Guidance Document for Consistent Reporting of 'Omics Data From Various Sources

-Transcriptomics Reporting Framework (TRF)
-Metabolomics Reporting Framework (MRF)

Joshua Harrill (US EPA) and Mark Viant (Univ. Birmingham) on behalf of TRF & MRF

Cosmetics Europe Toxicogenomics Meeting
October 30th 2020

OECD
BETTER POLICIES FOR BETTER LIVES

Disclaimer

- *The views expressed in this presentation are those of the author(s) and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or products represent endorsement for use.*

EAGMST Omics Reporting Frameworks

To develop frameworks for the standardisation of reporting of 'omics data generation and analysis, to ensure that all of the information required to understand, interpret and reproduce an 'omics experiment and its results are available.

Purpose: to ensure that sufficient information is available to enable an evaluation of the quality of the experimental data and interpretation, and support reproducibility.

NOT to stipulate the methods of data analysis or interpretation….**Rather**, provide guidance on reporting of information that fosters transparency and reproducibility.

Project Name	Project Leads
Metabolomics Reporting Framework (MRF)	Mark Viant (Univ. Birmingham, UK)
Transcriptomics Reporting Framework (TRF)	Joshua Harrill (USEPA) Carole Yauk (Health Canada)
Optimal Data Analysis Framework (ODAF)	Tim Gant (PHE, UK) Florian Caiment (Univ. Maastricht)
OECD Secretariat	Magda Sachana

TRF

TRF Document, Major Topic Areas

TOXICOLOGY EXPERIMENT MODULE:

- The experiment should be described in sufficient detail that would allow another researcher to replicate the experiment.
- Adapted from existing sources
- Information in this section is independent of ‘omics platform

PROCESSING AND ANALYSIS OF ‘OMICS DATA MODULES:

- The transcriptomics technology, sample processing procedures, methods used to collect raw data and methods used to generate processed data.
- Described in Gant et al. (2017).
- Information in this section is dependent on ‘omics platform

DOWNSTREAM ANALYSIS REPORTING MODULES [DA-RMs]

- Detail the steps and resources necessary to reproduce a computational analysis of the processed data.

TRF: Original Structure

Toxicology Experiment Reporting Module

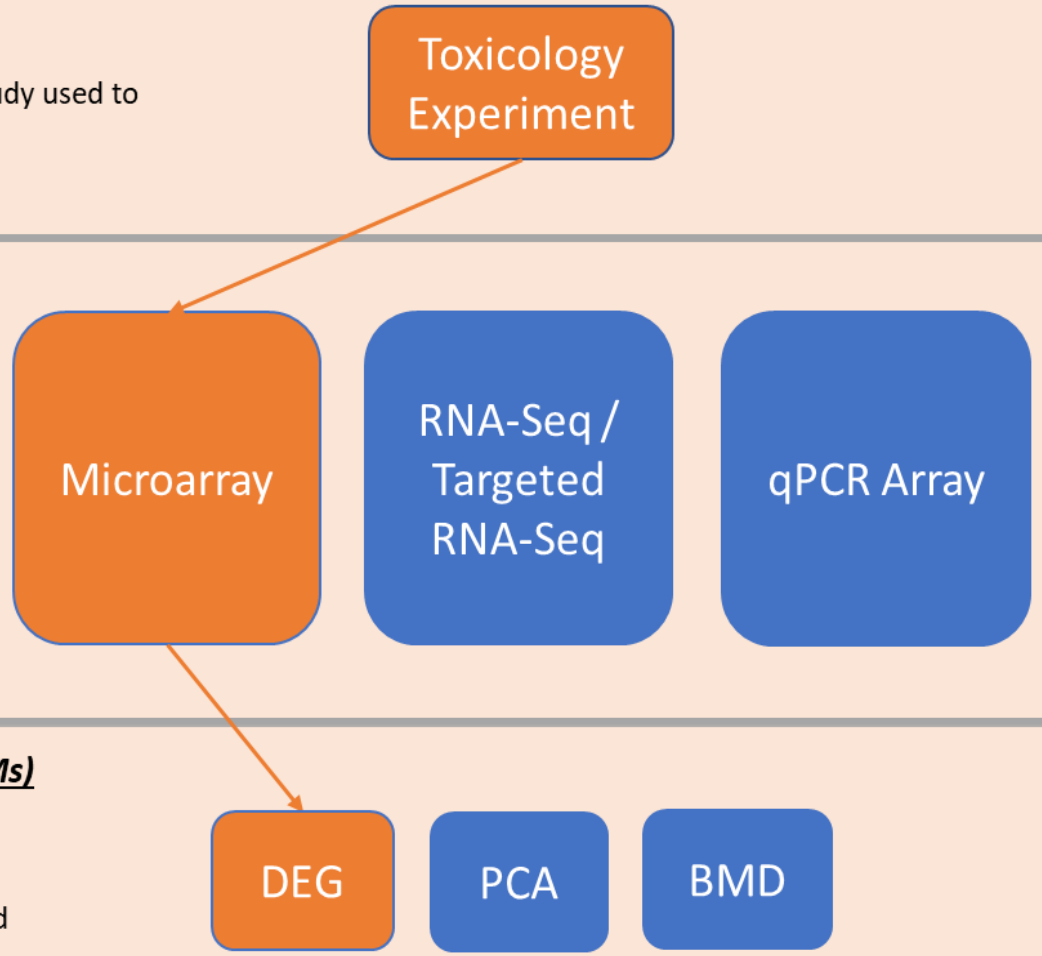
- Details the specifics of a laboratory-based toxicology study used to generate biological samples for transcriptomic analyses.

