METALS ENVIRONMENTAL RISK ASSESSMENT GUIDANCE

FACT SHEET

MERAG 08

CLASSIFICATION: CLASSIFICATION FOR EFFECTS ON THE AQUATIC ENVIRONMENT OF METALS/METAL COMPOUNDS AND ALLOYS

January 2007
Check you have the most recent fact sheet by visiting www.metalsriskassessment.org

defra
Department for Environment Food and Rural Affairs

euras

EuroMetaux

ICMM
International Council on Mining & Metals

CONTENTS

01. Introduction **01**
02. Transformation/dissolution test protocol **02**
2.1 24h screening transformation/dissolution test – sparingly soluble metal compounds 03
2.2 7-d full transformation/dissolution test – metals and sparingly soluble metal compounds 03
2.3 28-d transformation/dissolution test – metals and sparingly soluble metal compounds 03
03. Bioaccumulation and degradation **04**
04. Derivation of the ecotoxicity reference value **05**
4.1 Introduction 05
4.2 Data compilation and selection 05
4.2.1 Criteria for data reliability and data relevance 05
4.2.2 Conclusion 12
4.3 Aggregation / selection of L(E)C_{50} / NOEC data 12
4.4 Derivation of ecotoxicity reference values 14
4.4.1 Lowest species mean value method 15
4.4.2 Statistical extrapolation method 15
05. Application of classification criteria to metals, metal compounds and metallic alloys **17**
5.1 Introduction 19
5.2 Classification strategy for metals and metal compounds 20
5.2.1 Classification strategy for metal compounds 22
5.2.2 Classification strategy for metals 23
5.2.3 Ecotoxicity validation step 24
5.3 Classification of metal powders, metal massives (and alloys (powders and massives)) according to the Critical Surface Approach (CSA) concept 25
5.3.1 Calculation of the Critical Surface Loading (CSL) 27
5.3.2 Calculation of the Critical Reference Surface Area (CRSA) that triggers classification 28
5.3.3 Calculation of the Critical Particle Size (CPS) of a spherical particle 29
5.3.4 Classification strategy using the CRSA / CD concept (Critical Surface Concept) 30
5.3.5 Normal handling and use 30
5.4 Read-across for classification purposes between massives and metal powders and vice versa 31
5.5 Classification of preparations (mixtures of metal compounds) 31

Annex 1 – Overview and short description of the standard aquatic toxicity tests that can be used for classification purposes 32

Annex 2 – Thought starter on a specific classification strategy for metallic alloys in the powder and the massive form 40

References **46**

1. INTRODUCTION

The general strategy for classifying chemical substances for the aquatic environment (water column only) is hazard-based, and as such, intrinsic properties (aquatic toxicity, degradation, solubility[1] and bioaccumulation) of a substance are the sole basis for its hazard classification (OECD 2001a). For organic metal compounds, ecotoxicity potentials are to be evaluated on a case by case basis, as performed for organic substances. The strategy for environmental classification of metals and inorganic metal compounds, outlined below, accounts for the chemical and physical specificity of the inorganic metal compounds and metals to be considered.

Criteria on degradation behavior, as considered and used for organic substances, have limited or no meaning for metals (GHS, 2003). Mackay et al. (2003) indicated that non-degradability is even an appropriate metric of persistence for metal-containing substances. Nevertheless, the concepts that a substance may not be rapidly lost from the aquatic environment or may bioaccumulate are as applicable to metals and inorganic compounds as they are to organic substances. Of course metals are never really "lost from the environment" except on geological time scales. In fact metals can be shifted from one compartment to the other. With regard to measures of bioaccumulation, i.e. bioaccumulation factors (BAFs) these are not independent of concentration and therefore should be used with caution when applied for inorganic substances (McGeer, 2003, US-EPA, 2004).

The existing classification schemes for substances can be used in a straightforward manner for soluble metal compounds by applying the current classification criteria as described in the Globally Harmonized system of Classification and Labelling of Chemicals (GHS, 2003). In this scheme, the evaluation of both short term and long term aquatic hazard potentials of the metal under investigation are achieved using appropriate standard ecotoxicity data as determined with the soluble metal salt (acute and chronic values) and comparison with the classification cut-off values (1-100 mg/L). The ecotoxicity of soluble inorganic metal compounds is dependent on the physico-chemistry of the medium, irrespective of the original metal species released in the environment. Reading across metal compounds can therefore be done by comparison of the soluble metal ion concentration (µg Me/L) causing the appropriate standard ecotoxicity effect (acute, chronic) and translation of the results towards the compound under investigation using the molecular weight ratio (MW substance/MW metal ion) (GHS, 2003).

A number of problems are encountered when trying to use the same methodology developed for soluble substances to classify insoluble metals, sparingly soluble metal compounds (SSMC) (OECD, Ottawa workshop, 1995, GHS, 2003) and alloys, especially in relation to their toxicity. Therefore, a new approach has been developed for these compounds, in which the rate and extent to which metals, sparingly soluble metal compounds and alloys can produce soluble bioavailable ionic and other metal-bearing species in aqueous media is determined. This is done by using a transformation/dissolution test and evaluating whether the rate and extent of the formation of soluble metal ions is sufficiently rapid to be of concern. As such the T/D protocol is intended to facilitate estimation of the toxicity of sparingly soluble metal compounds (SSMCs), metals as well as alloys (GHS, 2003, Annex 9).

The unique chemical properties of metals, SSMCs and alloys that need to be taken into account in arriving at their hazard classification are outlined in more detail in the following sections. It should be noted that the proposed methodology in this fact sheet does deviates from the OECD guidance in some points and some of the proposed changes will still have to be discussed in a regulatory context. Currently, the metal specific guidance that is already been given in Annex A8.7 of the GHS document includes the use of the transformation/dissolution test protocol for sparingly soluble metal compounds and metals and make reference to the critical surface area approach. The GHS document has been formerly adopted by UN Economic and Social Council (ECOSOC) in July 2003 and countries have been encouraged to have the system fully operational by 2008. Derivation of the ecotoxicity reference value using species sensitivity distributions, normalization for bioavailability, removal through metal partitioning from the water column and extension of the concepts for the classification of the metallic alloys are some of the new elements discussed in this fact sheet that are not currently permitted in the OECD, GHS, or EU classification systems.

[1] The potential to release metal ions in solution

2. TRANSFORMATION/DISSOLUTION TEST PROTOCOL

Metal/ metal compounds and alloys undergo interactions with their surrounding media that affect the solubility of the metal ion, partitioning from the water column and the species of the metal ion that exists in the water column. It is thus necessary to consider whether metal ions are likely to be formed and if so whether they are likely to be formed rapidly enough to cause concern (GHS, 2003).

The level of the metal ion that may be released to a solution following the addition of the metal and/or its compounds, will largely be determined by the following processes/conditions: a) the corrosion rate, b) the extent to which it can reacts with the media to transform to water soluble forms and c) the pH and ionic composition of the media. The rate and the associated magnitude of the metal ion concentration at which this latter process, known as "transformation/dissolution (T/D)", can vary extensively between different compounds and the metal itself, and is an important factor in determining the appropriate hazard class. The evaluation of the removal rate of the metal species from the water column due to precipitation, partitioning to suspended solids, and binding to dissolved organic carbon, important for the determination of the chronic hazard class is described in section 3.

Metal transformation will be affected by a number of factors, not least of which is the properties of the media with respect to pH, ionic concentrations, temperature, O_2 concentration and CO_2 concentration. In addition to these properties, other factors such as the size and specific surface area of the particles that have been tested, exposure time to the test media and the mass or surface area loading of the metal in the media will play a part in determining the level of total dissolved metal[2] in the water. Transformation data, therefore, can generally only be considered as reliable for the purposes of classification if conducted according to the standard Transformation/Dissolution Protocol (T/Dp) as outlined in Annex 9 of the GHS (GHS, 2003). Further details on the composition of the T/D artificial media is also given in this annex. For example the concentration of total organic carbon in the medium should not exceed 2.0 mg/L.

A distinction has to be made between different forms of metal compounds (i.e. how to distinguish between soluble and sparingly soluble metal compounds), different metallic forms (i.e. massives and powders) and metallic alloys (massives and powders). For aquatic classification purposes, powders and massives are distinguished based on their particle size. Powders are categorized by default as a particle size < 1 mm and massives have a default particle size > 1 mm. For powders, the particle size chosen for the T/Dp should represent the smallest size used in normal handling and use. For massives 1 mm is recommended if only the default classification is investigated. There may, however, be cases where classification should be based on testing a representative particle size or surface area (see section 4).

Two types of T/D tests are available: a 24 hour "screening" test and a 7 or 28 day "full" test. The function of the screening test is to identify those metal compounds that undergo dissolution and/or transformation in such a way that they are indistinguishable from soluble forms. Metal compounds that do not behave in this way are then subjected to a "full" test. For metals and metallic alloys independent of the physical form only a "full" test is relevant. A further distinction in T/D tests can be found in particle size used, loading rate, pH and mixing speed. The recommended EU/OECD T/D test is presently undergoing validation by the OECD Transformation/Dissolution Validation Management Group (OECD T/D VMG). As pH has a significant influence on T/D, both the screening test and the full tests should in principle be carried out at a pH that maximizes the concentration of the dissolved metal ions in solution. If no relevant literature data exist a preliminary screening test may need to be carried out in order to ensure that the test is performed at a pH maximizing transformation/dissolution within the described pH ranges. With reference to the conditions generally found in the environment, a pH range of 6 to 8.5 must be used. For the 28 day full test, the pH range of 5.5 to 8.5 should be used to take into consideration possible long term effects on acidic lakes. However, the OECD Transformation/Dissolution Validation Management Group recommended restricting the pH range for the 28 full test to the pH range of 6.0-8.5 for the present time, since no suitable system could be

[2] Different definitions for the dissolved fraction exist. In the T/D protocol this refers to the fraction passing a filter of 0.2 µm and acidified with HNO_3)

recommended that could maintain the pH constant at the lower range (pH = 5.5) of the test, without influencing the transformation dissolution or ecotoxicity properties of the metal, alloy or the SSMC (Helsinki meeting June 2003).

2.1 24h screening transformation/dissolution test – sparingly soluble metal compounds

The function of the screening test is to identify those compounds that undergo either dissolution or transformation/dissolution such that their ecotoxicity potential is indistinguishable from soluble forms. Metal compounds, having the smallest representative particle size on the market are introduced into the aqueous medium at a single loading of 100 mg/L. Such dissolution as will occur is achieved by agitation at a standard rate during a 24 hour period. After 24 hours agitation, the dissolved metal ion concentration is measured.

The 24h screening T/D test can also be used as pre-testing phase before conducting a full 7-28 d T/D protocol. The purpose of the pre-testing phase is to insure accuracy and precision of the results obtained from the T/D protocol. Such a pre-test that may help laboratories that are not familiar with the protocol to successfully perform a T/D test and can include the following checks:

- efficiency of the filtration method
- efficiency of the shaking conditions
- issue of potential abrasion
- the measuring technique
- the mass balance

This pre-testing should be done with a reference material that has been well studied, stable in composition independent where you buy it from and have no particle size interference.

2.2 7-d full transformation/dissolution test - metals and sparingly soluble metal compounds

The 7-d full T/D test is intended to determine the rate and extent of transformation/dissolution of sparingly soluble metal compounds, metals and metallic alloys at different loadings of the aqueous phase. Normally massive forms and/or powders are introduced into the aqueous medium at three different loadings: 1, 10 and 100 mg/L. A single loading of 100 mg/L may be used if a significant release of dissolved metal species is not anticipated. Transformation/dissolution is accomplished by standardized agitation, without causing abrasion of the particles and maintaining the integrity of the surface of the test substance and of any solid reaction product coatings formed during the test. The 7-d T/D results finally used are the dissolved metal ion concentrations obtained after the 7 days T/D period.

2.3 28-d full transformation/dissolution test - metals and sparingly soluble metal compounds

A 28-d full T/D test is performed using a loading of 1 mg/L of test substance. The intent of this study is to determine whether or not sufficient metal ions will go into solution across a pH range of 6-8.5 that would result in chronic effects. The test conditions of the 28-d test are similar to those of the 7-d study (i.e., the 7-day study using 1 mg/L loading may be extended to 28 days to evaluate the potential for long term or chronic effects).

3. BIOACCUMULATION AND DEGRADATION

Chronic classification entries aim at identifying substances with a potential of causing effects to the aquatic environment on a large temporal and/or geographic scale because they are persistent or bioaccumulate. In accordance to the GHS & EU system, chronic classifications can be removed for substances with a chronic NOEC > 1mg/L. Although, the EU & GHS recognize that criteria on degradation & bioaccumulation, used for organic substances, should be used with caution for metals (GHS, 2003), the concept that a substance may/may not be rapidly lost from the environment or may/may not bioaccumulate remain valid. Chronic classification is therefore needed in absence of evidence on both, rapid partitioning from the water column and bioaccumulation. In the current EU classification system, because of lack of appropriate guidance, the default non-degradability of metals, an R53 is usually applied by default, unless there is evidence of no chronic toxicity at 1 mg/L loading[3].

With regards to the *removal from the water column* general guidance, provided by the GHS (2003) can, however, be applied. In this respect, it is useful to note that in the GHS (Para 277) rapid removal for organic substances is demonstrated if a substance degrades biotically or abiotically in the aquatic environment by > 70% in 28 days. In accordance to the GHS primary degradation clause, it is important to additionally demonstrate that the degradation products do not fulfill the criteria for classification as hazardous to the aquatic environment. The GHS further recognizes (para 308) that "as a result of naturally occurring geochemical processes, metal ions can partition from the water column. Data on water column residence time, the processes at the water-sediment interface (deposition & remobilization) are fairly extensive but have not been integrated into a meaningful database." Indeed, GHS (para 295) suggests that information on changes in bioavailability -of the metal ion from over a 28 days period should be carefully evaluated on a case by case basis. In absence of evidence on partitioning from the water column it is by default assumed that metals are not rapidly partitioned from the water column.

Applying the GHS guidance, rapid removal from the water column is to be assessed on a case by case basis from information on decreases in metal bioavailability/metal toxicity over 28 days. If through precipitation/partitioning processes,-rapid removal of the metal ions can be demonstrated in a broad range of environmental conditions, chronic classification entries can be removed, irrespective of the acute classification entry. Therefore for metals, if the removal of the metal ions from the water column over a 28 days period reduces the toxic metal ion concentration below the toxicity levels, than the metal ions are considered as rapidly removable. Such information can be provided from literature data and/or modeling approaches under realistic environmental conditions.

With regards to *bioaccumulation*, because the mechanism of uptake & depuration are complex and variable, the GHS proposes to consider information on bioaccumulation on a case by case basis. Indeed regulation of uptake/elimination/detoxification mechanisms may allow for internal metal homeostasis (leading to low bioaccumulation factors) and/or internal metal ion homeostasis (leading to accumulation but lack of toxicity due to internal immobilization). In this framework it is proposed to consider information on essentiality, homeostasis as well as acute to chronic ratio's for the chronic classification of metals/metal compounds.

[3] As a consequence of the above practice, all metals with R50 classification entry (LC_{50} <1mg/L), automatically obtain an R53 entry

4. DERIVATION OF THE ECOTOXICITY REFERENCE VALUE

4.1 Introduction

Hazard classification is based on comparing the T/D data with available toxicity data for fish, crustacea, and algae/aquatic plants, taxa generally accepted as representative of aquatic fauna and flora. Only those data that can be considered relevant and labeled high quality should be retained for the assessment and used for ecotoxicological benchmarking.

