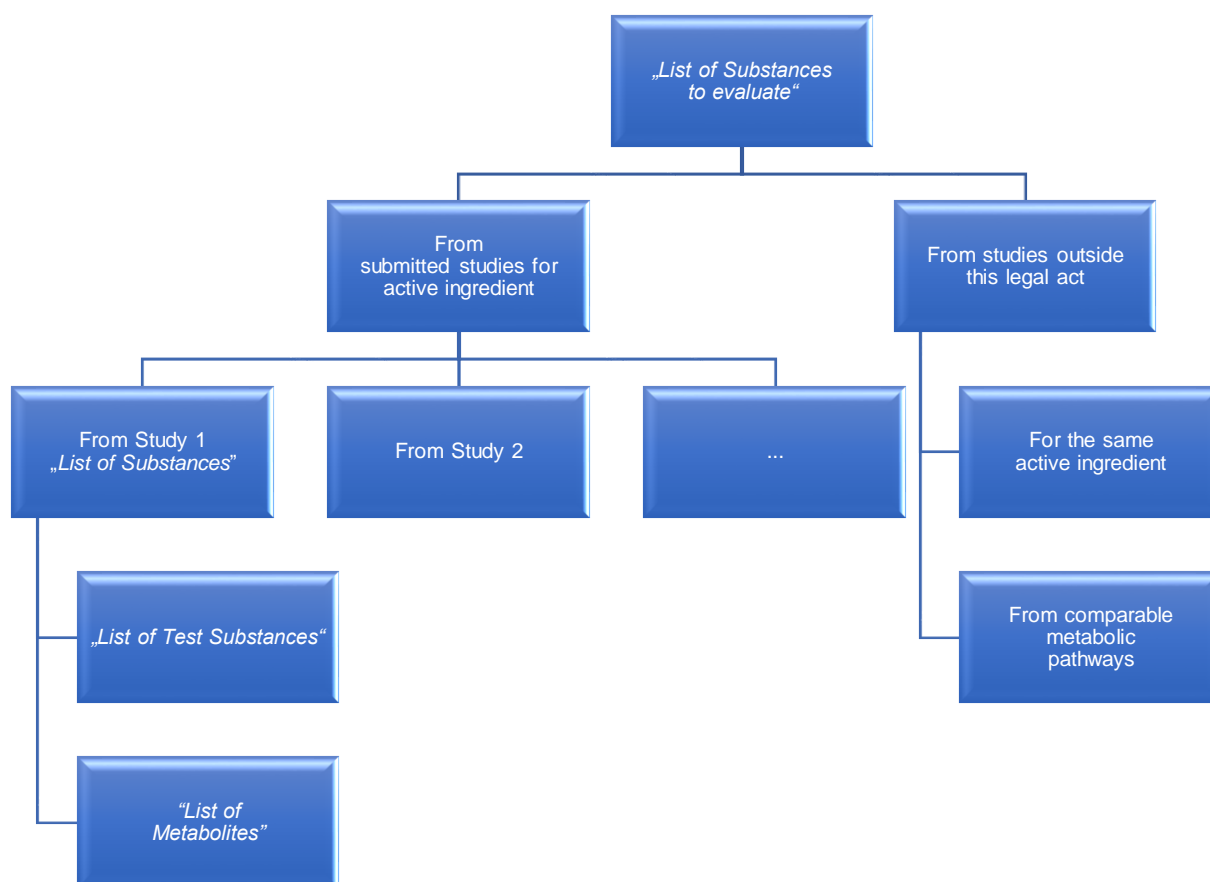
Figure 15:   Sources of the "View of Substances to evaluate"

R 6.3-129: "*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.

6.3.11.3 The user working stack "Set of Substances"

R 6.3-130: "*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). Evaluators can specify a name for the "*Set of Substances*" specific for this legal act.

R 6.3-131: "*Evaluators*" can use the "*Set of Substances*" for reports. The defined group names could be used for filtering or aggregations of results (see Figure 16).

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				

Example for user defined substance groups where results should be aggregated.

Figure 16: User defined aggregation groups on substance level

R 6.3-132: "*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 6.3-133: 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 6.3-134: 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 of the "*List of relevant studies*" for this detail view. Reports should be able for the current view.

There are the following open questions:

Q 6.3-135: Is it necessary to be able to add additional substances to the "*Set of Substances*"?

Q 6.3-136: Is an overall "*View of Substances to evaluate*" including the residue, toxicology and ground water perspective necessary?

6.3.12 Report

6.3.12.1 Listings

R 6.3-137: "*Evaluators*" should be able to create default reports specific for each type of the result list for all or for selected rows.

6.3.12.2 Default Study reports

R 6.3-138: The "*MetabolAS Tool*" contains a "*Report*" which is able to create a textual study summary over all study metadata information of the selected study.

R 6.3-139: The users should be able to recalculate results from one substance to another substance and to aggregate results according to his expert knowledge.

6.3.12.3 Pivot tables

R 6.3-140: "*Evaluators*" should be able to create <u>default</u> pivot tables from the "*Aggregated Raw Data*" of one selected study by choosing one of the "*Predefined Study Summary Tables*". Chapter 7.7 contains proposals for pivot tables.

R 6.3-141: The users should be able to create <u>flexible</u> pivot tables from the "*Aggregated Raw Data*" of one selected study by using the groups defined in the study or 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 "<u>flexible</u> pivot table" templates to use them in later similar problem settings.

R 6.3-142: The predefined substances groups ("known" and "unknown") could be used for the aggregation of the results.

R 6.3-143: "Evaluato*r*s" should be able to create additional substance groups by defined characteristics (e.g. according to functional groups, conjugates, …).

R 6.3-144: If mass balance data should be presented, the sum values of the Total radioactive residue (TRR) should be calculated automatically in the pivot tables.

R 6.3-145: It should be possible to recalculate "Analysed Values" from one substance to another "calculated as substance".

R 6.3-146: It should be possible to calculate mean, standard deviation and the count for individual data and to add these in additional columns.