Technology-Specific Reporting Modules

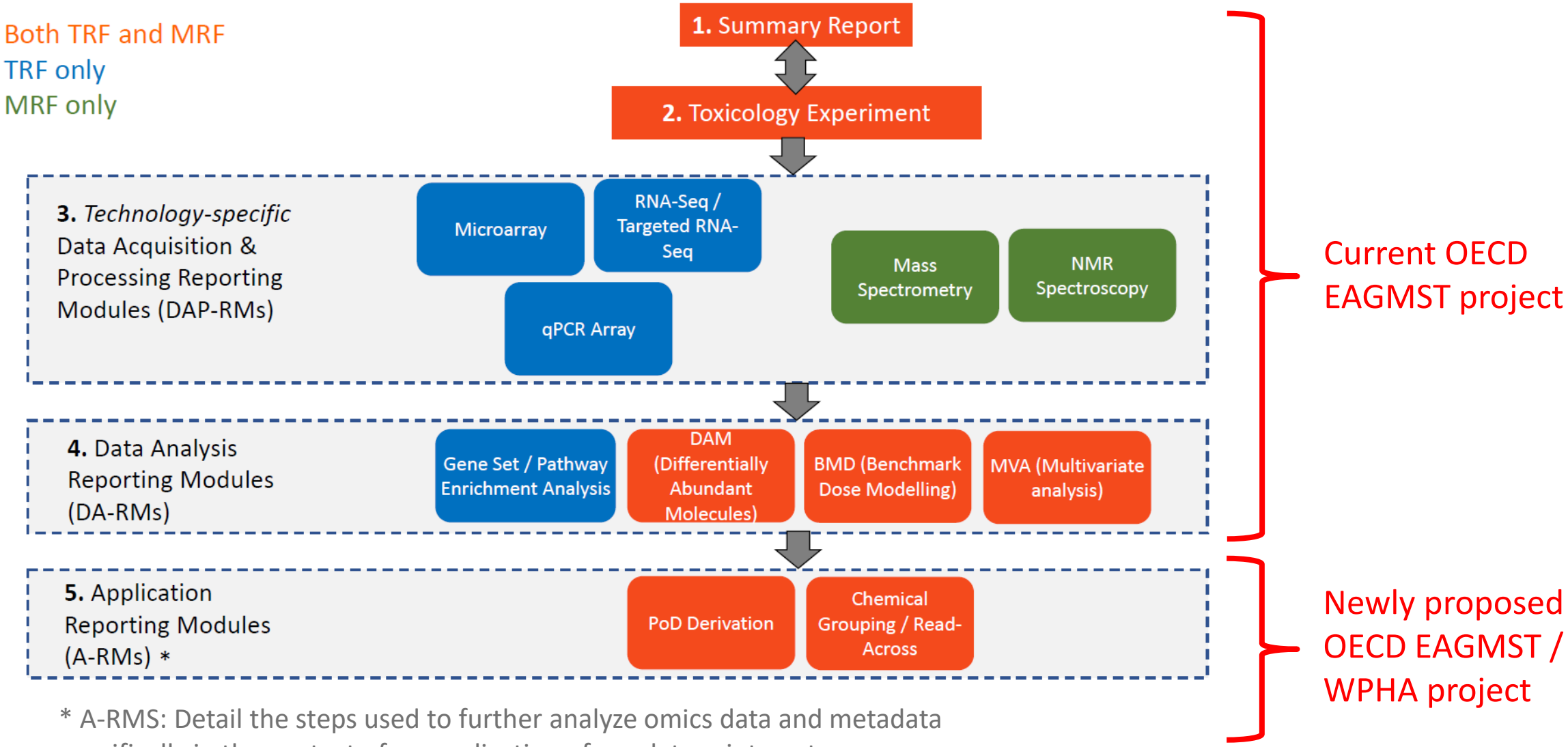
- Details technology-specific aspects of data generation and initial processing of transcriptomic data, including:
  - Transcriptomic Technology Description
  - Transcriptomics Experimental Design
  - Specification of Raw Data
  - Data Normalization
  - Data Filtering
  - Data QC

Downstream Analysis Reporting Modules (DARMs)

- Details the steps and resources necessary to reproduce computational analyses of transcriptomic data.
- Intended to be coupled with upstream experimental and technology-specific reporting modules.



Modular Structure of Omics Reporting Frameworks Harmonization of TRF and MRF



* A-RMS: Detail the steps used to further analyze omics data and metadata specifically in the context of an application of regulatory interest.

TRF Format

Table of Contents

- **Stylistic alignment:**
  - Previous OECD guidance in the biological sciences (where applicable)
  - MERIT Project / Metabolomics Reporting Framework (MRF) – *In Progress*
- **Reporting Format**
  - Narrative text followed by Reporting Fields
  - Excel spreadsheet for reporting
- **Consistent vocabulary across modules**

Reporting Structure

	A	B	C	D
1	**RELEVANT MODULE**	**REPORTING CATEGORY**	**REPORTING ELEMENT**	**DESCRIPTION OF ELEMENT**
2				
3	**Summary Report**	**Study identifier**	Unique study identifier	Unique code
4				
5		**Study rationale**	Objective of study	Controlled vocabulary: point-of-departure; hazard
6			Background information (supporting the objective)	Free text elaboration of above
7				
8		**Links to related study records**	Standardised toxicology dataset (e.g., linked to OECD	e.g. IUCLID ESR
9			Omics complete dataset (e.g., linked to MetaboLights,	e.g. MTBLSxx
10				
11	**Toxicology Experiment Module**	**Test item (chemical) and vehicle**	Test item name	
12			Test item identifier	SMILES or InChIKey
13			Vehicle name	
14				
15		**Test system characteristics**		
16		*EITHER in vivo*	a. Species	
17			b. Strain	
18			c. Sex	
19			d. Age	
20		*OR in vitro*	a. Cell type	
21			b. Species of origin	
22				
23		**Exposure conditions**	Test item concentration(s)	
24			Route of administration	
25			Schedule (single dose or repeated dosing)	
26			Frequency of repeated dosing	
27			Exposure duration(s)	
28				
29		**Biological samples**	Type (e.g., in vitro: cells, media, etc.; in vivo: cells, tissue, biofluid, whole organism, etc.)	
30			Sample preservation method (e.g., fresh, frozen, paraffin-embedded, etc.)	
31			Number of biological replicates per treatment	
32				
33	**TRF platform-specific data acquistion and processing Reporting Module**	**Sample preparation**	Sample preparation method (e.g., RNA extraction method, cell lysis, etc.)	
34				
35		**Technology**	Type (e.g., data acquisition and processing module used - DNA microarray, RNA-seq, etc.)	
36			Manufacturer(s) and model(s) (e.g., Agilent microarray, etc.)	

Study Summary revised | Experiment | Processing microarray data | DARM DEG

- **Executive summary included**
- **Modular**
  - **Flexible/expandable**
  - Add new modules as new platforms or analytical approaches released
- **Mandatory/optional reporting fields**
- **Database compatibility**
- **Summary, Experiment and DAM DARM harmonized with MRF**

TRF Module Status (Oct 2010)

Module Name	Module Development Lead	Status
Introduction	Joshua Harrill (US EPA) Carole Yauk (U Ottawa)	Complete
Study Summary	Carole Yauk (U Ottawa)	Complete
Toxicology Experiment Module (TEM)	Raffaella Corvi (JRC)	Complete
Technology Specific Data Acquisition and Processing Reporting Modules (DAP-RM)		
Microarray	Vikrant Vijay (NCTR)	Complete
RNA-Seq / Targeted RNA-Seq	Florian Caiment (U Maastricht)	Complete
qPCR Array	Jason O'Brien (ECCC)	In Process
Data Analysis Reporting Modules (DA-RM)		
Differentially Abundant Molecules (DAM)	Lyle Burgoon (ERDC)	Complete
Benchmark Dose Modeling (BMD)	Scott Auerbach (NIH DNTP)	In Process
Gene Set / Pathway Enrichment	TBD	Pending

TRF Case Studies

Submitter:

1) Identify dataset
2) Compute DEGs
3) Fill in TRF
   a) Technology specific DAP-RM
   b) DAM DA-RM
4) Ease-of-use commentary

At the discretion of the Submitter:

- Transcriptomics technology
- Specification of meaningful contrasts
- Computing environment
- Method for DEG determination

Referee:

Provide to End User:
1) Completed TRF
2) Accessory files
3) Blinded / truncated data output

Review the contents of the submission and generate truncated data outputs with blinded gene IDs to pass to End User.

