ZIRN - Interdisciplinary research unit on risk governance and sustainable technology development

University of Stuttgart
Institute for
social sciences
dep. for environmental
sociology

DIALOGIK
non-profit institute for
communication and
cooperation research

Stuttgart contributions to risk and sustainability research

# *Stakeholder Involvement – Results of two Workshops*

**OSIRIS - Optimized Strategies for Risk Assessment of Industrial Chemicals through Integration of Non-Test and Test Information**

**Christina Benighaus (Ed.)**

**No. 15 / October 2009**

Institute for Social Sciences
Department for Environmental Sociology
Prof. Dr. Dr. h.c. Ortwin Renn
University of Stuttgart

# *Stakeholder Involvement and Dissemination – Results of two Workshops*

**OSIRIS - Optimized Strategies for Risk Assessment of Industrial Chemicals through Integration of Non-Test and Test Information**

**Christina Benighaus (Ed.)**

**No. 15 / October 2009**

## Workshop reports

ISSN 1614-3035
ISBN 978-3-938245-14-9

Institute for Social Sciences
Department for Environmental Sociology
University of Stuttgart
Seidenstr. 36, 70174 Stuttgart, Germany
Tel: ++49 (0)711 685 83971, Fax: ++49 (0) 711 685 82487
Email: ortwin.renn@sowi.uni-stuttgart.de
Internet: http://www.uni-stuttgart.de/soz/tu

DIALOGIK, ltd.
Lerchenstr. 22, 70176 Stuttgart, Germany
Tel: ++49 (0)711 3585 216-4, Fax: ++49 (0)711 3585 216-0
Email: info@dialogik-expert.de
Internet: www.dialogik-expert.de

ZIRN
International Center for Cultural and Technological Studies
Interdisciplinary Research Unit on Risk Governance
and Sustainable Technology Development
Seidenstr. 36, 70174 Stuttgart, Germany
Tel: ++49 (0)711 685 83971, Fax: ++49 (0)711 685 82487
Email: ortwin.renn@sowi.uni-stuttgart.de
Internet: http://www.zirn-info.de

Contact: Christina Benighaus
Tel: ++49 (0)711 685 84295
christina.benighaus@sowi.uni-stuttgart.de

# Contents

# Preface

*Christina Benighaus, DIALOGIK Stuttgart, Germany*

This book describes different possibilities of stakeholder involvement in the nature science research process. It summarizes the outlines and results of two stakeholder workshops of the OSIRIS project. **OSIRIS** will develop integrated testing strategies (ITS) fit for REACH. These enable to significantly increase the use of non-testing information for regulatory decision making, and thus to minimise the need for animal testing.

The OSIRIS project aims to replace testing methods with non-testing strategies that provide results primarily based on computer modelling and simulation with a similar degree or even higher degree of validity and reliability than the results of experimental testing procedures. Using computer models and other non-testing methods help to optimise efficiency, reduce overall costs, match the ambitious time schedule of the REACH regime and improve public acceptance due to less animal testing. OSIRIS will work with statistical models and data obtained from testing methods that promise fast, accurate and targeted information. A potential drawback of these new procedures is the possibility that unexpected outlayers remain unnoticed because computer models can only deduce inferences from existing knowledge (reducing epistemic uncertainty) and random variations (reducing aleatory uncertainty).

To ensure optimal uptake of the results obtained in this project, end-users in industry and regulatory authorities (EU-stakeholders) have been invited to participate in the project, for example by becoming involved in monitoring and by providing specific technical contributions.

A central component of the stakeholder involvement strategy is the organisation of four workshops along the basic research steps. The workshops will highlight the different approaches and present the results of the project to key stakeholders. In this document, the outline and results of the first two workshops are described.

The first Expert Workshop took place in Stuttgart (Germany) at the Waldhotel Degerloch in November 2007 for two days. 24 scientific experts from industry, academia, and regulatory agencies attended the workshop. The first workshop aims at receiving feedback from the participants on the overall scientific approach and the main framing of the research. For this purpose, representatives of Pillar 4 coordinated by Dr. J.J.M. (Han) van de Sandt, (TNO Quality of Life in Zeist) presented their research plans and asked for feedback from experts. This consultation was organized in the form of a Group Delphi.

The second workshop was designed to provide an overview of preliminary results and was scheduled for November 2008 in Brussels. The main topic of this workshop was the match between REACH requirements and the proposed methods to meet these requirements with respect to human and ecological endpoints. Major stakeholders from industry and civil society as well as a group of interdisciplinary experts from academia and government have been invited to provide valuable input to the OSIRIS team and to discuss the contributions of the OSIRIS research for the REACH characterization and risk assessment process.

This paper reports the structure of the first two workshops and its results. The first chapter characterizes the Group Delphi method and describes the results of one individual and two group surveys in the context of the Group Delphi. The outcomes of the first workshop are summarized in the last section.

Chapter 2 characterizes the structure of the second workshop with presenters and opponents and World Café. The outcomes are decribed in the following sections.

The workshops were part of pillar 4 of the integrated EU-Project OSIRIS which is coordinated by Dr. J.J.M. (Han) van de Sandt, TNO Quality of Life, Utrechtseweg 48, 3704 HE Zeist, The Netherlands.

The project is funded by the European Commission within in the 6th Framework Programme under the theme "Global Change and Ecosystems", coordinated by Prof. Dr. Gerrit Schüürmann at the Helmholtz Centre for Environmental Research (UFZ) in Leipzig.