R 6.3-147: If concentration – time values are measured, corresponding summary tables of the results should be created.

There is the following open question:

Q 6.3-148: Is an adequate graphical output of measured concentration – time values of request R 6.3-147 needed?

6.3.13 Documentation

R 6.3-149: It should be possible to start a context sensitive User Documentation from all modules.

R 6.3-150: The User and the System Documentation should be part of the "*Open Source Project*".

6.3.14 Import / Export / Validation

R 6.3-151: The "*MetabolAS Tool*" could import and export data sets of "*Aggregated Raw Data*" of Metabolism studies.

R 6.3-152: The "*METABOL.XML*" file is one output which will be generated according the "*Scheme Definition*" by the "*MetabolAS Tool*".

R 6.3-153: The "*MetabolAS 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 6.3-154: A validation report could be generated with a detailed information about the found errors and warnings.

6.3.15 Assist the transport step via IUCLID

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 6.3-155: According to the approach of the "*MetabolAS Tool*" the GLP report plus the adequate attachment xml file should be submitted with the help of IUCLID (see chapter 6.6.2). Both documents have the same document life cycle.

R 6.3-156: All data which should be published should be included in the OECD harmonized template. "*Aggregated Raw Data*" as data duplicates in another format are not candidates for a publication.

R 6.3-157: Applicants should be able to store the output of the "*Report*" into the OECD harmonized template in the block "Applicant'' summary and conclusion".

R 6.3-158: Applicants have to add information in the section "Administrative data" and "Applicant'' summary and conclusion". They have to summarise the relevant aspects of the study including the conclusions reached in context of the regulatory context.

R 6.3-159: If needed, IUCLID should be able to import metadata information of the attached "*METABOL.XML*" file, which are necessary for other IUCLID user functions.

6.3.16 Management

6.3.16.1 System management

R 6.3-160: The database management system used provides all the necessary system management functions for a secure and effective running of the applications (system updates, backup, analysis tools, …).

6.3.16.2 User management

R 6.3-161: The "MetabolAS Tool" assists a user management in combination with a "*Role Concept*".

R 6.3-162: A privileged user can manage other users.

6.3.16.3 Substance management

R 6.3-163: A privileged user can manage the central list of "*Substances*".

6.3.16.4 Picklist management module

R 6.3-164: It should be transparent for the users which picklist elements could be used in which editing module in which input field. This module creates a list of all "*Picklists and Picklist elements*", grouped by the "*Picklists*".

There is the following open question:

Q 6.3-165: At the current point in time, is there a need for users to have a mechanism for requesting new picklist elements electronically at the "*Governance Body*" and to be able to use them? The need will arise eventually and to set it up from the start could be very helpful for the general acceptance of the tool.

6.3.16.5 Specialized administrator module

All content related management functions should be summarized in a specialized administrator module

R 6.3-166: Merge two substances and their references, because a these substances are duplicates.

R 6.3-167: Completion of pathways by related studies (same or other active substances).

R 6.3-168: Identification of structurally related compounds over all studies.

There are the following open questions:

Q 6.3-169: We need some more details for these management functions.

Q 6.3-170: Which other content related management functions should be provided?

6.3.17 Missing functions

There is the following open task / question:

Q 6.3-171: Which implemented "*MetaPath*" functions are essential but have been forgotten?

6.4 Internal Stakeholder MetabolAS Instances

R 6.4-172: Stakeholders can create their own internal instances of the MetabolAS for specific questions which are outside the "*MetabolAS ecosystem*".

There is the following open question:

Q 6.4-173: What other stakeholder-specific requirements should be considered?

6.5 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 6.5-174: The "*Authorities MetabolAS collection*" could be the official data source for the QSAR-Toolbox regarding the pesticide metabolism pathways and kinetics.

R 6.5-175: The "*Authorities MetabolAS collection*" should assists the interoperability with (Q)SAR Tools by an "*API*" (see section 6.2.6).

6.6 Transport Concepts for aggregated raw data of metabolism studies

Only technical solutions that are compatible with the IUCLID submission approach are considered for the transportation step. The following chapters compare different technical solution architectures for the transport of "*Aggregated Raw Data*" of metabolism studies within a legal act. The solutions should guarantee the needed information flow in the granularity required.

But the

- effort to initialize this information flow with all the needed IT-Systems and data interface and
- the effort required to maintain the systems

should be compared in a qualitative manner.

The *BfR* prefers the second technical transport mechanism of the "*Aggregated Raw Data*" of Metabolism studies (see chapter 6.6.2.) because of many obvious advantages of this solution. *BfR* proposes expanding the OECD house architecture with the new category "*OECD Attachment Type*" (see Figure 19).

The *BfR* prefers this approach because of the need for transporting "*Aggregated Raw Data*" for other endpoints e.g. "Genetic toxicity in vitro" (OHT 70) to improve the (Q)SAR models by broader training sets for agrochemicals in the future.

There is the following open question:

Q 6.6-176: For which other endpoints (outside of the metabolism studies and of the residue trials (OHT 85-5)) a generic transport approach using an "*OECD Attachment Type*" would make sense in the future?

