UN/SCEGHS/14/INF.3

COMMITTEE OF EXPERTS ON THE TRANSPORT OF DANGEROUS GOODS AND ON THE GLOBALLY HARMONIZED SYSTEM OF CLASSIFICATION AND LABELLING OF CHEMICALS

Sub-Committee of Experts on the Globally Harmonized System of Classification and Labelling of Chemicals

Fourteenth session,
Geneva, 12-14 December 2007
Item 2 (c) of the provisional agenda

UPDATING OF THE SECOND REVISED EDITION OF THE GLOBALLY HARMONIZED SYSTEM OF CLASSIFICATION AND LABELLING OF CHEMICALS (GHS)

Environmental hazards

Revision of chapter 4.1:
consequential amendments to Annex 9 (sections A9.1 to A9.3) and Appendix VI

Transmitted by the Organization for Economic Co-operation and Development (OECD)

Proposal for revision of Annex 9 as a consequence of proposed changes to Chapter 4.1, in order to accommodate chronic toxicity to aquatic organisms for assigning a chronic hazard category (with visible changes).This document is for information only.

ANNEX 9 (A9.1-A9.3 Appendix VI)

GUIDANCE ON HAZARDS TO THE

AQUATIC ENVIRONMENT

Annex 9

GUIDANCE ON HAZARDS TO THE AQUATIC ENVIRONMENT

CONTENTS

Annex 9

GUIDANCE ON HAZARDS TO THE AQUATIC ENVIRONMENT [1]

A9.1 Introduction

A9.1.1 In developing the set of criteria for identifying substances hazardous to the aquatic environment, it was agreed that the detail needed to properly define the hazard to the environment resulted in a complex system for which some suitable guidance would be necessary. Therefore, the purpose of this document is twofold:

(a) to provide a description of and guidance to how the system will work;

(b) to provide a guidance to the interpretation of data for use in applying the classification criteria.

A9.1.2 The hazard classification scheme has been developed with the object of identifying those chemical substances that present, through the intrinsic properties they possess, a danger to the aquatic environment. In this context, the aquatic environment is taken as the aquatic ecosystem in freshwater and marine, and the organisms that live in it. For most substances, the majority of data available addresses this environmental compartment. The definition is limited in scope in that it does not, as yet, include aquatic sediments, nor higher organisms at the top end of the aquatic food-chain, although these may to some extent be covered by the criteria selected.

A9.1.3 Although limited in scope, it is widely accepted that this compartment is both vulnerable, in that it is the final receiving environment for many harmful substances, and the organisms that live there are sensitive. It is also complex since any system that seeks to identify hazards to the environment must seek to define those effects in terms of wider effects on ecosystems rather than on individuals within a species or population. As will be described in detail in the subsequent sections, a limited set of specific properties of chemical substances have been selected through which the hazard can be best described: acute aquatic toxicity; chronic aquatic toxicity; lack of degradability; and potential or actual bioaccumulation. The rationale for the selection of these data as the means to define the aquatic hazard will be described in more detail in Section A9.2.

A9.1.4 This annex ~~The application of the criteria~~ is ~~also~~limited, at this stage, to the application of the criteria to chemical substances. The term substances covers a wide range of chemicals, many of which pose difficult challenges to a classification system based on rigid criteria. The following sections will thus provide some guidance as to how these challenges can be dealt with based both on experience in use and clear scientific rationale. While the harmonized criteria apply most easily to the classification of individual substances of defined structure (see definition in Chapter 1.2), some materials that fall under this category are frequently referred to as “complex mixtures”. In most cases they can be characterized as a homologous series of substances with a certain range of carbon chain length/number or degree of substitution. Special methodologies have been developed for testing which provides data for evaluating the intrinsic hazard to aquatic organisms, bioaccumulation and degradation. More specific guidance is provided in the separate sections on these properties. For the purpose of this Guidance Document, these materials will be referred to as “complex substances” or “multi-component substances”.

---

[1] *OECD Environment, Health and Safety Publications, Series on Testing and Assessment, No 27, Environment Directorate, Organization for economic Co-operation and Development, April 2001.*

A9.1.5 Each of these properties (i.e. acute aquatic toxicity, chronic aquatic toxicity, degradability, bioaccumulation) can present a complex interpretational problem, even for experts. While internationally agreed testing guidelines exist and should be used for any and all new data produced, many data usable in classification will not have been generated according to such standard tests. Even where standard tests have been used, some substances, such as complex substances, hydrolytically unstable substances, polymers etc, present difficult interpretational problems when the results have to be used within the classification scheme. Thus data are available for a wide variety of both standard and non-standard test organisms, both marine and freshwater, of varying duration and utilizing a variety of endpoints. Degradation data may be biotic or abiotic and can vary in environmental relevance. The potential to bioaccumulate can, for many organic chemicals, be indicated by the octanol-water partition coefficient. It can however be affected by many other factors and these will also need to be taken into account.

A9.1.6 It is clearly the objective of a globally harmonized system that, having agreed on a common set of criteria, a common data-set should also be used so that once classified, the classification is globally accepted. For this to occur, there must first be a common understanding of the type of data that can be used in applying the criteria, both in type and quality, and subsequently a common interpretation of the data when measured against the criteria. For that reason, it has been felt necessary to develop a transparent guidance document that would seek to expand and explain the criteria in such a way that a common understanding of their rationale and a common approach to data interpretation may be achieved. This is of particular importance since any harmonized system applied to the "universe of chemicals" will rely heavily on self-classification by manufacturers and suppliers, classifications that must be accepted across national boundaries without always receiving regulatory scrutiny. This guidance document, therefore, seeks to inform the reader, in a number of key areas, and as a result lead to classification in a consistent manner, thus ensuring a truly harmonized and self-operating system.

A9.1.7 Firstly, it will provide a detailed description of the criteria, a rationale for the criteria selected, and an overview of how the scheme will work in practice (Section A9.2). This section will address the common sources of data, the need to apply quality criteria, how to classify when the data-set is incomplete or when a large data-set leads to an ambiguous classification, and other commonly encountered classification problems.

A9.1.8 Secondly, the guidance will provide detailed expert advice on the interpretation of data derived from the available databases, including how to use non-standard data, and specific quality criteria that may apply for individual properties. The problems of data interpretation for "difficult substances", those substances for which standard testing methods either do not apply or give difficult interpretational problems, will be described and advice provided on suitable solutions. The emphasis will be on data interpretation rather than testing since the system will, as far as possible, rely on the best available existing data and data required for regulatory purposes. The ~~three~~ four core properties, acute and chronic aquatic toxicity (Section A9.3), degradability (Section A9.4) and bioaccumulation (Section A9.5) are treated separately.

A9.1.9 The range of interpretational problems can be extensive and as a result such interpretation will always rely on the ability and expertise of the individuals responsible for classification. However, it is possible to identify some commonly occurring difficulties and provide guidance that distils accepted expert judgement that can act as an aid to achieving a reliable and consistent result. Such difficulties can fall into a number of overlapping issues:

(a) The difficulty in applying the current test procedures to a number of types of substance;

(b) The difficulty in interpreting the data derived both from these "difficult to test" substances and from other substances;

(c) The difficulty in interpretation of diverse data-sets derived from a wide variety of sources.