4.2 Data compilation and selection

Ecotoxicity data can be drawn from data required for regulatory purposes (e.g. IUCLID) as well as from relevant literature and/or internationally recognized databases. Because the data quality of the extracted information may vary considerably between individual source documents it is very important to evaluate all ecotoxicity data with regard to their adequacy for classification purposes. In general, this evaluation involves a review of how well each study was conducted (see Table 1 below) and how the results are interpreted in order to accept (or reject) a study in accordance with the purpose of the assessment. The term adequacy covers here both the *reliability* of the available data and the *relevance* of the data for classification purposes in general and for metals/metal compounds/alloys in particular. These two basic elements are defined as follows:

- *Reliability*: covering the inherent quality of a test relating to test methodology and the way that the performance and results of a test are described.
- *Relevance*: covering the extent to which a test is appropriate to be used for the derivation of an ecotoxicity reference value

Only those data that can be considered of sufficiently high quality should be retained for the hazard classification. Guidance on how to screen and select the most appropriate data in the framework of this assessment is outlined in the next paragraphs. The proposed quality criteria are intended as guidance but should be used in a flexible manner, using expert judgment. Main acceptability and relevancy criteria are a clear concentration-relationship, proper statistics and proper analytics. For transparency reasons it should be clearly documented which studies are being rejected and on what ground.

4.2.1 Criteria for data reliability and data relevance

The term *reliable* or *relevant* can be assigned to a study if the study complies with a number of criteria.

4.2.1.1 Data reliability

A checklist for evaluating the general quality of ecotoxicity studies is provided in Table 1. These criteria are mostly not metal-specific: they simply adhere to the principles of good study conduct.

Table 1: Checklist of criteria for the evaluation of the reliability of ecotoxicity studies used for hazard classification.

Type of test • standard test or non-standard test • endpoint used reported • test duration reported • static or flow through
Description of test material and methods • test set-up, measuring chamber/device • test material (including purity), solutions, dilution water if applicable • test organism, including size (age), origin, number of organisms per replicate • test design (# replicates should be used) • type of food given (chronic tests)
Description of physico-chemical test conditions • proper description and control of physico-chemical conditions (e.g. pH, major cations and anions) that may influence the outcome of a test (validity criteria should be met at the end of the test)
Chemical analysis • test concentrations during the test are measured • test concentrations are not measured, but indication is given that the nominal concentrations are close to actual concentrations • evidence is given that concentrations were maintained during the test (< 30% variation)
Concentration-effect relationship • acceptable control mortality, reproduction, growth. • sound statistics used, 95 % confidence limits reported or data on the relationship given amenable to further analysis to derive a suitable L(E)C_x value • concentration range is given • at least 2 different concentrations must have been tested besides the control • a concentration related response should be clear (a progressive effect should be observed as a function of the dose) • hormesis effect observed or not

The criteria mentioned above should be met, for a study's results to be considered reliable. An experiment can be classified as *reliable (Q1)* if it has been carried out according to all criteria, or is missing one or two less important criteria. If one important criterion, or several less important criteria are missing the experiment should be classified as *less reliable (Q2),* while an experiment should be classified as *unreliable (R)* if several important criteria are missing. They are outlined in more detail further, together with some more metal specific focus points.

Type of test

- For the purpose of classification (GHS, 2003) aquatic toxicity studies carried out according to internationally harmonized test methods should be used by preference. Typically used standard species in this regard are fish, crustaceans and algae that are widely used as surrogate species covering a range of trophic levels and taxa found in the aquatic compartment. In addition test methods for these standard species are highly standardized and internationally accepted guidelines are available (acute e.g.: fish OECD Guideline 203, crustacean species 48h EC_{50} test OECD Guideline 202 and/or an algal species 72h EC_{50} OECD Guideline 201; chronic: e.g. crustaceans OECD guideline n° 211). In annex 1 a detailed list on the standard test species relevant for classification is provided. However, in absence of reliable toxicity data for the use for classification standard species classification may be considered using other non standard organisms provided that they represent equivalent species and test endpoints and were thoroughly checked on their compliance with reliability and relevance criteria before being used. Data derived from testing procedures that deviate significantly from standard guidelines and are considered unreliable, should not be used. However, it is always recommended that toxicity data on standard species are generated in order to come up with a proper basis for classification.

Description of test material and methods

- A detailed description of methods employed in the study should be provided. This description should include at least the method of test medium preparation, time of spiking, recorded observations. To calculate free ion concentrations with speciation codes the concentrations of dissolved major anion and cations, Fe, Mn, Al, dissolved organic carbon, pH are required. Furthermore the organisms used should be uniform in age and represent a sensitive life stage. The test results should allow a proper statistical analysis and the experimental design should provide sufficient replicates per test concentration to derive a high quality $L(E)C_x$/NOEC value[4].

Description of physico-chemical test conditions

- In Table 2 an overview is given of physico-chemical characteristics for each compartment that should preferably be reported and fall within the tolerance limits of the test organisms. If these limits are exceeded the test has to be considered not reliable.

Table 2: Physico-chemical parameters that should preferably be reported

Water
• temperature • oxygen • hardness • salinity • pH

- In addition to the above mentioned parameters, abiotic parameters, e.g. dissolved organic carbon concentration (DOC), hardness in the test water, that govern the speciation and hence

[4] $L(E)C_x$ = the concentration that causes x % change in response (e.g. mortality, immobility) during a specified time interval. NOEC = No Observed Effect Concentration is defined as the test concentration below the lowest concentration that did result in a significant effect (LOEC = Lowest Observed Effect Concentration) in the specific experiment

the bioavailability of some metals are to be considered in case correction for bioavailability is required. Furthermore, in the case of testing essential metals a proper description of the culture conditions, specifically related to the level of essential metals added or already present in the culture media could give valuable insight on issues such as acclimatization.

Chemical analysis

- There is a strong preference for using measured data. Analytical measurements of the metal concentrations in the test solution allow to (1) exclude human error related to the preparation/addition of test substance solutions; (2) since metals are a natural elements it is therefore important to know the total metal concentrations organisms are exposed to, including the metal background levels in the control/dilution test medium. In the case river waters are used, the metal levels in control waters can already be relatively elevated in comparison to the metal added as test solution. In this respect it is important to also consider that organisms adapt to the culture media not test media
- If it is not mentioned whether the reported toxicity values are based on measured or nominal concentrations, they should be considered as nominal concentrations. In cases where no measured data are available the use of nominal concentrations could be considered as long as soluble metal salts have been used and the reported effect levels are well above the background in the test medium. However, if the effect levels are close to reported metal background concentrations in the specific test medium (i.e. $\mu_{nominal} - 1.95\ \sigma_{nominal} \leq \mu_{bg} + 1.95\ \sigma_{bg}$) or the test concentrations are close to the essentiality levels only measured values should be used. For sparingly soluble metals (e.g. Sb_2O_3) measured data on the dissolved fraction[5] are always required. If the solubility is exceeded the test has to be considered as unreliable. Results from tests where a visual precipitation is observed should be discarded. The absence of a visual precipitation does not exclude that sometimes colloids may still be present that could still affect the test results.

Concentration-effect relationships

With regard to the acceptability of the test results the following recommendations can be formulated (these recommendations are not metal specific):

- Minimal requirements for endpoints such as mortality, growth, reproduction (e.g. control mortality < 10 %) are often given in standard procedures. When these requirements are not met studies should be considered as not reliable.
- When adverse effects are observed in the different treatment groups a clear and consistent (increasing effect with increasing dose) concentration-effect relationship should be present. If no concentration-effect relationship can be established the test should be considered not reliable.
- Sometimes a hormesis effect is observed (i.e. increased performance in for example growth, reproduction) at low metal doses. Such effects can be important especially for trace nutrients such as Fe, Zn, Cu In such cases, as positive effects should not be considered in the derivation of EC_x and often other models than the conventional log-logistic dose-response model should be used to fit the toxicity data. For example the linear-logistic model of Brain and Cousens (1989) has been extended to allow EC_{50} and EC_{10} calculations (Van Ewijk and Hoekstra, 1993; Schabenberger et al., 1999, Cedergreen et al, 2005) in the case of hormesis.

[5] Different definitions for the dissolved fraction exist. Most often the dissolved fraction in ecotoxicity tests refers to the fraction that passes through a filter of 0.45 µm. It should be noted, however, that this definition may not necessarily refer to the metals in solution. In the range of 0.01-0.45 µm colloid inert particles that remain suspended may exist and these could account for 50 % or more of the “dissolved” 0.45 µm fraction

- Because effect concentrations are statistically derived values, information concerning the statistics should also be used as a criterion for data selection. If no methodology is reported and no raw data are reported or if values are 'visually' derived, the data have to be considered unreliable. In absence of sound statistics or no $L(E)C_x$ or NOEC has been calculated or reported in the study itself, the study could still be used if data are available amenable to further analysis that allow to derive a suitable $L(E)C_x$ or NOEC/LOEC value.
- Test concentration intervals should bracket the NOEC with concentrations that are as closely spaced as practical. Increasing the size of the test concentration intervals leads to reduced statistical power for the test. Following new OECD guidelines (e.g. OECD, 2001) test concentrations should preferably differ by no more than a factor of 2.

With regard to the proper use of NOEC/LOEC values and $L(E)C_x$ values the following recommendations can be made (these recommendations are not metal specific):

- For acute studies $L(E)C_x$ values should be estimated using appropriate statistical analysis (e.g. probit analysis or linear regression).
- For chronic studies concentration-response modeling such as regression models to calculate $L(E)C_x$[6]) are generally preferred over the classical hypothesis testing ($p< 0.05$) used to derive NOEC values. The latter method has indeed a number of limitations (Moore and Caux, 1997). Since the NOEC is by definition an applied dose, its value is to some degree dependent upon the choice of the experimenter. Secondly the NOEC depends upon the variability of the organism to a single dose. Organisms which are particularly sensitive to small variations in their environment, and hence display a greater variability of response to a given dose, are likely to have higher NOECs than if they were less sensitive, independent of their sensitivity to the toxicant. The use of a regression based approach offers the advantage that all of the information in the concentration-response curve is used and furthermore precludes the use of poor quality information because in those cases an inadequate model fit will be obtained.
- In case a benchmark dose ($L(E)C_x$) is calculated using a regression based approach and this value is to be used as an equivalent for a NOEC value, then typically a cut-off level should be identified representing a low effect percentile. This cut off value to be used should be derived based on the plausibility to detect a statistical significant difference and is depending on the inherent variability observed in the control test. The choice of the appropriate effect level is still an area under discussion and more research is needed A concentration that causes a low level of reduction, such as an EC_5 or EC_{10}, is rarely statistically significantly different from the control treatment. Therefore in some guidance documents the EC_{20} is sometimes proposed as a compromise representing a low level of effect that is generally significantly different from the control treatment (US-EPA, 1999a). Whatever effect level is chosen it is recommended that the $L(E)C_x$ value should not be extrapolated below the lowest applied (non-zero) concentration. According to Reiley et al (2003) and the draft ISO document (ISO, 2004) estimation of $L(E)C_x$ values outside the concentration range tested introduces a great deal of uncertainty. Furthermore for metals/metal compounds it is imperative that this value should fall within the range of tested concentrations to avoid extrapolating $L(E)C_x$ values below the natural background. If the resulting $L(E)C_x$ value should be below the lowest applied control level (background level) or essentiality level, its reliability/relevance has to be questioned (another confounding factor in this respect is the hormesis phenomenon which for essential metals can be very important).. Before estimating the $L(E)C_x$ value it should also be checked, case-by-case if the experimental design is appropriate to be used for regression methods. The statistical design needed for a proper $L(E)C_x$ derivation are more doses with fewer replicates at each dose. For estimating an $L(E)C_x$ value three concentration groups, as well as the control group, is an absolute (theoretical) minimum. However, if there are only three treatment groups and one fails to show any (partial) effect the test would be considered inadequate. Therefore more

[6] Usually $L(E)C_{10}$ values are selected, but the use of other $L(E)C_x$ values (e.g. $L(E)C_{20}$) could also be warranted

concentration groups are recommended in practice (ISO, 2004). Many of the older toxicity data do not fulfill the statistical requirements in order to derive an $L(E)C_x$ value. In those cases the conventional NOEC and LOEC values should be used. NOEC values could be in the natural background range but LOEC values should not.

- If only a LOEC ≥ 20% effect is reported (i.e. no NOEC could be derived as the lowest test group produced a response significantly different from the control group) and a distinct concentration/effect relationship is apparent, the $L(E)C_x$ is calculated or extrapolated and should be evaluated if it can be regarded as the NOEC. If the effect percentage of the LOEC is unknown, no NOEC can be derived. Such an approach is only recommended if insufficient bounded NOECs are available.
- In general, the use of unbounded NOEC values is not recommended. Unbounded NOEC values should only be considered in specific cases. For example, if other toxicity values are not available for a particular species. In that case an unbounded NOEC could be used as a conservative estimate for the 'real' NOEC.

4.2.1.2 Data relevancy

After qualifying effect data as reliable, it also should be checked for relevancy for hazard/classification purpose. This is a step that is particularly important for metals/metal compounds. For example the pH of the test medium may be outside the boundaries of the boundaries of the T/D protocol (i.e. 6-8.5). These relevancy issues should be considered carefully. A summary of the main attention points is given hereunder.

Biological relevancy of the endpoint used

- For classification purposes the use of standardized endpoints is recommended (GHS, 2003). If not enough data are available, -the use of non-standardized endpoints (enzyme activity, morphological changes, etc.) may be considered relevant if for example a plausible link to the overall health of the organism can be established.

Relevancy of the test substance

- Since impurities can have an effect on the toxic properties of the substance under investigation or have toxic effects themselves, studies involving test substances in which impurity levels are >1% should not be used. Soluble metal salts should be used for the purpose of classification. For the classification of inorganic metals/metal compounds, ecotoxicity data from organic metal compounds exposures should not be used.

Relevancy of the species

- Hazard classification is based on the comparison of the toxicity of standard species under standard conditions to allow an equal and comparative ranking for all substances in respect to fixed criteria. Therefore for hazard classification purposes standard species commonly used in internationally harmonized test methods (e.g. OECD 201, 202, 203, 210, 211 or equivalent) are preferred. Non-standard species should not be used since this could lead to non-comparative ranking/hazard classification for data rich substances. However, non-standard species could be used as supporting data in a weight-of-evidence approach to validate the classification derived on the basis of standard species.

Relevancy of exposure duration

- Both acute and chronic data can be used for the derivation of ecotoxicity reference values. Acute/chronic exposure depends upon the exposure duration and is also a function of the life cycle of the test organisms. *A priori* fixed exposure durations are therefore not relevant and should instead be related to the species, their typical life cycle and to the recommended exposure duration as described in standard ecotoxicity protocols (e.g. acute: 24/48 h for daphnids (OECD n° 202), 96 h for fish (OECD n° 203); chronic exposure (e.g. 7 days for Ceriodaphnids (ASTM, 2004), 21 days for daphnids (OECD, 1998), 30 days for fish (OECD, 1992), The 72h algal growth inhibition test is a chronic test but the EC_{50} is treated as an acute value for classification purposes. Following the latest OECD requirements, relatively short-term studies, focusing on sensitive life stages rather than focusing on the full life stage are also deemed chronic studies.
- When there is a lack of chronic data it may be possible to use acute data in combination with appropriate acute to chronic ratios. Quantitative ion character-activity relationships (QICARs) or quantitative cationic-activity relationships (QCARs) can be used in the complete absence of experimental data (Owny and Newman 2003, Walker et al. 2003.) as is the case for some data poor inorganic substances. However, more research efforts are needed in this field to develop and validate appropriate models. If no appropriate models are available the ecotoxicity reference value has to be derived from acute data.