6.6.1 Use of 3rd Party Attachment Types

6.6.1.1 The concept
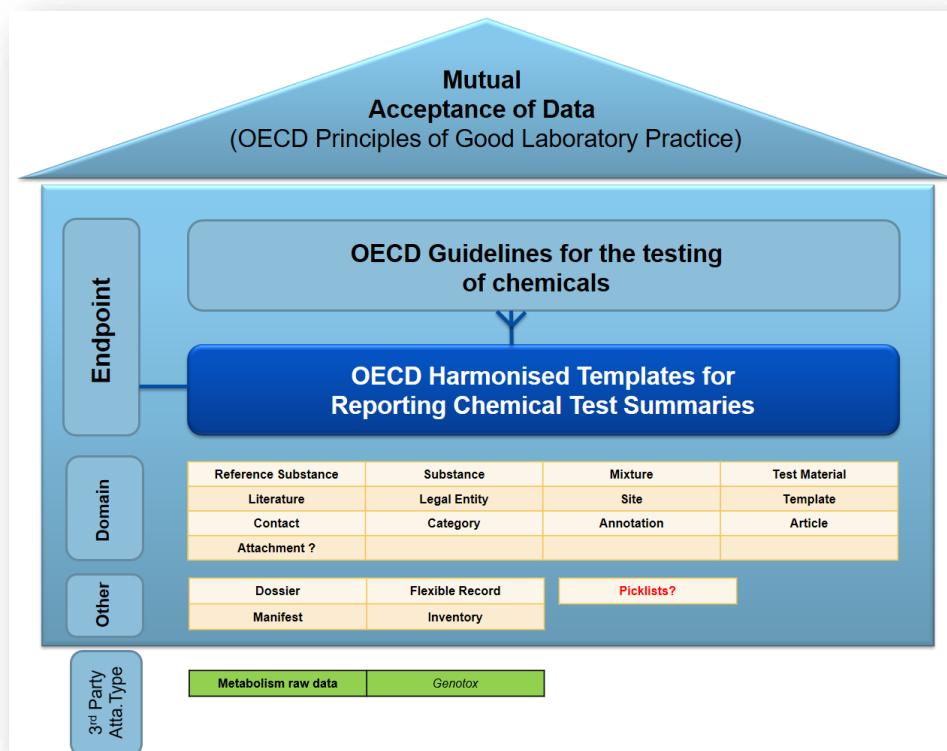
Figure 17: Add attachments with raw data to IUCLID dossiers according 3rd party schema definition outside of the OECD.

This solution is similar with the current European situation. Today the XML files of the DER/MSS-Composer Family are submitted in awareness of several weak points (see 7.5). However, this approach could be improved.

6.6.1.2 Principles

The main principles for attachments, which should transport "*Aggregated Raw Data*" are:

1. This solution is outside of the OECD regulatory scope. The schema description will be created and managed by 3rd Party stakeholders.
2. This is a generic approach. It should be possible to attach such a raw data attachment to all Metabolism study OHTs. The defined attachment structure should cover all needs for all "*Metabolism study*" endpoints.
3. On level of a legal act, it could be decided if such an attachment of "*Aggregated Raw Data*" is appropriate and should be mandatory.
4. The user interface for the OHTs will be free of "*Aggregated Raw Data*".
5. The provided machine readable information in the attachment is on the semantic level identical or a part of the human readable textual information in the OHT.
6. The attachments for "*Aggregated Raw Data*" of metabolism studies are of type XML.

7. All required data should be defined in the attachment file itself. So the delivered attachment is self-contained.

8. There is a possibility to decouple the creation of attachment and endpoint summary.

9. If references to already created IUCLID objects are known at the time of creating this attachment XML file, these UUID identifier should be integrated.
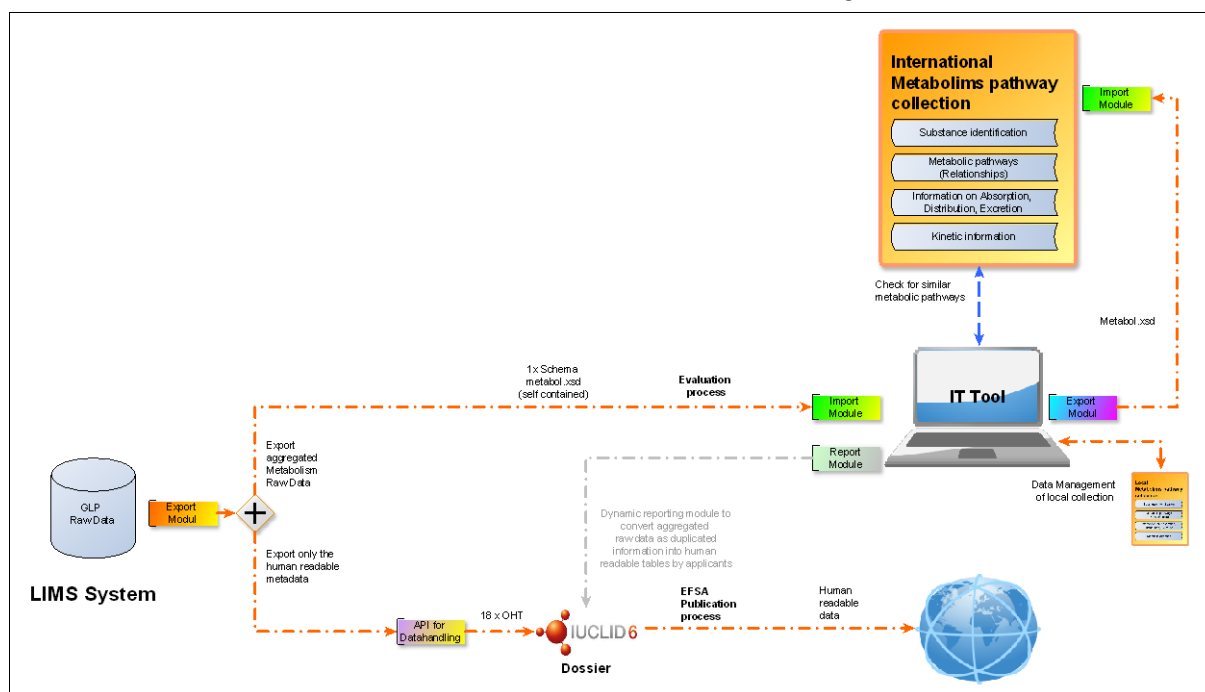
Figure 18: Needed modules if using the new 3rd Party attachment type "Metabolism raw data"

Table 5: Comparison of pro and contra for using the 3rd Party attachment type to transport aggregated raw data