# 1 First Stakeholder Workshop - Group Delphi

*Prof. Dr. Dr. h.c. Ortwin Renn, Dr. Michael Ruddat, Christina Benighaus, DIALOGIK Stuttgart, Germany*

*Dr. Han van de Sandt, Prof. Dr. Kees van Leeuwen, Dr. Dinant Kroese, TNO Zeist, the Netherlands*

## 1.1 Summary

The main aim of the first Expert Workshop of the OSIRIS project ("Optimized Strategies for Risk Assessment of Industrial Chemicals through Integration of Non-Test and Test Information") was to receive feedback from the participants on the overall scientific approach and the main framing of the research within Pillar 4. The principle results of the Group Delphi were the following:

- The participants agreed that it is highly important to develop integrated assessment strategies for (groups) of chemicals, using the different building blocks of ITS. Second place goes to the development and evaluation of individual ITS building blocks for physicochemical, (eco)toxicological and exposure data, followed by building a set of databases (with experimental and other data) that can be used for many purposes in and outside of OSIRIS.
- With respect to the output (products) of OSIRIS, operational ITSs for all endpoints, using a weight-of-evidence approach as well as operational overall ITS, also using a weight-of-evidence approach, should have priority. There was clear consensus that reproductive toxicity has the highest ranking because of the num-

ber of animals and costs involved. Local toxicity (skin, eye) got the lowest ranking from all four groups in the third Delphi round.

- Data obtained from in vivo studies were seen as very important for inclusion in the ITSs, but OSIRIS should not put effort in generating in vivo data. The innovation of ITSs should be realized through non-testing approaches. No consensus could be reached in prioritizing of the building blocks of ITS.
- Exposure and exposure categories, descriptions of categories of chemicals related to mode of toxic action and information of modes of toxic action (e.g. chemical reactivity) are the databases which were deemed most important in the future for (eco)toxicity and exposure assessment compared to today and should therefore be the focus of OSIRIS.
- The OECD (Q)SAR Toolbox as well as the RIP's should be considered by OSIRIS as being important ongoing international efforts and efforts should be made to link the various activities.
- There was high consensus among all four groups that data and model uncertainty should be included in a scoring system for data quality.

## 1.2 Methods of the Group Delphi

A Delphi process is aimed at obtaining a wide range of opinions among a group of experts (Turoff, 1970; Pill, 1971; Linstone and Turoff, 2002). The process is organized in four steps. In step 1, a questionnaire asks a group of distinguished scientists to rank or rate several items, in this case different methods for data collection, testing and verification. The scientists provide their best estimate and assign a confidence interval to their answers. In step 2, the organizing team feeds back to each participant the scores of the whole group, including medians, standard deviation and aggregated confidence intervals.

Each individual is then asked to perform the same task again, but now with the knowledge of the responses of all other participants. In step 3, this procedure is repeated until individuals do not change their assessment any more. In step 4, the organizer summarizes the results and articulates the conclusions.

A variation of the classic Delphi method is the group Delphi (Webler et al, 1991). During a group Delphi all participants meet face to face and make the assessments in randomly assigned small groups of three or four. The groups whose average scores deviate most from the median of all other groups are requested to defend their position in a plenary session. Then the small groups are reshuffled and perform the same task again. This process can be iterated three or four times until no further significant changes are made. At the end of a Delphi process, one receives either a normal distribution of assessments around a common median, a two- or three-peak distribution (signalling a majority and one or more minority votes) or a flat curve (which means that knowledge is insufficient to make any reliable assessment).

The advantage of Delphi is that a serious effort has been invested in finding the common ground among the experts and in finding the reasons and arguments that cause differences in assessments. The disadvantage is that Delphis depend upon the quality and completeness of the expertise and information brought into the process. In general, DIALOGIK has had mostly positive experiences with Delphi processes, particularly group Delphi.

**Literature**

Linstone, H. A. & Turoff, M. (Hrsg.). (2002). The Delphi Method: Techniques and Applications. New Jersey: Science and Technology University.

Pill, J. (1971). The Delphi method: Substance, context, a critique and an anotated bibliography. Socio-Economic Planning Sciences, 5.

Turoff, M. (1970). The Design of a Policy Delphi. Technological Forecasting and Social Change, 2, 149-171.

Webler, T., Levine, D., Rakel, H., Renn, O.: "The Group Delphi: A Novel Attempt at Reducing Uncertainty," Technological Forecasting and Social Change, 39 (1991), 253-263.

## 1.3 Results of first round (Individual Delphi)

In the first Delphi round, a questionnaire was sent to each participant nine days before the workshop. It consisted of six key topics including one or more key questions each. The research team of DIALOGIK received 17 questionnaires. The results were summarized in an excel sheet which indicated the median value as well as the distribution of answers. This sheet together with a short presentation of the results formed the basis for the deliberations at the Group Delphi. Figure 1-1 in Annex 2 on page 22 gives a short overview of the results.

There was some **clear consensus** among the respondents on most topics. For example, on key topic 3 (3.1: Databases of Chemicals - Kind of databases) almost all experts fully agreed that all of the available databases should be included in the future for (eco)toxicity and exposure assessment. Also, the OECD toolbox and the RIP's were rated as very important or important by a vast majority of the participants. Almost every one agreed on the high importance of the OECD principles as part of key topic 4 (4.3: Quality of data – Criteria) for OSIRIS. Key topic 5 was also addressed in a similar fashion by almost all respondents. This section dealt with public access of results and the installation of an open website for ITS. With the exception of one single expert voice, all of them agreed that all scientific results should be made public available.

