Overview of Concepts and Available Guidance related to Integrated Approaches to Testing and Assessment (IATA)

OECD

BETTER POLICIES FOR BETTER LIVES

Series on Testing and Assessment
No. 329

Overview of Concepts and Available Guidance related to Integrated Approaches to Testing and Assessment (IATA)

OECD

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 37 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways**. More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organisations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS
A cooperative agreement among **FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD**

Foreword

OECD member countries collaborate in developing and harmonising methods for assessing risk to human health and the environment, including methodologies for hazard and exposure assessment. Overall, the document supports the OECD's Working Party on Hazard Assessment (WPHA) mandate to harmonise hazard assessment methodologies across member countries, by promoting a common understanding of the IATA concept with reference to existing guidance documents and supporting materials.

The development of this document was led by the European Union's Joint Research Centre (JRC) and reviewed by the OECD Working Party on Hazard Assessment (WPHA). This document is published under the responsibility of the Joint Meeting of the Chemicals Committee and the Working Party on Chemicals, Pesticides and Biotechnology.

Table of contents

Tables

Figures

Abbreviation list

* For a definition see Annex A;

** For a definition see the glossary in Annex B.

3Rs	See Three Rs
ADME(T)	Absorption, distribution, metabolism, excretion (toxicity)
ANSES	French Agency for Food, Environmental and Occupational Health & Safety
AO	Adverse Outcome**
AOP	Adverse Outcome Pathway**
APCRA	Accelerating the Pace of Chemical Risk Assessment
BfR	German Federal Institute for Risk Assessment
BPR	Biocidal Products Regulation
CCA	Council of Canadian Academies
CDER	Centre for Drug Evaluation and Research (US FDA)
CLP	Classification, Labelling and Packaging Regulation
CMP	Canadian Chemicals Management Plan
CRED	Criteria for reporting and evaluating ecotoxicity data
CREM	US EPA's Council for Regulatory Environmental Modeling
CSAF	Chemical-Specific Adjustment Factor
DA	Defined Approach*
DIP	Data Interpretation Procedure *
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECHA	European Chemicals Agency
EDC	Endocrine disrupting chemical
EFSA	European Food Safety Authority
EMA	European Medicines Agency

GD	Guidance document
GIVIMP	Good *In Vitro* Methods Practice**
GLP	Good Laboratory Practice
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HAWC	Health Assessment Workplace Collaborative
HCI	High Content Imaging
HCS	High Content Screening
HESI	Health and Environmental Sciences Institute
HTS	High Throughput Screening**
HTTK	High Throughput Toxicokinetics
IARC	International Agency for Research on Cancer
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
IATA	Integrated Approaches to Testing and Assessment*
ILSI	International Life Sciences Institute
IPCS	International Programme on Chemical Safety (WHO)
ISA	Investigation (the project context), Study (a unit of research), Assay (analytical measurements) (metadata framework)
ITS	Integrated Testing Strategy*
IUCLID	International Uniform ChemicaL Information Database
IVIVE	*In vitro* to *in vivo* extrapolation**
JECFA	Joint FAO/WHO Expert Committee on Food Additives
JRC	European Commission's Joint Research Centre
KE	Key Event**
KER	Key Event Relationship**
MAD	Mutual Acceptance of Data**
MERIT	MEtabolomics standaRdss Initiative in Toxicology
MIAME	Minimum Information about a Microarray Experiment for Toxicogenomics
MIE	Molecular Initiating Event**
MoA	Mode of Action

NAFTA	North American Free Trade Agreement
NAM	New Approach Methodology*
NC3Rs	National Centre for the Replacement, Refinement and Reduction of Animals in Research
NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NOAEL	No observed adverse effect level
NRC	National Research Council
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
OHAT	Office of Health Assessment and Translation (US NTP)
OHT	OECD Harmonised Template
PBL	Netherlands Environmental Assessment Agency
PBPK	Physiologically-based pharmacokinetic model
PBTK	Physiologically-based toxicokinetic model**
PBTG	Performance Based Test Guideline**
QSAR	Quantitative structure-activity relationship**
RAAF	Read-Across Assessment Framework (ECHA)
REACH	Regulation on the Registration, Evaluation, Authorisation and Restriction of Chemicals
RTK	Reverse Toxicokinetics
SAR	Structure-activity relationship**
SCCS	Scientific Committee on Consumer Safety
SCENIHR	Scientific Committee on Emerging and Newly Identified Health Risk
SCHEER	Scientific Committee on Health, Environmental and Emerging Risks
SCHER	Scientific Committee on Health and Environmental Risks
SciRAP	Science in risk assessment and policy
SETAC	Society of Environmental Toxicology and Chemistry
STS	Sequential Testing Strategy*
SYRCLE	SYstematic Review Center for Laboratory animal Experimentation

TD	Toxicodynamic
TK	Toxicokinetic
TG	Test Guideline
Three Rs	Replacement, Reduction and Refinement of animal testing
TSCA	Toxic Substances Control Act
TTC	Threshold of Toxicological Concern
US EPA	US Environmental Protection Agency
US FDA	US Food and Drug Administration
UVCB	Unknown or Variable composition, Complex reaction products or Biological materials
WoE	Weight of Evidence*
WHO	World Health Organization

Executive Summary

This document gives an overview of existing guidance on Integrated Approaches to Testing and Assessment (IATA) and their component parts. While the number of documents, directly or indirectly related to guidance on IATA, is proliferating, the information is fragmented and hard to find.

The aims, characteristics and key concepts of IATA are explained, including an overview of possible IATA components (information sources). Basic definitions are provided and compared, with a view to identifying inconsistencies in the way terminology is used. A mapping exercise identified 153 guidance documents that are systematically described in a supporting file. Emphasis was given to OECD and documents from other international organisations, as well as those developed by member country agencies. The documents cover a range of topics, from overarching principles, to specific IATA components (methods and technologies), and cross-cutting issues such as data quality, uncertainty assessment, systematic review, and weight of evidence. In particular, the inclusion of considerations on uncertainty assessment was recorded for all documents described, as well as of practical tools such as templates or checklists. It is suggested that this supporting file be updated on a periodic basis, to enable easier access to the increasing body of guidance documentation.

In terms of the components of IATA, such as in vitro tests, computational models, defined approaches (DAs) and information based on read-across, there is an important distinction between those elements that can in principle be validated and standardised (DAs) and can thus be covered by the OECD principle of Mutual Acceptance of Data (MAD) and those elements that depend on expert judgement and can only be partially harmonised. This represents a challenge to the broader acceptance of IATA, since expert judgement can be difficult to translate to MAD, whereas standardised methods may be more accepted.

Another challenge relates to the rapid development of new methods, and the need to periodically update guidance as experience is gained in the practical application of the methods.

The mapping exercise revealed several gaps and inconsistencies in the literature, including:

a) a lack of harmonised definitions and interpretation of some frequently used terms, such as New Approach Methodology (NAM), Integrated Testing Strategy (ITS) and Sequential Testing Strategy (STS);

b) a need for practical guidance on the use of new approach methodologies, as well as guidance on how to integrate individual methods within IATA and how to use the results in an overall weight of evidence;

c) a need for overall guidance on uncertainty characterisation and documentation at the assessment level, including how this derives from the uncertainties associated with the individual components;

d) a need for an overarching roadmap of IATA-related guidance to navigate users;

e) a need for good modelling practices to support the mutual acceptance of QSAR predictions;

f) a need for guidance on toxicodynamic models, such as quantitative AOP models;

g) a need for high level principles to inform the design and application of IATA;

h) a need to further consider the relationship between IATA and MAD, and in particular the extent to which IATA and their components need to be MAD-compliant.

Chapter 1. Introduction

The aim of this document is to give an overview of existing guidance on Integrated Approaches to Testing and Assessment (IATA) and their component parts. While the number of documents, from different sources, directly or indirectly related to guidance on IATA, is proliferating, the information is fragmented and hard to find.

This overview document is expected to contribute to a common understanding of IATA, by explaining key concepts and providing basic definitions, and to support easier access to existing resources. Chapter 2 describes the aims and characteristics of IATA and gives an overview of possible IATA components (information sources). Chapter 3 describes the approach adopted for the mapping of guidance documents, including the inclusion and exclusion criteria applied. Chapter 4 summarises the current status of guidance documents, which may include overarching principles, guidance for specific IATA components, or cross-cutting topics such as data quality, assessment of uncertainty and weight of evidence. Based on the findings of Chapter 4, Chapter 5 identifies gaps, duplications or inconsistencies across the guidance landscape, which may inform the development of further guidance or tools.

Chapter 2. Clarification of IATA Terminology

2.1. IATA concept

According to OECD (2016a), an **Integrated Approach to Testing and Assessment (IATA)** is an "approach based on multiple information sources used for the hazard identification, hazard characterisation and/or safety assessment of chemicals. An IATA integrates and weights all relevant existing evidence and guides the targeted generation of new data, where required, to inform regulatory decision-making regarding potential hazard and/or risk. Within an IATA, data from various information sources (i.e. physicochemical properties, *in silico* models, grouping and read-across approaches, *in vitro* methods, *in vivo* tests and human data) are evaluated and integrated to draw conclusions on the hazard and/or risk of chemicals." These information sources typically include a variety of predictions or extrapolations. For example, *in silico* models and *in vitro* methods may be used to predict effects in test animals, just as animal tests have traditionally be used to predict effects in humans.

An IATA is thus designed to obtain and combine sufficient information to allow a decision to be made in the most efficient way, taking into account the context of use (problem formulation, regulatory possibilities and practical constraints). In some jurisdictions, an important principle is to respect the Principles of the Three Rs, i.e. replacing, reducing and refining animal testing (Russell and Burch 1959). In the EU, this is a legal requirement in accordance with Directive 2010/63/EU on the protection of animals used for scientific purposes.

IATA can be used in different regulatory decision-making contexts, including hazard identification, hazard characterisation, and risk assessment. IATA can be designed to provide definitive conclusions on which risk management decisions, including emergency responses, are based, or can be screening level assessments that serve the purpose of prioritising (chemicals and/or methods) for further testing.