End User:

1) Review completed TRF
2) Reproduce DEG analysis
3) Ease-of-use commentary

End User communicates with referee if difficulties are encountered reproducing the analysis.

Referee:

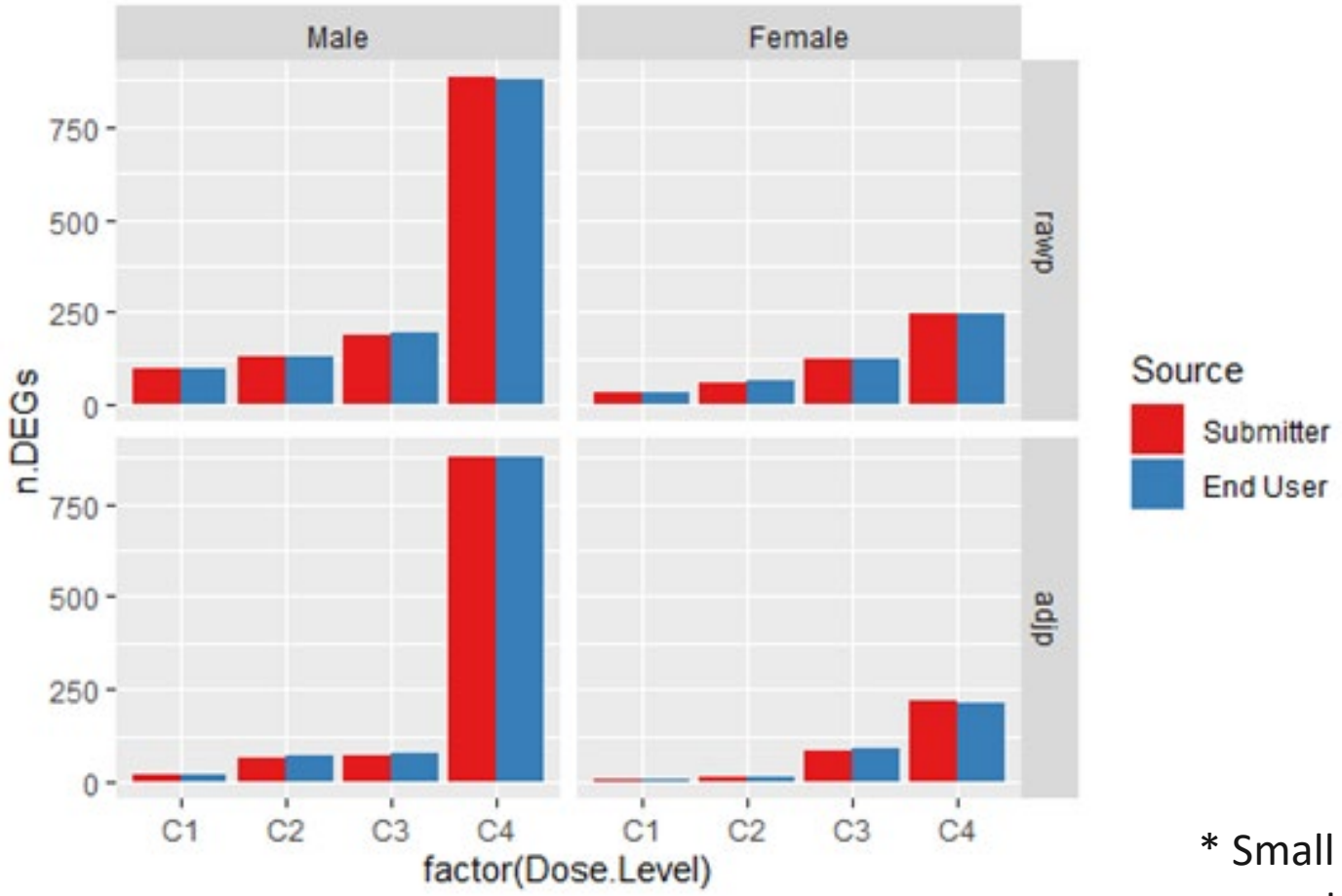
1) Concordance analysis
2) Summarize ease-of-use commentaries

TRF Case Study Descriptions

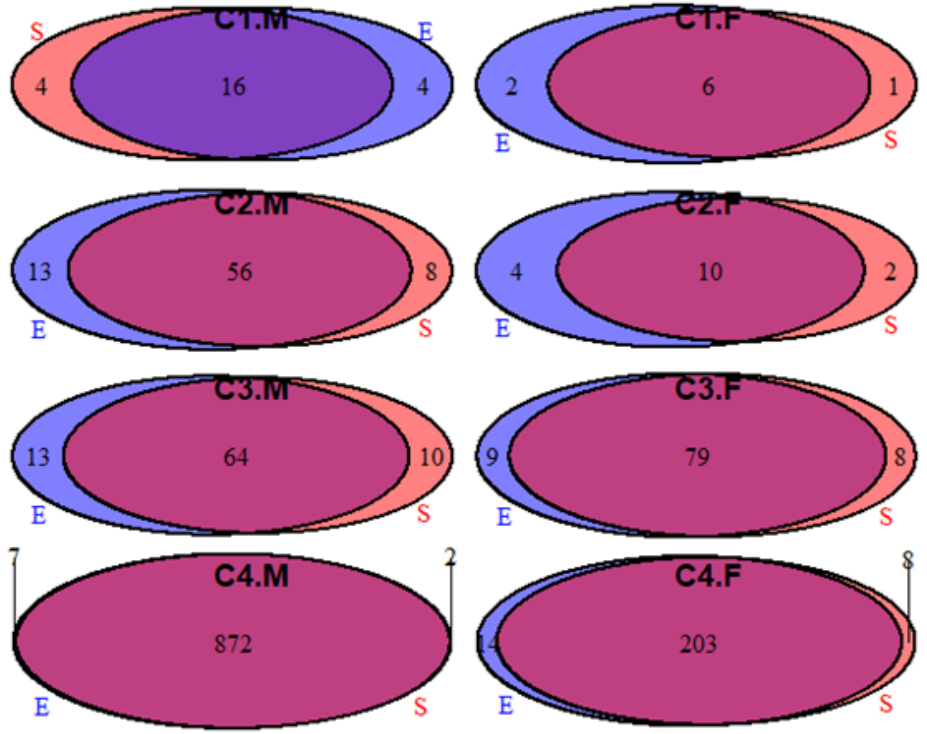
Platform	Study Description	DAM Method	Computing Environment	Submitter	End User	Status
Agilent Microarray	Four point concentration-response of furan in male and female Fisher rat liver (GEO GSE62805)	Submitter's Choice	R	Andrew Williams (Health Canada)	Leah Wehmas (US EPA)	Complete
Affymetrix Microarray	Comparison of PFOA responses in livers of 129S1/SvlmJ wild-type and PPAR-alpha null mice (GEO GSE9786)	Submitter's Choice	Partek Flow	Beena Vallanat (US EPA)	Alison Harrill (HHS DNTP	In Process
RNA-Seq	Three point concentration-response of hexabromocyclododecane in male and female Fisher rat liver (PRJNA395549)	ODAF	R	Matt Meier (Health Canada)	Brian Chorley (US EPA)	In Process
RNA-Seq	TBD	Submitter's Choice	R	Natalia Garcia-Reyero (MS State IGBB)	Andrew Williams (Health Canada)	In Process