A9.1.10 For many organic substances, the testing and interpretation of data present no problems when applying both the relevant OECD Guideline and the classification criteria. There are a number of typical interpretational problems, however, that can be characterized by the type of substance being studied. These are commonly called "difficult substances":

(a) poorly soluble substances: these substances are difficult to test because they present problems in solution preparation, and in concentration maintenance and verification during aquatic toxicity testing. In addition, many available data for such substances have been produced using "solutions" in excess of the water solubility resulting in major interpretational problems in defining the true $L(E)C_{50}$ or NOEC for the purposes of classification. Interpretation of the partitioning behaviour can also be problematic where the poor solubility in water and octanol may be compounded by insufficient sensitivity in the analytical method. Water solubility may be difficult to determine and is frequently recorded as simply being less than the detection limit, creating problems in interpreting both aquatic toxicity and bioaccumulation studies. In biodegradation studies, poor solubility may result in low bioavailability and thus lower than expected biodegradation rates. The specific test method or the choice of procedures used can thus be of key importance;

(b) unstable substances: such substances that degrade (or react) rapidly in the test system ~~again~~ present both testing and interpretational problems. It will be necessary to determine whether the correct methodology has been used, whether it is the substance or the degradation/reaction product that has been tested, and whether the data produced is relevant to the classification of the parent substance;

(c) volatile substances: such substances that can clearly present testing problems when used in open systems should be evaluated to ensure adequate maintenance of exposure concentrations. Loss of test material during biodegradation testing is inevitable in certain methods and will lead to misinterpretation of the results;

(d) complex or multi-component substances: such substances, for example, hydrocarbon mixtures, frequently cannot be dissolved into a homogeneous solution, and the multiple components make monitoring impossible. Consideration therefore needs to be given to using the data derived from the testing of water accommodated fractions (WAFs) for aquatic toxicity, and the utilization of such data in the classification scheme. Biodegradation, bioaccumulation, partitioning behaviour and water solubility all present problems of interpretation, where each component of the mixture may behave differently;

(e) polymers: such substances frequently have a wide range of molecular masses, with only a fraction being water soluble. Special methods are available to determine the water soluble fraction and these data will need to be used in interpreting the test data against the classification criteria;

(f) inorganic compounds and metals: such substances, which can interact with the media, can produce a range of aquatic toxicities dependant on such factors as pH, water hardness etc. Difficult interpretational problems also arise from the testing of essential elements that are beneficial at certain levels. For metals and inorganic metal compounds, the concept of degradability as applied to organic compounds has limited or no meaning. Equally the use of bioaccumulation data should be treated with care;

(g) surface active substances: such substances can form emulsions in which the bioavailablity is difficult to ascertain, even with careful solution preparation. Micelle formation can result in an overestimation of the bioavailable fraction even when "solutions" are

apparently formed. This presents significant problems of interpretation in each of the water solubility, partition coefficient, bioaccumulation and aquatic toxicity studies;

(h) ionizable substances: such substances can change the extent of ionization according to the level of counter ions in the media. Acids and bases, for example, will show radically different partitioning behaviour depending on the pH;

(i) coloured substances: such substance can cause problems in the algal/aquatic plant testing because of the blocking of incident light;

(j) impurities: some substances can contain impurities that can change in % and in chemical nature between production batches. Interpretational problems can arise where either or both the toxicity and water solubility of the impurities are greater than the parent substance, thus potentially influencing the toxicity data in a significant way.

A9.1.11 These represent some of the problems encountered in establishing the adequacy of data, interpreting the data and applying that data to the classification scheme. Detailed guidance on how to deal with these problems, as well as other issues related will be presented in the following sections. The interpretation of data on acute and on chronic aquatic toxicity will be covered in Section A9.3. This section will deal with the specific interpretational problems encountered for the above "difficult substances", including providing some advice on when and how such data can be used within the classification scheme. Also covered will be a general description of the test data used and the testing methodologies suitable for producing such data.

A9.1.12 A wide range of degradation data are available that must be interpreted according to the criteria for rapid degradability. Guidance is thus needed on how to use these data obtained by employing non-standard test methods, including the use of half-lives where these are available, of primary degradation, of soil degradation rates and their suitability for extrapolation to aquatic degradation and of environmental degradation rates. A short description of estimation techniques for evaluating degradability in relation to the classification criteria is also included. This guidance will be provided in Section A9.4.

A9.1.13 Methods by which the potential to bioaccumulate can be determined will be described in Section A9.5. This section will describe the relationship between the partition coefficient criteria and the bioconcentration factor (BCF), provide guidance on the interpretation of existing data, how to estimate the partition coefficient by the use of QSARs when no experimental data are available and in particular deal with the specific problems identified above for difficult substances. The problems encountered when dealing with substances of high molecular mass will also be covered.

A9.1.14 A section is also included which covers general issues concerning the use of QSARs within the system, when and how they may be used, for each of the three properties of concern. As a general approach, it is widely accepted that experimental data should be used rather than QSAR data when such data are available. The use of QSARs will thus be limited to such times when no reliable data are available. Not all substances are suitable for the application of QSAR estimations, however, and the guidance in Section A9.6 will address this issue.

A9.1.15 Finally, a section is devoted to the special problems associated with the classification of metals and their compounds. Clearly, for these compounds, a number of the specific criteria such as biodegradability and octanol-water partition coefficient cannot be applied although the principle of lack of destruction via degradation, and bioaccumulation remain important concepts. Thus it is necessary to adopt a different approach. Metals and metal compounds can undergo interactions with the media which affect the solubility of the metal ion, partitioning from the water column, and the species of metal ion that exists in the water column. In the water column, it is generally the dissolved metal ions which are of concern for toxicity. The interaction of the substance with the media may either increase or decrease the level of ions and hence toxicity. It is thus necessary to consider whether metal ions are likely to be formed from the substance and

dissolve in the water, and if so whether they are formed rapidly enough to cause concern. A scheme for interpreting the results from this type of study is presented in Section A9.7.

A9.1.16 While the Guidance Document provides useful advice on how to apply the criteria to a wide variety of situations, it remains a guidance only. It cannot hope to cover all situations that arise in classification. It should therefore be seen as a living document that in part describes the fundamental principles of the system, e.g. hazard based rather than risk based, and the fixed criteria. It must also, in part, be a repository for the accumulated experience in using the scheme to include the interpretations which allow the apparently fixed criteria to be applied in a wide variety of non-standard situations.

A9.2 The harmonized classification scheme

A9.2.1 *Scope*

The criteria were developed taking into account existing systems for hazard classification, such as EU- Supply and Use System, the Canadian and US Pesticide systems, GESAMP hazard evaluation procedure, IMO Scheme for Marine Pollutant, the European Road and Rail Transport Scheme (RID/ADR), and the US Land Transport. These systems include supply and subsequent use of chemicals, the sea transport of chemical substances as well as transport of chemical substances by road and rail. The harmonized criteria are therefore intended to identify hazardous chemicals in a common way for use throughout all these systems. To address the needs for all different sectors (transport, supply and use) it was necessary to create two different sub-classes, one ~~Acute~~ sub-class for acute aquatic hazards, consisting of three categories and one ~~Chronic~~ sub-class for long-term aquatic hazards, consisting of 4 categories. The Acute classification sub-class makes provision for two acute hazard categories (Acute 2 and 3) not normally used when considering packaged goods. For substances transported in bulk, there are a number of regulatory decisions that can uniquely arise because of the bulk quantities being considered. For these situations, for example where decisions are required on the ship type to be used, consideration of all acute hazard categories as well as the long-term ~~chronic~~ hazard categories are considered important. The following paragraphs describe in detail the criteria to be used in defining each of these hazard categories.

A9.2.2 *Classification categories and criteria*

The hazard categories for acute and chronic aquatic toxicity and their related criteria are set out in Chapter 4.1, para. 4.1.2. and ~~Figure~~ Table 4.1.1.