On other topics the answers of the respondents **differed, sometimes even dramatically.** For example, on key topic 2 (2.1: Tools and instruments for testing strategies of OSIRIS - Building blocks of ITS), hazard data from structurally related chemicals (read-across), (Q)SARs and Threshold of Toxicological Concern got very high rates of approval whereas in vitro methods, non-guideline animal data and animal data generated according to accepted guidelines were rated as

less important by some experts and highly important by others. Responses to "non-guideline animal data" and "in vitro methods generated according to 'suitable' methodologies" varied considerably among the experts. There were five participants who believed in the usefulness of "non-guideline animal data" as a building block of ITS and six participants who believed in the usefulness of "in vitro methods generated according to 'suitable' methodologies", while the others felt them to be less useful.

There was only little variation in key topic 1 (1.1: Focus of OSIRIS). The development of integrated assessment strategies for (groups) of chemicals and generation of ITS procedures (IT-Tools and Guidance Documents) were rated by a vast majority as more important than the contribution to the generation of databases and development and evaluation of individual ITS building blocks. On key topic 4 (4.1: Quality of data - Quality parameters - Reliability), the identity, purity and source of substance, the availability of information on structural analogues and the substantiation of deviations from guidelines were top priorities in the eyes of most experts. Statistics, analytical methods and publication in peer-reviewed journals were rated as less important. The availability of the complete test report or exposure considerations shows more expert dissent: Some experts see them as very important while others rate them as only partly important.

In essence, the first individual questionnaire demonstrated an astonishing degree of convergence for most issues. The variance was usually low and only few items were clearly controversial. The objective of the Group Delphi Rounds that were conducted after the individual data was displayed and explained was to investigate whether the response patterns remained stable under the condition of intensive discussions.

## 1.4 Results of next two rounds (Group Delphi)

The first Group Delphi round took place after Dr. Dr. Han van de Sandt, Dr. Dinant Kroese and Prof. Dr. Ortwin Renn gave a short introduction to OSIRIS and explained the Delphi Method. In addition, Prof. Dr. C. J. (Kees) van Leeuwen presented the results of the individual Delphi round.

Five small groups were formed by random selection consisting of four individuals. Each of the group was asked to fill out the same questionnaire that they had individually responded to in the individual survey. The groups were encouraged to discuss the meaning of the question and deliberate about the most suitable answers. As expected the group discussion revealed first that many participants associated different connotations with each question. Secondly, by looking more closely into each question more variance and disagreement was produced. The five groups had about 1,5 hours time to discuss and give their ratings as a group vote. They were also free to add comments or refine the wording of the question. Once the groups had completed their task, the research team of DIALOGIK processed the data from the filled out questionnaires and provided a summary of the results for presentation during the plenary session.

During the plenary, the moderator asked groups that deviated most from the median value on each question to justify their judgments. This way the discussion focused on the differences not the similarities. Often differences were due to unclear formulation of the question or to different connotations of the terms used in the questions or in the list of standardized response categories. Occasionally differences were the result of calibration problems with respect to response categories such as very important versus important. There were only few questions where respondents had polarized views or were at the opposite end of the response scale. Yet there were quite a few significant variations in the middle range of the answer categories.

Those questions that did not produce any significant disagreement were scraped from the next round of deliberation. For example, the central role of the OECD principles as a basis for (Q)SAR tools in OSIRIS was unchallenged in all groups. There was also unanimous support for the desired public availability of methodologies, databases and the webtool for ITS. Both questions were therefore left out in the third round. The questions that seemed ambiguous were reformulated and many response categories were further specified. Several rating scales (from not important to very important) were transformed into ranking scales to force respondents to set priorities. Finally, additional questions were added where needed and new scales introduced. Once the new questionnaire was completed, a second group round was organized. This time the composition of the group was permutated so that each new group consisted of at least one member of the four groups in the first round. Since the total number of participants was 20, a complete permutation was not possible but this objective could be widely met by composing a total of four groups with five participants each.

Overall, the ratio of consensus to dissent decreased from 1,4 to 0,5 between the individual and the first group round. This was mainly due to the discussion about the meaning of terms and the intention of the OSIRIS team. After carefully reformulating the ambiguous questions, generating new ones and introducing additional scales, the consensus/dissent-ratio increased to 3 indicating that consensual votes occurred three times more frequently than dissenting views. This comparison includes the consensual responses of round 1, which were deleted for the second Round of the Group Delphi. This is indicated by "– (▼)". Also important to notice is that dissent does not necessarily mean substantive dissent or difference in judgments. Very often misunderstanding, misinterpretations, different connotations of terminology or ambiguities in the response categories were most frequently the source of the differences. Those problems could be resolved in the plenary discussion.

One can briefly summarize the final results after round 3 as follows:

- The participants agreed that it is highly important to develop integrated assessment strategies for (groups) of chemicals, using the different building blocks of ITS. Second place goes to the development and evaluation of individual ITS building blocks for physicochemical, (eco)toxicological and exposure data, followed by building a set of databases (with experimental and other data) that can be used for many purposes in and outside of OSIRIS.
- No definite statement on the conditions for success of OSIRIS could be formulated, but the actual use of the OSIRIS tool in practice (by industry and regulators) should be the main aim of the project.
- With respect to the output (products) of OSIRIS, operational ITSs for all endpoints, using a weight-of-evidence approach as well as operational overall ITS, also using a weight-of-evidence approach, should have priority. There was clear consensus that reproductive toxicity has the highest ranking because of the number of animals and costs involved. Due to its toxicological complexity, it was realized that developing an innovative ITS for this endpoint is a challenge, both from a scientific and regulatory acceptance point of view. Local toxicity (skin, eye) got the lowest ranking from all four groups in the third Delphi round.
- No consensus could be reached in prioritizing of the building blocks of ITS, but some conditional remarks came up during the plenary discussion, which reduced the degree of dissent. For example, when high quality data is available for hazard data from structurally related chemicals (read-across), then all groups assigned a high degree of relevance to the building blocks; however if not, the relevance was seen as less pronounced. Data obtained from in vivo studies were seen as very important for inclusion in the ITSs, but OSIRIS should not put effort in generating in vivo data. The innovation of ITSs should be realized through non-testing approaches.