Historically the concept of IATA has evolved as a means of formalising the way hazard and risk assessments are carried out. The concept has also been referred to in various ways, including Intelligent or Integrated Testing Strategies (ITS) since the early 90s (Worth & Blaauboer 2018). In general, historical definitions have not distinguished between integrated approaches that are entirely prescriptive (rule-based) and those that are entirely or partially based on expert judgement. However, recent OECD guidance on IATA (OECD 2016a, 2016b) introduces a clear distinction between rule-based approaches that generate predictions, termed **"defined approaches" (DA)** to testing and assessment, and flexible, judgement-based approaches that lead to safety conclusions. A DA consists of a fixed set of data/information sources (data types) along with an algorithm (**data interpretation procedure; DIP**) that interprets the data, generally in the form of a toxicity prediction. An important feature of IATA is the need to explicitly define the problem formulation and context of use, since these will determine the acceptable level of uncertainty, the choice of methods (building blocks) and the approach to evidence integration. The distinction between DAs and the broader IATA definition is also motivated by the need to distinguish between those elements that can in principle be validated and standardised (DAs) and can thus be covered by the OECD principle of Mutual Acceptance of Data (MAD) and those elements that depend on expert judgement and can only be partially harmonised. An example would be the use of read-across within IATA, where expert choices are made in

terms of how analogues are sought (which tools and databases, and how suitable analogues are selected, including the choice of similarity metric). The types of uncertainties inherent in read-across have been outlined by Schultz et al (2019). IATA typically rely on a combination of rule-based approaches and expert judgement, but the balance varies case-by-case. This represents a challenge to the broader acceptance of IATA, since expert judgement can be difficult to translate to MAD, whereas agreed DAs may be more accepted. It also points to the need to justify conclusions based on expert judgement.

The OECD Guidance documents (2016a, 2016b) provide specific interpretations of terms such as Integrated Testing Strategy (ITS) and Sequential Testing Strategy (STS). The ITS term is used as a particular kind of IATA in which multiple lines of information / data sources are interpreted at the same time. In contrast, STS refers to an IATA in which different lines of information / data sources are used in sequence, with the possibility of a conclusion at each step. The use and interpretation of these terms is, however, not consistent, with some restricting the ITS and STS definitions to prescriptive approaches, while others include more flexible approaches.

A related concept that has been widely used is "**weight of evidence**" (WoE). According to Linkov et al (2009), this term can be found in the scientific literature with a variety of meanings, ranging from the purely colloquial use of the word to structured approaches to data integration and interpretation. Various definitions of WoE from different sources and overarching principles for carrying out WoE are given in OECD (2019a) (see also Annex A).

WoE refers to the approach taken to the synthesis, interpretation and weighing of multiple lines of evidence deriving from IATA components and resulting in a conclusion that informs decision making (see Figure 2.1 and Section 2.2.3). The various lines of evidence can consist of both existing pieces of information, as well as new pieces of information generated by a testing strategy. In this context, the testing strategy may be an existing DA or a purpose-built strategy informed by the initial weight of evidence. The generation of new data and the WoE are applied in an iterative process until a conclusion with the desired level of certainty is reached. While the WoE can be carried out in different ways (see list of guidance documents in section 3.2.1), the WoE approach should always be described in a clear and transparent way that can be easily followed, reproduced and reviewed by all stakeholders, as discussed in OECD (2019a).

Definitions of key IATA-related terms collated from existing OECD guidance documents are provided in Annex A.

2.2. IATA components: methods and other information sources

IATA represent a flexible framework for carrying out an assessment and reaching a regulatory conclusion. For the purpose of evaluating risk to human health and the environment from chemical exposure, an IATA constitutes a process that employs a comprehensive weight of evidence approach to weigh and integrate available information and empirical data on chemical hazard and exposure.

The building blocks (components) of IATA can be assembled in different ways, depending on the assessment question and protection goal, and can include a range of different methods and sources of information. Possible IATA components are listed in Table 2.1 and are briefly discussed below.

IATA can be informed by both toxicodynamic (TD) and toxicokinetic (TK) information. Sources of TD information for inclusion in an IATA can be informed by Adverse Outcome

Pathways (AOPs), which are structured representations of biological events leading to adverse effects and provide mechanistic information on the molecular initiating event (MIE), key events (KE) and key event relationships (KER) (see OECD 2016c). Further guidance on AOPs is given in other OECD guidance documents (OECD 2017a, 2018a) and in the scientific literature (e.g. Tollefsen et al 2014). Quantitative information on KEs and KERs can be derived from quantitative Adverse Outcome Pathways (qAOPs), which are increasingly being published in the scientific literature (Spinu et al, 2020). Sources of TK information can include *in vitro* methods as well as computational ones (e.g. QSAR and PBK models). Lines of evidence can also be integrated in grouping and read-across approaches or DAs, before being assessed in an overall weight of evidence (Figure 2.1).

Figure 2.1. The components (building blocks) of IATA
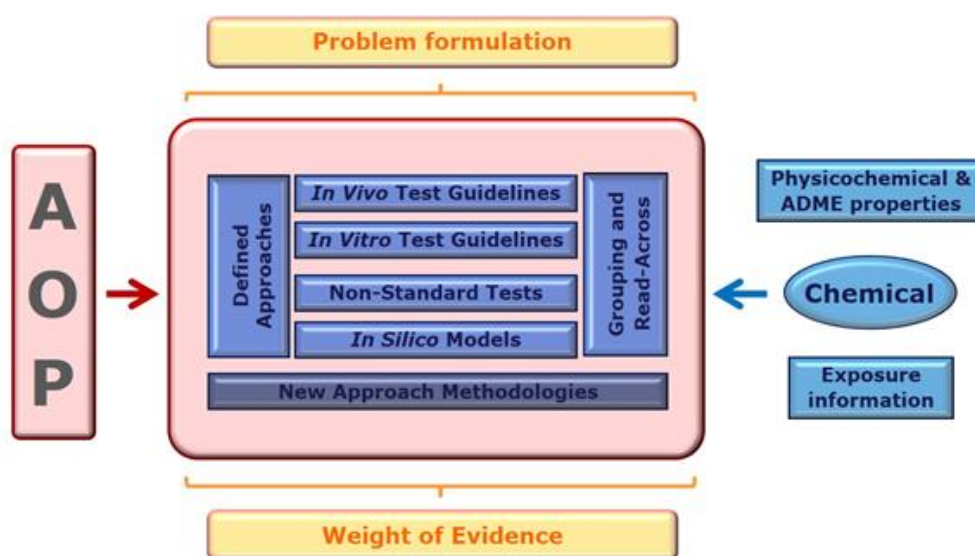
Note: New Approach Methodologies (NAMs) include grouping and read-across, defined approaches, *in vitro* test guidelines and *in silico* models. According to some definitions of NAMs, animal tests are also included, if they serve to reduce or refine another animal test.

2.2.1. Types of methods and information sources

In this document, information source refers to any source of data, experimental or predicted, existing or generated, that can be used as an input to the IATA. This includes traditional animal tests and new approach methodologies (NAMs), including computational methods that may be used for generating, interpreting and integrating data. A (non-exhaustive) list of information sources is given in Table 2.1.

The term **"test method"** is usually understood to include study results from *in vivo*, *ex vivo, in vitro* and *in chemico* **tests**. *In vitro* methods comprise a variety of different types of assays, which can be classified according to the (biological) basis of the assay, e.g. cell-based assays, or their technical implementation/output or underlying technology (e.g. high throughput screening (HTS), high content screening (HCS) and high content imaging (HCI), omics). These terms can also often be further broken down, for example "omics" into more specifically genomics, proteomics, metabolomics or transcriptomics approaches.

***In silico* approaches** include methods based on quantitative structure-activity relationship (QSAR) models, rule-based and knowledge-based prediction models, as well as mathematically-based biokinetic models. A variety of mathematical techniques can be used to develop *in silico* models, such as machine/deep learning in the case of QSARs and other

rule-based models (e.g. Bayesian networks). Biokinetic models are typically based on (ordinary) differential equations. *In silico* methods have sometimes been referred to as **"non-test methods"**; however, since these methods have been developed based on experimental test data this is a misleading expression. As computational methods increasingly provide a means not only of generating, but also of interpreting and extrapolating experimental data, the distinction between test and non-test method becomes less meaningful.

Grouping and read-across approaches are also often included in the group of *in silico* approaches since the workflow for grouping chemicals (based on toxicologically relevant criteria) and filling data gaps (by "reading across" the property from one or more analogues) can be carried out using a computational tool, such as the OECD QSAR Toolbox[1] or the US EPA Analog Identification Methodology (AIM) tool[2]. Read-across tools have been reviewed by Patlewicz et al (2017). Grouping and read-across can also be carried out manually, especially in cases where there are relatively few analogues. In addition, the approach is flexible, involving a number of expert choices and weight of evidence considerations.

Types of information that are typically taken into account within an IATA also include chemical structure, physicochemical properties, absorption, distribution, metabolism, excretion (ADME) and toxicokinetic properties, as well as exposure information. Information on internal exposure may be derived from biomonitoring studies or mathematical models such as physiologically-based toxicokinetic (PBTK) modelling.

The recently coined term **new approach methodology (NAM)** is becoming increasingly used, but there is no harmonised definition or common use of the term. In the context of a Topical Scientific Workshop held by the European Chemicals Agency (ECHA) in 2016 (ECHA 2016), "NAMs were taken in a broad context to include toxicological methods that serve as (replacement, reduction or refinement) alternatives to animal testing (e.g. *in silico*, *in chemico* and *in vitro* methods), as well as the inclusion of information from the exposure of chemicals in the context of hazard assessment. They also include a variety of new testing tools, such as "high-throughput screening" and "high-content methods" e.g. genomics, proteomics, metabolomics; as well as some "conventional" methods that aim to improve understanding of toxic effects, either through improving toxicokinetic or toxicodynamic knowledge for substances". The US Environmental Protection Agency (EPA) (2018) states in their "Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA Program" that NAM "has been adopted as a broadly descriptive reference to any technology, methodology, approach, or combination thereof that can be used to provide information on chemical hazard and risk assessment that avoids the use of intact animals"[3], referring to the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Strategic Roadmap (ICCVAM 2018). In the context of the US Toxic Substances Control Act (TSCA), NAM encompasses any "alternative test methods and strategies to reduce, refine or replace vertebrate animal testing" (US EPA 2018a). A "List of alternative test methods and strategies (or NAMs)", according to TSCA Section 4(h), has been published on the US EPA website (US EPA 2019). ICCVAM (2018) includes IATA and DAs in the NAM definition. NAMs are also intensively discussed in the context of the Canadian Chemicals Management Plan (CMP)[4] and include alternative methods that bridge the transition from conventional *in vivo* studies to in vitro assays, and that satisfy the 3Rs criteria, such as the zebrafish embryo and larval model. Overall, it can be stated that NAMs thus include both established methods, such as Test Guideline in vitro tests, as well as newly developed assays and technologies such as 3D-organoids and organ-on-a-chip devices. However, interpretations vary as to whether certain animal models (e.g. lower vertebrates), and whether grouping and read-across, are regarded as NAMs.