Module / aspect	Pro "3rd Party Attachment type"	Contra "3rd Party Attachment type"
Export module for each LIMS GLP System	Programmers can use the schema definition "*Scheme Definition*" to create this attachment type. No references to a IUCLID Reference Substances model is needed	Additional parallel export modules are needed for the OHTs
Flexibility per OHT and legal act	**No OHT change is needed**. On level of a legal act a validation rule could check if such an attachment of "Aggregated Raw Data" is mandatory or not.	-
OHT generated data input form	No input form is needed. No aggregated raw data is disturbing the IUCLID input form.	-
Report module	The "*MetabolAS Tool*" should be able to fulfil all requirements for reports	-
Import module - part: extracting the raw data	Not needed	-
Import module - part: import into the consuming IT-Tool	This module is needed for data exchange between different collections of metabolism information systems	-
Export module	This module is needed for data exchange between different collections of metabolism information systems	-

Governance of the new object type	If a 3rd party governance body was established, this body could react much more flexible.	This approach is outside of the OECD regulatory scope
Third party engagement	The new standard level opens the possibility for other third parties to develop compatible IT-Tools and data interfaces	-
Interoperability	The 3rd party standard level (schema definition) • could be based on IUCLID types • could contain references to IUCLID-Reference substances The delivered attachments are XML files. It is easy to produce such XML files with IT-Tools where these data are already stored (e.g. LIMS) It is easy to import these XML files into a "*MetabolAS Tool*"	The 3rd party standard level has to organize • the management of picklists and picklist items
Time point of creation	There is a possibility to decouple creation of attachment and endpoint summary. The "*GLP Study Report*" and the created attachment have the same lifetime.	-
Generalisation per knowledge sector	All metabolism studies could use this standard to submit aggregated raw data	-
Relationship to aggregated raw data of other knowledge sectors	It is assumed, that other scientific areas have similar problems to transport „Aggregated Raw Data" as an attachment of a study endpoint record	-
User interface for OHT where the attachment should be attached	To transfer the Metadata which could only be written by machines will make the user interface of the OHTs more manageable for humans.	-
Relationship to reports and the transparency regulation	The complete human readable information will be provided by the OHTs. This will not be changed. The „Aggregated Raw Data" are a new additional level of information.	Confidentiality aspects should be analysed.
Simplicity level	The delivered attachments are self-contained. It means all needed information is in one XML file. Needed third party data interfaces could be developed with low level budgets.	-
Reduction of duplicated information	-	The content of the aggregated raw data are provided in the Rich-Text elements of IUCLID as duplicates on the semantic level.
	The same information has two different formats: the human readable and the machine-readable format for different purposes.	-

6.6.2   Create an OECD Attachment Type

6.6.2.1   The concept
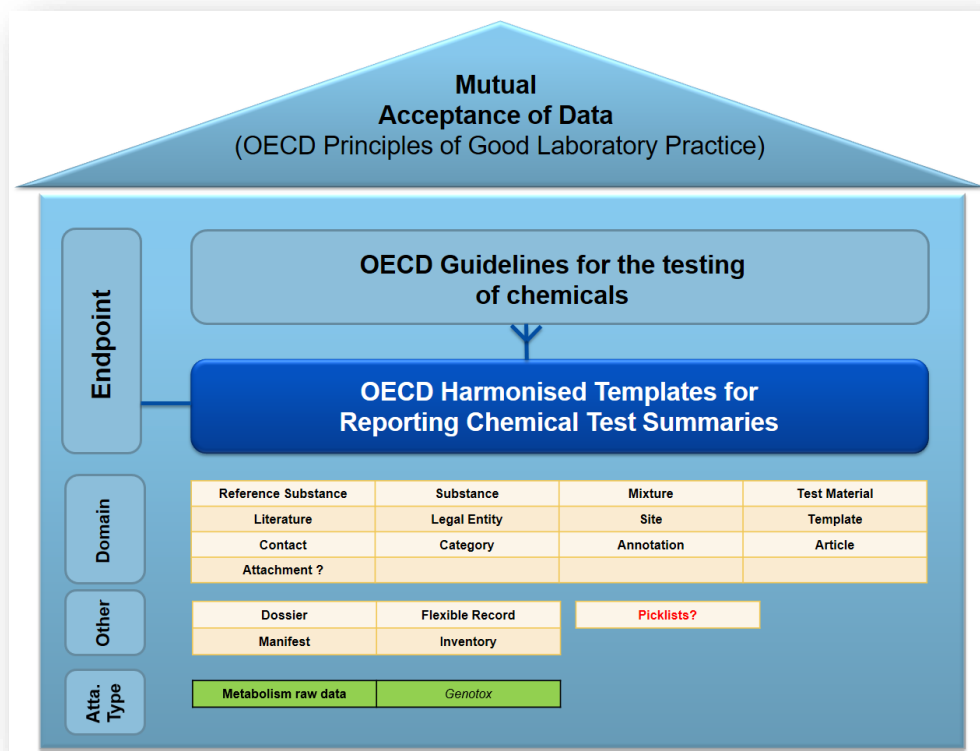
Figure 19: Expanding the OECD data architecture with the new category
            "OECD Attachment Type"

This solution creates a new category "*OECD Attachment Type*" with its first participant "Metabolism raw data" under the umbrella of the "OECD Harmonised Templates". Other attachment types for transporting "*Aggregated Raw Data*" would be useful, e.g. for data on genetic toxicity.

The only difference to the solution in chapter 6.6.1 is the difference in the governance body.

6.6.2.2   Principles

The main principles for attachments, which should transport "*Aggregated Raw Data*" are:

1. This is a generic approach. It should be possible to attach such a raw data attachment to all Metabolism study OHTs. The defined attachment structure should cover all needs for all "*Metabolism study*" endpoints.
2. On level of a legal act, it could be decided if such an attachment of "*Aggregated Raw Data*" is appropriate and should be mandatory.
3. The user interface for the OHTs will be free of "*Aggregated Raw Data*".
4. The provided machine readable information in the attachment is on the semantic level

identical or a part of the human readable textual information in the OHT.