Acclimatization/adaptation

- The fact that metal/metal compounds are naturally occurring substances should be taken into account when selecting toxicity data if phenomena such as acclimatization and adaptation are of importance. These concepts are described extensively in the MERAG background document. In short, due to the ubiquitous presence of metals in the natural environment, organisms have become conditioned to these backgrounds since they have evolved in the presence of the natural metal background concentrations. For this reason, exposure of organisms to the natural background level reflects in fact the theoretical lower limit of the predicted no effect concentration (PNEC) i.e. a concentration, which from an evolutionary perspective, does not present a risk to the survival of the species. This theory is applicable for all metals and is even more crucial for essential metals (EE[7]). As a result, the sensitivity of organisms to metals is determined to a large extent by the bioavailable concentration that the organism experienced before testing and their developed capability to cope with this concentration. Moreover, organisms cultured in media with a low essential metal concentration[8] may also exhibit an overall decreased fitness (deficiency issues) and become more sensitive to stress, including exposure to metals, even essential ones. Conversely, organisms cultured in media with elevated metal concentrations (both essential and non-essential metals, e.g. natural waters or contaminated waters) may become less sensitive. This phenomenon is related to the recently introduced “biogeochemical region” concept (Fairbrother and McLaughlin 2002).
- Consequently, ideally only those data sets should be used for classification purposes where (bioavailable) background concentrations in the culture medium (ideally both essential as non-essential metals) are representative for natural conditions suitable for the organism under testing. However, it is acknowledged that this type of information is rarely reported and hence difficult to use as a selection criteria. If the information is available (occurring especially for the major metals) the information can be used to consider not using test results where the organisms were cultured under natural background conditions that deviate from the conditions typically encountered in the environment. It is recommended that the essential metal concentration in the culture medium should be at least equal to the minimal concentration not

[7] An element is considered essential when (1) it is present in living matter; (2) it is able to interact with living systems; (3) a deficiency results in a reduction of a biological function, preventable or reversible by physiological amounts of the element (Mertz, 1974)

[8] This is especially the case in artificial media, since these media contain no or very little (essential) micronutrients.

causing deficiency for the test species used (lower boundary of OCEE, see background document). In case of multi-species tests (microcosm, mesocosm), the lower boundary of the No Risk Area (NRA) of the species tested could be used as the minimal concentration. Concentrations of non-essential metals should fall within the natural variation of these metals. Defining minimal levels of metal background for selection of relevant culture media should only be performed in case there is scientific evidence that acclimation/adaptation phenomena are relevant for the metal under investigation. If no direct information is available on the background concentrations of the metals in the culture medium, second line evidence (e.g. metal concentrations in river water used for maintaining the cultures could have been measured in other studies) can be used to support any decision taken on this issue.

- The concentration of essential metals in the culture medium could also have an effect on the tolerance of the test organisms for pH variation. For example Keating et al (1996) showed for *Daphnia magna* and *Daphnia pulex* that zinc-deprived animals in 96h acute tests had a narrower pH range of tolerance than non-deprived animals. These differences in sensitivity should be taken into account when selecting data outside the normal pH range (can be done if normalization models are available) for classification purposes.

4.2.2 Conclusion

Only ecotoxicity data that comply with the above-mentioned criteria can be considered valid and may be used for hazard/classification purposes for metals/metal compounds and alloys. However, the proposed quality criteria could be used in a flexible manner using expert judgment[9]. Main acceptability and relevancy criteria are a clear concentration-relationship, measured test concentrations, proper statistics, acceptable test performance and representativeness for the aquatic compartment. For transparency reasons, it should be clearly documented which studies are being rejected and on what ground.

4.3 Aggregation/selection of L(E)C_{50} /NOEC data

For data rich substances such as metals/metal compounds multiple data points can be available from reliable studies for a given species. These results will be subject to variability from several sources such as differences in geochemical characteristics of the test media, which can affect metal speciation and bioavailability, inter- and intra- laboratory variability, as well as inherent intra-specific heterogeneity in test organism sensitivity.

The most straightforward way to handle situations in which multiple data points exist for a given test species / endpoint, is to use the lowest value, e.g., the lowest NOEC/E(L)C_x. The use of the lowest value provides a conservative approach, especially when a wide range occurs between the lowest and highest data points for a given species. However, it should be realized that some of the lower toxicity values reported in literature may be the results of poor organism health, operational conditions or may just reflect differences in abiotic test conditions (bioavailability), and may therefore not reflect the intrinsic sensitivity of the organisms to a given toxicant.

When it is apparent from the data that the observed differences in test results for one species are due to differences in bioavailability in the test media then the use of the lowest toxicity value should be avoided whenever possible, and data aggregation approaches (grouping) should be used instead. In these approaches data are aggregated into geometric mean NOECs/EC_{10} values when multiple data are available from the same species, test duration and endpoints[10].

[9] For example information on metal concentrations in the culture medium will most often not be available. In those case and by lack of data the toxicity data of studies lacking this information could still be used.

[10] The Technical Guidance Document (2003) states that: “For equivalent data on the same endpoint and species, the geometric mean should be used as the input for the calculation.” In technical discussions, the meaning of “equivalent data” has been clarified to mean data collected from tests conducted under similar physical and geochemical conditions.

Prior to appropriate aggregation (grouping), it is recommended that intrinsic information relative to the effects of specific metals, e.g. mechanisms of toxicity and factors affecting bioavailability, be taken into account when aggregating data from multiple tests. Bioavailability differences should be taken into account by normalizing the data, prior to further processing, according to the best level of scientific knowledge available (e.g. organic carbon normalization, hardness correction, bioavailability models).

In case bioavailability models are available (e.g. Biotic Ligand Models) the scope of the data gathering can be broadened (provided that the models are validated over a broader range of conditions). General guidance on the principles and the way how bioavailability for the aquatic compartment can be incorporated are given in fact sheet 4.

When it is apparent from the data that the observed intra- species variability in toxicity test results can be assigned to differences in bioavailability and no bioavailability model is available to normalize the data, the effect data should be grouped by similar ranges of abiotic factors that control the bioavailability of metals. E.g. Toxicity data could still be split for reasons of comparison with the abiotic factors from the T/D protocol (e.g. pH 6-8, hardness).

If acclimation/adaptation is important test results should be grouped on the basis of the similarity of the background in the culture medium.

Summary grouping rules of selected data

In general, the following grouping rules can be applied:

- If for one species more than one[11] L(E)C_{50}/chronic NOEC values based on the same toxicological endpoint are available for a give species, these values are averaged by calculating the geometric mean, resulting in the "species mean" NOEC/L(E)C_{50}. In case of a flawed dataset: e.g. only two data points are available and one represents a very low value and another a high value it is recommended to repeat testing and take the geometric mean of all data.
- If for one species several acute L(E)C_{50}/chronic NOEC values based on different toxicological endpoints are available, the lowest value is selected. The lowest value is again determined on the basis of the geometric mean if more than one value for the same endpoint is available.
- In some cases, L(E)C_{50}/NOEC values for different life stages of a specific organism are reported in the same study. If from these data it becomes evident that a distinct life stage is more sensitive, the result for the most sensitive life stage is selected. The life stage of the organisms is to be indicated in the tables as the life stage at the start of the test (e.g. fish: yearlings) or as the life stage(s) during the test (e.g. eggs → larvae, which is a test including both the egg and larval stages).
- In case geometric means of different endpoints are given it is recommended to use the most sensitive endpoint.
- If acclimatization/adaptation is important test results should be grouped on the basis of the similarity of the background in the culture medium with the background found typically in the environment..
- When it is apparent from the data that the observed intra- species variability in toxicity test results can be assigned to differences in bioavailability and no bioavailability model is available to normalize the data, the effect data should be grouped by similar ranges of abiotic factors that control the bioavailability of metals. For example if the pH is driving the assessment toxicity data should be grouped in pH ranges around 6 (5.5-6.5) and 8 (7.5-8.5) relevant for pH 6 and 8 (8.5) used in the T/Dp for acute and chronic hazard classification respectively.

[11] In the GHS (2003) the use of a geometric mean is recommended for larger data sets (4 or more values)

4.4 Derivation of ecotoxicity reference values

It must be recognised that for hazard classification of new substances, acute ecotoxicity studies are to be carried out under very restricted conditions in laboratories with high quality standards (with a certificate for good laboratory practice (GLP)). The GLP certificate demands for standardization and control of the culture conditions, the test conditions and the species sensitivity. In practice, acute classification of new substances is based on the lowest ecotoxicity endpoint based on the results from one fish, and /or one invertebrate and/or one algae test, without any consideration of uncertainty. The EU/OECD demand for standardization and quality control of ecotoxicity tests to be used for hazard classification purposes is hence in contrast to the use of large ecotoxicity databases without quality and relevance control.

The choice of approach to assess the effect data set and derive acute and chronic aquatic reference values for metals, often with large ecotoxicity data sets, should therefore be based on a weight of evidence approach recognising aspects like the extent of the data set, the number of species covered, the reliability of the data, and appropriate statistical approach to assess the data. Care should be taken that the assessment is not more rigid for data rich substances (which is the case for some metals) which often contain one or two very low values not contained in small data sets.

The general scheme for the derivation of ecotoxicity reference values is outlined in Figure 1.

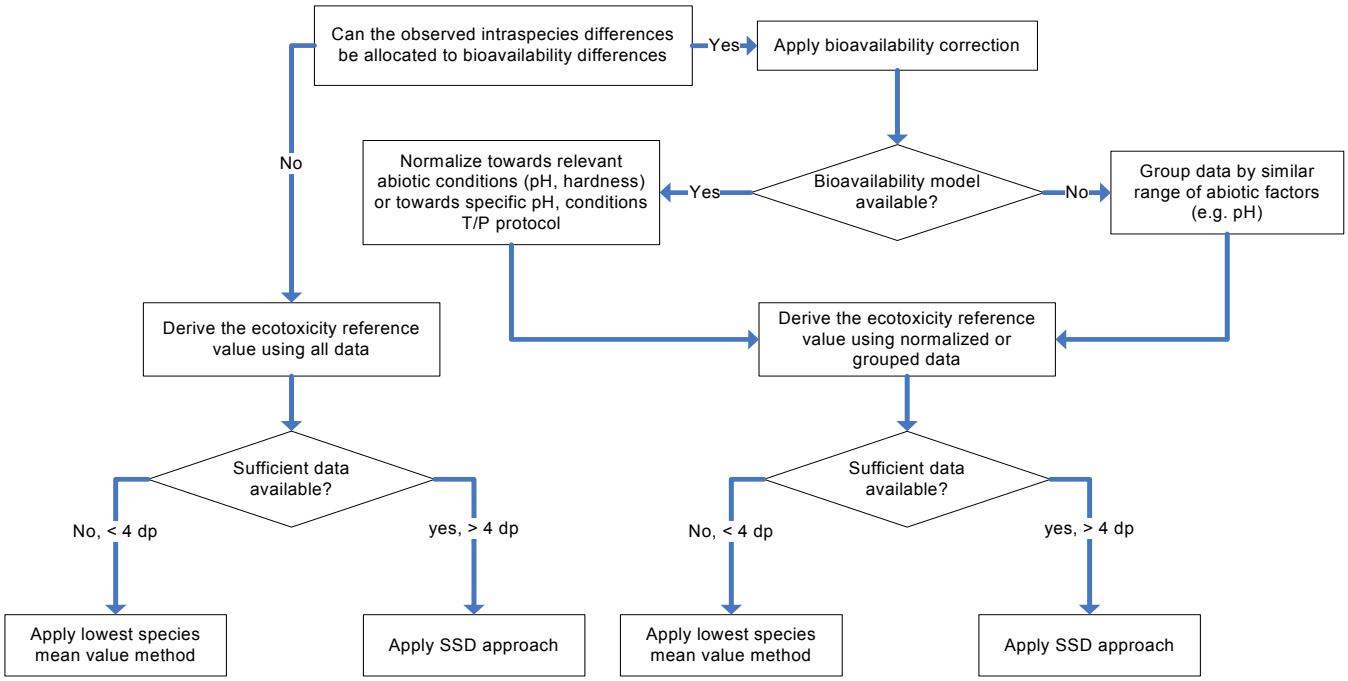


Figure 1: General scheme for the derivation of an ecotoxicity reference value for metal/metal compounds

When bioavailability is an issue and bioavailability tools (e.g. acute BLM) are available to normalize toxicity data towards a reference situation, pH specific reference values should subsequently be derived (e.g. for pH 6, 7 and 8 for acute classification purposes) prior to further treatment of the data. In case such a model is lacking, pH specific reference values could still be derived from grouped toxicity data generated at a specific pH value. If no data are available for a similar pH the comparison can be

made with data generated by a different pH. In case bioavailability is not an issue at all, the ecotoxicity reference value is derived using all data.

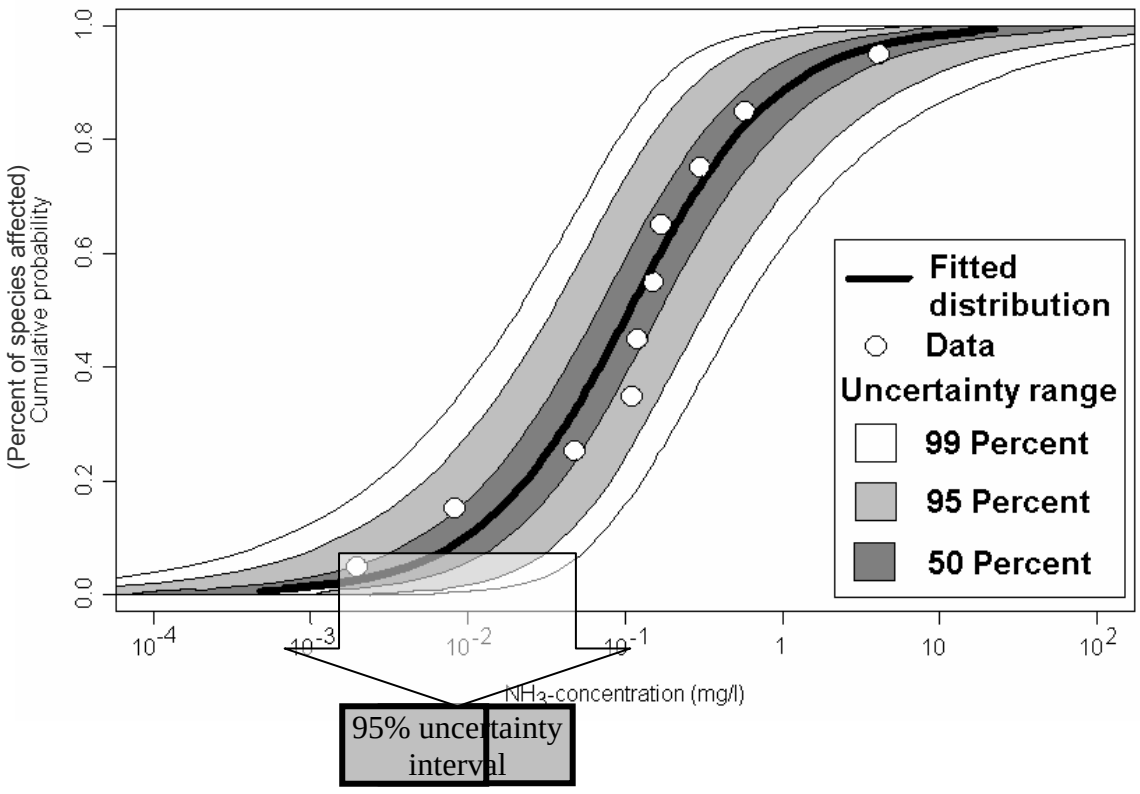
For small data sets (N < 4) the lowest acute geometric mean LC_{50} or EC_{50} or the lowest geometric mean chronic NOEC/L(E)C_x of the available data sets normally is used to define the hazard category. However, if on a case-by case evaluation it is apparent that enough data are available to define a species sensitivity distribution (SSD) with sufficient accuracy than a statistical extrapolation method using all available acute or chronic data could be preferred. Both methods to derive an ecotoxicity reference value are outlined below.