Table 2.1. List of possible IATA components (in alphabetical order)

General terms	*Ex vivo* method
	In chemico assay
	In silico approach
	In vitro assay
	In vivo method
	New approach methodology (NAM)
	Non-guideline method
	Test guideline method
Combinations of individual methods	Defined Approach (DA)
	Grouping and read-across
Examples of methods/methodologies/ technologies*	3D-organoids
	Absorption, distribution, metabolism, excretion (ADME) models
	Biokinetic model
	Cell-based assay
	Chemical structure information
	High Content Imaging (HCI)
	High Content Screening (HCS)
	High Throughput Screening (HTS)
	High Throughput Toxicokinetics (HTTK)
	In vitro to *in vivo* extrapolation (IVIVE)
	Machine learning
	Metabolite identification
	Omics (including e.g. genomics, proteomics, metabolomics, transcriptomics)
	Organ-on-a-chip
	Physicochemical properties
	Physiologically-based toxicokinetic (PBTK) model
	Prediction model
	Quantitative AOP model
	Quantitative structure-activity relationship (QSAR)
	Reporter gene assays
	Reverse Toxicokinetics (RTK)
	Structure-activity relationship (SAR)

Footnotes:

See the Glossary in Annex B for definition of some terms.
*This is a non-exhaustive list, encompassing a range of methods.

2.2.2. *Level of standardisation and acceptance*

Another distinction between methods is the degree of acceptance, generally distinguishing between **test guideline** (such as OECD Guidelines for health effects[5]) and performance-based test guideline (PBTG) methods or methods referring to performance standards, which are subject to Mutual Acceptance of Data (MAD)[6], and **non-guideline/non-standard methods,** which can be used within IATA but fall outside of the provisions of MAD.

Since DAs are standardised and rely on fixed data interpretation procedures (DIPs), they have the potential to become part of a test guideline (TG) and be covered by MAD, whereas
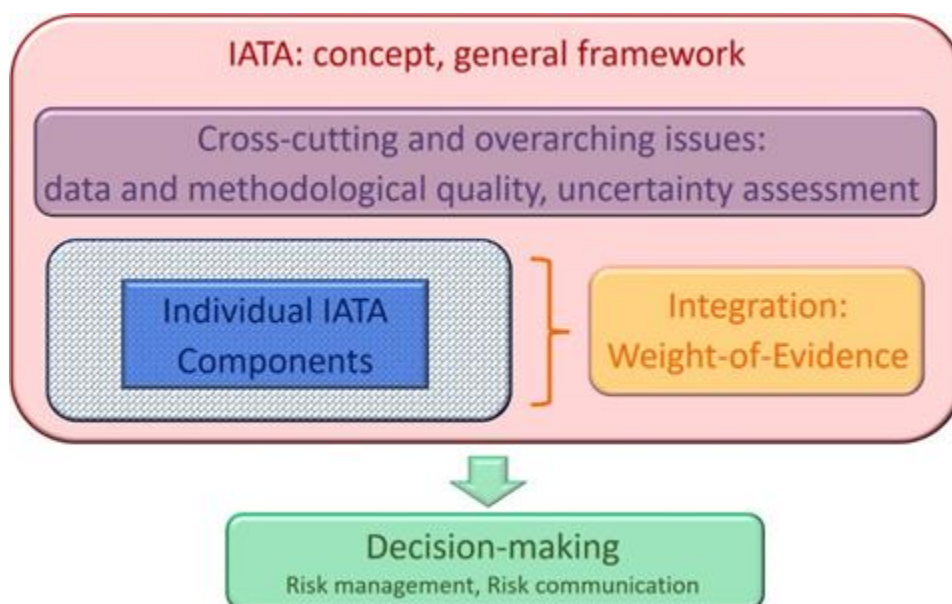
this is not necessarily the case for IATAs comprising WoE and expert judgment, which may only be partially covered by MAD.

It should also be noted that several OECD TGs include data interpretation that is subject to expert judgement, for example the single- or multi-generation reproductive studies, the two-year rodent cancer bioassay.

2.2.3. Generic IATA framework layers

For the purpose of this document, it is useful to distinguish between several layers of a generic IATA framework, as illustrated in Figure 2.2. The components (building blocks) of IATA comprise multiple methods and other information sources, selected according to the assessment question and protection goal. The components can be integrated in a variety of ways, and the final conclusion (on hazard or risk) is based on a weight of evidence of all relevant and reliable information. The focus here is not so much on guidance for individual building blocks, but rather on overarching concepts and approaches.

Figure 2.2. Different layers within a generic IATA framework

Notes

[1] https://qsartoolbox.org

[2] https://www.epa.gov/tsca-screening-tools/analog-identification-methodology-aim-tool

[3] The TSCA Section 4(h) list of NAMs is available from the US EPA website at https://www.epa.gov/sites/production/files/2019-12/documents/alternative_testing_nams_list_first_update_final.pdf

[4] See the November 2016 report of the CMP Science Committee regarding NAM (Health Canada 2016)

[5] https://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788

[6] http://www.oecd.org/env/ehs/mutualacceptanceofdatamad.htm

Chapter 3. Approach for the Mapping of Guidance Documents

3.1. Scope of the guidance mapping

A mapping exercise was conducted to identify relevant guidance documents, with emphasis on the "regulatory literature". The aim was to identify available guidance for overarching principles and approaches that can facilitate the development, application and harmonisation of IATA, as opposed to guidance on every individual method. Table 3.1 summarises the scope in terms of the layers of the IATA framework as shown in Figure 2.2.

Table 3.1. Different levels of guidance included in the mapping exercise

Level of generic IATA framework	In scope
• IATA concept/general considerations	Yes
• Individual IATA components (inputs, methods) including any interpretation / extrapolation steps	No
• Identification and characterisation of uncertainties	Yes
• Data and methodological quality	Yes
• Weight of evidence / integration leading to an assessment output / conclusion	Yes

3.2. Specific consideration of uncertainty assessment

The identification, characterisation, evaluation and, where possible, reduction of uncertainties is an essential part of hazard and risk assessment to allow informed decision making. Uncertainties need to be taken into consideration at different levels, from individual data sources, data interpretation steps, to regulatory conclusions.

Guidance for two main types of uncertainties are considered in this compilation: a) uncertainties related to the input data used; and b) uncertainties related to the extrapolations made.

The first type of uncertainty is related to the data and methodological quality, including relevance, reliability and completeness of the data. There is guidance available to support minimising these sources of uncertainty, e.g. good practice for method execution and reporting.

The second type of uncertainty comprises uncertainties in the interpretation, extrapolation and integration of available data, including knowledge about the phenomena of interest (e.g. adverse outcome pathways, exposure pathways), and methodological choices made.

Guidance on the identification and treatment of uncertainties is available as overarching stand-alone documents, or is sometimes also included in guidance for specific methodologies. The OECD IATA Case Studies Project[1] also considers uncertainty assessment and summarises the lessons learned stemming from the case studies in every

review cycle, including how the case studies captured and addressed their uncertainties (OECD 2016d, 2017c, 2018c and 2019b).

Uncertainty assessment is also an important element in the weight of evidence approach. It is a major aspect of the weight given to each relevant line of evidence (OECD 2019a).

3.3. Information extracted from the guidance documents

Table 2.2. Information extracted from the guidance documents

- Document title, year
- Sector, organisation, region
- Reference, hyperlink
- Purpose of the overall document
- Nature of the guidance in the document, link to layer within IATA framework
- Template included? Tool or checklist, flowchart included?
- Uncertainty evaluation included?

  If yes: Type(s) of uncertainty considered;
  Scoring system for uncertainties (qualitative/quantitative) included?

The information extracted from each document is listed in Table 3.2. It includes its overall aims and the scope of the guidance given, and the layer(s) it refers to in the IATA framework (see Figure 2.2; Table 3.1). Specific considerations include whether and how the characterisation of uncertainty is considered, and whether practical examples and reporting templates are provided.

3.4. Sources of guidance

Sources of guidance considered include in the first instance international (e.g. OECD, World Health Organization (WHO)) and national authority documents. If gaps of available guidance were noted, the search was extended to the scientific peer-reviewed literature for relevant input in selected cases.

OECD member countries were requested to identify country-specific guidance documents or to suggest any other relevant documents to include. References to in-house guidance received were not included in the published lists, but taken into account for the discussion.

The following inclusion/exclusion criteria were applied to focus on major and general principles relating to the use of IATA for chemical safety assessment:

- General guidance was preferred over endpoint-specific guidance, e.g., endpoint-specific OECD Test Guidelines (unless referring explicitly to an IATA framework) or ECHA endpoint-specific guidance were not included.

- Detailed methodological guidance was not included (such as how to derive no observed adverse effect levels (NOAELs).

- References from the scientific literature were included only when no official guidance was available for the topic and the article was "guidance-like" and taken up in regulatory risk assessment context.

It should be noted that the list of guidance documents is not exhaustive and the guidance landscape will change over time, with new guidance being adopted or updated and other documents becoming obsolete. It is planned to update the list of guidance appended to this document at regular intervals. For the first version of the appendices, guidance documents were searched until 20 January 2020.

3.5. Format of the guidance mapping

The compiled guidance documents are listed in an Excel Table. This format allows for a detailed search and filtering of the list, for example by country, or particular keywords (type of methods, documents with templates, etc).

In addition to the Excel table, an OECD-hosted website is foreseen to include links to all mapped guidance documents and tools. This will allow for an interactive and crosslinked search, direct access to the guidance documents or related websites and download of the related tools where applicable.

Notes

[1] http://www.oecd.org/chemicalsafety/risk-assessment/iata-integrated-approaches-to-testing-and-assessment.htm#Project

Chapter 4. Results of Guidance Document Mapping

The list and short description of the mapped guidance documents is included in Annex C of this document. It is planned to additionally link the compiled documents to an OECD-hosted website for easy access.

In the following, a brief overview of the mapped guidance is given, with some examples. These are not exhaustive, and only illustrative for the aspects discussed. The reader is referred to the list of guidance documents for the full overview. Cited scientific literature is not necessarily part of the guidance landscape.

4.1. IATA-related guidance document landscape

4.1.1. Availability of guidance

Generally, the compilation of IATA-related guidance has shown that there are many guidance documents available. However, the number varies with the layer of IATA considered, as does the format and level of detail. In "deeper" layers, e.g. for basic aspects of data such as reporting and quality, more guidance was found to be available.

The compiled guidance documents from national authorities focus on Europe and North America in the current version of the mapping, linked to the working group composition as well as the feedback received from OECD members. The list could be further extended in the future.

The focus has been otherwise on guidance documents from international organisations such as the OECD and WHO/ International Programme on Chemical Safety (IPCS), the Scientific Committee on Health, Environmental and Emerging Risks (SCHEER), aiming at international harmonisation.

4.1.2. Types of guidance related to the IATA framework

General IATA vs endpoint-specific guidance and for IATA components

There are not many documents with direct guidance relating to the concept of IATA frameworks (OECD 2015, OECD 2016a, b, c). Practical templates have been developed in the OECD IATA Case Studies project and are subject to further refinement based on the case study discussions (available e.g. in OECD 2019b).

Regarding endpoint-specific IATA guidance, the IATA for skin sensitisation (OECD 2016b) has been used to exemplify the reporting of DAs and information sources for use within IATA. There is guidance on IATAs for skin corrosion and irritation (OECD 2014a) as well as serious eye damage and eye irritation (OECD 2017b). Further endpoint-specific guidance documents will be available in the future, for example for non-genotoxic carcinogenicity and developmental neurotoxicity.