Case Study #1 Results

Number of Differentially Expressed Genes



Overlap of Differentially Expressed Genes



Comparison of DEGs identified by Submitter and End User (adjp)

* Small differences are expected because the analysis applied permutation-based *p*-value calculations

Case Study Lessons Learned

General findings relating to ability to reproduce analyses:

1. Analyses in **open source computing environments** (R, Python, etc.)
a) Much easier for an end user knowledgeable in coding languages to reproduce because they come with an "instruction manual" (i.e. the analysis script or notebook)
b) Details in the reporting fields become somewhat less critical for reproducing the analysis secondary to the scripts
c) There are also no financial or licensing barriers with regards to accessing the tools
d) Issue: users may not have sufficient expertise with open source computing environments

2. Analyses using **freeware analysis softwares or web applications** (BMDExpress; iDep; kallisto)
a) These types of software are more user friendly and require less technical or statistical expertise to use
b) No "pay wall" barrier that would prevent an end user from accessing such tools.
c) Reproducibility depends on clear and precise reporting in the TRF documentation as well as provision of a configuration file or some other configuration snapshot that the end user could follow. ***NEEDS TO BE TESTED***

3. Analyses using **proprietary software** (Partek, Ingenuity, etc.):
a) End user needs access to the same software (and maybe even version)
b) "pay wall" issues.
c) Reproducibility depends on precise reporting in the TRF documentation as well as provision of a configuration or workflow that the end user could follow.

Case Studies Next Steps

Additional case studies:

- More developer and user feedback and participation.

- Testing different analytical platforms:
  - Open computing environments versus
  - Web applications

- “Test Driving” reporting modules in development
  - qPCR case studies
  - BMD case studies

MRF

OECD MRF Expert Group

Title	Identity	Roles
Expert Group	Tripartite Industry Government / Regulator Academic ca. 15 very active members, ca. 10 'observers'	-Contribute expert knowledge wherever possible across the whole MRF guidance document -Ensure consistency of whole document
Facilitator	Mark Viant [Univ. Birmingham]	-Foster discussion -Monitor progress in accordance with project timeline -Ensure consistency with TRF
Administrator	David Epps [Univ. Birmingham]	-Meeting organisation
OECD Secretariat	Magda Sachana	-Project administration / OECD liaison

MRF builds on the Ecetoc MERIT project

MEtabolomics standaRds Initiative in Toxicology

nature COMMUNICATIONS

Published
July 2019

PERSPECTIVE

https://doi.org/10.1038/s41467-019-10900-y OPEN

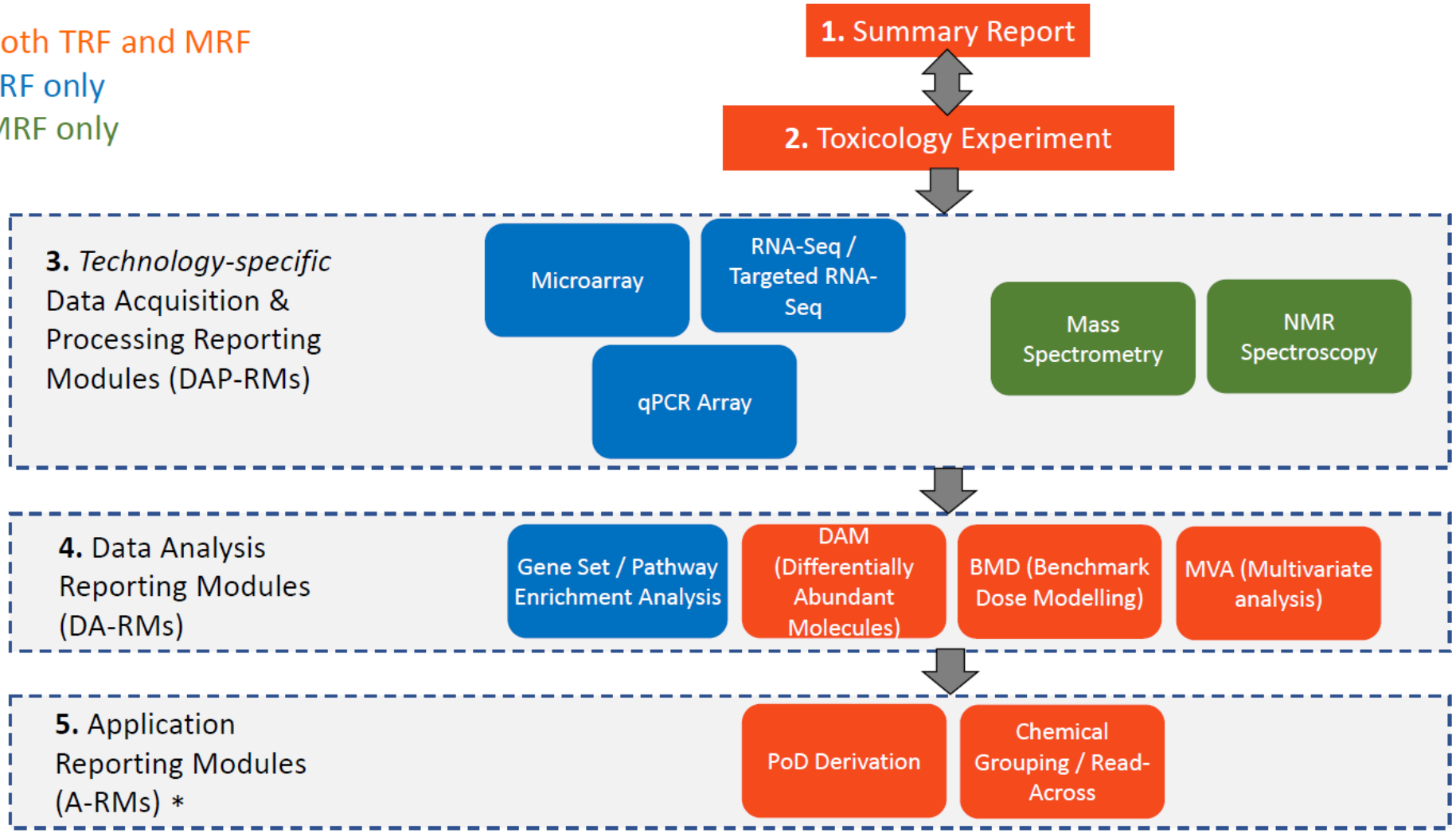
Use cases, best practice and reporting standards for metabolomics in regulatory toxicology