5. The attachments for aggregated raw of metabolism studies are of type XML.

6. All needed data should be defined in the attachment file itself. So the delivered attachment is self-contained.

7. There is a time decoupling of creating the attachment and the endpoint summary possible.

8. If references to already created IUCLID objects are known at time of creating this attachment XML file, these UUID identifier should be integrated.

9. The attachments for "*Aggregated Raw Data*" are using "*Picklists and Picklist elements*" of corresponding the OHTs.

If using the new attachment type "Metabolism raw data", the needed modules are identical with Figure 18.

Table 6:   Comparison of pro and contra for using the attachment type to transport aggregated raw data

Module / aspect	Pro "Attachment type"	Contra "Attachment type"
Export module for each LIMS "*GLP*" System	Programmers can use the schema definition "*Scheme Definition*" to create this attachment type. No references to a IUCLID Reference Substances model is needed	Additional parallel export modules are needed for the OHTs
Flexibility per OHT and legal act	**No OHT change is needed**. On level of a legal act a validation rule could check if such an attachment of "Aggregated Raw Data" is mandatory or not.	-
OHT generated data input form	No input form is needed. No aggregated raw data is disturbing the IUCLID input form.	-
Report module	The "*MetabolAS Tool*" should be able to fulfil all requirements for reports	-
Import module - part: extracting the raw data	Not needed	-
Import module - part: import into the consuming IT-Tool	This module is needed for data exchange between different collections of metabolism information systems	-
Export module	This module is needed for data exchange between different collections of metabolism information systems	-
Governance of the new object type	The new standard level is based on the OHT organizational procedures.	-
Third party engagement	The new standard level opens the possibility for third parties to develop compatible IT-Tools and data interfaces	-
Interoperability	The new standard level (schema definition) • is based on IUCLID types • Use picklist items created by IUCLID mechanisms • could be used for formal validations • could contain references to IUCLID-Reference substances The delivered attachments are XML files. It is easy to produce such XML files with IT-Tools where these data are already stored (e.g. LIMS) It is easy to import these XML files into a "*MetabolAS Tool*"	-

Time point of creation	There is a possibility to decouple creation of attachment and endpoint summary. The "*GLP Study Report*" and the created attachment have the same lifetime.	-
Generalisation per knowledge sector	All metabolism studies could use this standard to submit aggregated raw data	-
Relationship to aggregated raw data of other knowledge sectors	It is assumed, that other scientific areas have similar problems to transport „Aggregated Raw Data" as an attachment of a study endpoint record	-
User interface for OHT where the domain type should be used	To transfer the Metadata which could only be written by machines will make the user interface of the OHTs more manageable for humans.	-
Relationship to reports and the transparency regulation	The complete human readable information will be provided by the OHTs. This will not be changed. The „Aggregated Raw Data" are a new additional level of information.	Confidentiality aspects should be analysed.
Simplicity level	The delivered attachments are self-contained. This means all needed information is in one XML file. Needed third party data interfaces could be developed with low level budgets.	-
Reduction of duplicated information	-	The content of the aggregated raw data are provided in the Rich-Text elements of IUCLID as duplicates at the semantic level.
	The same information has two different formats: the human readable and the machine-readable format for different purposes.	-

If this solution will be realized, this would be a new generic approach to transport "*Aggregated Raw Data*" in addition to the OHT not only for metabolism studies.

6.6.3   Create an OECD Domain Type

6.6.3.1   The concept

As stated above, the main concept for this approach is to create an "*OECD Domain Type*" under the umbrella of the "OECD Harmonised Templates".

This domain type should then be generic enough to be included in all "OECD Harmonised Templates" for metabolism experiments, similar to the domain type "Literature" which can be included as an input frame for each "Data Source" section.
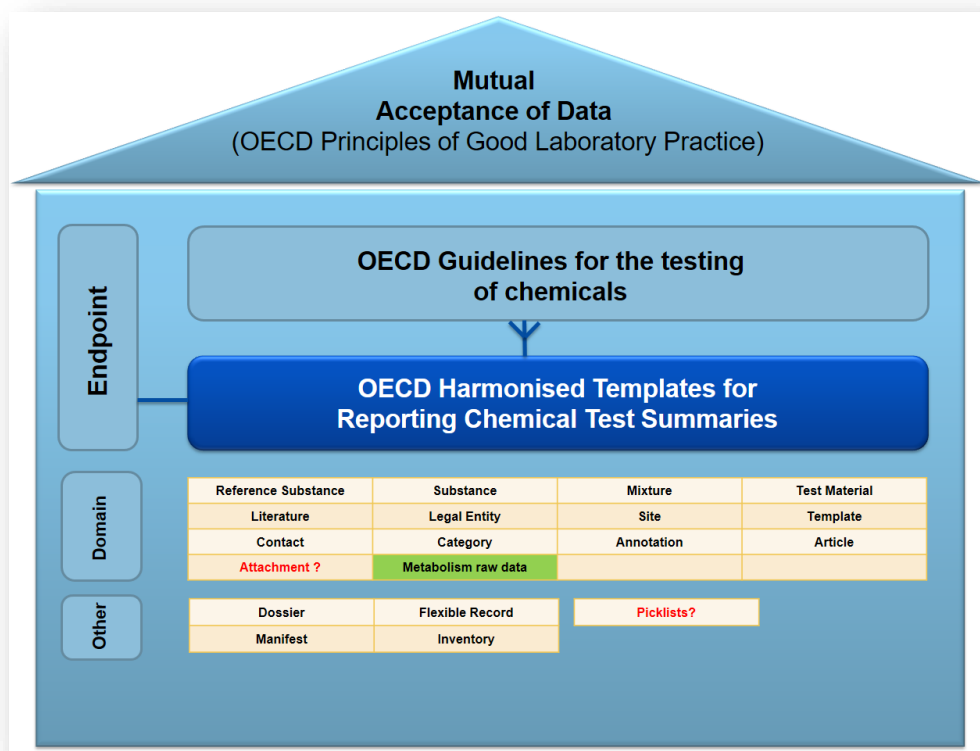
Figure 20: Expanding the OECD data architecture with a new domain type
"Metabolism raw data"

6.6.3.2   Principles

The main principles for this solution are:

1. This is a generic approach. All users interfaces for metabolism studies could refer to the new "*OECD Domain Type*" if "*Aggregated Raw Data*" should be submitted.