Generally, detailed guidance relating to the assessment of specific endpoints – some of them referring to integrated testing strategies – is available for example in the ECHA REACH endpoint-specific guidance documents. Individual test methods that can be information sources for IATA are described in the OECD Test Guidelines for the testing

of chemicals. These were out of the scope of this mapping exercise focussed on more general principles, but are available through the ECHA and OECD websites[1,2].

Much more guidance is available relating to the different IATA components and cross-cutting issues, as further detailed below.

Guidance on basic aspects of input data/methods vs integration for the assessment

More detailed guidance is available on the "technical level", i.e. general aspects related to the input data and methods rather than for the integration of the information for the chemical safety assessment in view of decision-making.

Data quality is indeed a fundamental and important aspect in chemical risk assessment, contributing to the allocation of weight to the available information and allowing informed and transparent decision-making. International efforts to define Good Laboratory Practice (GLP) (e.g. OECD 1998) were an important milestone in this area. Criteria to evaluate the quality of toxicological studies were mostly defined in scientific literature articles, which are widely applied akin to formal guidance. Most widely applied are the criteria set by Klimisch et al (1997), implemented in the Toxtool (Schneider 2009), complemented or replaced more recently by evaluation schemes such as SciRAP ("Science in risk assessment and policy", Molander et al 2015), CRED ("Criteria for reporting and evaluating ecotoxicity data", Moermond et al 2016). In particular the risk of bias has been recognised as an important aspect in guidance such as the OECD Weight of Evidence principles (OECD 2019a), and tools for the evaluation of bias have been provided (e.g. Hooijmans et al 2014, Sterne et al 2016 based on the Cochrane group). Bias is also taken into account in the evaluation of toxicity studies, where the use of a systematic review approach can be applied to consistently and transparently identify all relevant studies according to a defined set of inclusion criteria. To this end, guidance documents are available, both from regulatory authorities as well as other organisations via the scientific literature (e.g. EFSA 2010, NTP 2019, Moher et al 2015, US EPA 2018b). Initiatives to assess the certainty in evidence have been taken over from the clinical/health sector, e.g. the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach.

Similarly, standardisation and harmonisation of data and reporting formats helps to reduce the overall uncertainties and facilitate integration of results. They include clearly defining the endpoint or property of interest, units of measurement (where applicable), as well as important meta-data (including underlying protocol and experimental conditions) that should be recorded, as can be laid down in harmonised minimum information requirements. Harmonisation of terminology can be achieved by means of ontologies. These aspects have been addressed in diverse guidance and international initiatives, such as for example the OECD Harmonised Templates (OHT)[3], the MIAME "Minimum Information about a Microarray Experiment for Toxicogenomics" (NRC 2007), the ISA-Tab format and software (Rocca-Serra 2010), and the repositories of biomedical ontologies provided by the EMBL-EBI Ontology Lookup Service[4] and by the US NCBO BioPortal[5].

In vivo vs in vitro vs in silico

Apart from concrete methodological guidance on how to perform animal toxicological studies, such as OECD Test Guidelines, or in the form of Standard Operation Procedures (SOPs), guidance is related to how to report (and plan) the studies, for example the ARRIVE (Animal Research: Reporting of *In Vivo* Experiments) Guidelines and the Gold Standard Publication Checklist (GSPC) (Hooijmans et al 2010). As discussed above, the

systematic review approach can be used to evaluate animal studies, with several guidance documents and tools available.

For the reliability of *in vitro* methods, one guidance document (OECD 2014b) was applicable to the different types of *in vitro* test methods. A more conceptual, overarching guidance, analogous to GLP, has been published with the OECD Guidance Document on Good *In Vitro* Method Practices (GIVIMP, OECD 2018b). Technical guidance explaining the recently collected Tox21/ToxCast high throughput *in vitro* data has also been made available[6].

Guidance for *in silico* methods tends to be more general and overarching, for example the OECD Guidance Document on the validation of (Q)SAR models (2007) or the OECD Guidance on grouping of chemicals (2014c). More practical guidance is generally lacking, as discussed below, although for example the latter was complemented with more concrete guidance on how to use the OECD QSAR Toolbox (OECD 2009). Some software tools include practical guidance in their user guides or help files.

4.2. Analysis of the available IATA-related guidance

4.2.1. General Issues

Overall, there is a fragmentation of guidance documents, both in the extent of guidance for specific IATA-related aspects and the level of detail, with some topics being covered by several documents in parallel. An overarching guidance strategy is lacking, putting all this available guidance into perspective.

For several topics there already exist reviews, suggested frameworks or guidance in the scientific literature. However, it is a long process from discussions in the scientific and regulatory communities, building consensus on terminology and approaches, to the formal adoption by national authorities and international organisations. This also means there can be parallel and partial duplication of initiatives in different communities, including development of in-house guidance.

The adoption of guidance is a formal process. With the increasing pace of newly emerging methods and adaptations as well as technology developments, more flexibility and a more efficient cycle of guidance generation is needed. NAMs are also being developed or updated based on the increasing availability of data.

Overall, more practical tools are needed to complement the guidance documents for their actual application practice. Practical tools may include decision trees, checklists, and software that implements computational models.

4.2.2. Gaps

Generally, guidance is missing on how to report and how to assess the individual NAMs and their uncertainties. This is particularly the case with emerging methods, since the chemical assessment community does not have the benefit of a long practical experience with (all) those methods yet.

In the case of *in vitro* methods, for a long time the only guidance applicable to all types of new *in vitro* methods was the OECD Guidance Document for describing non-guideline *in vitro* test methods (OECD 2014b). In the meantime, there are several ongoing projects and discussions to develop guidance, for example in the context of omics the OECD project for the development of omics reporting frameworks, the ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals) MEtabolomics standaRds Initiative in

Toxicology MERIT, as well as suggestions in the scientific literature (e.g. Kauffmann et al 2017, Gantt et al 2017). Similarly, guidelines and guidance for new methods such as on organs-on-a-chip (Alepée et al 2014) and microphysiological systems (Marx et al 2016) will need to be developed at some point.

In addition to guidance on the methods themselves, the role of current and emerging NAMs and their contribution to the overall assessment in an IATA framework needs to be (re-)considered on a regular basis. A particular challenge is to assess how uncertainties in the relevance and reliability of the individual NAMs propagates into the overall uncertainty of the assessment. This could include, for example, the impact of false positive and negative predictions, or the impact of inaccuracies in potency predictions. Another consideration is whether the overall uncertainty can be reduced by incorporating additional NAMs into the IATA, i.e. relying on batteries of NAMs at various steps rather than individual NAMs. An example of the latter is the use of multiple QSARs for genotoxicity prediction, rather than single QSAR models.

Furthermore, taking the example of in silico methods, even though there are several international guidance documents available on QSARs, there is still room for more concrete guidance on good modelling practice. Guidance on the systematic evaluation of the model predictions within an uncertainty assessment framework would be helpful.

Similarly, although there is guidance available on read-across from different authorities and several approaches have been suggested to facilitate uncertainty assessment and reporting in a structured way, an internationally harmonised template for practical use would still be useful.

In addition to existing guidance related to toxicokinetics (such as WHO/IPCS 2010, ECHA 2017d), an OECD Guidance on the characterisation, validation and documentation of PBK models is under development.

There is also a lack of guidance on toxicodynamic models, such as quantitative AOP models. However, there is ongoing research and discussions in the scientific literature (e.g. Spinu et al 2020).

4.2.3. Overlaps and duplication

There are overlaps of guidance on the same topics published by different institutions, such as different national authorities as well as international organisations. There are for example several guidance documents on Weight of Evidence (e.g. from OECD, WHO/IPCS, the Scientific Committee on Health, Environmental and Emerging Risks (SCHEER) and national authorities) as well as on uncertainty assessment (e.g. from WHO, US EPA, the German Federal Institute for Risk Assessment (BfR) and EFSA), even several from the same organisations. In the Excel table () the most recent versions were included, or multiple examples were included where these relate to different assessment fields or aspects. This demonstrates that there is a multitude of guidance documents being available. However, these documents can be complementary, as was found for example in a case study carried out to compare the application of the guidelines on uncertainties from BfR and EFSA (EFSA 2018c). They also cross-reference each other. The conclusions drawn and decisions based upon the assessment may however be different, depending on the specific regulatory context in different regions or legislations. Overall, the multitude of documents, in the absence of an overarching guide, makes it difficult to navigate the guidance landscape.

In parallel, there are ongoing discussions and suggestions for standards and guidance in the scientific and regulatory communities, published in the scientific literature, but not translated into official guidance yet.

Therefore, overall, the guidance landscape is complex and difficult to navigate. It might not be clear for a user which guidance to use and methods to follow in a particular case in question, or not even that there is other/more suitable guidance available. Therefore, this mapping of available guidance was undertaken with the aim to contribute to a better awareness and understanding of the guidance landscape. The WHO has developed an overview document for their methodology guidance documents, putting the different key methodology documents into context. A similar approach could have merit in the context of putting OECD IATA related guidance into perspective for the potential user.

4.2.4. Divergence

Even though there is some overlap in content covered by the guidance documents, no striking inconsistencies could be found. The guidance adheres to commonly-accepted principles inherent to hazard/risk assessment. Differences might be due to different regulatory legislations and practices and thus different assessment needs. For example, either quantitative or qualitative uncertainty assessment might be preferred. Similarly, accepted performance standards vary for different applications/legislations. Even if the same guidance is used, the interpretation of results or resulting decisions might be different depending on the context. This needs to be taken into account while trying to harmonise guidance across different regions and legislations and legislative practices. In other words, there is a limit to what can usefully be harmonised at an international level.

It also became evident that there is divergence in the use of terms such as for NAM, ITS, STS. A common understanding of the nature of new approach methodologies, in particular to what extent they are "non-animal" methods, as well as of the level of flexibility vs prescriptiveness in integrated approaches, would be helpful for international harmonisation efforts to be consistently applied. It is recommended that a clear and comprehensive definition of NAM is adopted at OECD level.

4.3. Guidance related to the characterisation and reporting of uncertainties

4.3.1. Availability of uncertainty assessment guidance

The emphasis on uncertainty evaluation in chemical risk assessment continues to increase, i.e. the importance to characterise, transparently and consistently document and communicate uncertainties to allow for informed decision making. All IATA components are associated with uncertainty at different levels, including basic issues of data representation and reporting, data and methodological quality, uncertainties of extrapolation as well as integration and weight of evidence. Consequently, different levels of guidance for the evaluation of these different uncertainties are needed.