4.4.1 Lowest species mean value method

In this approach the lowest geometric mean value (acute L(E)C_{50} or chronic NOEC/L(E)C_x) for a species at a specific pH value should be used for classification.

4.4.2 Statistical extrapolation method

When sufficient acute/chronic data are available the use of the statistical extrapolation method could be considered. In the statistical extrapolation method the susceptibility of a set of species for a given toxicant can be described by some statistical distribution (i.e. Species Sensitivity distribution or SSD). A SSD can be visualized as a cumulative distribution function (Figure 2). The cumulative distribution function curve follows the distribution of the sensitivity data obtained from ecotoxicological testing, plotting effect concentrations derived from acute or chronic toxicity tests, for example LC_{50} values and No Observed Effect Concentrations (NOECs), respectively.



15^{th}-25^{th} percentile

Figure 2: Example of a SSD (Species Sensitivity Distribution - log logistic distribution) with uncertainty band and an HC_{15}- HC_{25} (Hazardous Concentration at 15-25 %) as possible ecotoxicity threshold for classification purposes.

As for the lowest species mean value method if multiple data are available for one species/endpoint and it is apparent from the data that the observed difference in test results for one species is due to differences in bioavailability the effect data should preferentially be normalized prior to further processing. If no bioavailability tools have yet been developed for the metal/metal compound under consideration, a pre-selection of the aquatic toxicity data according to pH should be performed.

An appropriate question to consider while evaluating the data set as a candidate for the statistical extrapolation approach is 'how many data are needed?' to fit a Species Sensitivity Distribution (SSD) model with sufficient confidence using all available acute/chronic NOEC values as input. Note that if multiple test results on the same species is available the geometric 'species mean' can be used as input in the SSD or with regard to the loss of relevant information by using a SSD constructed with species geometric means it could be deemed more appropriate to use the individual data in a weighted approach for environmental classification purposes.

Generally, the larger the sample size, the greater one's confidence in the choice of a probability distribution and the corresponding estimates. Conversely, for small sample sizes, goodness-of-fit statistics will often fail to reject any of the hypothesized probability distribution function. In general, there is no rule of thumb for the minimum sample size needed to specify a SSD. Increasing sample size may however be an important consideration when making decisions about uncertainty related to the use and relevance of the SSD (US EPA, 1999).

4.4.2.1 Choice of the appropriate distribution model

Both parametric (e.g. Log-normal, Weibull distributions, etc.) and nonparametric distributions can be used to characterize the SSD. Although it is often impossible to exclude *a priori* any distribution there may be mechanistic reasons that may dictate the shape of the "real" distribution. For example from a conceptual viewpoint the use of threshold models can be considered in the case of natural elements such as metals. Indeed assessment of metal SSDs requires consideration of several unique aspects, such as background concentrations, which organisms have evolved with, and essentiality for normal metabolic functions. Metals taken up by active transport have a threshold metal concentration below which the organism cannot uptake the metal from the environment. Accordingly, the *a priori* use of a model such as the normal (or log-normal) distribution, with tails extending to infinity, may result in unrealistically low HC_5 estimates that are within the range of typical background concentrations or, in the case of essential metals, potentially HC_5 estimates that may lie within the range of metal deficiency for some organisms.

Distribution types largely driven by central tendency and variability in toxicity values, such as the mean and standard deviation for the normal distribution, can also result in very long SSD tails if the toxicity data are highly variable. This is also an important issue for metals because many metals are regulated by aquatic organisms differently. Some organisms are highly tolerant because they can store the metals in non-toxic forms and other organisms are very sensitive because they do not have the same detoxifying mechanisms, thereby resulting in a range in toxicity values for aquatic organisms that can be very large. Practical assessment of the resulting elongated lower tail is further confounded by the fact that many metals tend to exhibit a threshold response in the lower tail, as discussed above. For this reason, it has been suggested that a threshold model for SSD development may be more appropriate for metals in general, and essential elements in particular (Brix et al. 2001; Van Straalen 2002; Van Sprang et al. 2005). Van Straalen (2002), for example, found that the triangular distribution provided the best fit of four finite distributions fit to zinc toxicity data, while Brix et al. (2001) and Van Sprang et al. (2005) used a Pareto model to characterize the threshold response observed in chronic copper and zinc toxicity data, respectively.

Other factors could also constrain the choice of a distribution. For example, as unrealistic values (e.g. NOEC values above the solubility product of the considered metal) may bias the estimation of the

threshold value, truncating the tails of a distribution should be considered. In all cases, it is essential to explain clearly and fully the reasoning underlying the choice of a specific distribution.

It is preferable to select functions based in goodness-of-fit or other statistical comparisons of alternative functions. Goodness-of-fit tests (e.g. Anderson-Darling and Kolgomorov-Smirnov tests) are formal statistical tests of the hypothesis that the data represent an independent sample from an assumed distribution. These tests involve a comparison between the actual data and the theoretical distribution under consideration. The calculated goodness-of-fit statistic measures how good the fit is: critical values are calculated and used in order to determine whether a fitted distribution should be accepted or rejected at a specific level of confidence. Typically, these values depend on the type of distribution fit, the number of data points and the confidence interval. The level at which one distinguishes between likely and unlikely values of the test statistic is a matter of judgement. A significance level of 0.05 is most often used, implying that a value of the test statistic below the 95th percentile of the distribution for the statistic is acceptable and leads to the inability to reject the hypothesis. A value of the calculated A-D/K-S statistic above the 95th percentile of the distribution leads to the rejection of the null hypothesis, i.e. the distribution is not a good fit (Cullen & Frey, 1999). In case of lack of fit at the 95% confidence level, the statistical extrapolation method should not be used.

The (A-D) test places most emphasis on tail values whereas the (K-S) test investigates the data fit for the whole distribution curve to the same extent. Care must be taken when evaluating results of best-fit analyses, since one goodness of fit test statistic (e.g. A-D) may indicate that one distribution offers the best fit, while another goodness of fit test statistic (e.g. K-S) may indicate that a different distribution has the best fit. This can influence the choice of the distribution, and also the derivation of the ecotoxicity threshold. Anyway it is recommended that SSD functions should not be too complex (2-3 parameters functions are preferred)[12].

4.4.2.2 Incorporation of bioavailability in the SSD

Improving the comparability of the reference value with the conditions (e.g. hardness, dissolved organic carbon (DOC)) or the pH ranges used to generate transformation/dissolution data could be achieved by considering bioavailability. Abiotic factors including alkalinity, ionic strength, and pH can influence the toxicity of metals in two ways: (i) by influencing the level of dissolved free ions of the metal in water (chemical speciation) and (ii) by influencing the uptake and binding of available metal by biological tissues (competition effects).

Interaction of metals with organic and inorganic ligands in test media and natural environments can be assessed from metal speciation models (for example WHAM, PHREEQ etc.). Such models calculate the uncomplexed and complexed fractions of the metal ions. Recently, more advanced bioavailability models have been developed such as the Biotic Ligand Model (BLM) allowing for the calculation of the concentration of metal ion responsible for the toxic effect at the level of the organism. The biotic ligand model (BLM) integrates data from speciation models with the interactions between metal ions and other competing ions at binding sites on the organism-water interface (e.g. epithelial cells of gill tissue. The BLM model has at present been developed and validated for a limited number of metals, organisms, and endpoints. Acute BLM models exist for Cu, Ag, Cd, Zn, Ni and Pb. Chronic BLM models have been developed for Cu, Zn, Ni. An extensive review on the concepts of the different existing BLM models and the model parameters can be found in Niyogi and Wood, (2004). A more concise overview on the conceptual framework, the general uptake principles and limitation of the BLM model is given in the background document.

The tools mentioned above can be used to normalize toxicity data towards the hardness/DOC or to the pH conditions dictated by the dissolution/transformation protocol. If used, the models and formulae used

[12] In statistics, overfitting is fitting a statistical model that has too many parameters. An absurd and false model may fit perfectly if the model has enough complexity by comparison to the amount of data available. A perfect fit can therefore always be obtained by using for example a high degree polynomial distribution. However, one should not forget that the NOECs in a SSD represent only a small sample of all sensitivities encountered in an ecosystem and as such the true distribution of sensitivities will always be unknown.

for the characterization of metal bioavailability in the test media should always be clearly reported, allowing for their translation back to natural environments.

According to GHS (2003), both ecotoxicity and transformation should be evaluated over the pH range 6-8 and 6-8.5 for acute and chronic hazard classification, respectively. In cases where data are available, ecotoxicity and transformation should be compared at the same pH. When data sets are not complete for the different pH values, data obtained at different pHs can be compared. With regard to the other physico-chemical parameters it is recommended that values are chosen to be similar to realistic worst case conditions found in the aquatic environment. For example, for cationic metals, the hardness levels of the OECD transformation medium and dissolved organic carbon (DOC) as encountered in the OECD transformation/dilution medium (DOC max 2 mg/L) could be used.

4.4.2.3 Derivation of the ecotoxicity reference value using a SSD approach

The use of a SSD approach has already shown its benefits in the context of deriving safe environmental thresholds in a risk assessment framework and environmental quality standard setting (fact sheet 3). However, its use for classification purposes is still new. For ranking purposes the choice of the 5^{th} percentile as an ecotoxicity reference value could be to stringent in comparison with the lowest value method. The 20th percentile value of a SSD (HC_{20}) as a reference value corresponds to the lowest ecotoxicity value of 3 available, ecotoxicity values. This percentile is based on a comparison between different plotting positions (i.e. methods to determine percentiles of data points). The 20th percentile is the average of the most common plotting system (mean) and the scientifically recommended most accurate method (Hazen). Therefore, a cut-off value of HC_{20} values is recommended for classification purposes.

5. APPLICATION OF CLASSIFICATION CRITERIA TO METALS, METAL COMPOUNDS AND METALLIC ALLOYS

5.1 Introduction

Metals, metal compounds and metallic alloys can exist in a variety of physical forms, most commonly as powders (metallic compounds) or as "massive" forms (metals and metallic alloys). These different physical forms may have different transformation/dissolution rates in aqueous media and hence different potential to cause aquatic toxicity. Therefore the potential environmental (aquatic) hazards may need to be assessed for both powders and massives separately which might result in different classifications for each type.

With regard to classification in general the following approaches as outlined in Figure 3 can be followed.

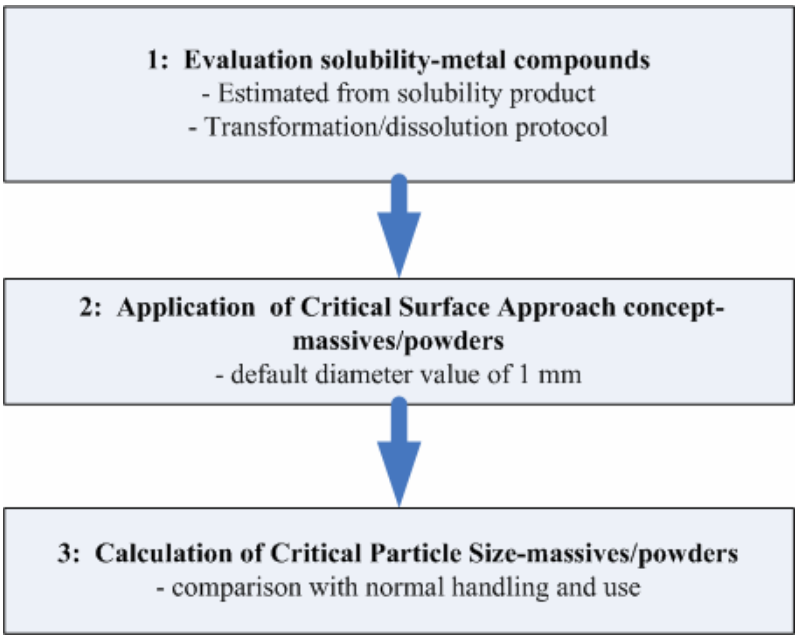


Figure 3: Generic approaches for the classification of metals, metal compound and alloys

In Tier 1 the solubility of the metal/metal compound/alloy is evaluated. Depending on the identification as soluble, sparingly soluble or massive different classification strategies can be developed. For soluble metal compounds, the reference toxicity values are obtained from testing of soluble salts and read across classification of soluble inorganic metal compounds can be done by comparing the soluble metal ion concentration (µg Me/L) causing the appropriate standard ecotoxicity effect (acute, chronic) and translation of the results towards the compound under investigation using the molecular weight ratio (MW substance/MW metal ion). After substance-specific molecular weight adjustments, comparison with the classification cut-off values (1-100 mg/L) allows for appropriate classification of different soluble inorganic metal compounds. Information on bioaccumulation and partitioning from the water column can be used to alter the chronic classification entries.

In case the solubility is limited (metals/alloys and SSMC) the classification strategy is based on the comparison of the T/D data with the selected ecotoxicity reference value. The crucial question that has to be answered is: at which loading rate is the concentration of the dissolved metal ion greater or equal to the derived ecotoxicity reference value (L(E)C_{50}-NOEC), based on the soluble metal ion concentration, adjusted for molecular weight as needed. In order to allow reading across physical metal forms, the Critical Surface Area concept can be used to enable self classification of powders and massives. For appropriate classification of massives/powders, further refinement consists of the comparison of the calculated critical particle size with normal handling and use. The subsequent sections cover in detail the proposed testing strategies and the classification route for soluble-,

sparingly soluble metal compounds, metals and metallic alloys, respectively. The presented schemes are built around the EU classification strategy for metals and metal compounds (67/548/EEC, annex 6 w L225/263) but GHS terminology has already been introduced to broaden the scope and to facilitate the comparison with the GHS scheme. Please note that due to the lack of clarity of the GHS classification text with regard to the implementation there are some items with option for discussion. This is in particularly so with the removal of the R53. At the moment the methodology as is used in the EU has been adapted in this fact sheet.

5.2 Classification Strategy for metals and metal compounds

The strategy for the classification of metals and metal compounds is summarized in Figure 4.