Mark R. Viant[1,12], Timothy M.D. Ebbels[2,12], Richard D. Beger[3], Drew R. Ekman[4], David J.T. Epps[1], Hennicke Kamp[5], Pim E.G. Leonards[6], George D. Loizou[7], James I. MacRae[8], Bennard van Ravenzwaay[5], Philippe Rocca-Serra[9], Reza M. Salek[10], Tilmann Walk[11] & Ralf J.M. Weber[1]

https://www.nature.com/articles/s41467-019-10900-y

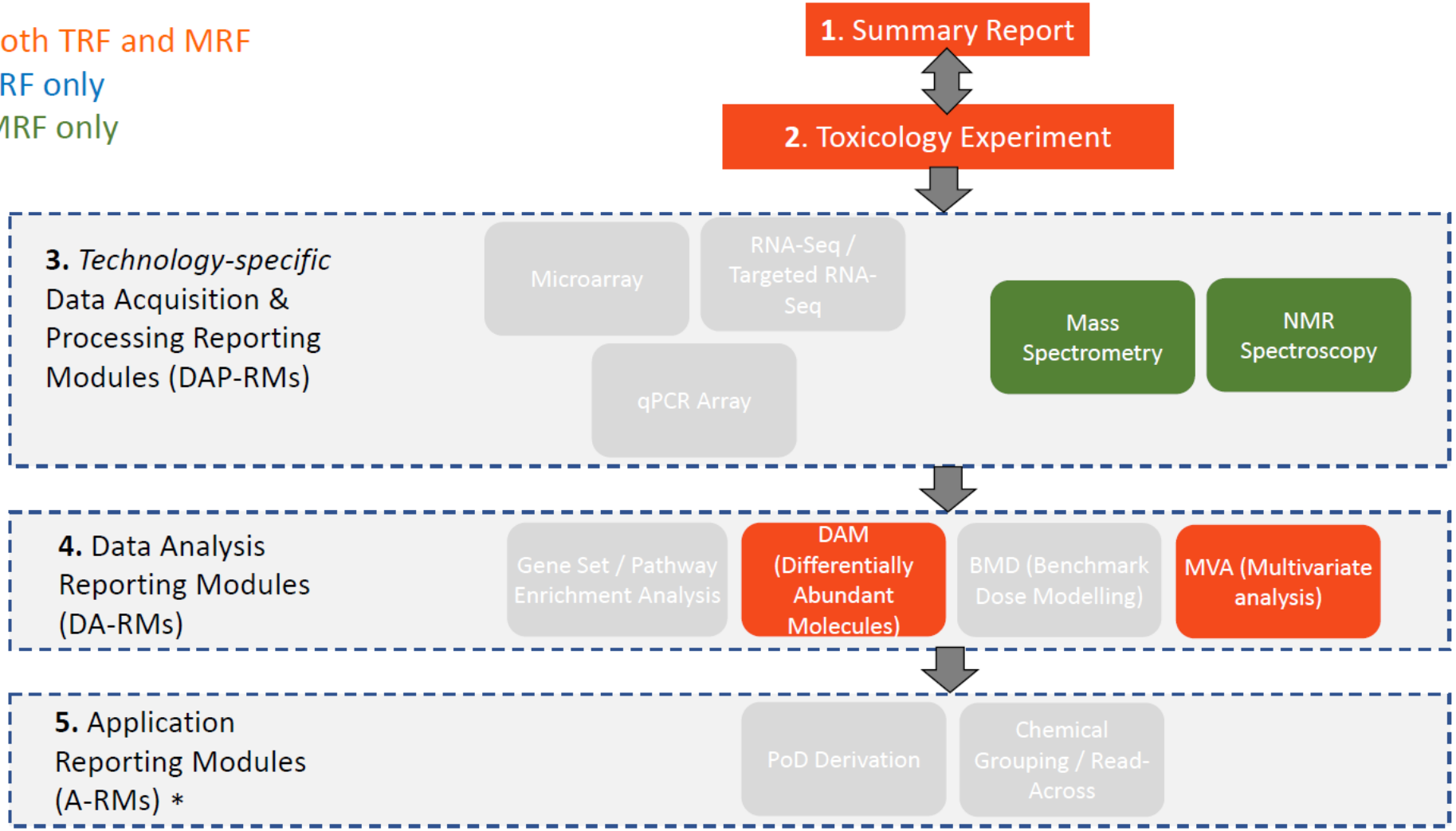
Modular Structure of Omics Reporting Frameworks