Uncertainty evaluation considerations are available as part of many overarching guidance documents related to chemical risk assessment from international organisations and national authorities, e.g. US EPA (2011), WHO/IPCS (2008), WHO/ICPS (2018) containing a practical spreadsheet tool (APROBA), and ECHA (2017b) including a template (ECHA 2017c). The OECD IATA guidance on the reporting of DAs based on multiple information sources (OECD 2016a) also lists the consideration of known uncertainties as an important principle. The consideration of uncertainties is also part of guidance for weight of evidence (e.g. OECD 2019a, EFSA 2017). In addition, a comprehensive systematic guidance on uncertainty evaluation has been compiled by EFSA (2018a,b).

Historically, uncertainty, for example from animal studies, has been taken into account by uncertainty factors, such as intra- and inter-species extrapolation, and has been included as

such in risk assessment guidance. The concept has been extended by chemical-specific adjustment factors (CSAFs) (WHO/IPCS 2005, US EPA 2014, Bhat et al 2017).

For some methods and approaches, i.e. components of IATA, there is more method-specific guidance related to uncertainties available. For example, the ECHA Read-Across Assessment Framework (RAAF) (ECHA 2017a) guides through a structured assessment evaluating the confidence in the assessment for read-across. However, for some NAMs, in particular for emerging methods, guidance for the specific uncertainties related to the method is lacking. Therefore, it is not clear how to take these methods into account in the overall IATA framework when weighing information to reach a hazard/risk assessment conclusion. In this respect, a number of case studies recently published or being developed, e.g. under the OECD IATA Case Studies project and APCRA initiative (Kavlock et al 2018; see also[7]), should be informative.

It has become increasingly apparent that an important aspect of uncertainty evaluation is the appropriate communication of the uncertainties, e.g. to risk assessors, risk managers and the general public, including how uncertainties have been reduced in the assessment or taken into consideration in decision making (e.g. via the application of assessment factors). A guidance document on the communication of uncertainties has been published by EFSA (EFSA 2019).

4.3.2. Specific aspects of uncertainty guidance and needs

The concepts of uncertainty and uncertainty assessment are generally consistently explained, however different organisations place different emphasis on different aspects, e.g. on quantitative or qualitative uncertainty assessment.

As discussed above, more guidance is available on basic aspects related to the input data than for the integration of the results for the hazard/risk assessment goal. These aspects relate to the quality of the data and used methods, and thus are related to the uncertainty associated with the overall assessment. Guidance to improve study and data quality and reliability contributes to improving the quality of the IATA component results and consequently the conclusions by reducing overall uncertainties and increasing confidence. Several international initiatives are working on guidance for these aspects such as standardisation of data formats and reporting formats, consensus on the study/method content to be reported (minimum information requirements, metadata), standardisation of terminology in the form of ontologies (see above). The evaluation of study quality to identify associated uncertainties and possible bias (such as by systematic reviews) is another focus of guidance development.

However, there is less guidance on the evaluation of uncertainties at the assessment level, integrating the different (types of) uncertainties from different methodologies and putting them into perspective for decision-making. This is also discussed in the OECD guidance on WoE in relation to the confidence scoring of lines of evidence (OECD 2019a). No overarching framework exists to comprehensively reconcile the different levels of guidance and types of uncertainty assessment within the IATA framework. Such a framework, also aiming at exploiting the uncertainty assessment practices emerging in different (scientific) communities would help to increase confidence in risk assessment results as well as to support the aim of mutual regulatory acceptance. A balance would need to be struck between being comprehensive while not being too complex.

Notes

[1] https://echa.europa.eu/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment
[2] http://www.oecd.org/env/ehs/testing/oecdguidelinesforthetestingofchemicals.htm
[3] http://www.oecd.org/ehs/templates/#d.en.192217
[4] https://www.ebi.ac.uk/ols/index (European Molecular Biology Laboratory (EMBL) - European Bioinformatics Institute (EBI); The Ontology Lookup Service is part of the ELIXIR infrastructure)
[5] http://bioportal.bioontology.org/ (National Center for Biomedical Ontology (NCBO))
[6] https://lri.americanchemistry.com/Users-Guide-for-Accessing-and-Interpreting-ToxCast-Data.pdf
[7] https://news.bloombergenvironment.com/environment-and-energy/insight-new-approaches-to-chemical-assessment-a-progress-report

Chapter 5. Conclusions and Recommendations

5.1. Conclusions

The mapping of guidance documents has shown that there is a wealth of guidance available, in different forms and levels of detail, for different IATA components or related cross-cutting issues, such as data quality and uncertainty assessment.

In general, more guidance is available on data generation and the reporting and interpretation of the resulting data, rather than the integration and use of information (via defined or flexible approaches). Many initiatives are supporting basic steps towards standardisation and improvement of data quality, with a view to increasing confidence in the quality and applicability of the results.

However, guidance is fragmented and sometimes duplicated across sectors, scientific areas, countries or pieces of legislation.

5.1.1. Gaps

In terms of guidance gaps, there is a general need for more guidance on new approach methodologies, especially practical guidance on their use, but also on how to integrate individual methods within IATA and how to use the results in an overall weight of evidence.

Gaps have been identified for example for *in silico* models, with room for more concrete guidance on good modelling practice as well as on the systematic evaluation of the model predictions within an uncertainty assessment framework. Such an initiative would support the mutual acceptance of QSAR predictions, for example in the context of Defined Approaches. Similarly, for read-across, there might still be a need for an internationally harmonised template for practical use. To this end, valuable experience is being gained in the OECD IATA Case Studies Project.

Overall, there could be tools and templates developed to apply guidance in practice.

Regarding the lack of guidance on toxicodynamic models, such as quantitative AOP models, ongoing developments in the scientific literature should form the basis of eventual regulatory guidance at the international level.

In relation to uncertainty assessment, guidance is available in different forms and detail, some including practical tools to help the user. However, most of the guidance concentrates on specific components within IATA, namely the individual methods and the reliability and relevance of the data they generate. Thus, there is a need for guidance on uncertainty characterisation and documentation at the overall assessment level, including how this derives from the uncertainties associated with the individual components (methods). An overall uncertainty framework could guide through and integrate all aspects across IATA components.

Overlaps or duplication

Overall, the guidance landscape might be confusing for the user in terms of which guidance to choose for a given assessment question and decision-making context.

For example, several guidance documents exist on Weight of Evidence and uncertainty assessment. However, these guidance documents can be complementary and often cross-reference each other.

In terms of duplication, there are also relevant efforts and approaches described in the scientific literature, including different suggestions how to address one particular issue. The results obtained have not yet been taken up in official guidance.

Divergence

Apart from the inevitable overlap in content, there are no obvious inconsistencies between guidance documents, which generally rely on the same principles. Moreover, any discrepancies between guidance documents are often the result of different organisations addressing their specific assessment needs and regulatory practices. For example, guidance documents on uncertainty assessment have a different level of emphasis for the need of quantitative vs qualitative uncertainty assessment, and performance standards might vary for different applications and pieces of legislation.

An important issue to resolve, however, is the lack of harmonised definitions and usage of some frequently used terms, for example NAM, ITS, STS, as well as the understanding of the level of prescriptive vs flexible / expert judgement-based nature of the IATA (components).

5.2. Future needs and way forward

In view of the extensive and fragmented guidance landscape, a high level **harmonisation of concepts and principles**, as well as overarching advice on the use of different guidance documents and associated tools, would be beneficial.

For example, harmonisation of the definition of frequently used **terms** such as new approach methodologies, integrated testing strategies, sequential testing strategies would promote more consistent usage and common understanding.

There may also be merit in **formulating some high level principles for the design and application of IATA**. A suitable starting point would be the OECD principles and key elements for WoE evaluation (OECD 2019a), and the Principles for Next Generation Risk Assessment formulated by the International Cooperation on Cosmetics Regulation (ICCR; Dent et al. 2018).

The following aspects may be taken into consideration in this context in particular:

- The amount of evidence required from IATA components to be proportionate to the need (assessment question, protection goal, resource availability);

- The types of evidence selected for toxicological effects to be mechanistically informed, wherever possible, taking advantage of knowledge of AOPs and exposure pathways;

- IATA to be flexible in principle, to adapt to the specific risk assessment question, but able to bridge the gap to international mutual acceptance;

- Uncertainty assessment to be considered and integrated across all IATA-components in a transparent uncertainty framework.

In addition, the development of an **overarching roadmap of IATA-related guidance** would further help to place the available guidance documents in the context of IATA

workflows. This would go beyond the scope of the current overview document, but would draw upon the results of the mapping exercise and the planned online version of cross-linked documents.

The **evaluation, reporting and communication of uncertainties** in chemical safety assessment are being discussed in many international initiatives, especially for new methodologies. An overall high-level guidance, linking the uncertainties at the assessment level to those at the level of individual IATA components may be needed. This **uncertainty framework** would reconcile the fact that existing evaluation frameworks have evolved somewhat independently for different types of method (e.g. QSARs, non-standard *in vitro* tests, omics). It would also bridge the gap with MAD-compliant Test Guidelines, while retaining the major benefit of IATA, namely their flexibility and applicability to different endpoints, assessment questions and decision-making contexts.

References

Alepee N, Bahinski A, Daneshian M, De Wever B, Fritsche E, Goldberg A, Hansmann J, Hartung T, Haycock J, Hogberg H, Hoelting L, Kelm JM, Kadereit S, McVey E, Landsiedel R, Leist M, Lübberstedt M, Noor F, Pellevoisin C, Petersohn D, Pfannenbecker U, Reisinger K, Ramirez T, Rothen-Rutishauser B, Schäfer-Korting M, Zeilinger K, Zurich MG (2014) State-of-the-art of 3D cultures (organs-on-a-chip) in safety testing and pathophysiology. ALTEX 31: 441-477.

Bhat V, Meek MEB, Valcke M, English C, Boobis A, Brown R (2017) Evolution of chemical-specific adjustment factors (CSAF) based on recent international experience; increasing utility and facilitating regulatory acceptance. Crit Rev Toxicol 47: 729-749.

Dent M, Amaral RT, Da Silva PA, Ansell J, Boisleve F, Hatao M, Hirose A, Kasai Y, Kern P, Kreiling R, Milstein S, Montemayor B, Oliveira J, Richarz A, Taalman R, Vaillancourt E, Verma R, Posada NVOC, Weiss C, Kojima H (2018) Principles underpinning the use of new methodologies in the risk assessment of cosmetic ingredients. Computational Toxicology 7: 20-26.