A9.2.3 *Rationale*

A9.2.3.1 The harmonized system for classification recognizes that the intrinsic hazard to aquatic organisms is represented by both the acute and chronic or long~~er~~-term toxicity of a substance, the relative importance of which is determined by the specific regulatory regimes in operation. Distinction can be made between the acute hazard and the chronic hazard and therefore hazard classes are defined for both properties representing a gradation in the level of hazard identified. Clearly the hazard identified by Chronic Category 1 is more severe than Chronic Category 2. Since the acute hazard and long-term ~~chronic~~ hazard represent distinct types of hazard, they are not comparable in terms of their relative severity. Both hazard sub-classes should be applied independently for the classification of substances to establish a basis for all regulatory systems.

A9.2.3.2 The principal hazard classes defined by the criteria relate largely to the potential for chronic hazard. This reflects the overriding concern with respect to chemicals in the environment, namely that the effects caused are usually sub-lethal, e.g. effects on reproduction, and caused by longer-term exposure. While recognizing that the ~~chronic~~ long-term hazard represents the principal concern, particularly for packaged goods where environmental release would be limited in scope, it must also be recognized that chronic toxicity data are expensive to generate and generally not readily available for most substances. On the other hand, acute toxicity data are frequently readily available, or can be generated to highly standardised

protocols. It is this acute toxicity which has therefore been used as the core property in defining both the acute and the long-term ~~chronic~~ hazard if no adequate chronic test data are available. Nevertheless, it has been recognized that~~, where~~ chronic toxicity data ~~are~~, if available, ~~it~~ should be ~~possible to use these~~ preferred in defining the ~~appropriate~~long-term hazard category~~. The development of specific criteria using such data is thus a high priority in the future development of the scheme~~

A9.2.3.3 The combination of chronic toxicity and intrinsic fate properties reflects the potential hazard of a chemical. Substances that do not rapidly degrade have a higher potential for longer term exposures and therefore should be classified in a more severe category than substances which are rapidly degradable (see 9.3.3.2.2).

A9.2.3.4 (old A9.2.3.3) While recognizing that acute toxicity itself is not a sufficiently accurate predictor of chronic toxicity to be used solely and directly for establishing hazard, it is considered that, in combination with either a potential to bioaccumulate (i.e. a log K_{ow} ≥4 unless BCF <500) or potential longer-term exposure (i.e. lack of rapid degradation) it can be used as a suitable surrogate for classification purposes. Substances that show acute toxicity and also bioaccumulate to a significant degree will normally show chronic toxicity at a significantly lower concentration. ~~Precise acute: chronic ratios are difficult to predict and thus the surrogate data are generally precautionary.~~Equally substances that do not rapidly degrade have a higher potential for giving rise to longer term exposures which again may result in long-term toxicity being realized. Thus, for example, in absence of adequate chronic test data Category Chronic 1 should be assigned if either of the following criteria are met:

(a) $L(E)C_{50}$ for any appropriate aquatic species ≤1 mg/l and a potential to bioaccumulate (log K_{ow} ≥4 unless BCF <500);

(b) $L(E)C_{50}$ for any appropriate aquatic species ≤1 mg/l and a lack of rapid degradation.

A9.2.3.5 (old A9.2.3.4) The precise definitions of ~~acute toxicity~~ the core elements of ~~an appropriate species, lack of rapid degradation and potential to bioaccumulate~~ this system are ~~detailed~~described in detail in Sections A9.3, A9.4 and A9.5 respectively.

A9.2.3.6 (old A9.2.3.5) For some poorly soluble substances, which are normally considered as those having a water solubility < 1 mg/l, no acute toxicity is expressed in toxicity tests performed at the solubility limit. If for such a substance, however, the BCF ≥ 500, or if absent, the log K_{ow} ≥ 4 (indicating a bioaccumulating potential) and the substance is also not rapidly degradable, a safety net classification is applied, Chronic Category 4. For these types of substance the exposure duration in short term tests may well be too short for a steady state concentration of the substance to be reached in the test organisms. Thus, even though no acute toxicity has been measured in a short term (acute) test, it remains a real possibility that such non-rapidly degradable and bioaccumulative substances may exert chronic effects, particularly since such low degradability may lead to an extended exposure period in the aquatic environment.

A9.2.3.7 (old A9.2.3.6) In defining ~~acute~~ aquatic toxicity, it is not possible to test all species present in an aquatic ecosystem. Representative species are therefore chosen which cover a range of trophic levels and taxonomic groupings. The taxa chosen, fish, crustacea and aquatic plants that represent the “base-set” in most hazard profiles, represent a minimum data-set for a fully valid description of hazard. The lowest of the available toxicity values will normally be used to define the hazard category. Given the wide range of species in the environment, the three tested can only be a poor surrogate and the lowest value is therefore taken for cautious reasons to define the hazard category. In doing so, it is recognized that the distribution of species sensitivity can be several orders of magnitude wide and that there will thus be both more and less sensitive species in the environment. Thus, when data are limited, the use of the most sensitive species tested gives a cautious but acceptable definition of the hazard. There are some circumstances where it may not be appropriate to use the lowest toxicity value as the basis for classification. This will usually only arise where it is possible to define the sensitivity distribution with more accuracy than would normally be possible, such as when large data-sets are available. Such large data-sets should be evaluated with due caution.

A9.2.4 ***Application***

A9.2.4.1 Generally speaking, in deciding whether a substance should be classified, a search of appropriate databases and other sources of data should be made for the following data elements:

(a) water solubility;

~~(b) octanol/water partition coefficient (log K_{ow});~~ (b) acute aquatic toxicity (L(E)C50s);

~~(c) fish bioconcentration factor (BCF);d) acute aquatic toxicity (L(E)C50s);~~

(c) chronic aquatic toxicity (NOECs and/or equivalent ECx);

(~~f~~ d) available degradation (and specifically evidence of ready biodegradability);

(~~g~~ e) stability data, in water.

(f) fish bioconcentration factor (BCF);

(g) octanol/water partition coefficient (log K_{ow});

The water solubility and stability data, although not used directly in the criteria, are nevertheless important since they are a valuable help in the data interpretation of the other properties (see A9.1.10).

A9.2.4.2 To classify, a review should first be made of the available aquatic toxicity data. It will be necessary to consider all the available data and select those which meet the necessary quality criteria for classification. If there are no data available that meet the quality criteria required by the internationally standardized methods, it will be necessary to examine any available data to determine whether a classification can be made. If the data indicate that the acute aquatic toxicity L(E)C_{50} >100 mg/l for soluble substances and the chronic aquatic toxicity > 1mg/l, then the substance is not classified as hazardous. There are a number of cases where no effects are observed in the test and the aquatic toxicity is thus recorded as a >water solubility value, i.e. there is no acute toxicity within the range of the water solubility in the test media. Where this is the case, and the water solubility in the test media is ≥1 mg/l, again, no classification need be applied.

A9.2.4.3 If chronic aquatic toxicity data are available cut-off values will depend on whether the chemical is rapidly degradable or not. Therefore, for non-rapidly degradable substances and those for which no information on degradation is available, the cut-off levels are higher than for those substances where rapid degradability can be confirmed (see Chap. 4.1 Tab. 4.1.1 and Tab. 4.1.2) .

A9.2.4.4 (old A9.2.4.3) Where the lowest acute aquatic toxicity data are below 100 mg/l and no adequate chronic toxicity data are available, it is necessary to first decide which hazard category the toxicity falls in, and then to determine whether the chronic and/or the acute sub-class should be applied. This can simply be achieved by examining the available data on the partition coefficient, log K_{ow} and the available data on degradation. If either the log K_{ow}≥4 or the substance cannot be considered as rapidly degradable, then the appropriate long-term ~~chronic~~ hazard category and the corresponding acute hazard category are applied independently. It should be noted that, although the log K_{ow} is the most readily available indication of a potential to bioaccumulate, an experimentally derived BCF is preferred. Where this is available, this should be used rather than the partition coefficient. In these circumstances, a BCF ≥500 would indicate bioaccumulation sufficient to classify in the appropriate ~~chronic~~ long-term hazard class. If the substance is both rapidly degradable and has a low potential to bioaccumulate (BCF <500 or, if absent log K_{ow} <4) then it

should not be assigned to a long-term ~~chronic~~ hazard category, ~~only the acute hazard categories need be applied (see A9.2.1~~ unless the chronic toxicity data indicate otherwise (A 9.2.4.3).