- Exposure and exposure categories, descriptions of categories of chemicals related to mode of toxic action and information of modes of toxic action (e.g. chemical reactivity) are the databases which were deemed most important in the future for (eco)toxicity and exposure assessment compared to today and should therefore be the focus of OSIRIS. This was the common judgment of tall four groups as the result of the third Delphi round.
- The OECD (Q)SAR Toolbox as well as the RIP's should be considered by OSIRIS as being important ongoing international efforts and efforts should be made to link the various activities.
- There was high consensus among all four groups that data and model uncertainty should be included in a scoring system for
- data quality.

The participants added many comments and specifications, which were all recorded by the research team. The group results and comments are shown in Figures 1-3 and 1-4 in Annex 2 on page 33 and 41, respectively. Please note that for priority judgments the first digit indicates the number of the group, while the second digit indicates priority. For example "3,1" means that group number 3 gave this item first priority.

## 1.5 Concluding remarks

The Delphi exercise demonstrated the importance of structured discussion about terms and categories as the individual responses indicated a degree of consensus that did not reflect the true representation of the respondents' views. The process of small group judgment and plenary justification for explaining group differences lead to a more precise wording of the issues and topics and helped the OSIRIS research team to gain a better view of the priorities that the partici-

pants assigned to the different tasks and activities. Furthermore new topics were introduced in the group discussions that added more depth to the analysis.

The results of the Group Delphi will have an impact on the protocol and agenda of Pillar 4. Since the respondents made clear choices with respect to priorities, the research team can concentrate on those tasks that all respondents felt of having high importance to the OSIRIS overall objectives. In addition, the team has gained a better understanding of the preferences of the respondents and are better informed about their needs. Finally, the respondents approved of the main tasks and planned activities of Pillar 4, which represents a powerful message to the team that they are on the right track.

## 1.6 Annexes: Additional comments, Tables & Agenda

**Annex 1: Additional comments**

*A: PD Dr. Jan Ahlers, Federal Environmental Agency, Dessau, Germany, member of the OSIRIS Advisory Board*

- The target of OSIRIS was defined in "substitution of vertebrate testing", which in my understanding is far too narrow.

Although definitely substitution of vertebrate tests is - beside protection of man and the environment - an important issue of REACH, I interpret the objectives of OSIRIS

- "OSIRIS will undertake distinct research …, and their integration in a decision theory framework" or
- "The OSIRIS project will develop ITS ….to significantly increase the use of non-testing information for regulatory decision making", (flyer of the workshop)

- …"minimizing the need for new testing in risk assessment procedures" (Objective Pillar 4)

In a broader way. At least two additional items are rather important:

1. substitution of non-vertebrate testing (e.g. tests for soil and sediment organisms) will help to save time and costs and even more important integrating alternative information enables us to perform a more comprehensive and faster assessment of these compartments and thus will certainly be an important contribution to environmental protection.
2. The information obtained from alternative methods should not only be used for testing strategies, but should also be introduced in risk assessment. It can contribute considerably in reducing uncertainty in regulatory decision making.

*B: Dr. Monika Nendza, Analytical Laboratory Luhnstedt, Germany*

As agreed, I comment on the QSAR principles:

1. I agree with the OECD criteria for QSARs, I only have same reservations about their practical use.
2. 'Unambiguous algorithm' is a good idea with regard to transparency, but may not be realizable (i) with modern statistics, e.g., multivariate procedures with continuous update of databases or (ii) in case of proprietary models (with independent external validation).
3. 'Defined applicability domain' is currently often restricted to chemical domain and as such may be misleading: It may pretend confidence in a model but that is not substantiated because other (more) important aspects of similarity / dissimilarity are neglected (e.g. toxicological domain, mode of (inter)action).

## Annex 2: Tables

**Fig. 1-1:** Results of first Delphi round (n =17)

| **Key Topic 1 (KT1): Focus of OSIRIS (1.1)** | | | | |
|---|---|---|---|---|
| *1.1 What should OSIRIS do?* | | | | |
| | **Very important** | **Important** | **Partly important** | **Not important at all** |
| Contribute to the generation of databases | 6 | 5 | 5 | 1 |
| Develop integrated assessment strategies for (groups) of chemicals | 11 | 4 | 2 | 0 |
| Develop and evaluate individual ITS building blocks | 7 | 9 | 1 | 0 |
| Generate ITS procedures (IT-Tools and Guidance Documents) | 11 | 5 | 1 | 0 |
| **KT1: Output of OSIRIS (1.1)** | | | | |
| *1.1 What should OSIRIS do?* | | | | |
| | **Very important** | **Important** | **Partly important** | **Not important at all** |
| Summaries of evaluated experimental data following the OECD Guidelines for the testing of chemicals. | 5 | 6 | 4 | 2 |
| Estimates of individual fate and (eco)toxicity data including information about the uncertainty of the predictions. | 5 | 8 | 3 | 1 |
| Generation of PNEC and DNEL information. | 3 | 4 | 7 | 3 |