ECHA (European Chemicals Agency) (2016) New Approach Methodologies in Regulatory Science, Proceedings of a scientific workshop Helsinki, 19–20 April 2016, ECHA-16-R-21-EN

ECHA (2017a) Read-across Assessment Framework (RAAF). ECHA-17-R-01-EN, https://echa.europa.eu/documents/10162/13628/raaf_en.pdf

ECHA (2017b) Weight of Evidence/Uncertainty in Hazard Assessment - Background Document & Examples, available at https://echa.europa.eu/documents/10162/17169198/wo_eu_uncertainty_background_en.docx/4f2b49ab-ade0-6ee3-e977-8abe00c21c23

ECHA (2017c) Template for Weight of Evidence/Uncertainty in hazard assessment, available at https://echa.europa.eu/documents/10162/17169198/template_for_weight_of_evidence_en.docx/eb183c2e-c360-cbce-7a58-ad2d1270e5bd

ECHA (2017d) Guidance on Information Requirements and Chemical Safety Assessment, Chapter R.7c: Endpoint specific guidance. ECHA-17-G-11-EN

EFSA (European Food Safety Authority) (2010) Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA Journal, 8(6), 163

EFSA Scientific Committee, Hardy A et al (2017) Guidance on the use of the weight of evidence approach in scientific assessments. EFSA Journal, 15(8) 4971

EFSA Scientific Committee, Benford D et al (2018a) Guidance on uncertainty analysis in scientific assessments. EFSA Journal 16(1), 5123

EFSA Scientific Committee, Benford D et al (2018b) The principles and methods behind EFSA's Guidance on Uncertainty Analysis in Scientific Assessment. EFSA Journal 16(1), 5122

EFSA, Schendel T, Jung C, Lindtner O, Greiner M (2018c) Guidelines on Uncertainty Analysis: Comparison of the respective Documents of EFSA and BfR for Exposure Assessments. EFSA supporting publication 2018: EN-1472

EFSA, Hart A et al (2019) Guidance on Communication of Uncertainty in Scientific Assessments, EFSA Journal, 17(1), 5520

Federal Institute for Risk Assessment (BfR) (2015) Guidance document on uncertainty analysis in exposure assessments

Flaumenhaft R (2007) in: Taylor JB, Triggle DJ (eds) Comprehensive Medicinal Chemistry II, Vol 3, p. 129-149

Gantt TW, Sauer UG, Zhang SD, Chorley BN, Hackermüller J, Perdichizzi S, Tollefsen KE, van Ravenzwaay B, Yauk C, Tong W, Poole A (2017) A generic Transcriptomics Reporting Framework (TRF) for 'omics data processing and analysis. Regul Toxicol Pharmacol 91 Suppl 1, S36-S45

Health Canada (2016) Chemicals Management Plan Science Committee report - November 16–17, 2016, available at: https://www.canada.ca/en/health-canada/services/chemical-substances/chemicals-management-plan/science-committee/meeting-records-reports/committee-report-november-16-17-2016.html

Hooijmans CR, Leenaars M, Ritskes-Hoitinga M (2010) A gold standard publication checklist to improve the quality of animal studies, to fully integrate the Three Rs, and to make systematic reviews more feasible. Altern Lab Anim 38: 167-182.

Hooijmans CR, Rovers MM, de Vries RBM, Leenaars M, Ritskes-Hoitinga M, Langendam MW (2014) SYRCLE's risk of bias tool for animal studies. BMC Med Res Methodol 14: 43.

ICCVAM (Interagency Coordinating Committee on the Validation of Alternative Methods) (2018) A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States. Available: https://ntp.niehs.nih.gov/go/iccvam-rdmp. https://dx.doi.org/10.22427/NTP-ICCVAM-ROADMAP2018

Kauffmann H-M, Kamp H, Fuchs R, Chorley BN, Deferme L, Ebbels T, Hackermüller J, Perdichizzi S, Poole A, Sauer UG, Tollefsen KE, Tralau T, Yauk C, van Ravenzwaay B (2017) Framework for the quality assurance of 'omics technologies considering GLP requirements. Regul Toxicol Pharmacol 91, Supplement 1, S27-S35.

Kavlock RJ, Bahadori T, Barton-Maclaren TS, Gwinn MR, Rasenberg M, Thomas RS (2018) Accelerating the Pace of Chemical Risk Assessment, Chem. Res. Toxicol. 31: 287−290.

Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG (2010) Animal research: Reporting in vivo experiments: The ARRIVE guidelines. Br J Pharmacol 160: 1577–1579.

Klimisch H-J, Tillmann MA (1997) A Systematic Approach for Evaluating the Quality of Experimental Toxicological and Ecotoxicological Data. Regul Toxicol Pharmacol 25: 1-5.

Linkov I, Loney D, Cormier S, Satterstrom FK, Bridges T (2009). Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. Science of the Total Environment 407: 5199-5205.

Marx et al (2016) Biology-inspired microphysiological system approaches to solve the prediction dilemma of substance testing. ALTEX 33: 272-321.

Moermond CT, Kase R, Korkaric M, Ågerstrand M (2016) CRED: Criteria for reporting and evaluating ecotoxicity data. Env Tox and Chem 35(5), 1297–1309.

Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA & PRISMA-P Group (2015) Preferred Reporting Items for Systematic Reviews and Meta-Analyses protocols (PRISMA-P) 2015 statement. Systematic Reviews 4, Article number: 1

Molander L, Ågerstrand M, Beronius A, A Hanberg & Christina Rudén (2015) Science in risk assessment and policy (SciRAP): an online resource for evaluating and reporting in vivo (Eco)Toxicity studies. Human and Ecological Risk Assessment: An International Journal 21: 753-762.

National Research Council (NRC) (2007) Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment, Appendix B: Minimum Information about a Microarray Experiment for Toxicogenomics (MIAME). The National Academies Press, Washington, DC

National Toxicology Program (NTP) (2019) Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Office of Health Assessment and Translation (OHAT), available at https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf

OECD (1998) OECD Principles on Good Laboratory Practice (as revised in 1997). ENV/MC/CHEM(98)17, available at http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/mc/chem(98)17

OECD (2005) Series on Testing and Assessment No. 34: Guidance Document on the validation and international acceptance of new or updated test methods for hazard assessment. ENV/JM/MONO(2005)14, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2007) Series on Testing and Assessment No. 69: Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship (Q)SAR Models. ENV/JM/MONO(2007)2, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2009) Series on Testing and Assessment No. 102: The Guidance Document for Using the OECD (Q)SAR Application Toolbox to Develop Chemical Categories According to the OECD Guidance on Grouping Chemicals. ENV/JM/MONO(2009)5, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2014a) Series on Testing and Assessment No. 203: Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Skin Corrosion and Irritation. ENV/JM/MONO(2014)19, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2014b) Series on Testing and Assessment No. 211: Guidance Document for Describing Non-Guideline In Vitro Test Methods. ENV/JM/MONO(2014)35, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2014c) Series on Testing and Assessment No. 194: Guidance on Grouping of Chemicals, Second Edition. ENV/JM/MONO(2014)4, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2015) Series on Testing and Assessment No. 215: Report of the Workshop on a framework for the development and use of Integrated Approaches to Testing and Assessment. ENV/JM/MONO(2015)22, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2016a) Series on Testing and Assessment No. 255: Guidance Document on the Reporting of Defined Approaches (DAs) to Be Used within IATA. ENV/JM/MONO(2016)28, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2016b) Series on Testing and Assessment No. 256: Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to Be Used within IATA for Skin Sensitisation. ENV/JM/MONO(2016)29, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2016c) Series on Testing and Assessment No. 260: Guidance Document for the use of Adverse Outcome Pathways in developing Integrated Approaches to Testing and Assessment (IATA). ENV/JM/MONO(2016)67, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2016d) Series on Testing and Assessment No. 250: Report on considerations from case studies on Integrated Approaches for Testing and Assessment (IATA). First Review Cycle (2015) Case Studies on Grouping Methods as a Part of IATA. ENV/JM/MONO(2016)48, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2017a) Series on Testing and Assessment No. 184: Revised Guidance Document on Developing and Assessing Adverse Outcome Pathways. ENV/JM/MONO(2013)6, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2017b) Series on Testing and Assessment No. 263: Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Serious Eye Damage and Eye Irritation. ENV/JM/MONO(2017)15, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2017c) Series on Testing and Assessment No. 270: Report on considerations from case studies on Integrated Approaches for Testing and Assessment (IATA). Second Review Cycle (2016). ENV/JM/MONO(2017)22, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2018a) Series on Testing and Assessment No. 233: Users' handbook supplement to the Guidance Document for developing and assessing AOPs. ENV/JM/MONO(2016)12, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2018b) Series on Testing and Assessment No. 286: Guidance Document on Good In Vitro Method Practices (GIVIMP). ENV/JM/MONO(2018)19, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2018c) Series on Testing and Assessment No. 289: Report on considerations from case studies on Integrated Approaches for Testing and Assessment (IATA). Third Review Cycle (2017). ENV/JM/MONO(2018)25, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2018d) Test No. 442D: In Vitro Skin Sensitisation: ARE-Nrf2 Luciferase Test Method, OECD Guidelines for the Testing of Chemicals, Section 4. https://doi.org/10.1787/9789264229822-en

OECD (2019a) Series on Testing and Assessment No. 311: Guiding principles and key elements for establishing a Weight of Evidence for chemical assessment. ENV/JM/MONO(2019)31, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2019b) Series on Testing and Assessment No. 307: Report on considerations from case studies on Integrated Approaches for Testing and Assessment (IATA). Fourth Review Cycle (2018). ENV/JM/MONO(2019)26, available at http://www.oecd.org/chemicalsafety/testing/series-testing-assessment-publications-number.htm

OECD (2019c) Test No. 492: Reconstructed human Cornea-like Epithelium (RhCE) test method for identifying chemicals not requiring classification and labelling for eye irritation or serious eye damage. https://doi.org/10.1787/9789264242548-en

Patlewicz G, Helman G, Pradeep P, Shah I (2017) Navigating through the minefield of read-across tools: A review of in silico tools for grouping. Computational Toxicology 3: 1-18.

Rocca-Serra P, Brandizi M, Maguire E, Sklyar N, Taylor C, Begley K, Field D, Harris S, Hide W, Hofmann O, Neumann S, Sterk P, Tong W, Sansone SA (2010) ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. Bioinformatics 26: 2354–2356.

Russell WMS, Burch RL (1959) The Principles of Humane Experimental Technique, Methuen, London

Schultz TW, Richarz A-N, Cronin MTD (2019) Assessing uncertainty in read-across: Questions to evaluate toxicity predictions based on knowledge gained from case studies. Computational Toxicology 9: 1-11.

Spinu N, Cronin MTD, Enoch S, Madden JC, Worth AP (2020) Quantitative adverse outcome pathway (qAOP) models for toxicity prediction. Archives of Toxicology, in press. https://doi.org/10.1007/s00204-020-02774-7

Sterne AC et al (2016) ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 355: i4919.

Tollefsen KE, Scholz S, Cronin MT, Edwards SW, de Knecht J, Crofton K, Garcia-Reyero N, Hartung T, Worth A, Patlewicz G (2014) Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA), Regulatory Toxicology and Pharmacology 70: 629-640.