2. Such a generic approach should only be done for OHTs where some data consumers are waiting for these "*Aggregated Raw Data*". Somebody has to decide which OHTs should be modified. Otherwise, a data grave will be organized!

3. Manual data input is not appropriate for "*Aggregated Raw Data*". The way to input the data is the most important and crucial step (compare Figure 21).

4. The user, who is reading the "*Applicants Study Summary*" can switch to the referred object "*OECD Domain Type*". It would be possible to read raw databut has to be whether this is requested or not.

5. There are parallel consumption processes. The publication process is based only on the human readable data. All "*Aggregated Raw Data*" should be excluded. The 2nd process is the evaluation process which is the intended consumer of the raw data.

6. Each object which is referenced in the "*OECD Domain Type*" will be exported into an IUCLID-6 file (XML). The requested information on referenced substances will be included in a set of additional XML files.
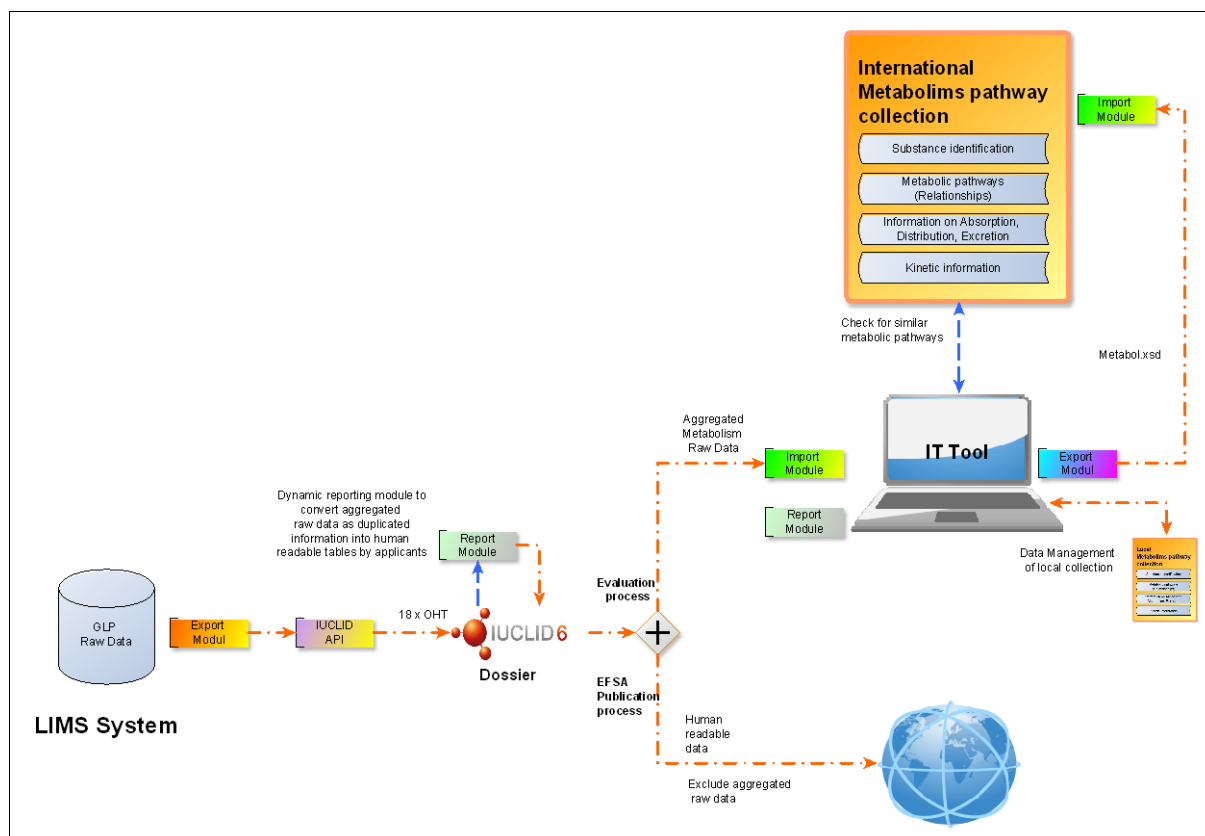
Figure 21: Needed modules if using the new domain type "Metabolism raw data"

Table 7: Comparison of pro and contra for using the domain type to transport aggregated raw data

Module / aspect	Pro "Domain type"	Contra "Domain type"
Export module for each LIMS "**GLP**" System	Programmers can use the existing IUCLID "*API*" to create the needed IUCLID records. One export module should be able to assist all OHTs where the new domain type "Metabolism raw data" was included because of the possibility of using radiolabelled test substances.	All LIMS "**GLP"** Systems have to structure the maintenance of the references to the IUCLID Reference Substances model. The origin problem of one metabolism study will grow up to a global problem of managing an IUCLID substance backbone for all metabolism studies.
Flexibility per OHT and legal act	-	**For all relevant OHTs this domain type will be included**. On level of a legal act it should be defined if such a block of "Aggregated Raw Data" is mandatory or not.
OHT generated data input form	All needed data input forms are generated by IUCLID in an automatic way. No additional effort is needed.	No human is able to fill out these aggregated raw data. There is no need to visualize these aggregated raw data.
Report module	Existing report generator could be used. For different purposes different report templates should be developed.	Report module should assist **dynamic reporting templates to create pivot tables.** There would be an for IUCLID: "Should be IUCLID and the reporting module the mechanism to create the needed texts blocks /tables to include them into the applicant's summary?" If yes, the time points are hurdles in the process!
Import module - part: extracting the raw data	Programmers can use the existing IUCLID "*API*" to create the needed exporting module.	All users who are interested in using the submitted aggregated raw data for own recalculations / reports have to