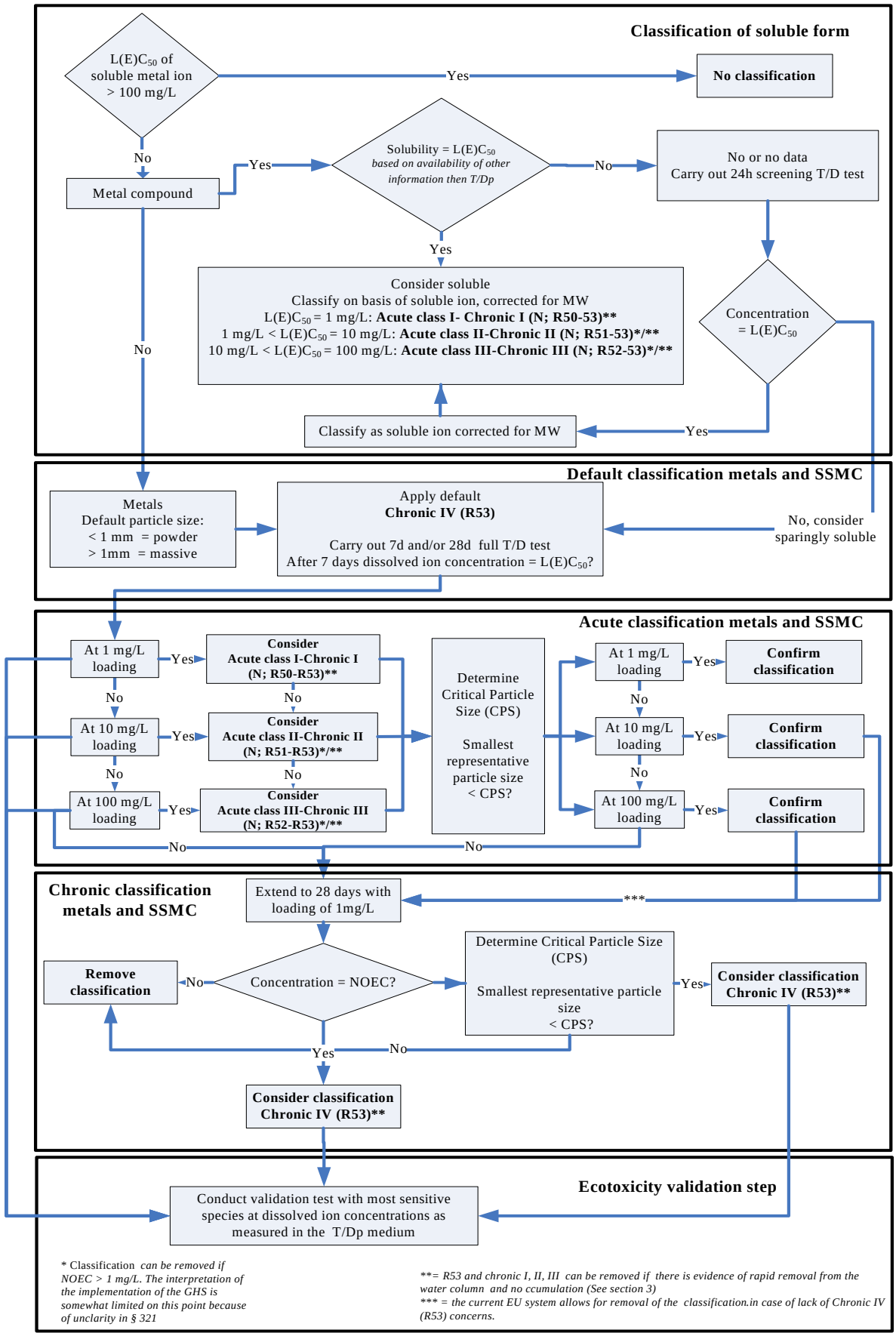


Figure 4: Classification strategy for metals and metal compounds

5.2.1 Classification Strategy for metal compounds

- Where the L(E)C_{50} for the metal ions of concern is greater than 100 mg/L, the metal compounds need not be considered further in the classification scheme (Figure 2).
- All metal compounds with a water solubility (either measured, e.g. through 24h Dissolution Screening test or estimated, e.g. from the solubility product) greater or equal to the L(E)C_{50} of the dissolved metal ion concentration[13] are considered as soluble metal compounds (a MW translation from the dissolved metal ion to the metal compound is to be done[14]). Care should be exercised for compounds whose solubility is close to the acute toxicity value as the conditions under which solubility is measured could differ significantly from those of the acute toxicity test. In these cases the results of the Dissolution Screening Test are preferred. Where data are available from the screening test detailed in the T/Dp, the maximum solubility obtained over the tested pH range should be used. Where data are not available over the full pH range, a check should be made that this maximum solubility has been achieved by reference to suitable thermodynamic speciation models or other suitable methods. If other solubility data are available to show that the dissolution concentration would not exceed the L(E)C_{50} across the entire pH range then the substance should not be classified on its soluble form.

5.2.1.1 Soluble metal compounds (SMC)

SMC are classified on the basis of the acute L(E)C_{50} and/or Chronic NOEC (corrected where necessary for molecular weight). Under the GHS scheme the following classification classes are distinguished (the classification classes under the EU system are indicated between brackets).

(i) If the L(E)C_{50} of the dissolved metal ion is less than or equal to 1 mg/L then classify Acute class I. Classify also as chronic I unless there is evidence of both rapid partitioning from the water column and no bioaccumulation[15]; (similar to R50-R53 in the EU system)

(ii) If the L(E)C_{50} of the dissolved metal ion is greater than 1 mg/L but less than or equal to 10 mg/L then classify Acute class II. Classify also Chronic II unless (1) there is evidence of both rapid partitioning from the water column and no bioaccumulation; (similar to R51-R53 in the EU system) or (2) the chronic NOEC>1mg/L.

(iii) If the L(E)C_{50} of the dissolved metal ion is greater than 10 mg/L and less than or equal to 100 mg/L then classify Acute class III. Classify also as chronic III unless (1) there is evidence of both rapid partitioning from the water column and no bioaccumulation (similar to R52-R53 in the EU system) or (2) the chronic NOEC>1mg/L.

(iv) If solubility < L(E)C_{50} classify default as Chronic IV (similar to R53 in the EU system) unless chronic NOEC>1mg/L.

5.2.1.2 Sparingly soluble metal compounds (SSMC)

In the context of the classification criteria, sparingly soluble compounds of metals are defined as those with a known solubility (either measured e.g. through 24-hour Dissolution Screening test or estimated e.g. from the solubility product) less than the pH specific L(E)C_{50} of the soluble metal ion. In those cases

[13] T/D data and L(E)C_{50} should be compared at equal pH level. A default procedure using the lowest EC_{50} value independent from the pH may be applied if ecotoxicity information is lacking to derive pH specific EC_{50} values

[14] Thus L(E)C_{50} metal compounds = L(E)C_{50} of metal x (Molecular weight of metal compound/Atomic weight of metal)

[15] The interpretation of the implementation of the GHS system is somewhat limited on this point because of unclarity provided by § 321. However, applying the principle described in the GHS of rapid removal from the water column it is suggested to include a parallel reasoning for metals, i.e.see section 3.

when the soluble forms of the metal of sparingly soluble metal compounds have a pH specific $L(E)C_{50}$ less than or equal to 100 mg/L and the substance can be considered as sparingly soluble the default safety net classification (Chronic IV or R53) should be applied or the substance should be evaluated using the 7-d and 28-d T/D protocol (see classification strategy metals below) and applying the molecular weight correction.

5.2.2 Classification Strategy for metals[16]

The strategy for the classification of metals is summarized here:

- Where the $L(E)C_{50}$ for the metal ions of concern is greater than 100 mg/L, the metals need not be considered further in the classification scheme (Figure 2).
- Where the $L(E)C_{50}$ for the metal ions of concern is less than or equal to 100 mg/L, consideration must be given to the data available on the rate and extent to which these ions can be generated from the metal. Such data, to be valid and useable should have been generated using the T/Dp (Annex 9 GHS, 2003).
- Where such data are unavailable, i.e. there are no clear data of sufficient validity to show that the transformation to metal ions will not occur; the safety net classification (Chronic IV or R53) should be applied since the known classifiable toxicity of these soluble forms is considered to produce sufficient concern. Or T/D testing should be conducted.
- If it can be demonstrated that the metal ions are readily removed from the water column and not bioaccumulated (cf section 3), than the chronic classification entry can be removed.
- Where data from the T/Dp are available, the results should be used to aid classification according to the following rules (note T/Dp results can also be used for classification of sparingly soluble metals):

[16] Guidance given on the interpretation T/D test results is also applicable to the classification of sparingly soluble metal compounds

7-day Transformation Test

If the dissolved metal ion concentration after a period of 7 days (or earlier) exceeds that of the $L(E)C_{50}$, then the default classification for the metals is replaced by the following classification as given under the GHS scheme (the classification classes under the EU system are indicated between brackets):

(i) if the dissolved metal ion concentration at the low loading rate (1 mg/L) is greater than or equal to the pH specific $L(E)C_{50}$, then classify Acute class I. Classify also as chronic I unless there is evidence of both rapid partitioning from the water column and no bioaccumulation[17]; (similar to R50-R53 in the EU system)

(ii) If the dissolved metal ion concentration at the medium loading rate (10 mg/L) is greater than or equal to the pH specific $L(E)C_{50}$, then classify Acute class II. Classify also Chronic II unless there is evidence of both rapid partitioning from the water column and no bioaccumulation; (similar to R51-R53 in the EU system)

(iii) If the dissolved metal ion concentration at the high loading rate (100 mg/L) is greater than or equal to the pH specific $L(E)C_{50}$, then classify Acute class III. Classify also as chronic III unless there is evidence of both rapid partitioning from the water column and no bioaccumulation (similar to R52-R53 in the EU system)

28- day Transformation Test

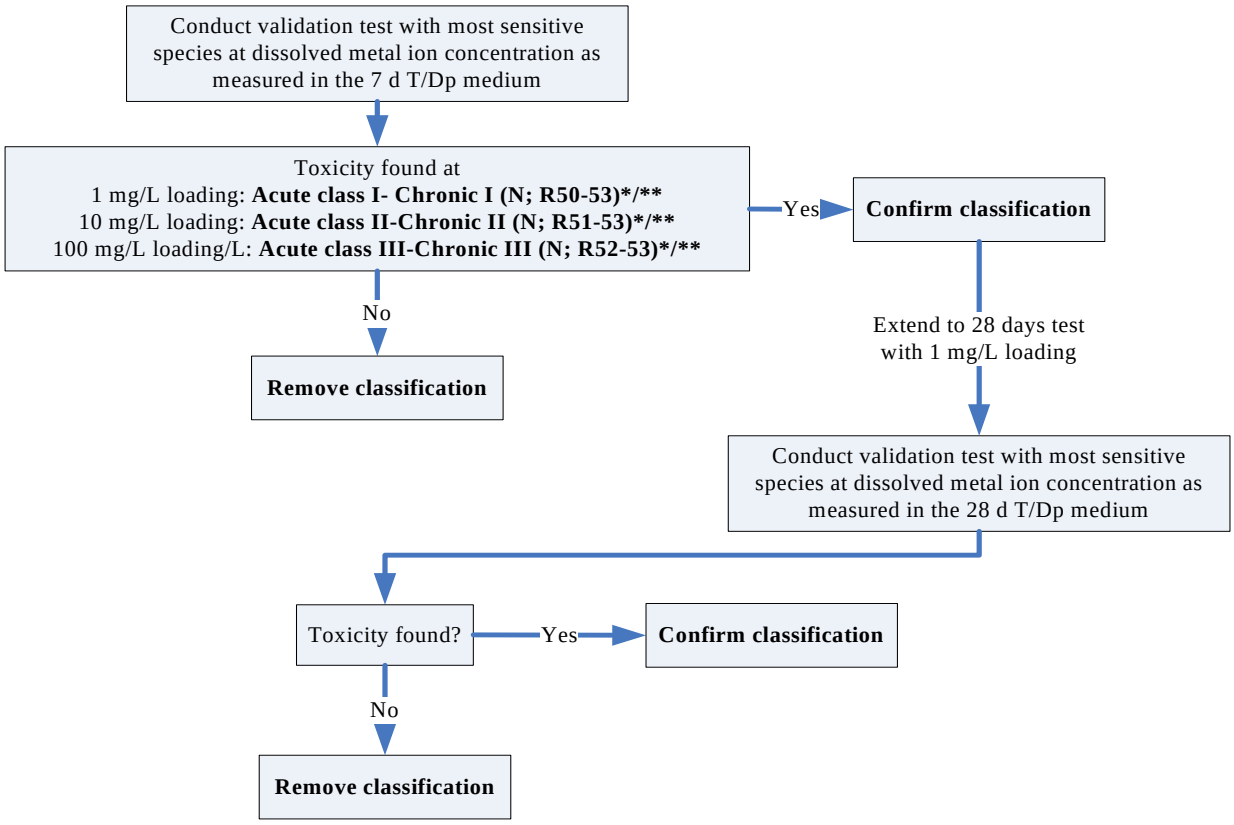
If the process described in previous section results in the classification of chronic 1 (R50-R53 in EU), no further assessment is required as the metal/metal compound will be classified irrespective of any further information. In all other cases, further data may have been generated through the dissolution/transformation test for 28 days in order to show that the classification may be amended. Under the GHS system and EU system if for poorly soluble metal compounds classified as chronic II (similar to R51-R53 in EU), chronic III (similar to R52-R53 in EU) or chronic IV (similar to R53 in EU) the dissolved metal ion concentration at the 1 mg/L loading rate after a total period of 28 days is less than or equal to the long-term NOECs, then classification is removed.

Within the GHS classification scheme all acute classes (I-III) can exist as stand alone classifications. For the EU scheme only R50 and R51 may exist as stand alone acute classification.

5.2.3 Ecotoxicity validation step

Finally an ecotoxicity validation step is suggested in cases where a significant uncertainty is associated with the existing toxicity data. This ecotoxicity validation should be conducted with the most sensitive species at dissolved ion concentrations equivalent to those measured in the T/Dp medium. However, ecotoxicity testing directly in the T/Dp medium is not recommended because the composition of this medium is unlikely to meet the requirements for standard test media to ensure proper survival and/or reproduction (e.g. for algae, pH 6 will already cause a reduced growth).. Therefore, ecotoxicity tests should be conducted in standard media dosed at metal concentration equivalent to the concentration level actually measured in the T/Dp medium. These tests can be initiated for metals and sparingly soluble metal compounds classified as Acute I-chronic I (similar to R50-R53), Acute II-chronic II (similar to R51-R53 in EU), Acute III-chronic III (similar to R52-R53 in EU). When toxicity is found, the metal/metal compounds are classified accordingly. If no toxicity is found classification is removed.

[17] see section 3



* Classification *can be removed if NOEC > 1 mg/L. The interpretation of the implementation of the GHS is somewhat limited on this point because of unclarity in § 321*
***= R53 and chronic I, II, III can be removed if there is evidence of rapid removal from the water column and no ccumulation (See section 3)*

Figure 5: Ecotoxicity validation step

5.3 Classification of metal powders, metal massives (and alloys (powders and massives)[18]) according to the Critical Surface Approach (CSA) concept

The ***critical surface approach*** concept as originally introduced by Skeaff et al (2000) is aimed at enabling self-classification for different powders/massive forms of metals and SSMCs. The concept is based on establishing a correlation between *the dissolved concentrations of the metal ion,* as measured in a T/D test after a specified time interval, and the surface area loadings of the physical forms tested. From this relationship the ***critical surface loading (CSL)*** can be determined. CSL is defined as *the surface area of the solid per liter of aqueous medium that, after a given period of time, will generate the reference L(E)C_{50} of the metallic (bioavailable) ion*. This reference toxicity value is determined from ecotoxicity tests using one of the metal's soluble salts and the most appropriate (sensitive) standard test organism.

A schematic overview of the different steps is given in Figure 6.

[18] A thought starter on a potential specific classification Strategy for metallic alloys is provided under Annex 2

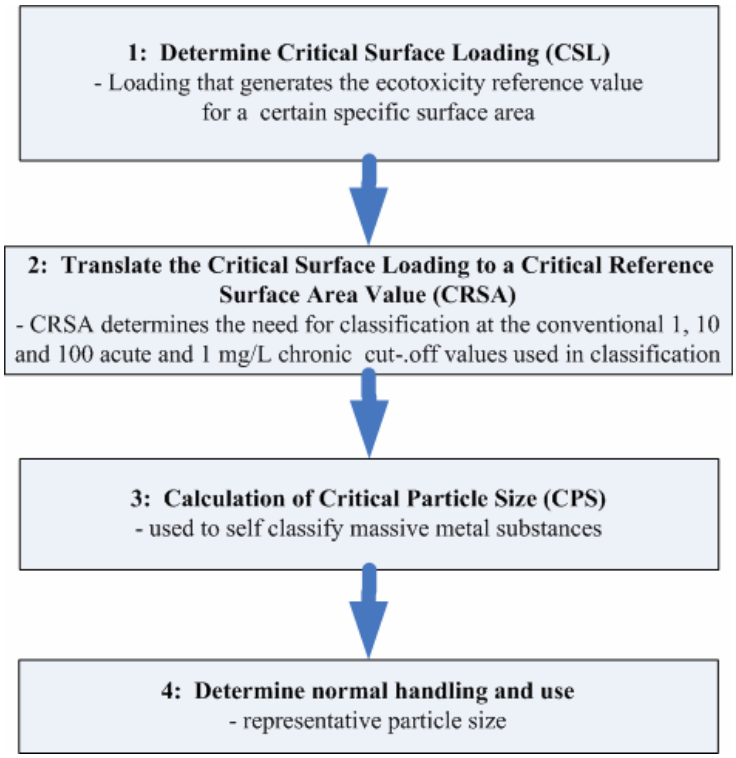


Figure 6: General overview of the different steps involved in the Critical Surface Approach

Dissolution kinetics of metals and SSMC's are driven by the mass loading of the particles. As such, the mass loading that generates the ecotoxicity reference value (section 4.4) for a certain specific surface area needs to be determined as a first step. Subsequently this Critical Surface Loading (CSL) can be used to calculate the ***Critical Reference Surface Area (CRSA)*** that triggers classification at the conventional 100 and 1 mg/L cut-.off points used in acute and chronic hazard classification. These CRSA values can be further translated into a ***Critical Particle Size (CPS)*** which can be used to establish self-classification levels. This self-classification should be done on a particle size representative for normal handling and use.