Both TRF and MRF
TRF only
MRF only



Modular Structure of Omics Reporting Frameworks

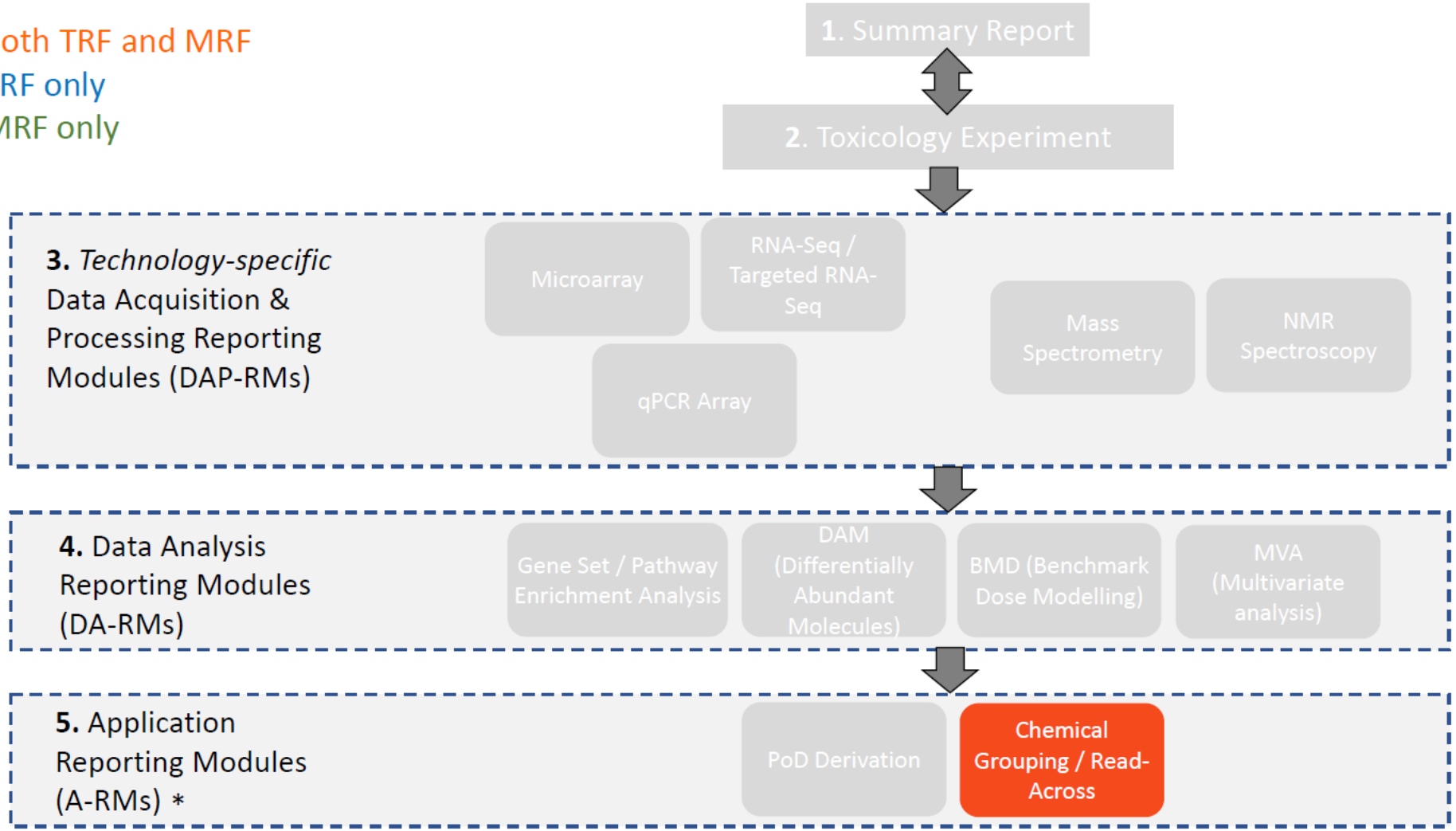
Both TRF and MRF
TRF only
MRF only



6 modules finished and trialing has begun

Modular Structure of Omics Reporting Frameworks

Both TRF and MRF
TRF only
MRF only



Grouping/RA module has been drafted, to be continued as part of new EAGMST / WPHA project

First MRF

Table of Contents

- **Reporting Format**
  - Narrative text followed by Reporting Fields
  - Excel spreadsheet for reporting
- **Consistent vocabulary across modules**
- **Database compatibility**
- **ca. 80 page document**

Noteworthy Progress

MASS SPECTROMETRY METABOLOMICS MODULE:

- Describes the acquisition and data processing for mass spectrometry based metabolomics studies.

- For the first time:
  - Integrated untargeted and targeted metabolite analysis into consistent reporting framework
  - Integrated a range of platforms (LC-MS, GC-MS, direct infusion MS) into one module
  - Developed consistent terminology describing relative quantification, semi-quantification and absolute metabolite quantification, and the description of reference standards
  - Defined new terminology for MS assay names including “hybrid” assays that combine targeted and untargeted analyses.

Noteworthy Progress

MASS SPECTROMETRY METABOLOMICS MODULE:

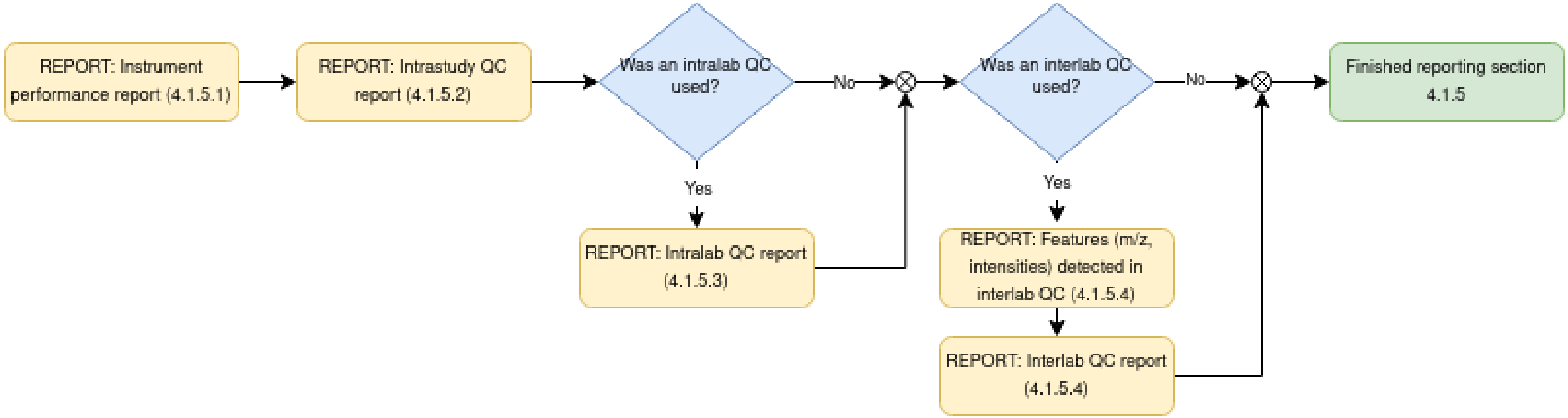


Figure 5: Workflow for mass spectrometry metabolomics reporting - Demonstration of quality of metabolomics analysis

Trialling the MRF - Case Studies

Data submitter:

1) Identify dataset
2) Process / analyse dataset
3) **Fill in MRF: SR, MS, NMR, DAM, MVA modules**
4) Write ease-of-use commentary