A9.2.4.5 (old A9.2.4.4) For poorly soluble substances, generally speaking, those with a water solubility in the test media of < 1 mg/l, for which no aquatic toxicity has been found, should be further examined to determine whether Chronic Category 4 needs to be applied. Thus, if the substance is both not rapidly degradable and has a potential to bioaccumulate (BCF ≥500 or, if absent log K_{ow} ≥4), the Chronic Category 4 should be applied.

A9.2.5 ***Data availability***

The data used to classify a substance can be drawn from data required for regulatory purposes as well as the relevant literature, although a number of internationally recognized data-bases exist which can act as a good starting point. Such databases vary widely in quality and comprehensiveness and it is unlikely that any one database will hold all he information necessary for classification to be made. Some databases specialize in aquatic toxicity and others in environmental fate. There is an obligation on the chemical supplier to make the necessary searches and checks to determine the extent and quality of the data available and to use it in assigning the appropriate hazard category.

A9.2.6 ***Data quality***

A9.2.6.1 The precise use of the available data will be described in the relevant section but, as a general rule, data generated to standard international guidelines and to GLP is to be preferred over other types of data. Equally, however, it is important to appreciate that classification can be made based on the best available data. Thus if no data is available which conforms to the quality standard detailed above, classification can still be made provided the data used is not considered invalid. To assist this process, a quality scoring guide has been developed and used extensively in a number of fora and generally conforms to the following categories:

(a) Data derived from official data sources that have been validated by regulatory authorities, such as EU Water Quality Monographs, USEPA Water Quality Criteria. These data can be considered as valid for classification purposes. No assumption should be made that these are the only data available, however, and due regard should be given to the date of the relevant report. Newly available data may not have been considered;

(b) Data derived from recognized international guidelines (e.g. OECD Guidelines) or national guidelines of equivalent quality. Subject to the data interpretation issues raised in the following sections, these data can be used for classification;

(c) Data derived from testing which, while not strictly according to a guideline detailed above, follows accepted scientific principles and procedures and/or has been peer reviewed prior to publication. For such data, where all the experimental detail is not recorded, some judgement may be required to determine validity. Normally, such data may be used within the classification scheme;

(d) Data derived from testing procedures which deviate significantly from standard guidelines and are considered as unreliable, should not be used in classification;

(e) QSAR data. The circumstances of use and validity of QSAR data are discussed in the relevant sections;

(f) Data derived from secondary sources such as handbooks, reviews, citation, etc. where the data quality cannot be directly evaluated. Such data should be examined where

data from quality 1, 2 and 3 are not available, to determine whether it can be used. Such data should have sufficient detail to allow quality to be assessed. In determining the acceptability of these data for the purposes of classification, due regard should be given to the difficulties in testing that may have affected data quality and the significance of the reported result in terms of the level of hazard identified (see A9.3.6.2.3).

A9.2.6.2 Classification may also be made on incomplete toxicity data-sets, e.g. where data are not available on all three trophic levels. In these cases, the classification may be considered as "provisional" and subject to further information becoming available. In general, all the data available will need to be considered prior to assigning a classification. Where good quality data are not available, lower quality data will need to be considered. In these circumstances, a judgement will need to be made regarding the true level of hazard. For example, where good quality data are available for a particular species or taxa, this should be used in preference to any lower quality data which might also be available for that species or taxa. However, good quality data may not always be available for all the basic data set trophic levels. It will be necessary to consider data of lower quality for those trophic levels for which good quality data are not available. Consideration of such data, however, will also need to consider the difficulties that may have affected the likelihood of achieving a valid result. For example, the test details and experimental design may be critical to the assessment of the usability of some data, such as that from hydrolytically unstable chemicals, while less so for other chemicals. Such difficulties are described further in Section A9.3.

A9.2.6.3 Normally, the identification of hazard, and hence the classification will be based on information directly obtained from testing of the substance being considered. There are occasions, however, where this can create difficulties in the testing or the outcomes do not conform to common sense. For example, some chemicals, although stable in the bottle, will react rapidly (or slowly) in water giving rise to degradation products that may have different properties. Where such degradation is rapid, the available test data will frequently define the hazard of the degradation products since it will be these that have been tested. These data may be used to classify the parent substance in the normal way. However, where degradation is slower, it may be possible to test the parent substance and thus generate hazard data in the normal manner. The subsequent degradation may then be considered in determining whether an acute or long-term ~~chronic~~ hazard class should apply. There may be occasions, however, when a substance so tested may degrade to give rise to a more hazardous product. In these circumstances, the classification of the parent should take due account of the hazard of the degradation product, and the rate at which it can be formed under normal environmental conditions.

A9.3 Aquatic toxicity

A9.3.1 *Introduction*

The basis for the identification of hazard to the aquatic environment for a substance is the aquatic toxicity of that substance. Classification is predicated on having toxicity data for fish, crustacea, and algae/aquatic plant available. These taxa are generally accepted as representative of aquatic fauna and flora for hazard identification. Data on these particular taxa are more likely to be found because of this general acceptance by regulatory authorities and the chemical industry. Other information on the degradation and bioaccumulation behaviour is used to better delineate the aquatic hazard. This section describes the appropriate tests for ecotoxicity, provides some basic concepts in evaluating the data and using combinations of testing results for classification, summarizes approaches for dealing with difficult~~y~~ substances, and includes a brief discussion on interpretation of data quality.

A9.3.2 *Description of tests*

A9.3.2.1 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 types of substances, e.g. ionizable organic chemicals or organometallic substances 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.

A9.3.2.2 The GHS criteria for determining health and environmental hazards should be test method neutral, allowing different approaches as long as they are scientifically sound and validated according to international procedures and criteria already referred to in existing systems for the endpoints of concern and produce mutually acceptable data. According to the proposed system (OECD 1998):

> "*Acute toxicity would normally be determined using a fish 96 hour LC50 (OECD Test Guideline 203 or equivalent), a crustacea species 48 hour EC50 (OECD Test Guideline 202 or equivalent) and/or an algal species 72 or 96 hour EC50 (OECD Test Guideline 201 or equivalent). These species are considered as surrogate for all aquatic organisms and data on other species such as the duckweed Lemna may also be considered if the test methodology is suitable.*"

Chronic testing <u>generally</u> involves an exposure that is lingering or continues for a longer time; the term can signify periods from days to a year, or more depending on the reproductive cycle of the aquatic organism. Chronic tests can be done to assess certain endpoints relating to growth, survival, reproduction and development.

> "*Chronic toxicity data are less available than acute data and the range of testing procedures less standardised. Data generated according to the OECD Test Guidelines 210 (Fish Early Life Stage), 202 Part 2 or 211 (Daphnia Reproduction) and 201 (Algal Growth Inhibition) can be accepted. Other validated and internationally accepted tests could also be used. The NOECs or other equivalent L(E)Cx should be used.*"

<u>An OECD document describes the main statistical methods for the analysis of data of standardised ecotoxicity tests (OECD 2006).</u>

A9.3.2.3 It should 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, the expert group that developed the harmonized criteria for classification intended some flexibility in test duration or even species used.