The different steps are explained in more detail in the following sections. The concepts can in principle easily be extended to metallic alloys (see annex 2).

5.3.1 Calculation of the Critical Surface loading (CSL)

Derivation of the critical surface loading (dissolved metal concentration) is obtained according to the T/Dp conducted at a certain pH and is plotted as a function of the corresponding surface area loadings on a log-log basis (Figure 7). These loadings expressed as mm^2/L can be obtained by multiplying the particle specific surface area (mm^2/g) and the loadings tested (g/L) (Eq-2):

$$SL_a = SSA \times SL_m \qquad \text{(Eq-2)}$$

SL_a = surface loading (mm^2/L)

SSA = specific surface area (mm^2/g)

SL_m = substance mass loading (g/L)

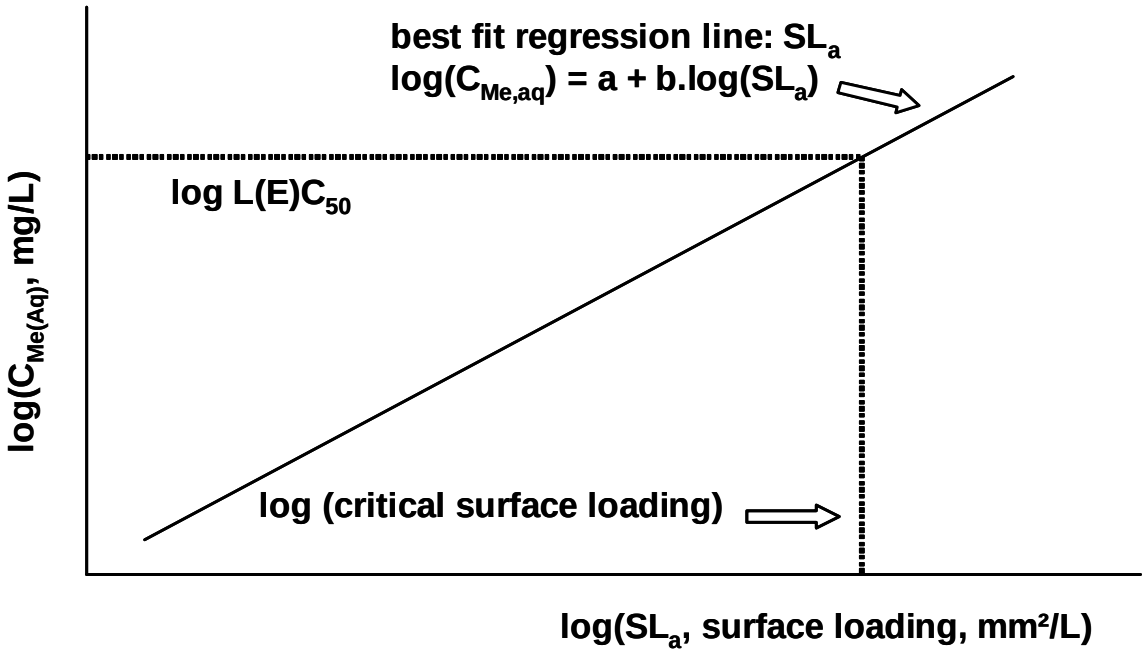


Figure 7: Determination of critical surface loading from regression line of log concentration of dissolved metal versus log surface loading (adapted from Skeaff et al, 2000)

From the best-fit regression line (Eq-3)[19] and by putting the reference ecotoxicity value into the function the Critical Surface Loading can be calculated (i.e. the surface loading at which the reference ecotoxicity value is just reached (Eq-4).

$$\log(C_{Me(aq)}) = a + b\log(SL_a) \qquad \text{(Eq-3)}$$

$C_{Me(aq)}$ = dissolved metal concentration (mg/L)

a,b = regressions coefficients

[19] Transformation relationships in which the regression parameter r^2 is greater than 0.9 and the standard errors in the regression coefficients are less than about 5 % can be used with confidence.

$$CSL_a = \left(\frac{-\log(L(E)C_{50} - a}{b} \right) \quad \text{(Eq-4)}$$

CSL_a = Critical surface loading (mm^2/L)

It must be noted that, Skeaff et al (2000) initially observed log-log relationships between the surface loading (mm^2/L) and the dissolved metal concentrations when evaluating powder transformation/dissolution data. More recent investigations of T/D of massives & powders demonstrated that the transformation/dissolution function was often linear and could hence be described as:

$$C_{Me(aq)} = a + b(SL_a) \quad \text{(Eq-5)}$$

5.3.2 Calculation of the Critical Reference Surface Area (CRSA) that triggers classification

Since particle size, or more specifically, *surface area*, is a crucial dissolution parameter any variation in the particle size or surface area tested may cause a significant change in the levels of metals ions released in a given time. Hence, the development of the critical particle size defines the potential for a given substance under a given set of experimental conditions. For the purposes of the transformation test, the particle size or surface area for a given metal substance is fixed (i.e., it is a property of that substance). This allows for classification to be based *solely on the loading levels* of the tested metal substance that delivers the metal in solution in excess of the 1, 10 and 100 mg/L cut-off values..

Therefore the critical surface loading can be converted into a ***Critical Reference Surface Area*** that correspond to the conventional 100, 10 and 1 mg/L cut-off points used in acute and chronic hazard identification. In general Eq-6 can be used to calculate the Critical Reference Surface Area (CRSA)

$$CRSA = \left(\frac{10 \times CSL_a}{CP} \right) \quad \text{(Eq-6)}$$

CRSA = Critical Reference Surface Area(cm^2/g) that triggers classification

CSL_a = Critical Surface Loading (mm^2/L)

CP = classification cut-off point (1, 10, 100 mg/L)

5.3.3. Calculation of the Critical Particle Size (CPS) of a spherical particle

The Critical Reference Surface Area of a particle can be further translated into a ***Critical Particle Size (CPS),*** i.e. the size of a particle (massive, powder or alloy) above which an insufficient amount of ions is produced in an aqueous medium at the highest loading of the metal (i.e. is 100 mg/L for acute[20] and 1 mg/L for chronic) to reach the ecotoxicity reference value (i.e. above which a downgrading in classification or a no classification result is obtained).

For standardization it is assumed that the particle size is spherical so that the CPS can be expressed as a diameter. The benefit of this approach is that it allows for comparison on an equal basis between different metals and particle sizes of the same metal.

Hence, the CPS corresponds with a ***Critical Diameter (CD)*** according to the following reasoning:

The surface (SA_{sphere}) and volume (V) of a sphere[21], is given by Eq-7 and Eq-8, respectively:

$$SA_{sphere} = \prod \times D^2 \qquad \text{(Eq-7)}$$

$$V = \left(\frac{\prod \times D^3}{6} \right) \qquad \text{(Eq-8)}$$

Combining these equations yields Eq-9:

$$\frac{SA_{sphere}}{V} = \left(\frac{6}{D} \right) \qquad \text{(Eq-9)}$$

Similar the SA/V ratio for the respective Critical Reference Surface Area can be calculated using Eq-10

$$\frac{SA_{sphere}}{V} = CRSA \times \rho_{Me} = \frac{6}{CD} \qquad \text{(Eq-10)}$$

SA_{sphere} = surface area sphere (cm^2)

V = volume sphere (cm^3)

CRSA = Critical Reference Surface Area (cm^2/g) that triggers classification

ρ_{Me} = density metal (g/cm^3)

CD =critical diameter (cm)

By solving Eq-10 the critical diameter (CD) can be obtained.

[20] A subsequent refinement can be to calculate the CPS also at the 10 and 1 mg/L loadings

[21] Although there is no assumption made in the T/D protocol that massives are spherical the cut-off for massive is defined by default as 1 mm diameter particle. Therefore a sphere is the most logic interpretation of a particle with a diameter of 1 mm as used in Skeaff et al. (2000). If other particle shapes are used other specifications could be required. In that case the more general term Critical Particle Size (CPS) is used

5.3.4 Classification strategy using the CRSA/CD concept (Critical Surface Concept)

Dir. 67-548 as well as REACH require that producers and importers self-classify the substances they place on the market. Both the specific surface area as the critical diameter can be used to establish self-classification levels for the metal under investigation as shown in Figure 8.

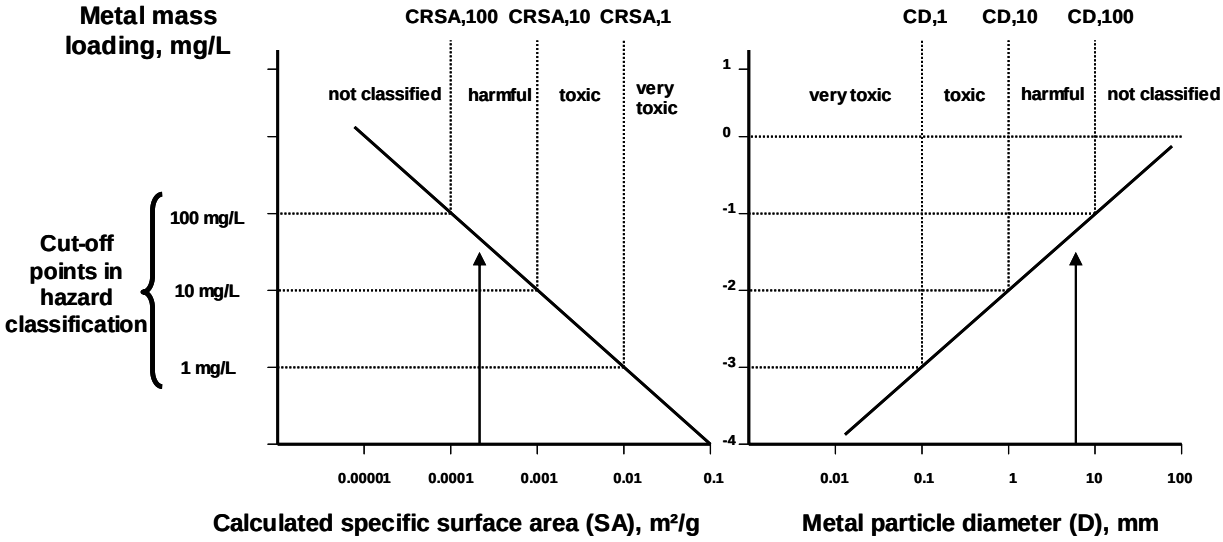


Figure 8: Illustration of hazard identification line and resulting self-classification levels for a metal with (a) various calculated critical reference surface area values and (b) corresponding diameters (CD) (adapted from Skeaff et al, 2000)

Massive forms/powders that have a surface area larger than the critical reference surface area at a certain pH at the specific acute loadings (1, 10, 100 mg/L) should be classified acute Class I, II and III (R50-R51 or R52) respectively. Or similar massive forms/powders that have a diameter smaller than the critical diameter at the specific acute loading (1, 10, 100 mg/L) should be classified acute Class I, II and III (R50-R51 or R52), respectively.

For chronic classification the surface area or diameter should be compared with the respective chronic critical reference surface area or the chronic critical diameter.

5.3.5 Normal handling and use

The classification of metals or metallic alloys in massive form should be done on a particle size representative for normal handling and use. As indicated before in the absence of information, a 1 mm default size should be used for massives. However, a detailed investigation on the particle sizes typically released during the production of the massives may lead to the selection of a more representative particle size to base the classification of the massive form on. In deriving this particle size special attention should be given to these processes that may generate small particle sizes, for example, processes that involve grinding, polishing, sawing etc. Particle sizes of cuttings that are recycled in the process should not be taken into account. Once the typical particle size representative for normal handling and use is selected this size should be compared with the critical particle size (see Figure 5).

5.4 Read-across for classification purposes between massives and metal powders and vice versa

Read across may reduce the need for Transformation/Dissolution testing considerably. Contrary as for organics, read across should be understood here as a tool to predict the TD properties of different physical forms (different powder sizes or massives form). The Read across principle is based upon the Critical Surface concept. As explained in Figure 5 the TD properties of any particle type can be estimated from the derived critical reference surface area values -based on the surface loading or critical diameter of the testing material under investigation. Read across can be applied in two ways in this respect:

- from different powder sizes to the massive avoiding the sometimes very difficult testing conditions for massive forms OR
- from the massive to the powder form on condition that the surface area of the powders have been well characterized (by measuring the specific surface). This approach may avoid some of the artifacts seen with powders like conglomeration or abrasion.

5.5 Classification of preparations (mixtures of metal compounds)

Classification of mixtures of soluble metals should be done according to the preparation directive.

The same default rule would apply also for mixtures including of sparingly soluble metal compounds. A better method however is to measure the different T/D kinetics and equilibriums for the individual metals included in the mixture. The classification for SSMC mixtures can be estimated by dividing the reference toxicity value by the mass-specific release ratio (MSRR), expressed as mg dissolved/mg metal compound added. The latter can be calculated as the slope of the regression Equation 5.

It should be noted that alloys can not be considered as simple mixtures. The preparation directive and the UN GHS system describes them as special preparations since the alloy has clearly distinctive properties compared to a classical mixture of its metal components. A more suitable assessment strategy is therefore being required to take account for this distinctive different behavior.

An initial thought starter on a specific environmental classification strategy for alloys has been developed in Annex 2.

ANNEX 1: OVERVIEW AND SHORT DESCRIPTION OF THE STANDARD AQUATICTOXICITY TESTS THAT CAN BE USED FOR CLASSIFICATION PURPOSES

For classifying substances in the harmonized system, freshwater and marine species toxicity data can be considered as equivalent data. It should be noted that some metals may express different toxicities in freshwater and marine environments. Since the purpose of classification is to characterize hazard in the aquatic environment, the result showing the highest toxicity should be chosen. It should also be noted that several of the OECD guidelines cited as examples for classification are being revised or are being planned for updating. Such revisions may lead to minor modifications of test conditions. Therefore, some flexibility in test duration or species used is allowed.

Fish tests

- *Acute testing*

Acute tests are generally performed with young juveniles 0.1-5 g in size for a period of 96 hours. The observational endpoint in these tests is mortality. Tests consistent with OECD Test Guideline 203 or equivalent should be used for classification.

- *Chronic testing*

Chronic or long term tests with fish can be initiated with fertilized eggs, embryos, juveniles, or reproductively active adults. Tests consistent with OECD Test Guideline 210 (Fish Early Life Stage), the fish life-cycle test (USEPA 850.1500) or equivalent can be used in the classification scheme. Durations can vary widely depending on the test purpose (anywhere from 7 days to over 200 days). Observational endpoints can include hatching success, growth (length and weight changes), spawning success, and survival. Technically, the Fish Early Life Stage Test is not a “chronic” test but a sub-chronic test on sensitive life stages. It is widely accepted as a predictor of chronic toxicity and is used as such for purposes of classification in the harmonized system. Fish early life stage data are much more available than fish life cycle or reproduction studies.