		have an extracting module.
		The effort of programming such an extracting module is high. A solution would be that IUCLID offers such an extracting module.
Import module - part: import into the consuming IT-Tool	-	One module which could import the extracted raw data into a metabolism information system is needed.
		This import module is also needed for data exchange between different collections of metabolism information systems
Export module	-	This export module is needed for data exchange between different collections of metabolism information systems
Governance of the new object type	The new domain type is based on the OHT organizational procedures. No change of the Governance model is needed.	-
Third party engagement	The new elements could be assisted by the provided IUCLID "*API*"	The usage of the "*API*" for aggregated raw data is a hurdle because of the IUCLID substance model.
Interoperability	The new domain type is based on IUCLID. The creation of a new domain type could be done by a regular IUCLID update.	This approach makes the "*MetabolAS Tool*" completely dependent on IUCLID. Interoperability is to be welcomed; a 100% dependency limits the development capability of the "*MetabolAS Tool*".
		To improve LIMS to be compatible with IUCLID is a complex project.
		The metabolism raw data information is not self-contained. Each import system has to analyse and resolve all references in the Dossier.
Time point of creation	-	It is impossible to guarantee that the "*GLP Study Report*" and the prepared OHT have the same lifetime.
		All needed IUCLID references should be known at time point of creation of this object.
		What happens with metabolites which are unknown at time of the **"*GLP*"** Report but analysed afterwards?
		What happens with metabolite (reference substance) IUCLID records when it is later realised that e.g. "Unknown 1" and "Unknown 20" are actually identical?
Generalisation per knowledge sector	All Metabolism study OHTs could use this additional domain type to include aggregated raw data.	-
Relationship to aggregated raw data of other knowledge sectors	-	-
User interface for OHT where the domain type should be used	-	User interfaces for aggregated raw data makes no sense. Manual data input is not possible. Users does not want to see these data.
Relationship to reports and the transparency regulation	Human readable data and additional aggregated raw data where published by the same procedure.	It is much more difficult to generate human readable reports from the additional aggregated raw data.
Simplicity level	-	The development of needed third party data interfaces bind relevant resources.
Reduction of duplicated information	-	The content of the aggregated raw data, which are provided in the Rich-Text elements has a high degree of

		overlapping of information on the semantic level.
	The same information has two different formats: the human readable and the machine-readable format for different purposes.	-

If this solution will be put into practice, this would be a generic approach to transport "*Aggregated Raw Data*" embedded in the OHT concept.

7 Appendix

7.1 Bibliography

<mark>To-do</mark>

7.2 Abbreviations

Short	Meaning
AD	**A**dministered **D**ose
ADME	**A**bsorption, **D**istribution, **M**etabolism and **E**xcretion
API	**A**pplication **P**rogrammable **I**nterface
BfR	**B**undesinstitut **f**ür **R**isikobewertung
CR	**C**urrent high level user **R**equirement
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
EFSA	**E**uropean **F**ood **S**afety **A**uthority
EPA	United States **E**nvironmental **P**rotection **A**gency
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
JDBC	**J**ava **D**ata**B**ase **C**onnectivity
MetaPath	Software and knowledge base for the purpose of archiving, sharing and analysing experimental data on metabolism and metabolic pathways; Copyright by OASIS LMC
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
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**armonized **T**emplates
QA	**Q**uality **A**ssurance
(Q)SAR	(**Q**uantitative) **S**tructure **A**ctivity **R**elationship
R	User **R**equirement
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

7.3 List of Harmonized Templates where radioactive labelled test material could be used

OHT Group	OHT No	ENDPOINT_STUDY_REC-ORD name	R [1]	L-TP [2]	RD [1]	Test Guide-line	Test Guideline Name	Definition terms regarding metabolism	Labelled test substances
Environmental fate & behaviour	OHT 24	PhototransformationInAir	-	+	D	none	-	-	-
Environmental fate & behaviour	OHT 25	Hydrolysis	-	+	-	TG111	Hydrolysis as a Function of pH	Transformation products; Hydrolysis products	Yes
Environmental fate & behaviour	OHT 26	Phototransformation	-	+	D	TG316	Phototransformation of Chemicals in Water – Direct Photolysis	Transformation (biodegradation, mineralization) and parameter of the transformation process	Yes
Environmental fate & behaviour	OHT 27	PhotoTransformationInSoil	-	+	D	none	-	-	-
Environmental fate & behaviour	OHT 28	BiodegradationInWater-ScreeningTests	-	-	D	TG301	Ready Biodegradability	No separate definition block but biodegradation is explicitly named	Yes
						TG302A	Inherent Biodegradability: Modified SCAS Test	No separate definition block but biodegradation is explicitly named	Yes
						TG302B	Inherent Biodegradability: Zahn-Wellens/ EVPA Test	No separate definition block but biodegradation is explicitly named	No
						TG302C,	Inherent Biodegradability: Modified MITI Test (II)	No separate definition block but biodegradation is explicitly named	No
						TG306	Biodegradability in Seawater	No separate definition block but biodegradation 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 biodegradation is explicitly named	No
Environmental fate & behaviour	OHT 29	BiodegradationInWater-AndSedimentSimulationTests	-	+	D	TG303A TG303B	Simulation Test - Aerobic Sewage Treatment -- A: Activated Sludge Units; B: Biofilms	No separate definition block but biodegradation is explicitly named	Yes
						TG308	Aerobic and Anaerobic Transformation in Aquatic Sediment Systems	Transformation products and parameter of the transformation process	Yes