A9.3.2.4 Guidelines for conducting acceptable tests with fish, crustacea, and algae can be found in many sources (OECD, 1999; EPA, 1996; ASTM, 1999; ISO EU). The OECD monograph No.11, Detailed Review Paper on Aquatic Toxicity Testing for Industrial Chemicals and Pesticides, is a good compilation of pelagic test methods and sources of testing guidance. This document is also a source of appropriate test methodologies.

A9.3.2.5 *Fish Tests*

A9.3.2.5.1 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. Fish larger than this range and/or durations shorter than 96 hours are generally less sensitive. However, for classification, they could be used if no acceptable data with the smaller fish for 96 hours are available or the results of these tests with different size fish or test durations would influence classification in a more hazardous category. Tests consistent with OECD Test Guideline 203 (Fish 96 hour LC_{50}) or equivalent should be used for classification.

A9.3.2.5.2 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 (US EPA 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 OECD 210 Guideline (Fish Early Life Stage) 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 toxicity data are much more available than fish life cycle or reproduction studies.

A9.3.2.6 *Crustacea Tests*

A9.3.2.6.1 Acute testing

Acute tests with crustacea generally begin with first instar juveniles. For daphnids, a test duration of 48 hours is used. For other crustacea, such as mysids or others, a duration of 96 hours is typical. The observational endpoint is mortality or immobilization as a surrogate to mortality. Immobilization is defined as unresponsive to gentle prodding. Tests consistent with OECD Test Guideline 202 Part 1 (Daphnia acute) or USA-EPA OPPTS 850.1035 (Mysid acute toxicity) or their equivalents should be used for classification.

A9.3.2.6.2 Chronic testing

Chronic tests with crustacea 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 US-EPA 850.1350 (Mysid chronic) or their equivalents be used in the classification scheme.

A9.3.2.7 *Algae/Plant Tests*

A9.3.2.7.1 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 algal test is a short-term test that ~~and, although it~~ provides both acute and chronic endpoints, ~~only the acute EC50 is used for classification in the harmonized system~~. The preferred observational endpoint in this study is algal growth rate inhibition because it is not dependent on the test design, whereas biomass depends both on growth rate of the test species as well as test duration and other elements of test design. If the endpoint is reported only as reduction in biomass or is not specified, then this value may be interpreted as an equivalent endpoint.

A9.3.2.7.2 Tests in aquatic macrophytes

The most commonly used vascular plants for aquatic toxicity tests are duckweeds (*Lemna gibba* and *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 the 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.

A9.3.3 ***Aquatic toxicity concepts***

This section addresses the use of acute and chronic toxicity data in classification, and special considerations for exposure regimes, algal toxicity testing, and use of QSARs. For a more detailed discussion of aquatic toxicity concepts, one can refer to Rand (1996).

A9.3.3.1 *Acute toxicity*

A9.3.3.1.1 Acute toxicity for purposes of classification refers to the intrinsic property of a substance to be injurious to an organism in a short-term exposure to that substance. Acute toxicity is generally expressed in terms of a concentration which is lethal to 50% of the test organisms (LC_{50}), causes a measurable adverse effect to 50% of the test organisms (e.g. immobilization of daphnids), or leads to a 50% reduction in test (treated) organism responses from control (untreated) organism responses (e.g. growth rate in algae).

A9.3.3.1.2 Substances with an acute toxicity determined to be less than one part per million (1 mg/l) are generally recognized as being very toxic. The handling, use, or discharge into the environment of these substances poses a high degree of hazard and they are classified in Chronic and/or Acute Category 1. Decimal bands are accepted for categorizing acute toxicity above this category. Substances with an acute toxicity measured from one to ten parts per million (1 - 10 mg/l) are classified in Category 2 for acute toxicity, from ten to one hundred parts per million (10 - 100 mg/l) are classified in Category 3 for acute toxicity, and those over one hundred parts per million are regarded as practically non-toxic.

A9.3.3.2 *Chronic toxicity*

A9.3.3.2.1 Chronic toxicity, for purposes of classification, refers to the ~~potential or actual properties~~ intrinsic property of a substance to cause adverse effects to aquatic organisms during exposures which are determined in relation to the life-cycle of the organism. Such chronic effects usually include a range of sublethal endpoints and are generally expressed in terms of a No Observable Effect Concentration (NOEC), or an

equivalent ECx. Observable endpoints typically include survival, growth and/or reproduction. Chronic toxicity exposure durations can vary widely depending on test endpoint measured and test species used.

~~A9.3.3.2.2 Since chronic toxicity data are less common in certain sectors than acute data, for classification schemes, the potential for chronic toxicity is identified by appropriate combinations of acute toxicity, lack of degradability, and/or the potential or actual bioaccumulation. Where such data exist and show long-term NOECs > 1 mg/l, this can be taken into account when deciding whether the classification based on the acute data should be applied. In this context, the following general approach should be used. In order to remove a chronic classification, it must be demonstrated that the NOEC used would be suitable in removing the concern for all taxa which resulted in classification. This can often be achieved by showing a long-term NOEC >1 mg/l for the most sensitive species identified by the acute toxicity. Thus, if a classification has been applied based on a fish acute LC50, it would generally not be possible to remove this classification using a long-term NOEC from an invertebrate toxicity test. In this case, the NOEC would normally need to be derived from a long-term fish test of the same species or one of equivalent or greater sensitivity. Equally, if classification has resulted from the acute toxicity to more than one taxa, it is likely that NOECs > 1 mg/l from each taxa will need to be demonstrated. In case of classification of a substance as Chronic Category 4, it is sufficient to demonstrate that NOECs are greater than the water solubility of the substances under consideration.~~

A.9.3.3.2.2 For the classification based on chronic toxicity a differentiation is made between rapidly degradable and non-rapidly degradable substances. Substances that do rapidly degrade are classified in Chronic Category 1 when a chronic toxicity determined to be ≤ 0.01 mg/l. Decimal bands are accepted for categorizing chronic toxicity above this category. Substances with a chronic toxicity measured from 0.01 to ≤ 0.1 mg/l are classified in Category 2 for chronic toxicity, from 0.1 to ≤ 1.0 mg/l are classified in Category 3 for chronic toxicity, and those over 1.0 mg/l are regarded as practically non-toxic. For substances that do not rapidly degrade or where no information on rapid degradation is available two Chronic Categories are used: Category 1 when a chronic toxicity determined to be ≤ 0.1 mg/l and Category 2 when chronic toxicity is measured from 0.1 to ≤ 1.0 mg/l.

A9.3.3.2.3 Since chronic toxicity data are less common in certain sectors than acute data, for classification schemes, the potential for chronic toxicity is, in absence of adequate chronic toxicity data, identified by appropriate combinations of acute toxicity, lack of degradability, and/or the potential or actual bioaccumulation. However, where adequate chronic toxicitydata exist, this shall be used in preference over the classification based on the combination of acute toxicity with degradability, and/or bioaccumulation. In this context, the following general approach should be used.

(a) If adequate chronic toxicity data are available for all three trophic levels this can be used directly to determine an appropriate chronic hazard category.

(b) If adequate chronic toxicity data are available for one or two trophic levels, it should be examined if acute toxicity data are available for the other trophic level(s). A potential classification is made for the trophic level(s) with chronic data and compared with that made using the acute toxicity data for the other trophic level(s). The final classification shall be made according to the most stringent outcome.

(c) In order to remove or lower a chronic classification using chronic toxicity data, it must be demonstrated that the NOEC(s) (or equivalent Ecx) used would be suitable to remove or lower the concern for all taxa which resulted in classification based on acute data in combination with degradability, and/or bioaccumulation. This can often be achieved by using a long-term NOEC for the most sensitive species identified by the acute toxicity. Thus, if a classification has been applied based on a fish acute LC50, it would generally not be possible to remove or lower this classification using a long-term NOEC from an invertebrate toxicity test. In this case, the NOEC would normally need to be derived from a long-term fish test of the same species or one of equivalent or greater sensitivity. Equally, if classification has resulted from the acute toxicity to more than one taxa, it is likely that NOECs from each taxa will be needed. In case of classification of a substance as Chronic Categorie 4, sufficient evidence

should be provided that the NOEC or equivalent EC_x for each taxa is greater than 1 mg/l or greater than the water solubility of the substances under consideration.