MRF referees:

1) Review MRF report for completeness
2) Remove statistical results and sends to End User

End User:

1) Use the partially completed MRF to **reproduce the data processing and stats**
2) Write ease-of-use commentary

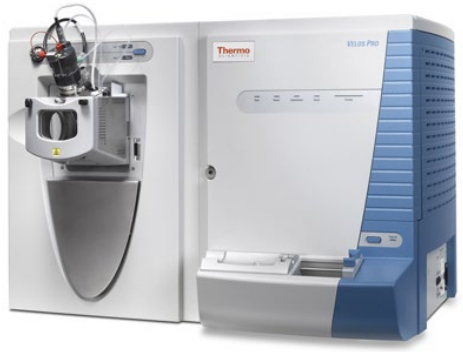
MRF referees:

1) Review MRF report for completeness
2) Concordance analysis of 2 completed MRF reports

MRF expert group:

1) Review concordance analysis and two ease-of-use commentaries
2) Update MRF accordingly

Trialling the MRF - Case Studies



Mass spectrometry metabolomics trial - Underway

- **Data submitter:** David Crizer (National Toxicology Program, US)
    - 5-day rodent assay, plasma samples, thujone exposure
- **MRF referees:** Oliver Schmitz (BASF, DE), Pim Leonards (VU University, NL), Aniko Kende (Syngenta, UK)
- **End user:** Tom Lawson (Michabo Health Science, UK)

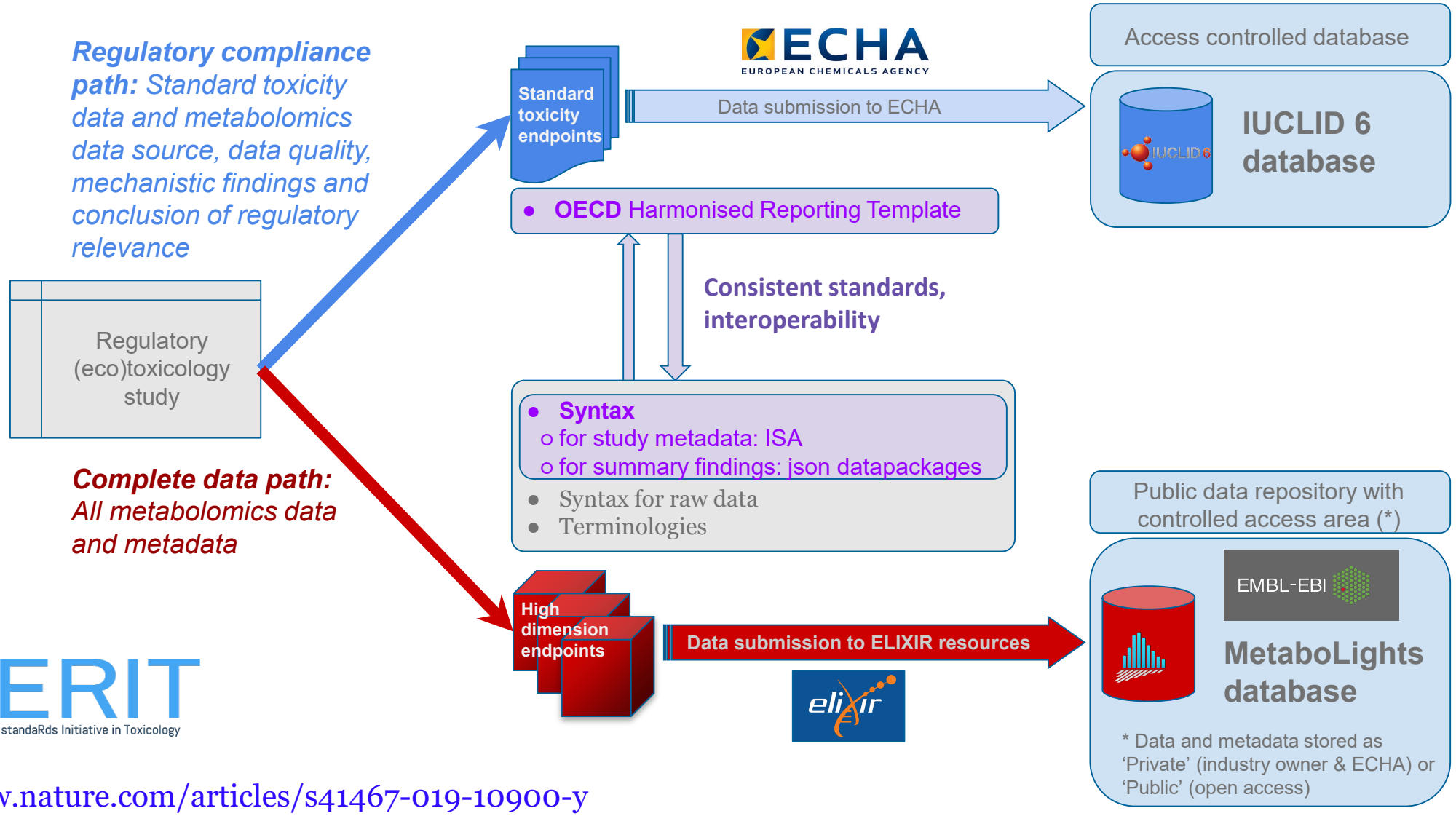
NMR spectroscopy metabolomics trial - Now starting

- **Data submitter:** Fabien Jourdan, Nicolas Cabaton, Cécile Canlet (INRA, FR)
    - Mouse study, brain tissue, bisphenol A exposure
- **MRF referees:** Drew Ekman (EPA, US), Mark Viant (University of Birmingham, UK)
- **End user:** Tracey Schock (NIST, US)

How and where to report (metabol)omics data from a regulatory toxicology study in Europe?