**KT2: Tools and instruments for testing strategies of OSIRIS - Building blocks of ITS (2.1)**

*2.1 Existing tools in OSIRIS: Which of the tools below for generating fate and (eco)toxicity information are important for the testing strategies in the REACH process and should be included in Pillar 4?*

| | Very important | Important | Partly important | Not important at all |
|---|---|---|---|---|
| Animal data generated according to accepted guidelines | 6 | 7 | 3 | 0 |
| Non-guideline animal data | 5 | 5 | 5 | 1 |
| In vitro methods generated according to validated methodologies | 7 | 6 | 2 | 1 |
| In vitro methods generated according to 'suitable' methodologies | 6 | 5 | 5 | 0 |
| Hazard data from structurally related chemicals (read-across) | 9 | 5 | 1 | 1 |
| (Q)SARs | 9 | 6 | 1 | 0 |
| Threshold of Toxicological Concern | 10 | 4 | 2 | 0 |

**KT3: Databases of Chemicals - Kind of databases (3.1)**

*3.1 Which databases should be included in the future for (eco)toxicity and exposure assessment?*

| | Fully agree | Partly agree | Partly disagree | Fully disagree |
|---|---|---|---|---|
| Exposure and exposure categories | 12 | 3 | 1 | 1 |
| Descriptions of categories of chemicals related to mode of toxic action | 14 | 3 | 0 | 0 |
| Physicochemical properties | 13 | 3 | 1 | 0 |
| Toxicity data | 14 | 3 | 0 | 0 |
| Ecotoxicity data | 13 | 4 | 0 | 0 |
| Information of modes of toxic action | 12 | 5 | 0 | 0 |
| Estimates of fate and (eco)toxicity data including an estimate of their uncertainty | 11 | 4 | 1 | 1 |

**KT3: Databases of Chemicals - International activities (3.2)**

*3.2 OSIRIS aims at being complementary to ongoing international efforts, such as the REACH Implementation Projects (RIP's) and further activities. How do you rate the importance of the RIP's, OECD QSAR toolbox and others? What other international activities are important for OSIRIS?*

| | Very impor-tant | Important | Partly important | Not import-ant at all |
|---|---|---|---|---|
| Reach imple-mentation Projects RIP's | 13 | 3 | 1 | 0 |
| OECD QSAR toolbox | 13 | 3 | 1 | 0 |

**KT3: Databases of Chemicals - Databases needed (3.3)**

*3.3 What kind of databases has priority?*

| | Very impor-tant | Important | Partly important | Not import-ant at all |
|---|---|---|---|---|
| Local toxicity (skin, eye) | 4 | 5 | 4 | 0 |
| Reproductive toxicity | 13 | 4 | 0 | 0 |
| Aquatic toxic-ity | 7 | 7 | 3 | 0 |
| Chronic toxic-ity | 11 | 5 | 1 | 0 |
| Carcinogenicity | 8 | 5 | 2 | 0 |
| Repeated dose toxicity (90-d) | 10 | 4 | 1 | 0 |

**KT 4: Quality of data - Quality parameters (Reliability, 4.1)**

*4.1 What information about the data in OSIRIS is, according to you, needed from a scientific and regulatory point of view?*

| | Very important | Important | Partly important | Not important at all |
|---|---|---|---|---|
| Availability of complete test report | 6 | 6 | 5 | 0 |
| Substantiation of deviations from guidelines | 8 | 8 | 1 | 0 |
| In case of non-guideline studies: publication in a peer-reviewed journal | 2 | 11 | 2 | 1 |
| In case of non-guideline studies: interpretation of results (alone or in combination)? | 7 | 5 | 3 | 0 |
| Performance of the study according to GLP | 1 | 7 | 9 | 0 |
| Identity, purity and source of substance | 10 | 7 | 0 | 0 |
| Exposure considerations | 4 | 10 | 3 | 0 |
| Analytical methods | 2 | 10 | 5 | 0 |
| Statistics | 1 | 11 | 5 | 0 |

| | | | | |
|---|---|---|---|---|
| Availability of information on structural analogues | 9 | 4 | 3 | 1 |

**KT 4: Quality of data - Quality parameters (Relevance, 4.1)**

*4.1 What information about the data in OSIRIS is, according to you, needed from a scientific and regulatory point of view?*

| | Very important | Important | Partly important | Not important at all |
|---|---|---|---|---|
| Animal species | 3 | 13 | 1 | 0 |
| Route of administration | 6 | 10 | 1 | 0 |
| Effect (with regard to target population) | 6 | 10 | 0 | 0 |

**KT 4: Quality of data - Quality aspect (4.2)**

*4.2 Do you believe that Klimisch (1997) is useful for this purpose? Which aspect do you consider important for a scoring system?*

| | Very important | Important | Partly important | Not important at all |
|---|---|---|---|---|
| Data uncertainty | 8 | 5 | 0 | 0 |
| Model uncertainty | 5 | 7 | 1 | 0 |
| Representativeness of results (generalisability) | 6 | 7 | 1 | 0 |
| Stochastic effects | 0 | 6 | 5 | 0 |
| Outlayers and surprises | 2 | 9 | 3 | 0 |

| | | | | |
|---|---|---|---|---|
| Effects on specific ecosystems | 2 | 4 | 4 | 1 |

**KT 4: Quality of data - Criteria (OECD principles, 4.3)**

*4.3 OSIRIS aims at basing the (Q)SAR tools on the OECD principles. Do you agree?*

| | **Fully agree** | **Partly agree** | **Partly disagree** | **Fully disagree** |
|---|---|---|---|---|
| "a defined endpoint" | 17 | 0 | 0 | 0 |
| "an unambiguous algorithm"; | 15 | 2 | 0 | 0 |
| "a defined applicability domain"; | 15 | 0 | 2 | 0 |
| "goodness-of-fit, robustness and prediction power" | 16 | 1 | 0 | 0 |
| "a mechanistic interpretation, if possible" | 14 | 3 | 0 | 0 |