US EPA (US Environmental Protection Agency) (2011) Risk Assessment Guidance for Superfund: Volume III - Part A, Process for Conducting Probabilistic Risk Assessment

US EPA (2014) Guidance for applying quantitative data to develop data-derived extrapolation factors for interspecies and intraspecies extrapolation. Risk Assessment Forum. EPA/100/R-14/002F. Washington (DC) 20460

US EPA (2018a) Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA Program, U.S. Environmental Protection Agency, Office of Chemical Safety and Pollution Prevention, Washington, DC, available at https://www.epa.gov/sites/production/files/2018-06/documents/epa_alt_strat_plan_6-20-18_clean_final.pdf

US EPA (2018b) Application of Systematic Review in TSCA Risk Evaluations, https://www.epa.gov/sites/production/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf

US EPA (2019) 2019 List of Alternative Test Methods and Strategies (or New Approach Methodologies [NAMs]), available at https://www.epa.gov/sites/production/files/2019-12/documents/alternative_testing_nams_list_first_update_final.pdf

Worth A, Blaauboer B (2018) Integrated Approaches to Testing and Assessment, in Balls M, Combes R, Worth A (Eds) *The History of Alternative Test Methods in Toxicology*, New York, NY, USA, Academic Press, Elsevier p.301-306

WHO/IPCS (2005) Chemical-Specific Adjustment Factors for Interspecies Differences and Human Variability: Guidance Document for the Use of Data in Dose/Concentration - Response Assessment

WHO/IPCS (2008) Uncertainty and Data Quality in Exposure Assessment, ISBN 9789241563765

WHO/IPCS (2010) Characterization and application of physiologically based pharmacokinetic models in risk assessment, ISBN 9789241500906

WHO/IPCS (2018) Guidance document on evaluating and expressing uncertainty in hazard characterization, 2nd ed, ISBN 9789241513548

Annex A. Definitions of key IATA-related terms in OECD Guidance documents, as well as definitions of NAM and WoE from different sources.

Term	OECD (2015) STA 215, ENV/JM/MONO(2015)22	OECD (2016a) STA 255, ENV/JM/MONO(2016)28	OECD (2016b) STA 256, ENV/JM/MONO(2016)29	GIVIMP (OECD 2018b) STA 286, ENV/JM/MONO(2018)19	Other sources
Integrated Approach to Testing and Assessment (IATA)	A structured approach that strategically integrates and weights all relevant data to inform regulatory decisions regarding potential hazard and/or risk and/or the need for further targeted testing and therefore optimising and potentially reducing the number of tests that need to be conducted. An IATA may be comprised of one or more elements. These elements can be informed by an AOP, e.g. SAR / QSAR, testing assays etc., or could also contain elements that are not informed by an AOP, such as exposure, ADME, use profiling, etc.	An Integrated Approach to Testing and Assessment is an approach based on multiple information sources used for hazard identification, hazard characterisation and/or safety assessment of chemicals. An IATA integrates and weights all relevant existing evidence and guides the targeted generation of new data, where required, to inform regulatory decision-making regarding potential hazard and/or risk. Within an IATA, data from various information sources (i.e. physicochemical properties, in silico models, grouping and read-across approaches, *in vitro* methods, in vivo tests and human data) are evaluated and integrated to draw conclusions on the hazard and/or risk of chemicals. Within this process, the incorporation of data generated with non-animal testing and non-testing methods is expected to contribute considerably to a reduction of testing in animals. The output of an IATA is a conclusion that, along with other considerations, informs regulatory decision-making.		An IATA is an approach that integrates and weighs all relevant existing evidence and guides the targeted generation of new data, where required, to build up a hazard or risk assessment acceptable in regulatory decision-making. Ideally, IATA should be informed by mechanistic understanding of the underlying toxicokinetics and toxicodynamics. A framework for capturing the toxicodynamic information is provided by Adverse Outcome Pathways (AOP). IATA are pragmatic, science-based approaches for chemical hazard characterisation that rely on an integrated analysis of existing information coupled with the generation of new information using testing strategies.	**OECD (2018d) TG 442D** A structured approach used for hazard identification (potential), hazard characterisation (potency) and/or safety assessment (potential/potency and exposure) of a chemical or group of chemicals, which strategically integrates and weights all relevant data to inform regulatory decision regarding potential hazard and/or risk and/or the need for further targeted and therefore minimal testing.
Defined		A defined approach to testing and	Defined approaches to testing and	A defined approach is a formalised	

| Approach (DA) | | | assessment consists of a fixed data interpretation procedure (DIP) used to interpret data generated with a defined set of information sources, that can either be used on its own, or together with other information sources within an IATA, to satisfy a specific regulatory need. A defined approach consists of a fixed data interpretation procedure (DIP) (e.g. statistical, mathematical models) applied to data (e.g *in silico* predictions, *in chemico*, *in vitro* data) generated with a defined set of information sources to derive a prediction. In contrast to the assessment process within Integrated Approaches to Testing and Assessment (IATA), that necessarily involves some degree of expert judgment, predictions generated with defined approaches are rule-based and can either be used on their own if they are deemed to fit-for-purpose or considered together with other sources of information in the context of IATA. | assessments, in most cases, are designed to predict an existing line of evidence (i.e. responses in animal models or in humans). Within such defined approaches data generated with selected sources of information (i.e. phisicochemical properties, *in silico*, *in chemico*, *in vitro* data etc.) are converted into predictions by applying a DIP. Examples of DIP include mathematical and statistical models. As defined in the OECD guidance document 255, defined approaches to testing and assessment are based on a fixed set of information sources and a fixed data interpretation procedure (DIP) to convert inputs from the different information sources into a prediction. In contrast to the WoE process, in a defined approach the weighting of the different information is fixed and does not leave room for subjective interpretation. The final prediction can be used on its own if fit-for-purpose to satisfy a specific regulatory need or can be used as a component within IATA and thus considered in the WoE assessment together with other relevant information. A defined approach consists of a fixed data interpretation procedure (DIP) (e.g. statistical, mathematical models) applied to data (e.g *in silico* predictions, in chemico, in vitro data) generated with a defined set of information sources to derive a prediction. A defined approach to testing and assessment can be designed in different ways, and may take for example the form of a Sequential Testing Strategy (STS) or an Integrated Testing Strategy (ITS). In contrast to the assessment process within Integrated Approaches to Testing and Assessment (IATA), that necessarily involves some degree of expert judgment, predictions generated with defined | decision-making approach consisting of a fixed data interpretation procedure used to interpret data from a defined set of information elements. | |

		approaches are rule-based and can either be used on their own if they are deemed to fit-for-purpose or considered together with other sources of information in the context of IATA.		
Data Interpretation Procedure (DIP)		The concept of DIP, taken from OECD guidance document 34 (OECD, 2005), is defined here as any algorithm for interpreting data from one or more information sources. The output of a DIP is typically a prediction (e.g. prediction of skin sensitisation potential from peptide binding data and/or chemical structure). Is defined here as any fixed algorithm for interpreting data from one or typically several information sources. The output of a DIP is typically a prediction of a biological effect of interest. A DIP is rule-based in the sense that is based for example on a formula or an algorithm (e.g. decision criteria, rule or set of rules) that do not involve expert judgment. This definition has been taken and adapted from OECD guidance document 34.	The DIP within defined approaches can range from simple rule-based decision steps to mathematical and statistical models		**OECD (2005) STA 34, ENV/JM/MONO(2005)14** An interpretation procedure used to determine how well the results from the test predict or model the biological effect of interest.
Integrated Testing Strategy (ITS)		An ITS is an approach in which multiple sources of data or information are assessed at the same time by applying a variety of specific methodologies to convert inputs from the different information sources into a prediction. For this purpose, a variety of specific methodologies can be applied, such as statistical and mathematical models.		Guidance on how various types of available data (including those obtained from in vitro testing methods or assays) should be evaluated, and addresses additional aspects on some elements such as the use of other toxicity data or weight of evidence analysis of existing and relevant data.	

| Sequential Testing Strategy (STS) | | A STS is a fixed stepwise approach for obtaining and assessing test data, involving interim decision steps, which, depending on the test results obtained, can be used on their own to make a prediction or to decide on the need to progress to subsequent steps. At each step, information from a single source/method is typically used by applying a prediction model associated with that source/method. | | | |

	OECD (2014c) STA 194 ENV/JM/MONO(2014)4	OECD (2016a) STA 255, ENV/JM/MONO(2016)28	ECHA (2017b)	EFSA (2017)	OECD (2019a)
Weight of evidence (WoE)	Weight of evidence refers to a positive expert opinion that considers available evidence from different independent sources and scientific viewpoints on a particular issue, coming to a considered view of the available, oftentimes conflicting, data. It is preferred when every source does not provide sufficient information individually.	A WoE determination means that expert judgement is applied on an ad hoc basis to the available and scientifically justified information bearing on the determination of hazard or risk. The overall assessment process within a WoE approach involves an assessment of the relative values/weights of different pieces of the available information	Weight of Evidence approach can be generally described as a stepwise process/approach of collecting evidence, assessing, integrating and weighing them to reach a conclusion on a particular problem formulation with (pre)defined degree of confidence. The following proposed 6 steps are considered to constitute the backbone of the WoE approach: 1. Problem formulation, 2. Collection and Documentation of all information, 3. Assessment of quality of individual evidence (reliability, relevance, adequacy), 4. Integration & Weighing of Evidence, 5. Application of levels of confidence, 6. Uncertainty Analysis, 7. Conclusion.	Weight of evidence assessment is a process in which evidence is integrated to determine the relative support for possible answers to a question. The guidance document considers the weight of evidence assessment as comprising three basic steps: (1) assembling the evidence into lines of evidence of similar type, (2) weighing the evidence, (3) integrating the evidence.	WoE can be generally understood to mean a method for decision-making that involves consideration of known lines of evidence where a "weight" is assigned to each line of evidence based on the confidence associated with the evidence. Evidence is combined and the overall strength of evidence determined to support or refute a hypothesis question posed during a problem formulation stage. The ultimate goal of WoE is to provide a transparent means for communicating decision-making such that decisions can be clearly understood and questioned by all stakeholders.

	ECHA Topical Scientific Workshop (2016)	US EPA (2018)	ICCVAM (2018)	Canadian Chemicals Management Plan (Health Canada 2016)
New approach methodology (NAM)	NAM as an overarching term for all methods including toxicological methods that serve as	NAM has been adopted as a broadly descriptive reference to any technology, methodology, approach (including computational/in silico models	NAM has been adopted as a broadly descriptive reference to any alternative test method or methodology that can be used to provide information on chemical hazard and risk assessment. These new	NAMs include alternative methods that bridge the transition from conventional *in vivo* studies to *in vitro* assays, and that

(replacement, reduction or refinement) alternatives to animal testing, as well other sources of information such as exposure data. Thus, NAMs may include *in silico* approaches, *in chemico* and *in vitro* assays.	(i.e., QSARs)), or combination thereof that can be used to provide information on chemical hazard and risk assessment that avoids the use of intact animals. In the context of TSCA, NAM encompasses any alternative test methods and strategies to reduce, refine or replace vertebrate animal testing.	approaches include IATAs, defined approaches for data interpretation, and performance-based evaluation of test methods. In this context, alternative test methods include non-animal test systems and phylogenetically lower species, methods that reduce the number of animals required for a specific test, or refine animal use to lessen or avoid pain and distress.	satisfy the 3Rs criteria, such as the zebrafish embryo and larval model.