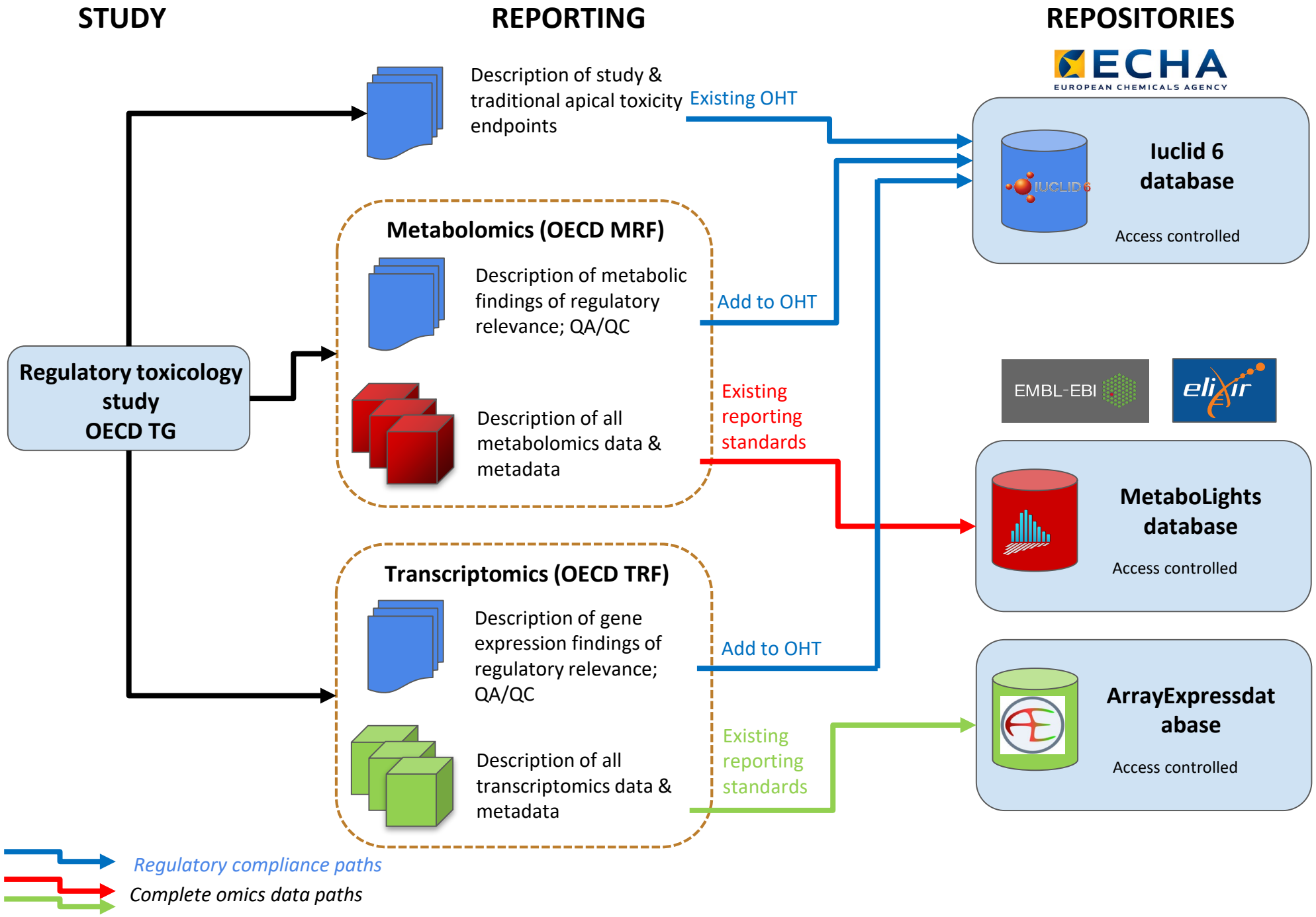
Chemical Safety Study & Assessment

Sponsor: Chemical company
Process: In-house or outsourced study encompassing several data acquisition modalities
Output: Dataset [Study Metadata Descriptions, Raw Data, Processed Data, Findings and Reports]



MERIT
MEtabolomics standaRds Initiative in Toxicology

Figure 5
https://www.nature.com/articles/s41467-019-10900-y


STUDY
REPORTING
REPOSITORIES
ECHA
EUROPEAN CHEMICALS AGENCY
Regulatory toxicology study OECD TG
Description of study & traditional apical toxicity endpoints
Existing OHT
Iuclid 6 database
Access controlled
Metabolomics (OECD MRF)
Description of metabolic findings of regulatory relevance; QA/QC
Add to OHT
Description of all metabolomics data & metadata
Existing reporting standards
EMBL-EBI
elixir
MetaboLights database
Access controlled
Transcriptomics (OECD TRF)
Description of gene expression findings of regulatory relevance; QA/QC
Add to OHT
Description of all transcriptomics data & metadata
Existing reporting standards
ArrayExpressdatabase
Access controlled
Regulatory compliance paths
Complete omics data paths

Michabo Health Science – ECHA project

Mapping to IUCLID and MetaboLights

REPORTING ELEMENT	MANDATORY / OPTIONAL	DESCRIPTION OF ELEMENT AND CONTROLLED VOCABULARY	REPORT TO SPECIALIST REPOSITORY? (termed 'Complete Data path', e.g. to EBI MetaboLights or NIH Metabolomics Workbench)	POTENTIAL ISA COMPONENT FOR METABOLIGHTS	REPORT TO REGULATOR? (termed 'Regulatory Compliance path', e.g. to IUCLID)	POTENTIAL COMPONENT FOR IUCLID
Extraction method general description	Mandatory	Free text	Y	Extraction protocol	N (included in Summary	
Solvent(s) used,	Mandatory	Free text	Y	Extraction protocol	N	
Means of agitation/maceration	Mandatory	Free text	Y	Extraction protocol	N	
Temperatures and times	Mandatory	Free text	Y	Extraction protocol	N	
Post extraction handling, e.g., storage	Mandatory	Free text	Y	Extraction protocol	N	
Derivatization method general description	Mandatory if used	Free text	Y	Extraction protocol	N	
Reagents and reaction (including incubation	Mandatory if used	Free text	Y	Extraction protocol	N	
Clean-up/partitioning (if used).	Mandatory if used	Free text	Y	Extraction protocol	N	
Evaporation and reconstitution method general	Mandatory	Free text	Y	Extraction protocol	N	
Final reconstitution solvent(s) and final volume (if	Mandatory if used	Free text	Y	Extraction protocol	N	
Storage temperature (if relevant)	Mandatory if used	Free text	Y	Extraction protocol	N	
Duration of reconstituted extracts (if relevant)	Mandatory if used	Free text	Y	Extraction protocol	N	
Quality assessment reference standard general	Mandatory if used	col	Y	Extraction protocol	N	
Report in Table A4.1(1): Quality assessment	Mandatory if used	Table	Y	Separate file	N	

MRF Project Timeline

- MRF draft (version 1): completed
- MRF trialling: on-going, deadline April 2021
- Revised MRF submitted to OECD EAGMST for formal review: June 2021
- Extension of TRF + MRF to include *Application Reporting Modules*: new OECD EAGMST-WPHA proposal asap

Further information

- TRF: Josh Harrill (Harrill.Joshua@epa.gov),
  Carole Yauk (Carole.Yauk@uottawa.ca)

- MRF: Mark Viant (M.Viant@bham.ac.uk)

- OECD: Magda Sachana (Magdalini.Sachana@oecd.org)