A9.3.3.2.4 (old A9.3.3.2.3) Testing with algae/Lemna cannot be used for ~~de-classifying chemicals~~ removing or lowering a classification because (~~1~~i) the algae and Lemna tests are not long-term studies, (~~2~~ii) the acute to chronic ratio is generally narrow and (~~3~~iii) the endpoints are more consistent with the acute endpoints for other organisms.

However where classification is applied solely due to the acute toxicity ($L(E)C_{50}$) observed in single algae/aquatic plant tests, but there is evidence from a range of other algae tests that the chronic toxicity (NOECs) for this taxonomic group is in the toxicity band corresponding to a less stringent classification category or above 1mg/l, this evidence could be used to consider ~~de~~removing or lowering a classification. At present this approach cannot be applied to aquatic plants since no standardized chronic toxicity tests have been developed.

~~A9.3.3.2.4 The GHS is intended to contain a specific value of chronic toxicity below which substances would be classified as chronically toxic, but the criteria are not yet set.~~

A9.3.3.3 *Exposure regimes*

Four types of exposure conditions are employed in both acute and chronic tests and in both freshwater and saltwater media: static, static-renewal (semi-static), recirculation, and flow-through. The choice for which test type to use usually depends on test substance characteristics, test duration, test species, and regulatory requirements.

A9.3.3.4 *Test media for algae*

Algal tests are performed in nutrient-enriched media and the use of one common constituent, EDTA, or other chelators, should be considered carefully. When testing the toxicity of organic chemicals, trace amounts of a chelator like EDTA are needed to complex micronutrients in the culture medium; if omitted, algal growth can be significantly reduced and compromise test utility. However, chelators can reduce the observed toxicity of metal test substances. Therefore, for metal compounds, it is desirable that data from tests with high concentration of chelators and/or tests with stoichiometrical excess of chelator relative to iron should be critically evaluated. Free chelator may mask heavy metal toxicity considerably, in particular with strong chelators like EDTA. However, in the absence of available iron in the medium the growth of algae can become iron limited, and consequently data from tests with no or with reduced iron and EDTA should be treated with caution.

A9.3.3.5 *Use of QSARs*

For purpose of classification, and in the absence of experimental data, QSARs can be relied upon to provide predictions of acute toxicity for fish, daphnia, and algae for non-electrolyte, non-electrophilic, and otherwise non-reactive substances (See Section A9.6 on *Use of QSAR*). Problems remain for substances such as organophosphates which operate by means of special mechanisms such as functional groups which interact with biological receptors, or which can form sulfhydryl bonds with cellular proteins. Reliable QSARs have been derived for chemicals acting by a basic narcosis mechanism. These chemicals are nonelectrolytes of low reactivity such as hydrocarbons, alcohols, ketones and certain aliphatic chlorinated hydrocarbons which produce their biological effects as a function of their partition coefficients. Every organic chemical can produce narcosis. However, if the chemical is an electrolyte or contains specific functional groups leading to non-narcotic mechanisms as well, any calculations of toxicity based on partition coefficient alone would severely underestimate the toxicity. QSARs for acute aquatic toxicity of parent compounds cannot be used to predict the effects of toxic metabolites or degradates, when these arise after a longer time period than the duration of acute tests.

A9.3.4 ***Weight of evidence***

A9.3.4.1 The best quality data should be used as the fundamental basis for classification. Classification should preferably be based on primary data sources. It is essential that test conditions ~~be~~ are clearly and completely articulated.

A9.3.4.2 Where multiple studies for a taxonomic group are available, a decision on what is the most sensitive and highest quality must be made. A judgement has to be made on a case by case basis whether a non-GLP study with a more sensitive observation is used in lieu of a GLP study. It would appear that results that indicate a high toxicity from tests performed according to non-standard or non-GLP guidelines should be able to be used for classification, whereas studies, which demonstrate negligible toxicity, would require more careful consideration. Substances, which are difficult to test, may yield apparent results that are more or less severe than the true toxicity. Expert judgement would also be needed for classification in these cases.

A9.3.4.3 Where more than one acceptable test is available for the same taxonomic group, the most sensitive (the one with the lowest $L(E)C_{50}$ or NOEC) is generally used for classification. However, this must be dealt with on a case-by-case basis. When larger data sets (4 or more values) are available for the same species, the geometric mean of toxicity values may be used as the representative toxicity value for that species. In estimating a mean value, it is not advisable to combine tests of different species within a taxa group or in different life stages or tested under different conditions or duration.

A9.3.5 ***Difficult to test substances***

A9.3.5.1 Valid aquatic toxicity tests require the dissolution of the test substance in the water media under the test conditions recommended by the guideline. In addition, a bioavailable exposure concentration should be maintained for the duration of the test. Some chemical substances are difficult to test in aquatic systems and guidance has been developed to assist in testing these materials (DoE 1996; ECETOC 1996; and US EPA 1996). OECD is in the process of finalizing a Guidance Document on Aquatic Toxicity testing of Difficult Substances and Mixtures (OECD, 2000). This latter document is a good source of information on the types of substances that are difficult to test and the steps needed to ensure valid conclusions from tests with these materials.

A9.3.5.2 Nevertheless, much test data exist that may have used testing methodologies which, while not in conformity with what might be considered best practice today, can still yield information suitable for application of the classification criteria. Such data require special guidance on interpretation, although ultimately, expert judgement must be used in determining data validity. Such difficult to test substances may be poorly soluble, volatile, or subject to rapid degradation due to such processes as phototransformation, hydrolysis, oxidation, or biotic degradation. When testing algae, coloured materials may interfere with the test endpoint by attenuating the light needed for cell growth. In a similar manner, substances tested as cloudy dispersions above solubility may give rise to false toxicity measurements. Loading of the water column with test material can be an issue for particulates or solids such as metals. Petroleum distillate fractions can also pose loading problems, as well as difficult interpretational problems when deciding on the appropriate concentrations for determining $L(E)C_{50}$ values. The draft Guidance Document on Aquatic Toxicity Testing of Difficult Substances and Mixtures describes the more common properties of many types of substances which are likely to pose testing difficulties.

(a) Stability: If test chemical concentrations are expected to fall below 80% of nominal, testing, in order to be valid, may require exposure regimes which provide for renewal of the test material. Semi-static or flow-through conditions are preferred. Special problems arise, therefore, with respect to testing on algae, where the standard guidelines generally include static tests to be conducted. While alternative exposure regimes are possible for crustacea and fish, these tests are frequently conducted on static conditions as included in the internationally agreed guidelines. In these tests, a certain level of degradation as well as other relevant factors ~~has~~ have to be tolerated and appropriate account must be taken in calculations of toxic concentrations. Some approaches on how this can be dealt with are covered in A9.3.5.6. Where degradation occurs, it is also important to consider the

influence of the toxicity of the degradation products on the recorded toxicity in the test. Expert judgement will need to be exercised when deciding if the data can be used for classification;

(b) Degradation: When a compound breaks down or degrades under test condition, expert judgement should be used in calculating toxicity for classification, including consideration of known or likely breakdown products. Concentrations of the parent material and all significant toxic degradates are desirable. If degradates are expected to be relatively non-toxic, renewable exposure regimes are desirable in order to ensure that levels of the parent compounds are maintained.