**KT 5: Public availability**

*5.1 OSIRIS aims at developing a webtool which is publicly available. Do you agree? Which elements need special attention with respect to confidentiality and ownership?*

| | **Publicicly available** | **Should not be available** | | |
|---|---|---|---|---|
| Methodologies | 17 | 0 | | |
| Databases | 17 | 0 | | |
| Webtool | 16 | 1 | | |

| **KT 6: Support for industry and regulation** | | | | |
|---|---|---|---|---|
| *6.1 How can we support industry and regulators in providing effective and efficient testing methods and procedures in a timely manner?* | | | | |
| | **Very important** | **Important** | **Partly important** | **Not important at all** |
| Demo version of the OSIRIS webtool | 10 | 5 | 0 | 1 |

**Fig. 1-2 - Overview over consensus and dissent in Delphi rounds 1 – 3 (consensus (⮟) and dissent (✖))**

**Legend:** Individual Delphi (ID, round (rou.) 1) and Group Delphi (GD), round (rou.) 2 and 3, n =17 persons for individual Delphi, n = 5 groups for first Group Delphi, n = 4 groups for second Group Delphi, ⮟=consensus, ✖ = dissent, - = question not posed in this round, - (⮟) = consensus from round 2, because of that question not posed again in round 3.

| | | | **ID rou. 1** | **GD rou. 2** | **GD rou. 3** |
|---|---|---|---|---|---|
| **Key Topic 1: Focus and Output of OSIRIS** | **Focus of OSIRIS (1.1)** | *What should OSIRIS do?* | ✖ | ✖ | - |
| | | *What should be the priorities for OSIRIS?* | - | - | ⮟ |
| | | *What are the conditions for the success of OSIRIS?* | - | - | ✖ |
| | **Output of OSIRIS (1.1)** | *What should OSIRIS do?* | ✖ | ✖ | - |
| | | *What should have priority with respect to the output (products) of OSIRIS?* | - | - | ⮟ |

| | | | | | |
|---|---|---|---|---|---|
| **Key Topic 2: Building blocks of ITS** | **Building blocks of ITS (2.1) / Existing tools in OSIRIS:** | *Existing tools in OSIRIS: Which of the tools below for generating fate and (eco)toxicity information are important for the testing strategies in the REACH process and should be included in Pillar 4?* | × | × | - |
| | | *What building blocks should be developed preferentially for inclusion in ITS?* | - | - | × |
| **Key Topic 3: Databases of Chemicals** | **Kind of databases (3.1)** | *Which databases should be included in the future for (eco)toxicity and exposure assessment?* | ✓ | × | - |
| | | *Which databases will be more important in the future for (eco)toxicity and exposure assessment than today?* | - | - | ✓ |
| | **International activities (3.2)** | *How do you rate the importance of the RIP's, OECD QSAR toolbox and others? What other international activities are important for OSIRIS?* | ✓ | × | - |
| | | *Which ongoing international efforts should be considered by OSIRIS?* | - | - | ✓ |

| | | | | | |
|---|---|---|---|---|---|
| | **Databases needed (3.3)** | *What kind of databases has priority?* | ➘ | × | - |
| | | *On what databases/endpoints should OSIRIS focus in order to reduce or replace vertebrate testing?* | - | - | × |
| **Key Topic 4: Quality of data** | **Quality parameters (Reliability, 4.1):** | *What information about the data in OSIRIS is, according to you, needed from a scientific and regulatory point of view?* | × | ➘ | - (➘) |
| | **Quality parameters (Relevance, 4.1):** | *What information about the data in OSIRIS is, according to you, needed from a scientific and regulatory point of view?* | ➘ | ➘ | - (➘) |
| | **Quality aspect (4.2):** | *Which aspect do you consider important for a scoring system?* | × | × | - |
| | | *Risk assessors use scoring systems to assess the quality of the available information. Which aspect do you consider important with respect to a scoring system?* | - | - | ➘ |

| | | | | | |
|---|---|---|---|---|---|
| | **Criteria (OECD principles, 4.3):** | *OSIRIS aims at basing the (Q)SAR tools on the OECD principles. Do you agree?* | ➘ | ➘ | - (➘) |
| **Key Topic 5: Public availability** | *OSIRIS aims at developing a webtool which is publicly available. Do you agree? Which elements need special attention with respect to confidentiality and ownership?* | | ➘ | ➘ | - (➘) |
| **Key Topic 6: Support for industry and regulation** | *How can we support industry and regulators in providing effective and efficient testing methods and procedures in a timely manner?* | | ➘ | ✖ | - |
| | *How can we support industry and regulators in providing effective and efficient testing methods and procedures in a timely manner? Please be as specific as possible.* | | - | - | ➘ |
| | | | **ID rou. 1** | **GD rou. 2** | **GD rou. 3** |
| **Ratio of consensus / dissent total** | | | **7 / 5 = 1,4** | **4 / 8 = 0,5** | **10 / 3 = 3** |