Annex B. Glossary of selected relevant terms.

Absorption, distribution, metabolism, excretion (ADME)	Describes the disposition of a toxicological compound in an organism.
Adverse Outcome (AO)	A specialised type of key event that is generally accepted as being of regulatory significance on the basis of correspondence to an established protection goal or equivalence to an apical endpoint in an accepted regulatory guideline toxicity test. (OECD 2018a)
Adverse Outcome Pathway (AOP)	An AOP describes a sequence of events commencing with initial interaction(s) of a stressor with a biomolecule within an organism that causes a perturbation in its biology (i.e., molecular initiating event, MIE), which can progress through a dependent series of intermediate key events (KEs) and culminate in an adverse outcome (AO) considered relevant to risk assessment or regulatory decision-making. Importantly, AOPs do not describe every detail of the biology but instead focus on describing critical steps or check-points along the path to adversity, which are both measurable and have potential predictive value. (OECD 2018a)
Biokinetics	Time-course of a chemical (substance and mixture) and its metabolites in a living organism, i.e., increase or decrease of substance concentration at the site of measurement due to transport or due to formation or breakdown. (OECD 2018b)
Chemical category	see group of substances
Data interpretation procedure (DIP)	See Annex A
Defined Approach (DA)	see Annex A
Good *In vitro* Methods Practice (GIVIMP)	Good scientific, technical and quality practices from *in vitro* method development to *in vitro* method implementation for regulatory use. (OECD 2018b)
Good Laboratory Practice (GLP)	A quality system concerned with the organisational process and the conditions under which non-clinical health and environmental safety studies are planned, performed, monitored, recorded, archived and reported. (OECD 2018b). See for example the OECD Principles of GLP (OECD 1998).
Group of substances	Substances that have physicochemical, toxicological and ecotoxicological properties that are likely to be similar or follow a regular pattern as a result of structural similarity may be considered as a group or 'category' of substances. (ECHA 2017a)
Harmonised Templates (OHTs), OECD	Standard data formats for reporting information used for the risk assessment of chemicals, mainly studies done on chemicals to determine their properties or effects on human health and the environment, but also for storing data on use and exposure. (OECD 2018b)
High Content Screening (HCS)	High-content screening involves the use of simultaneous recording of multiple parameters in cell-based assays. Automated microscopy is the most frequently used approach to high-content screening. (Flaumenhaft 2007)
High Throughput Screening (HTS)	A scientific approach relevant to chemistry and biology in which a very large number (e.g., tens of thousands per day) of experimental samples are subjected to testing under given conditions in a prescribed procedure. (OECD 2018b)
In chemico	The use of abiotic chemical reactivity methods as replacements to animal experiments.
In silico models	The technique of performing experiments via computer simulations. Examples are Structure-Activity Relationships (SAR) and Quantitative Structure-Activity Relationships (QSAR). (OECD 2018b)

Integrated Approaches to Testing and Assessment (IATA)	see Annex A
In vitro test	The technique of performing a given experiment in a test tube, or, more generally, in a controlled environment outside of a living organism. (OECD 2018b)
In vitro to *in vivo* extrapolation (IVIVE)	The qualitative or quantitative transposition of experimental results or observations made *in vitro* to predict phenomena *in vivo*, i.e. in whole organisms. (OECD 2018b)
In vivo test	Experimentation using a whole, living organism as opposed to a partial or dead organism, or an *in vitro* controlled environment. Animal testing and clinical trials are two forms of *in vivo* research. (OECD 2018b)
Integrated Testing Strategy (ITS)	see Annex A
Key Event (KE)	A change in biological or physiological state that is both measurable and essential to the progression of a defined biological perturbation leading to a specific adverse outcome. (OECD 2018a)
Key event relationship (KER)	A scientifically-based relationship that connects one key event to another, defines a causal and predictive relationship between the upstream and downstream event, and thereby facilitates inference or extrapolation of the state of the downstream key event from the known, measured, or predicted state of the upstream key event. (OECD 2018a)
Line of evidence	Set of data/evidence with common properties (e.g., same type of test or directed to the same endpoint) of scientific or regulatory relevance to the hypothesis. (OECD 2019a)
Molecular Initiating Event (MIE)	A specialised type of key event that represents the initial point of chemical/stressor interaction at the molecular level within the organism that results in a perturbation that starts the AOP. (OECD 2018a)
Mutual Acceptance of Data (MAD)	The OECD MAD is a multilateral agreement which states that test data generated in any member country in accordance with OECD Test Guidelines and Good Laboratory Practice (GLP) shall be accepted in other member countries for assessment purposes and other uses relating to the protection of human health and the environment. The application of MAD avoids unnecessary and costly duplication of testing as well as nontariff barriers to trade. In addition, it reduces the number of laboratory animals used for in vivo testing. (OECD 2018b)
New approach methodology (NAM)	see Annex A
Omics	A general term for a broad discipline of science and engineering for analysing the interactions of biological information objects in various omes (these include genome, transcriptome, proteome, metabolome, expressome, and interactome). Some examples of 'Omics' technologies: genomics, proteomics, metabolomics, transcriptomics. (OECD 2018b)
Performance based test guidelines (PBTG)	A test guideline that contains one or more *in vitro* methods that are mechanistically and functionally similar. A PBTG defines the important components of the *in vitro* method and describes in detail characteristics and performance standards that a new *in vitro* method should meet in order to be considered as an additional method. (OECD 2018b)
Performance Standards	The purpose of performance standards is to provide the basis by which new or modified in vitro methods, both proprietary (i.e. copyright, trademarked, registered) and non-proprietary, can be deemed to be structurally and mechanistically similar to a validated reference method and demonstrate to have sufficient reliability and relevance for specific purposes (i.e. in accordance with the principles to OECD GD 34). (OECD 2018b)
Physiologically-based toxicokinetic (PBTK) model	Physiologically based toxicokinetic, or alternatively referred to as physiologically based pharmacokinetic or biokinetic models, are quantitative descriptions of absorption, distribution, metabolism, and excretion (ADME, possibly including toxicity as ADMET) of synthetic or natural chemical substances in humans and other animal species. PBTK models are increasingly being used as an effective tool for designing toxicology experiments and for conducting extrapolations essential for risk assessments (e.g., in pharmaceutical research and drug development, and in health risk assessment for cosmetics or general chemicals).

	(OECD 2018b)
Prediction model	A method by which the *in vitro* endpoint value(s) is used to predict the *in vivo* equivalent activity (i.e., degree of toxicity). (OECD 2018b)
(Quantitative) Structure-Activity Relationship ((Q)SAR)	Structure-activity relationships and quantitative structure-activity relationships based on the chemical structure of a compound, collectively referred to as (Q)SARs, are simplified mathematical representations of complex chemical-biological interactions that can be used to predict the physicochemical and biological properties of molecules. (OECD 2018b)
Read-across	Physicochemical, human health and/or environmental properties are predicted from information from tests conducted on reference substance(s) within a group of substances, referred to as source substance(s), by interpolation to other substances in the group, referred to as target substance(s). (ECHA 2017a)
Relevance	Description of the relationship of the test to the effect of interest and whether it is meaningful and useful for a particular purpose. It is the extent to which the test correctly measures or predicts the biological effect of interest. Relevance incorporates consideration of the accuracy (concordance) of a test method. (OECD 2019c) The degree of correspondence of scientific or regulatory evidence to the hypothesis. (OECD 2019a)
Reliability	The confidence assigned to evidence based on the assessment of data quality, sufficiency (quantity), plausibility and uncertainty. (OECD 2019a)
Replace, Reduce, Refine (3Rs)	A term describing current internationally accepted strategies for minimising use and suffering of laboratory animals used in experimental research. The optimal solution is to replace the test method requiring animal experiments with one or several *in vitro* methods; if this is not possible at least it might be possible to modify the methods in order to reduce the number of animals being used in each study without compromising data quality; if this is also not possible it might at least be possible to refine the test method so that experiments are conducted in a way minimising stress and other impact on the animals. (OECD 2018b)
Sequential Testing Strategy (STS)	see Annex A
Standard Operation Procedure (SOP)	A documented procedure which describes how to perform testing methods or assays or activities normally not specified in detail in study plans or test guidelines. (OECD 2018b)
Test Guideline (TG), OECD	OECD Test Guidelines are harmonised test methods included in the OECD Council Decision on Mutual Acceptance of Data. This means that "data generated in the testing of chemicals in an OECD Member country (or some non-member economies) in accordance with OECD Test Guidelines and OECD principles of Good Laboratory Practice shall be accepted in other Member countries (or non-member economies) for purposes of assessment and other uses relating to the protection of man and the environment". (OECD 2018b)
Tiered testing strategy	A stepwise testing strategy, which uses test methods in a sequential manner. All existing information on a test chemical is reviewed at each tier, using a weight-of-evidence process, to determine if sufficient information is available for a hazard classification decision, prior to progression to the next tier in the strategy. If the hazard potential/potency of a test chemical can be assigned based on the existing information at a given tier, no additional testing is required. (OECD 2019c)
Uncertainty	The combination of lack of knowledge (true uncertainty) and data variability OR according to the EFSA Guidance on Uncertainty "a general term referring to all types of limitations in available knowledge that affect the range and probability of possible answers to an assessment question " (OECD 2019a, EFSA 2018)
Validation	The process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose. (OECD 2005)
Weight of evidence (WoE)	see Annex A

Annex C. List and short description of the mapped guidance documents.

[Click here to access the Annex C](#)

The aim of this document is to give an overview of existing guidance on Integrated Approaches to Testing and Assessment (IATA) and their component parts. While the number of documents, from different sources, directly or indirectly related to guidance on IATA, is proliferating, the information is fragmented and hard to find.

This overview document is expected to contribute to a common understanding of IATA, by explaining key concepts and providing basic definitions, and to support easier access to existing resources.

Chapter 1 presents the aim of this document. Chapter 2 describes the aims and characteristics of IATA and gives an overview of possible IATA components (information sources). Chapter 3 describes the approach adopted for the mapping of guidance documents, including the inclusion and exclusion criteria applied. Chapter 4 summarises the current status of guidance documents, which may include overarching principles, guidance for specific IATA components, or cross-cutting topics such as data quality, assessment of uncertainty and weight of evidence. Based on the findings of Chapter 4, Chapter 5 identifies gaps, duplications or inconsistencies across the guidance landscape, which may inform the development of further guidance or tools.

https://oe.cd/iata