(c) Saturation: For single component substances, classification should be based only on toxic responses observed in the soluble range, and not on total chemical loading above solubility. Frequently, data are available which indicate toxicity at levels in excess of water solubility and, while these data will often be regarded as not valid, some interpretation may be possible. These problems generally apply when testing poorly soluble substances, and guidance on how to interpret such data is included in A9.3.5.7 (see also the Guidance Document on Aquatic Toxicity testing of Difficult Substances and Mixtures);

(d) Perturbation of test media: Special provisions may be needed to ensure dissolution of difficult to test substances. Such measures should not lead to significant changes in the test media when such changes are likely to lead to an increase or decrease in the apparent toxicity and hence the classification level of the test substance;

(e) Complex substances: Many substances covered by the classification scheme are in fact mixtures, for which measurement of exposure concentrations is difficult, and in some cases impossible. Substances such as petroleum distillate fractions, polymers, substances with significant levels of impurities, etc can pose special problems since the toxic concentration is difficult to define and impossible to verify. Typical testing procedures often rely on the formation of a Water Soluble Fraction (WSF) or Water Accommodated Fraction (WAF) and data are reported in terms of loading rates. These data may be used in applying the classification criteria.

A9.3.5.3 For classification of organic compounds, it is desirable to have stabilized and analytically measured test concentrations. Although measured concentrations are preferred, classification may be based on nominal concentration studies when these are the only valid data available under certain circumstances. If the material is likely to substantially degrade or otherwise be lost from the water column, care must be taken in data interpretation and classification should be done taking the loss of the toxicant during the test into account, if relevant and possible. Additionally, metals present their own set of difficulties and are discussed separately. Table A9.3.1 lists several properties of difficult to test substances and their relevance for classification.

A9.3.5.4 In most difficult to test conditions, the actual test concentration is likely to be less than the nominal or expected test concentration. Where acute toxicities ($L(E)C_{50}s$) are estimated to be less than 1 mg/l for a difficult to test substance, one can be fairly confident the classification in the Acute Category 1 (and Chronic Category 1 if appropriate) is warranted. However, if the estimated acute toxicity is greater than 1 mg/l, the estimated toxicity is likely to under-represent the toxicity. In these circumstances, expert judgement is needed to determine the acceptability of a test with a difficult to test substance for use in classification. Where the nature of the testing difficulty is believed to have a significant influence on the actual test concentration when acute toxicity is estimated to be greater than 1 mg/l and the test concentration is not measured, then the test should be used with due caution in classification.

A9.3.5.5 The following paragraphs provide some detailed guidance on some of these interpretational problems. In doing so it should be remembered that this is guidance and hard and fast rules cannot be applied.

The nature of many of the difficulties mean that expert judgement must always be applied both in determining whether there is sufficient information in a test for a judgement to be made on its validity, and also whether a toxicity level can be determined suitable for use in applying the classification criteria.

A9.3.5.6 *Unstable substances*

A9.3.5.6.1 While testing procedures should ideally have been adopted which minimized the impacts of instability in the test media, in practice, in certain tests, it can be almost impossible to maintain a concentration throughout the test. Common causes of such instability are oxidation, hydrolysis, photodegradation and biodegradation. While the latter forms of degradation can more readily be controlled, such controls are frequently absent in much existing testing. Nevertheless, for some testing, particularly acute and chronic fish toxicity testing, a choice of exposure regimes is available to help minimize losses due to instability, and this should be taken into account in deciding on the test data validity.

A9.3.5.6.2 Where instability is a factor in determining the level of exposure during the test, an essential prerequisite for data interpretation is the existence of measured exposure concentrations at suitable time points throughout the test. In the absence of analytically measured concentrations at least at the start and end of test, no valid interpretation can be made and the test should be considered as invalid for classification purposes. Where measured data are available, a number of practical rules can be considered by way of guidance in interpretation:

(a) where measured data are available for the start and end of test (as is normal for the acute Daphnia and algal tests), the $L(E)C_{50,}$ for classification purposes, may be calculated based on the geometric mean of the start and end of test concentrations. Where the end of test concentrations are below the analytical detection limit, such concentrations shall be considered to be half that detection limit.

(b) where measured data are available at the start and end of media renewal periods (as may be available for the semi-static tests), the geometric mean for each renewal period should be calculated, and the mean exposure over the whole exposure period calculated from these data.

(c) where the toxicity can be attributed to a degradation breakdown product, and the concentrations of this are known, the $L(E)C_{50}$ for classification purposes, may be calculated based on the geometric mean of the degradation product concentration, back calculated to the parent substance.

(d) similar principles may be applied to measured data in chronic toxicity testing.

A9.3.5.7 *Poorly soluble substances*

A9.3.5.7.1 These substances, usually taken to be those with a solubility in water of <1 mg/l, are frequently difficult to dissolve in the test media, and the dissolved concentrations will often prove difficult to measure at the low concentrations anticipated. For many substances, the true solubility in the test media will be unknown, and will often be recorded as < detection limit in purified water. Nevertheless such substances can show toxicity, and where no toxicity is found, judgement must be applied to whether the result can be considered valid for classification. Judgement should err on the side of caution and should not underestimate the hazard.

A9.3.5.7.2 Ideally, tests using appropriate dissolution techniques and with accurately measured concentrations within the range of water solubility should be used. Where such test data are available, they should be used in preference to other data. It is normal, however, particularly when considering older data, to find such substances with toxicity levels recorded in excess of the water solubility, or where the dissolved levels are below the detection limit of the analytical method. Thus, in both circumstances, it is not possible to verify the actual exposure concentrations using measured data. Where these are the only data available on which to classify, some practical rules can be considered by way of general guidance:

(a) where the acute toxicity is recorded at levels in excess of the water solubility, the $L(E)C_{50}$ for classification purposes, may be considered to be equal to or below the measured water solubility. In such circumstances it is likely that Chronic Category 1 and/or Acute Category 1 should be applied. In making this decision, due attention should be paid to the possibility that the excess undissolved substance may have given rise to physical effects on the test organisms. Where this is considered the likely cause of the effects observed, the test should be considered as invalid for classification purposes;

(b) where no Acute toxicity is recorded at levels in excess of the water solubility, the $L(E)C_{50}$ for classification purposes may be considered to be greater than the measured water solubility. In such circumstances, consideration should be given to whether the Chronic Category 4 should apply. In making a decision that the substance shows no acute toxicity, due account should be taken of the techniques used to achieve the maximum dissolved concentrations. Where these are not considered as adequate, the test should be considered as invalid for classification purposes;

(c) where the water solubility is below the detection limit of the analytical method for a substance, and acute toxicity is recorded, the $L(E)C_{50}$ for classification purposes, may be considered to be less than the analytical detection limit. Where no toxicity is observed, the $L(E)C_{50}$ for classification purposes, may be considered to be greater than the water solubility. Due consideration should also be given to the quality criteria mentioned above;

(d) where chronic toxicity data are available, the same general rules should apply. ~~In principle, only data showing no effects at the water solubility limit, or greater than 1 mg/l need be considered.~~ Again, where these data cannot be validated by consideration of measured concentrations, the techniques used to achieve the maximum dissolved concentrations must be considered as appropriate.

A9.3.5.8 *Other factors contributing to concentration loss*

A number of other factors can also contribute to losses of concentration and, while some can be avoided by correct study design, interpretation of data where these factors have contributed may, from time to time, be necessary.

(a) sedimentation: this can occur during a test for a number of reasons. A common explanation is that the substance has not truly dissolved despite the apparent absence of particulates, and agglomeration occurs during the test leading to precipitation. In these circumstances, the $L(E)C_{50}$ <u>or NOEC</u> for classification purposes, may be considered to be based on the end of test concentrations. Equally, precipitation can occur through reaction with the media. This is considered under instability above;

(b) adsorption: this can occur for substances of high adsorption characteristics such as high log Kow substances. Where this occurs, the loss of concentration is usually rapid and exposure may best be characterized by the end of test concentrations.