							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 Biodegradability of Chemicals Discharged in Wastewater](#)	No separate definition block but degradation products are explicitly named	Yes
Environmental fate & behaviour	OHT 30	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 parameter of the transformation process	Yes
Environmental fate & behaviour	OHT 32	BioaccumulationAquaticSediment	-	-	-		TG305,	[Bioaccumulation in Fish: Aqueous and Dietary Exposure](#)	Bioaccumulation, Bioconcentration, Biomagnification	Yes
							TG315	[Bioaccumulation in Sediment-dwelling Benthic Oligochaetes](#)	Bioaccumulation, Bioconcentration, Biomagnification	Yes
Environmental fate & behaviour	OHT 33	BioaccumulationTerrestrial	-	-	-		TG317	[Bioaccumulation in Terrestrial Oligochaetes](#)	Bioaccumulation, Bioconcentration, Biomagnification	Yes
Environmental fate & behaviour	OHT 34	AdsorptionDesorption	-	+	-		TG106,	[Adsorption - Desorption Using a Batch Equilibrium Method](#)	Looks primarily at physical phenomena, but points to a possible transformation	Yes
							TG121	[Estimation of the Adsorption Coefficient (Koc) on Soil and on Sewage Sludge using High Performance Liquid Chromatography (HPLC)](#)	Looks only at physical phenomena	Yes
Effects on biotic systems	OHT 56	BiotransformationAndKinetics	-	+	-		none	-	-	-
Health effects	OHT 58	BasicToxicokinetics	-	+	-		TG417	[Toxicokinetics](#)	Biotransformation, Metabolism	Yes
							TG319A, TG319B	[Determination of in vitro intrinsic clearance …](#)	No separate definition block but biotransformation and bioaccumulation is explicitly named	No
Health effects	OHT 59	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	OHT 85-1	MigrationOfResidues	-	+	-		none	-	-	-
Pesticide residue chemistry	OHT 85-2	MetabolismInLivestock	+	+	+		TG503	[Metabolism in Livestock](#)	No separate definition block but metabolism is explicitly named	Yes

			R	L-TP	RD				
Pesticide residue chemistry	OHT 85-3	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	OHT 85-8	NatureResiduesInProcessed-Commod	-	+	+	TG507	Nature of the Pesticide Residues in Processed Commodities - High Temperature Hydrolysis	No separate definition block but metabolism is explicitly named	Yes
Pesticide residue chemistry	OHT 85-10	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)

7.4 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"[30]

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 of improvements needed.	IT-Tool
S 3.3.2-3	It seems that the current flow of information is connected with a relevant amount of duplication of work.	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 QSAR models are of sufficient quality for predicting metabolism pathways" suggests that the QSAR-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
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,

---

[30] https://www.bfr.bund.de/en/analysis_of_the_information_flow_in_metabolism_studies_on_pesticides-272198.html

		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

7.5   List of further weak points of the DER/MSS-Composer Family and MetaPath

To-do

Internal number	Weak point	Group
S 7.4-1	OASIS LMC supports "custom versions" of MetaPath also. Therefore it doesn't exist one version of the tool MetaPath.	
S 7.4-2	No system documentation with the used data model of MetaPath is publicly available.	
S 7.4-3	MetaPath works with Firebird as the data management system. Firebird presents a major hurdle for a direct data access via JDBC or ODBC driver.	
S 7.4-4	The DER/MSS-Composer Family and MetaPath assist that one metabolic pathway could have different bibliographic citations. This is in contrast to the OECD understanding of a "Study summary".	
S 7.4-5	The user can edit many data fields in the DER/MSS-Composers which are not imported into MetaPath e.g. the physicochemical properties. Therefore, a "loss of data" is expected. No documentation exists on which data fields will be included into MetaPath.	
S 7.4-6	MetaPath has input functions which are already implemented in the DER/MSS-Composer Family. It is not clear if the same rules for validation are applied/integrated.	
S 7.4-7	No separate validation function is implemented in MetaPath to check if the XML file is according to the corresponding MSS-Composers schema before import.	
S 7.4-8	Via non-publicly documented functions, a data transfer to the QSAR toolbox is possible. No documentation exist which MetaPath data will be included into the OECD QSAR Toolbox	
S 7.4-9	The MetaPath database uses duplicate data structure according to the DER Composer and the MSS Composer. No attempt was made to implement a uniform, expandable concept. This presents a technological hurdle for the integration of other types of metabolism experiments in future. However, this also means that in the actual version all interfaces and reports have to be provided twice.	
S 7.4-10	DER and MSS Composer aren't optimized for a manual data input of tables of raw data.	
S 7.4-11	The MSS-Composer Family and "MetaPath" are storing analysed values in a cell of a complex table structure.	
	The predefined tables of the MSS-Composer Family are not suitable for "freestyle" studies not according the newest OECD TG. But the	

amount of such studies is relevant (approximately > 1000 studies)	
The predefined tables of the MSS-Composer Family are limited to 7 columns per radiolabel. This is not enough and a reason for improvisations.	
Limitation 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).	
The export/ import of complex search queries is not possible.	
All initialisation phases are too slowly (for 715 maps): • To open Metapath needs 90 seconds • To open a search for transformations, chemicals, similarity needs 30 seconds • To open a search for table needs 60 seconds	
The initialisation phase for similarity search is too long (	
Metapath assists no common short keys like <Ctl><v> for paste the content of the clipboard into the Metapath field	
Metapath assists no function to copy messages into the clipboard e.g. the textual representation of a query like Q1: Chemical name containing "Meco"	
The Metabolic pathway frame to compose a search query is too difficult. The message "The search clause is not correctly complete!" is not helpful.	
Metapath: No integrated documentation	
Metapath: Additional non-guideline experiments hardly possible to code (e.g. stem injection, cell cultures); can principally be coded by free-text fields (but: character restrictions!)	
Metapath. tentative results difficult to handle	

7.6 Proposal for the schema Metabol.xsd

To-do: If it is desired at this stage of the project

7.7 Standard tables for the presentation of metabolism studies

Table 8: <Generic title depending on the filter rules of the "List of analysed Values">.

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					

Table 9:   Total radioactive residues (TRR) in excretion product (time interval _____) or in organs after ____ hours of application.

Dose group / and or ID for animal	DG1	DG2	…
Excretion product 1			
Excretion product 2	TRR per excretion product or organ		
Excretion product …			
Organ 1			
Organ 2			
Organ …			
Sum of all TRR			

To-do: Inclusion of further standard tables as pivot tables