Crustacea tests

- *Acute testing*

Acute tests with crustacean generally begin with first instar juveniles. For daphids, a test duration of 48 hours is used. For other crustacean, such as mysids or other, a duration of 96 hours is typical. Tests consistent with OECD Test Guideline 202 Part I (Daphnia acute) or US-EPA OPPTS 850.1035 (Mysid acute toxicity) or their equivalents should be used for classification.

- *Chronic testing*

Chronic tests with crustacean also generally begin with first instar juveniles and continue through maturation and reproduction. For daphnids, 21 days is sufficient for maturation and the production of 3 broods. For mysids, 28 days is necessary. Observational endpoints include time to first brood, number of offspring produced per female, growth and survival. It is recommended that tests consistent with OECD Test Guideline 202 part 2 (Daphnia reproduction) or 211 or US-EPA 850.1350 (Mysid chronic) or their equivalents be used in the classification scheme.

Algae/plant tests

- *Tests in algae*

Algae are cultured and exposed to the test substance in a nutrient-enriched medium. Tests consistent with OECD Test Guideline 201 (Algal growth inhibition) should be used. Standard test methods employ a cell density in the inoculum in order to ensure exponential growth through the test, usually 3 to 4 days duration. The preferred observational endpoint in this study is algal growth rate inhibition (EC50). If the endpoint is reported only as reduction in biomass or is not specified, then this value may be interpreted as an equivalent endpoint.

- *Tests with aquatic macrophytes*

The most common vascular plants for aquatic toxicity tests are duckweeds (*Lemna gibba* or *Lemna minor*). The *Lemna* test is a short-term test and although it provides both acute and sub-chronic endpoints, only the acute EC_{50} is used for classification in the harmonized system. The tests last for up to 14 days and are performed in nutrient enriched media similar to that used for algae, but may be increased in strength. The observational endpoint is based on change in number of fronds produced. Tests consistent with OECD Test Guideline on *Lemna* (in preparation) and US-EPA 850.4400 (aquatic plant toxicity, *Lemna*) should be used.

Table A.1.1: Overview of the standard aquatic toxicity tests that can be used for classification purposes.

Test type (acute/chronic)	**Recommended test species**	**Recommended life stage and/or size**	**Test duration (hours or days)**	**pH medium /test**	**Test endpoint**	**Test guidelines**
ALGAE						
Acute, chronic	Freshwater green algae: e.g. *Pseudokirchneriella subcapitata*, (formerly *Selenastrum capricornutum, Raphidocelis subcapitata*), *Scenedesmus subspicatus, Chlorella vulgaris* *P. subcapitata* (freshwater green alga), *Navicula pelliculosa* (freshwater diatom), *Anabaena flos-aquae* (blue-green algae) *Skeletonema costatum* (marine diatom),	Initial cell concentration: 10^4 cells/ml Initial cell concentration: 10^4 cells/ml $7.7x10^4$ cells/ml	72h 96h	8 pH deviation during test not more than 1 unit	Growth inhibition (biomass, growth rate) Growth inhibition (biomass, growth rate)	- OECD Test Guideline 201: Alga growth inhibition test (1984) - EC C.3 Algal inhibition test (1992) - OPPTS 850.5400 Algal toxicity, Tiers I and II
HIGHER PLANTS						
Sub-chronic	*Lemna sp.*: *Lemna gibba* and *L. minor*	9-12 fronds per test vessel (2-4 fronds per plant)	7d		Growth inhibition (biomass, growth rate)	- OECD Test Guideline 221: *Lemna* sp. growth inhibition test (in prep.)
Sub-chronic	*Lemna gibba G3 and L. minor*	12-16 fronds per test chamber (3 or	7d	pH of nutrient medium:	Growth inhibition, frond mortality	- OPPTS 850.4450 Aquatic plant toxicity test using *Lemna* sp. Tiers I and II

Test type (acute/chronic)	Recommended test species	Recommended life stage and/or size	Test duration (hours or days)	pH medium /test	Test endpoint	Test guidelines
		4-frond per plant)		4.8-5.2 (M. Hoagland's medium) & 7.5±0.1 for 20X-AAP medium		

Table A.1.1 continued:

Test type (acute/chronic)	Recommended test species	Recommended life stage and/or size	Test duration (hours or days)	pH medium /test	Test endpoint	Test guidelines
CRUSTACEA						
Acute	*Daphnia* sp.: e.g. *Daphnia magna*	Juveniles (<24h)	48h	7-7.5	Immobilization	- OECD Test Guideline 202: *Daphnia* sp., acute immobilisation test (1984) - EC C.2 Acute toxicity for *Daphnia* (1992)
	D. magna, D. pulex	First instar daphnids (<24h)	48h		Immobilization	- OPPTS 850.1010 Aquatic invertebrate acute toxicity, test, freshwater daphnids
Acute	Amphipods *Gammarus fasciatus, G. Pseudolimnaeus* and *G. lacustris*	Similar age or size from the same source or culture population (not further specified)	96h	pH dilution water: variation ± 0.4 unit; no range provided	Mortality	- OPPTS 850.1020 Gammarid acute toxicity test
Acute	Mysids	Juvenile (<24h old) or young	96h	-	Mortality	- OPPTS 850.1035 Mysid acute toxicity test

Test type (acute/chronic)	Recommended test species	Recommended life stage and/or size	Test duration (hours or days)	pH medium /test	Test endpoint	Test guidelines
	Americamysis bahia (formerly *Mysidopsis bahia*) marine	adults (5-6d old) (most sensitive life stage)				
Acute	Penaeid: *Penaeus aztecus, Penaeus duorarum, Penaeus setiferus*	Post-larval juvenile shrimp	96h	-	Mortality	- OPPTS 850.1045 Penaeid acute toxicity test
Chronic	*Daphnia magna* *D. magna, D. pulex*	Juveniles (<24h) First instar daphnids (<24h)	21d 21d	6-9; no more variation than 1.5 units in test	Immobilisation, reproduction Immobilisation, reproduction	- OECD Test Guideline 211: *Daphnia magna* reproduction test (1998) - EC C.20 Daphnia magna reproduction test (2001) - OPPTS 850.1300 Daphnid chronic toxicity test
Chronic	Mysids *Americamysis bahia* (formerly *Mysidopsis bahia*)	Juvenile (<24h old)	28d	-	Mortality, reproduction, growth	- OPPTS 850.1350 Mysid chronic toxicity test

Table A.1.1 continued:

Test type (acute/chronic)	Recommended test species	Recommended life stage and/or size	Test duration (hours or days)	pH	Test endpoint	Test guidelines
FISH						
Acute	Fish: freshwater species: *Oncorhynchus*	Total length of fish: 5 ± 1	96h	6-	Mortality	- OECD Test Guideline 203: Fish, acute toxicity test

Test type (acute/chronic)	Recommended test species	Recommended life stage and/or size	Test duration (hours or days)	pH	Test endpoint	Test guidelines
	mykiss, Danio rerio, *Pimephalis promelas, Cyprinus carpio,* *Oryzias latipes,* *Poecilia reticulate, Lepomis macrochirus* Fish: freshwater species *Salmo salar, Lepomis macrochirus, Salvelinus fontinalis, Ictalurus punctatus, Oncorhynchus kisutch, Cyprinus carpio, Pimephales promelas, Poecilia reticulata, Oncorhynchus mykiss, Oryzias latipes, Gasterosteus aculeatus, Danio rerio* Saltwater species: *Menidia menidia, Cyprinodon variegatus, Menidia penisulae*	cm 2 ± 1 cm 2 ± 1 cm 3 ± 1 cm 2 ± 1 cm 2 ± 1 cm 2 ± 1 cm juvenile fish <3.0g	96h	8.5	Mortality	(1992) - EC C.1: Acute toxicity for fish (1992) - OPPTS 850.1075 Fish acute toxicity test, freshwater and marine
Sub-chronic	Fish: freshwater species: *Oncorhynchus mykiss, Danio rerio,* *Pimephalis promelas,*	Total length of fish: 5 ± 1 cm	14d	6-8.5	Mortality, behaviour	- OECD Test Guideline 204: Fish, prolonged toxicity test: 14-day study (1984)

Test type (acute/chronic)	Recommended test species	Recommended life stage and/or size	Test duration (hours or days)	pH	Test endpoint	Test guidelines
	Cyprinus carpio, *Oryzias latipes,* *Poecilia reticulate, Lepomis macrochirus*t	2 ± 1 cm 2 ± 1 cm 3 ± 1 cm 2 ± 1 cm 2 ± 1 cm 2 ± 1 cm				
Sub-chronic	Fish: freshwater species: *Danio rerio, Oncorhynchus mykiss, Pimephales promelas, Oryzias latipes* Fish: freshwater species: cfr. OECD guideline Saltwater species: *Cyprinidon variegatus*	Embryonic stage Embryonic stage	30-60 d post hatch dependent on species 30-60 d post hatch dependent on species		Hatching, survival, growth, abnormal appearance, abnormal behaviour Hatching, survival, growth, abnormal appearance, abnormal behaviour	- OECD Test Guideline 210: Fish, early-life stage toxicity test (1992) - OPPTS 850.1400 Fish early-life stage toxicity test
Sub-chronic	Fish: freshwater species: *Oncorhynchus mykiss, Danio rerio, Cyprinus carpio, Oryzias latipes,* *Pimephalis promelas*	Embryo and sac-fry stages	8-55d dependent on species		Hatching, survival, growth, abnormal appearance, abnormal behaviour	- OECD Test Guideline 212: Fish, short-term toxicity test on embryo and sac-fry stages (1998) - EC C.15: Fish, short-term toxicity test on embryo and sac-fry stages (2001)
Sub-chronic	Fish: freshwater species: *Oncorhynchus mykiss,*	Juvenile Weight/fish: 1-5 g	≥28d		Growth (weight), abnormalities	- OECD Test Guideline 215 (2000) EC C.14: Fish juvenile growth test (2001)

Test type (acute/chronic)	Recommended test species	Recommended life stage and/or size	Test duration (hours or days)	pH	Test endpoint	Test guidelines
	Danio rerio, *Oryzias latipes*	0.05-0.10 g 0.05-0.10 g				
Chronic	Fish: freshwater species e.g. *Pimephales promelas* Saltwater species: *Cyprinidon variegatus*	Fertilised egg	From one stage of the life cycle to at least the same stage of the next generation (e.g. egg to egg)		Mortality, reproduction, behavioural, physiological and pathological effects	- OPPTS 850.1400 Fish life-cycle toxicity test

OECD Guidelines for the testing of chemicals. OECD, Paris, 1993 with regular updates (Homepage: http://www.oecd.org/ehs/test/testlist.htm)

EC guidelines: European Commission (1996). Classification, packaging and labelling of dangerous substances in the European Union. Part 2 – Testing methods. European Commission, 1997. ISBN92-828-0076-8. (Homepage: http://ecb.ei.jrc.it/testing-methods/)

OPPTS guidelines: US-EPA homepage:http://www.epa.gov/opptsfrs/home/guideline.htm and (htpp://www.epa.gov/OPPTS_Harmonized/850_Ecological_Effects_Test_)

ANNEX 2: THOUGHT STARTER ON A SPECIFIC CLASSIFICATION STRATEGY FOR METALLIC ALLOYS[22] IN THE POWDER AND THE MASSIVE FORM

A.2.1 Introduction

Recital 10 of the present EU preparation directive (1999/45/EC) indicates the need to develop specific classification guidance for alloys based on the consideration that alloys have distinctively different properties than their constituents. However, no attempt has been made so far to develop such a strategy for metallic alloys . This thought starter provides a first constructive attempt in this respect aiming for catalyzing the discussion on alloys classification at technical level.

The schemes in current legislation consider alloys to be mixtures (OECD) or preparations (EU) and these schemes are based on the intrinsic properties of the individual metal constituents, ignoring the unique properties of metallic alloys. The definition internationally (OECD and UN) used for an alloy is *a metallic material, homogeneous on a macroscopic scale, consisting of two or more elements so combined that they cannot be readily separated by mechanical means*. As such alloys should not be considered as simple mixtures. Rather alloys have specific behaviors and intrinsic properties that are usually very different from those of their constituents. Therefore classification based in the hazardous properties of its constituents may be incorrect and alloys should be classified on the basis of their own intrinsic properties rather than those of their alloying elements. This annex provides a framework proposal to achieve appropriate classification of alloys. The basis of the proposal is that classification of alloys should normally be made on the basis of evidence for the alloy itself.

A.2.2 Proposed approach for the classification of metallic alloys

A.2.2.1 Classification strategy for metallic alloys

Most of the concepts developed for the classification and labeling of metals and sparingly soluble metal compounds such as the use of transformation/dissolution tests and critical surface area/critical diameter, can be used and extended to assess alloys. . To this end, a tiered, iterative classification strategy similar) as depicted in Figure A.2.1, can provide the basis towards reliable, evidence-based alloy classification. This scheme builds further to the scheme used for the classification of metals (Figure 2)

[22] Throughout the rest of the fact sheet the term “alloy” always mean “metallic alloy”

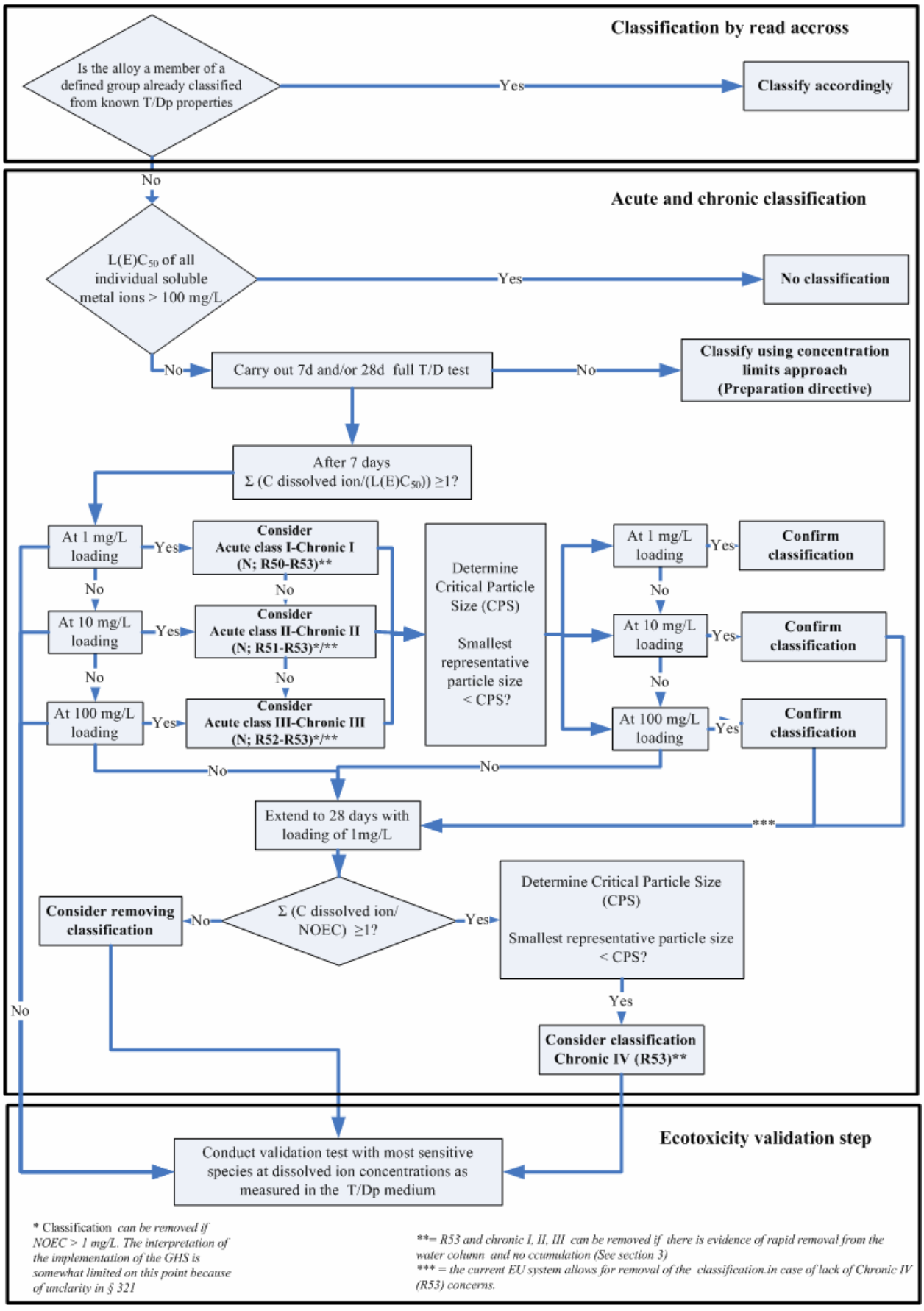


Figure A.2.1: Classification strategy for metallic alloys in powder and massive form

BOX 1: classification by read across

- Collect and assess existing data on the alloy for each ecotoxicological endpoint under consideration.
- Assess if structural relationships can be used to classify the alloy based on comparison with alloys already classified for the same end-point.
- Relevant information and data that can be used for this purpose include:
    - Concentration of the (metal) constituents
    - Physico-chemical properties of the alloy and constituents
    - Crystal structure of the alloy
    - Relevant ecotoxicological data on the metal(s) and alloy
    - Speciation and bioavailability parameters
    - Surface properties of the alloy
    - Corrosion data, metal release and run-off studies
- The adequacy and quality of the data should be reviewed and if the alloy belongs to a well defined group for which a classification has already been derived and there is no evidence that the alloy under consideration would behave differently, it is suggested to classify accordingly.