(c) bioaccumulation: losses may occur through the bioaccumulation of a substance into the test organisms. This may be particularly important where the water solubility is low and log K_{ow} correspondingly high. The $L(E)C_{50}$ <u>or NOEC</u> for classification purposes, may be calculated based on the geometric mean of the start and end of test concentrations.

A9.3.5.9 *Perturbation of the test media*

A9.3.5.9.1 Strong acids and bases may appear toxic because they may alter pH. Generally however

changes of the pH in aquatic systems are normally prevented by buffer systems in the test medium. If no data are available on a salt, the salt should generally be classified in the same way as the anion or cation, i.e. as the ion that receives the most stringent classification. If the effect concentration is related to only one of the ions, the classification of the salt should take the molecular weight difference into consideration by correcting the effect concentration by multiplying with the ratio: $MW_{salt}/MW_{ion.}$

A9.3.5.9.2 Polymers are typically not available in aquatic systems. Dispersible polymers and other high molecular mass materials can perturb the test system and interfere with uptake of oxygen, and give rise to mechanical or secondary effects. These factors need to be taken into account when considering data from these substances. Many polymers behave like complex substances, however, having a significant low molecular mass fraction which can leach from the bulk polymer. This is considered further below.

A9.3.5.10 *Complex substances*

A9.3.5.10.1 Complex substances are characterized by a range of chemical structures, frequently in a homologous series, but covering a wide range of water solubilities and other physico-chemical characteristics. On addition to water, an equilibrium will be reached between the dissolved and undissolved fractions which will be characteristic of the loading of the substance. For this reason, such complex substances are usually tested as a WSF or WAF, and the $L(E)C_{50}$ recorded based on the loading or nominal concentrations. Analytical support data are not normally available since the dissolved fraction will itself be a complex mixtures of components. The toxicity parameter is sometimes referred to as LL_{50}, related to the lethal loading level. This loading level from the WSF or WAF may be used directly in the classification criteria.

A9.3.5.10.2 Polymers represent a special kind of complex substance, requiring consideration of the polymer type and their dissolution/dispersal behaviour. Polymers may dissolve as such without change, (true solubility related to particle size), be dispersible, or portions consisting of low molecular weight fractions may go into solution. In the latter case, in effect, the testing of a polymer is a test of the ability of low molecular mass material to leach from the bulk polymer, and whether this leachate is toxic. It can thus be considered in the same way as a complex mixture in that a loading of polymer can best characterize the resultant leachate, and hence the toxicity can be related to this loading.

Table A9.3.1 Classification of difficult test substances

Property	Nature of difficulty	Relevance for classification
Poorly water soluble	Achieving/maintaining required exposure concentration. Analysing exposure.	When toxic responses are observed above apparent solubility, expert judgement is required to confirm whether effects are due to chemical toxicity or a physical effect; if no effects are observed, it should be demonstrated that full, saturated dissolution has been achieved.
Toxic at low concentrations	Achieving/maintaining required exposure concentration. Analysing exposure.	Classified based on toxicity < 1 mg/l
Volatile	Maintaining and measuring exposure concentration.	Classification should be based on reliable measurement of concentrations.
Photo-degradable	Maintaining exposure concentrations. Toxicity of breakdown products.	Classification requires expert judgement and should be based on measured concentrations. Toxicity of significant breakdown products should be characterized.
Hydrolytically unstable	Maintaining exposure concentrations. Toxicity of breakdown products. Comparison of degradation half-lives to the exposure regimen used in testing.	Classification requires expert judgement, should be based on measured concentrations, and needs to address the toxicity of significant breakdown products.
Oxidizable	Achieving, maintaining and measuring exposure concentration. Toxicity of modified chemical structures or breakdown products. Comparison of degradation half-lives to the exposure regimen used in testing.	Classification requires expert judgement, should be based on measured concentrations, and needs to address the toxicity of significant breakdown products.
Subject to corrosion/ transformation (this refers to metals /metal compounds)	Achieving, maintaining and measuring exposure concentration. Comparison of partitioning from the water column half-lives to the exposure regimen used in testing.	Classification requires expert judgement, should be based on measured concentrations, and needs to address the toxicity of significant breakdown products.
Biodegradable	Maintaining exposure concentrations. Toxicity of breakdown products. Comparison of degradation half-lives to the exposure regimen used in testing.	Classification requires expert judgement, should be based on measured concentrations, and needs to address the toxicity of significant breakdown products.
Adsorbing	Maintaining exposure concentrations. Analysing exposure. Toxicity mitigation due to reduced availability of test substance.	Classification should use measured concentration of available material.
Chelating	Distinguishing chelated and non-chelated fractions in media.	Classification should use measurement of concentration of bioavailable material.
Coloured	Light attenuation (an algal problem).	Classification must distinguish toxic effects from reduced growth due to light attenuation.
Hydrophobic	Maintaining constant exposure concentrations.	Classification should use measured concentration.
Ionized	Maintaining exposure concentrations. Toxicity of breakdown products. Comparison of degradation half-lives to the exposure regime used in testing.	Classification requires expert judgement, should be based on measured concentrations, and needs to address the toxicity of significant breakdown products.
Multi-component	Preparing representative test batches.	Considered same as complex mixture.

A9.3.6 ***Interpreting data quality***

A9.3.6.1 *Standardization*

Many factors can influence the results of toxicity tests with aquatic organisms. These factors include characteristics of the test water, experimental design, chemical characteristics of the test material, and biological characteristics of the test organisms. Therefore, it is important in conducting aquatic toxicity tests to use standardized test procedures to reduce the influence of these sources of extraneous variability. The goal of test standardization and international harmonization of these standards is to reduce test variability and improve precision, reproducibility, and consistency of test results.

A9.3.6.2 *Data hierarchies*

A9.3.6.2.1 Classification should be based on primary data of good quality. Preference is given to data conforming to OECD Test Guidelines or equivalent and Good Laboratory Practices (GLP). While data from internationally harmonized test methods performed on standard test species are preferred, results of tests performed using widely recognized international or national methods or their equivalent may also be used, e.g. ISO or ASTM methods. Data from tests that appear to conform to accepted guidelines but which lacks provisions for GLP can be used in the absence of pertinent GLP data.

A9.3.6.2.2 Pedersen et al (1995) provides a data quality-scoring system, which is compatible with many others in current use, including that, used by the US-EPA for its AQUIRE database. See also Mensink et al (1995) for discussions of data quality. The data quality scoring system described in Pedersen *et al*. includes a reliability ranking scheme, which can be a model for use with in classifying under the harmonized scheme. The first three levels of data described by Pedersen are for preferred data.

A9.3.6.2.3 Data for classification under the harmonized scheme should come from primary sources. However, since many nations and regulatory authorities will perform classification using the globally harmonized scheme, classification should allow for use of reviews from national authorities and expert panels as long as the reviews are based on primary sources. Such reviews should include summaries of test conditions, which are sufficiently detailed for weight of evidence and classification decisions to be made. It may be possible to use the reviews, which were made by a well-recognized group such as GESAMP for which the primary data are accessible.

A9.3.6.2.4 In the absence of empirical test data, validated Quantitative Structure Activity Relationships (QSARs) for aquatic toxicity may be used. Test data always take precedence over QSAR predictions, providing the test data are valid.

Proposed revision of Appendix VI:

Add in Section 1 the following reference:

OECD 2006. Current Approaches in the Statistical Analysis OF Ecotoxicity Data: A Guidance to Application, OECD Environment? Health and Safety Publications, Series Testing and Assessment N) 54