**Fig. 1-3:** Results of second Delphi round (n = 5)

| **Key Topic 1: Focus and output of OSIRIS, What should OSIRIS all about?** | | | | |
|---|---|---|---|---|
| **Focus of OSIRIS** | **Very important** | **Important** | **Partly important** | **Not important at all** |
| to contribute to the generation of databases (experimental and estimated data) | 2 (?) | 3/1/ 4 | 5 | 4 |
| the development of integrated assessment strategies for (groups) of chemicals, using the different building blocks of ITS | 3/1/2/ 4/ 5 | 1 | | |
| to develop and evaluate individual ITS building blocks for physicochemical, (eco)toxicological and exposure data | 1/ 2/ 4 | 3/1/ 5 | | |
| to generate ITS procedures (IT-Tools and Guidance Documents) for integrating these building blocks into integrated strategies to estimate fate, effects and exposure information. | 1/ 2/ 4/ 5 | 3 | | |

| Output of OSIRIS | Very important | Important | Partly important | Not important at all |
|---|---|---|---|---|
| We want to have summaries of evaluated experimental data following the OECD Guidelines for the testing of chemicals | | 5 | 1/2/ 4 | 3/2 |
| We want to generate estimates of individual fate and (eco)toxicity data including information about the uncertainty of the predictions | 2/ 4/ 5 | 3/1 | | |
| We want to generate PNEC, and DNEL information | | 4 | 3/1/ 5 | 2 |
| Comment: yes (3), yes (2), yes (4) | | | | |

**Note:** Numbers indicate group positions.

| **Key Topic 2: Tools and instruments for testing strategies of OSIRIS** | | | | |
|---|---|---|---|---|
| **Existing tools in OSIRIS:** Which of the following tools for generating fate and (eco)toxicity information are important for the testing strategies in the REACH process and should be included in Pillar 4? | | | | |
| **Building blocks of ITS** | **Very important** | **Important** | **Partly important** | **Not important at all** |

| | | | | |
|---|---|---|---|---|
| Animal data generated according to accepted guidelines | 3,1/2/ 5 | | | 1 |
| Non-guideline animal data | 3,2/2 | 1 | | |
| *In vitro* methods generated according to validated methodologies | 3,4/2 | 1 | | |
| *In vitro* methods generated according to 'suitable' methodologies | 3,6/2 | 1 | | |
| Hazard data from structurally related chemicals (read-across) | 3,3/2/ 5 | 1 | | |
| (Q)SARs | 3,7/2 | 1 | | |
| Threshold of Toxicological Concern | 3,5/2 | 1 | | |
| Comment: yes (3), yes (2), yes (4), yes (5) | | | | |

**New tools:** Should new tools be developed for the REACH process? And if so which ones do you have in mind? Exposure tools for low exposure situation (3), Additions to OECD toolbox (3), TTC (3)

Focus on sensitivity of tests (3)

| **Key Topic 3: Databases of Chemicals** | | | | |
|---|---|---|---|---|
| Which databases should be included in the future for (eco)toxicity and exposure assessment? | | | | |
| **Kind of databases** | **Fully agree** | **Partly agree** | **Partly disagree** | **Fully disagree** |
| Exposure and exposure categories | 3,1/2/4/ 5 | 1 | | |
| Descriptions of categories of chemicals related to mode of toxic action | 3,2/1/2 /4/5 | | | |
| Physicochemical properties | 3,3/1/2 /4/5 | | | |
| Toxicity data | 3,4/1/2/4/5 ,1 | 1 | | |
| Ecotoxicity data | 3,5/1/2/4/5 | 1 | | |
| Information of modes of toxic action | 3,2 (?)/1/2/4/5 | | | |
| Estimates of fate and (eco)toxicity data including an estimate of their uncertainty | 3,6/2/4/5 | 1 | | |
| Comment: yes (4) | | | | |
| OSIRIS aims at being complimentary to ongoing international efforts, such as the OECD QSAR toolbox. Do you agree and which efforts are relevant to OSIRIS according to you? | | | | |
| **International activities** | **Very important** | **Important** | **Partly important** | **Not important at all** |

| | | | | |
|---|---|---|---|---|
| OECD QSAR tool-box (steering group 3) | 3/1/2/4/5 | | | |
| RIPs | 3/1/2/5 | | | |
| Fobig ECVAM data quality | | 3 | | |
| *Health Canada* | 1/ 2 | 3 | | |
| Comment:Toxcase partly important (3), EU CESAR (2), Predictonics (2), EU projects (2), EPAA (2), industry projects (2), US PMV activities (4), Testguidelines OECD (5) | | | | |
| What kind of databases has priority, considering that the OECD Toolbox is already strong in aquatic toxicity and some mammalian toxicity endpoints (e.g. Ames test and sensitization)? | | | | |

| **Databases needed** | **Very im-portant** | **Important** | **Partly important** | **Not im-portant at all** |
|---|---|---|---|---|
| Local toxicity (skin, eye) | 4 | | 3/1/2 | |
| Reproductive tox-icity | 3/1/2/4/5 | | | |
| Carcinogenicity | 4/5 | 1/2 | 3 | |
| Repeated dose toxicity (90-d) | 4/5 | 3/1/2 | | |
| Sensititation | 3/1 | 2 | | |
| Comment: Aquatic tox (BCF, 3), Aquatic tox (1,very important), chronic tox (1, very important), mutagenicity (2), yes (4) | | | | |

| **Key Topic 4: Quality of data** | | | | |
|---|---|---|---|---|
| What information about the data in OSIRIS is, according to you, needed from a scientific and regulatory point of view? | | | | |
| **Quality parameters** | **Very important** | **Important** | **Partly important** | **Not important at all** |
| **Reliability** | | | | |
| Availability of complete test report | 2 | 3/5 | 1/ 4 | |
| Substantiation of deviations from guidelines | 3/2 | 1/ 4/5 | | |
| In case of non-guideline studies: publication in a peer-reviewed journal | | 1/5 | 3/ 4 | |
| In case of non-guideline studies: interpretation of results (alone or in combination)? | 3 | 4/5 | 1 | |
| Performance of the study according to GLP | | 5 | 3/1/ 2/ 4 | |
| Identity, purity and source of substance | 3,1/ 4,1 | 1/5 | | |
| Exposure considerations | 3,1/ 4 | 1/5 | 1 | |
| Analytical methods | | 3/1/ 4/5 | | |