BOX 2: Acute and chronic classification

- Where the acute $L(E)C_{50}$ of all metal ions of concern is greater than 100 mg/L, the alloy need not be considered further in the classification scheme (Figure 3).
- If the alloy does not belong to a well defined group and the previous condition is not fulfilled, then the classification is assigned to the alloy based on the conventional method of classification i.e. percentage of the classified constituents in accordance with the Preparations Directive (1999/45/EC)
- Alternatively transformation/dissolution testing of the alloy can be initiated
- Alloys that are only used in the massive form and that not give rise to fine particles during their normal handling and use should only be tested for transformation dissolution kinetics at the default particle size of 1mm (diameter of a spherical particle)
- The classification of an individual alloys powder size can be achieved by conducting Transformation Dissolution testing on the concerned alloys particle size
- The Critical surface concept as discussed for metals under section 5.3.4 could be a useful tool to assess alloys that are produced in different powder sizes or for alloys that give rise to significant powder generation during normal handling and use.

As for metals, transformation/dissolution tests on the alloy (massive and/or powder) can be conducted to determine the levels of the constituent metal ion release into the medium. The loading rates would be 1, 10 and 100 mg/L. The individual metal concentrations in the T/D medium should then be compared with the individual ecotoxicity reference values for each metal derived from the soluble metal species. Exceeding one of these reference values would lead to the respective classification of the alloy. In case that all of the constituent substances released in to the T/D medium are below the respective individual ecotoxicity reference values, the *potential for toxicity of the mixture still has to be evaluated*. As a conservative assumption, it has been assumed that the joint effect of the metals act is additive. As

such, the toxicity of the mixture can be calculated on a theoretical basis using the concentration-addition model proposed by Anderson and Weber (1975).

In this model, the dissolution concentrations for the individual metals are compared with the ion toxicity of each single metal for the most sensitive species and expressed as a fraction of its $L(E)C_{50}$ (or NOEC) (i.e. toxic units). The expected toxicity (toxic strength) of the mixture (based on the additive hypothesis and expressed as toxic units (T.U.) is given by Eq-A1, i.e. the sum of the ratios of actual metal concentration to their effective concentrations (LC_{50}):

$$\text{Mixture toxicity} = \sum \frac{C_{mi,a\ dissolved}}{C_{mi,\ e\ dissolved}} \qquad \text{Eq-A1}$$

$C_{mi,a\ dissolved}$ = *actual dissolved metal concentration metal i*

$C_{mi,e\ dissolved}$ = *effective dissolved metal concentration metal i: i.e.($L(E)C_{50}$) or NOEC value*

In case of the use of $L(E)C_{50}$ values and assuming complete concentration addition the 50% response of a mixture of chemicals is obtained when the sum of T.U. of all individual constituents equals unity. Therefore if Σ T.U. ≥ 1 the alloy should be classified or further evaluated (e.g. ecotoxicity testing) (see validation step).

When appropriate models are available, a further level of refinement is to normalize effects to standard abiotic conditions of the T/Dp medium by using a bioavailability model such as BLM. Mixture BLM models are, however, not yet available to assess the impact of the metal mixture released by metallic alloys. Therefore the option to perform an ecotoxicity test has been embedded in the classification strategy for metallic alloys to really account for possible synergistic or antagonistic effects of the metal mixture.

Where data from the T/Dp are available, the results should be used to aid classification according to the following rules:

<u>7-day Transformation Test</u>

If the $\sum$ (dissolved metal ion concentration/$L(E)C_{50}$) after a period of 7 days (or earlier) exceeds one, then the default classification for the metals is replaced by the following classification as given under the GHS scheme (the classification classes under the EU system are indicated between brackets):

(iv) if the summation of the toxic units of the mixture at the low loading rate (1 mg/L) is greater than or equal to one, then classify Acute class I. Classify also as chronic I unless there is evidence of both rapid partitioning from the water column and no bioaccumulation; (similar to R50-R53 in the EU system)

(v) If the summation of the toxic units of the mixture at the medium loading rate (10 mg/L) is greater than or equal to one, then classify Acute class II. Classify also Chronic II unless there is evidence of both rapid partitioning from the water column and no bioaccumulation; (similar to R51-R53 in the EU system)

(vi) If the the summation of the toxic units of the mixture at the high loading rate (100 mg/L) is greater than or equal to one, then classify Acute class III. Classify also as chronic III unless there is evidence of both rapid partitioning from the water column and no bioaccumulation (similar to R52-R53 in the EU system).

The 7d-transformation test could be extended to 28 days in order to remove the chronic IV (R53) default classification.

28- day Transformation Test

If the process described in previous paragraph results in the classification of chronic 1 (R50-R53 in EU), no further assessment is required as the alloy will be classified irrespective of any further information. In all other cases, further data may have been generated through the dissolution/transformation test for 28 days in order to show that the classification may be amended. If ∑ (dissolved metal ion concentration/NOEC) after a period of 28 days (or earlier) is higher or equal to one classify as Chronic IV (R53) unless antagonistic effects can be expected. In that case, a validation test is run with the most sensitive species and classification should be removed when no toxicity is found. If ∑ (dissolved metal ion concentration/NOEC) is smaller than one a validation test has to be conducted to exclude synergistic effects. If toxicity is found keep the classification (Chronic IV (R53)). Otherwise the classification is removed.

BOX 3: Ecotoxicity validation step

- Finally an ecotox validation step is suggested especially when from existing information antagonistic or synergistic effects may be expected.
- This validation test should be conducted with the most sensitive species at dissolved ion concentrations equivalent to those measured in the T/Dp medium. However, ecotoxicity testing directly in the T/Dp medium is not recommended because the composition of this medium is unlikely to meet the requirements for standard test media to ensure proper survival and/or reproduction (e.g. for algae, pH 6 will already cause a reduced growth).. Therefore, ecotoxicity tests should be conducted in standard media dosed at metal concentrations equivalent to the concentration levels actually measured in the T/Dp medium. These tests can also be initiated for alloys classified as Acute I-chronic I (similar to R50-R53 in EU) ,Acute II-chronic II (similar to R51-R53 in EU), Acute III-chronic III (similar to R52-R53 in EU) when antagonistic effects are expected. When toxicity is found, the alloys are classified accordingly (Figure A-2.2)

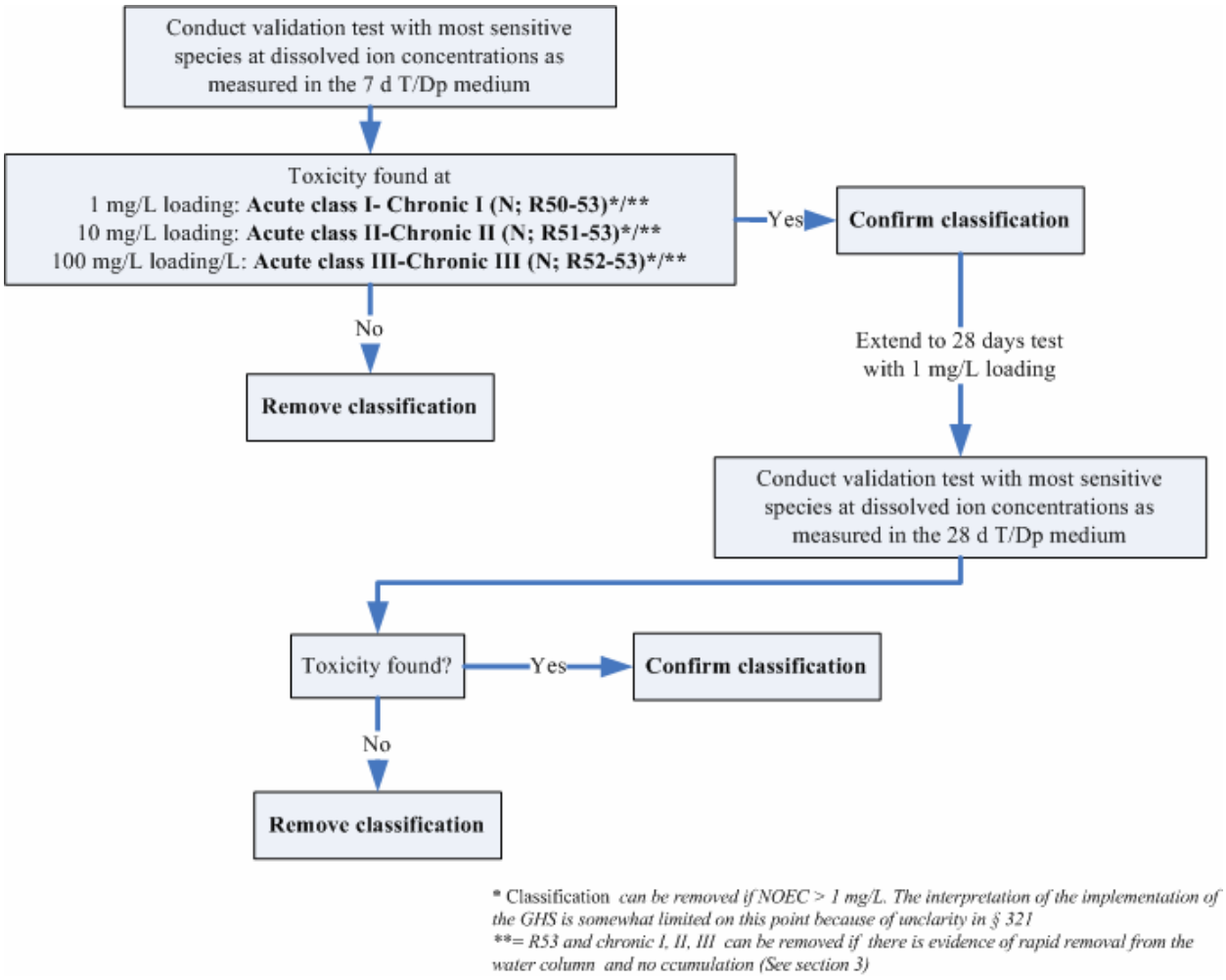


Figure A.2.2: Ecotoxicity validation step

A.2.2.2 Alloy grouping and read across

Since there are a very large number of alloys in existence it is worthwhile to consider grouping alloys based on chemical compositions, microstructure and properties (e.g. nickel-containing stainless steels (austenitic) based on 16-30% Cr and 6-22% Ni; low alloy steels containing 15-30% Cr and up to 3% Ni; copper-nickel alloys). Within an alloy group it can be assumed that these alloys display similar behavior (e.g. release rate of metals assessed by dissolution tests) which can be related to the potential biological impact. Work is ongoing to develop a strategic approach to alloy grouping for hazard classification purposes. Grouping for environmental classification has value to facilitate read across of the data available for a well-defined representative alloy to those alloys for which data are limited in order to bridge data gaps and avoiding unnecessary testing. Preferentially, grouping should be applied in a very careful way recognizing the "*mechanism of action*" (biological and chemical factors) for the endpoint under consideration (e.g. for aquatic toxicity the free metal ion interacting with the gills). In this regard the bioavailable fraction would be the ideal comparative level but this type of information is not always available. In this respect, "solubility" under a given pH condition may in those cases be a valid alternative.

REFERENCES

Aldenberg T. & Jaworska J. S. 2000. Estimation of the hazardous concentration and fraction affected for normally distributed species sensitivity distributions. Ecotoxicology and Environmental Safety, 46, 1-18.

Brix K., D.K DeForest and W.J. Adams, 2001. Assessing acute and chronic copper risks to freshwater aquatic life using species sensitivity distributions for different taxonomic groups. Environmental Toxicology and Chemistry, 20 (8), 1846-1856.

Duboudin C., P. Ciffroy and H. Magaud, 2004. Acute-to-chronic species sensitivity distribution extrapolation. Environmental Toxicology and Chemistry, 23 (7), 1774-1785.

http://ecb.jrc.it/classification-labelling

GHS, 2003. Globally harmonized system of classification and labeling of chemicals (GHS). United Nations, New York and Geneva, 2003.

ISO, 2004. Working draft, Water Quality- Guidance document on the statistical analysis of ecotoxicity data, March 2004. ISO/CD20281.

Keating I.K., P.B. Caffrey and B.C. Dagbusan, 1996. Buffers in daphnid culture and bioassay. Environmental Toxicology and Chemistry, 15 (3), 348-352.

London, workshop (2001). Report of the expert consultation workshop on statistical extrapolation techniques for environmental effects assessments. Londen, VK, European Chemicals Bureau, 17-18 januari 2001.

Mackay, D., E. Webster, D. Woodfine, T. Cahill, P. Doyle, Y. Couillard and D. Gutzman, 2003. Towards consistent evaluation of the persistence of organics, inorganic and metallic substances. Human and ecological risk assessment 9, 1445-1475.

OECD, 1995. Organization for Economic Cooperation and Development. Aquatic toxicity testing of sparingly soluble metals, inorganic metal compounds and minerals. Draft Workshop Report. September 5–8, 1995. Natural Resources Canada, Ottawa, ON.

OECD, 2001a. Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances. OECD, Paris.

OECD 2001b. OECD guidelines for the testing of chemicals proposal for a new guideline 218 sediment-water Chironomid toxicity test using spiked sediment.

Skeaff J., K. Delbeke, F. Van Assche and B. Conard (2000). A critical surface area concept for acute hazard classification of relatively insoluble metal-containing powders in aquatic environments. Environmental Toxicology and Chemistry, 19 (6), 1681-1691.

US-EPA (1999). Draft Risk assessment guidance for Super Fund. Volume 3-Part A, process for conducting probabilistic risk assessment.