| | | | | |
|---|---|---|---|---|
| Statistics | | | 3/1/ 4 | |
| Availability of information on structural analogues | 3/1 | 4/5 | | |
| **Relevance** | | | | |
| Animal species | 2/ 4,1 | 3,3/1/5 | | |
| Route of administration | 2/ 4 | 3,2/1/5 | | |
| Effect (with regard to target population) | 3,1/ 2 | 1/ 4/5 | | |
| Comment: yes (3) | | | | |
| In practice, risk assessors in industry, academia and governmental organizations may not have time to read all the details about the data sources and quality and may wish to use simple scoring systems for reliability. Do you believe that the Klimisch (1997) is useful for this purpose? Which aspect you consider important with respect to a scoring system? | | | | |
| **Quality aspect** | **Very important** | **Important** | **Partly important** | **Not important at all** |
| Data uncertainty | 3,1/ 2/5 | 1/ 4 | | |
| Model uncertainty | 3,3/ 2 | 1/ 4/5 | | |
| Representativeness of results (generalisability) | 3,4/ 2/ 4 | 1 | 5 | |
| Stochastic effects | ?? 3, 2/ 4 | 1 | 1 | |
| Outlayers and surprises | 3,2/ 2/ 4/5 | 1 | 1 | |
| Effects on specific eco- | 4 | | 1/5 | 3 |

| systems | | | | |
|---|---|---|---|---|
| OSIRIS aims at basing the (Q)SAR tools on the OECD principles. Do you agree? | | | | |
| **Criteria (OECD principles)** | **Fully agree** | **Partly agree** | **Partly disagree** | **Fully disagree** |
| "a defined endpoint" | 3/1/2,1/ 4/5 | | | |
| "an unambiguous algorithm"; | 1/ 2/ 4/5 | 3 | | |
| "a defined applicability domain"; | 1/ 2,1/ 4/5 | | 3* | |
| "goodness-of-fit, robustness and prediction power"; | 3/2/5 | 1/ 4 | | |
| "a mechanistic interpretation, if possible". | 3/2/ 4/5 | 1 | | |
| Comment: yes (3) | | | | |

| **Key Topic 5: Public availability** | | |
|---|---|---|
| OSIRIS aims at developing a webtool which is publicly available. Do you agree? Which elements need special attention with respect to confidentiality and ownership? | | |
| **Information** | **Publicicly available** | **Why not?** |
| Methodologies | 3/1/ 2/ 4/5 | |
| Databases | 3/1/ 4/5 | Industrial restrictions |
| Webtool | 3/1/ 2/ 4/5 | |
| Comment: yes (2) | | |

| **Key Topic 6: Support for industry and regulation** | | | | |
|---|---|---|---|---|
| How can we support industry and regulators in providing effective and efficient testing methods and procedures in a timely manner? | | | | |
| **Effort** | **Very important** | **Important** | **Partly important** | **Not important at all** |
| Demo version of the OSIRIS webtool to get feedback and enhance implementation | 1/ 2/5 | 3/ 4 | | |
| Training | 3/ 2 | | | |
| Communication | 3 | | | |

**Fig.1-4:** Results of Third Delphi Round (n = 4)

| **Key Topic 1: Main Focus of OSIRIS: What should be the priorities for OSIRIS?** | | | | |
|---|---|---|---|---|
| **Priorities of OSIRIS** | **Rank 1** | **Rank 2** | **Rank 3** | |
| To build a set of databases (with experimental and other data) that can be used for many purposes in and outside of Osiris | | | 3/ 2/ 4/ 1 | |
| To develop and evaluate individual ITS building blocks for physicochemical, (eco)toxicological and exposure data | | 3/ 2/ 4/ 1 | | |

| **Priorities of OSIRIS** | **Rank 1** | **Rank 2** | **Rank 3** | |
|---|---|---|---|---|
| To develop integrated assessment strategies for (groups) of chemicals, using the different building blocks of ITS | 3/ 2/ 4/ 1 | | | |
| Others: yes (3), yes (2)<br>Identify existing databases and see if they are suitable for the purpose of Osiris<br>Comment: sequential ranking | | x | | |
| **What are the conditions for the success of OSIRIS?** | **Rank 1** | **Rank 2** | **Rank 3** | **Rank 4** |
| To have existing testing data on relevant endpoints (in database format) | 3/2 | | 4/1 | |
| To have existing databases operational which are relevant for IST | 3/4/1 | 1 | 2 | |
| To have harmonized templates for quality assessment of data | 1 | 3/2/4 | | |
| Others: yes (3)<br>• Here conditions related to data<br>• More success criteria such as acceptability, timeliness, easy implementability, communicability<br>• Integrate item 1 and 2 (very high rank) | | | | |

| **What should have priority with respect to the output (products) of Osiris?** | **Rank 1** | **Rank 2** | **Rank 3** | **Rank 4** |
|---|---|---|---|---|
| Operational ITSs for all endpoints, using a weight-of-evidence approach. | 3/ 2/ 1 | 4 | | |
| Operational overall ITS, using a weight-ofevidence approach. | 3/2/4 | 1 | | |
| Summaries of assessment of adequacy of available information based on OECD guidelines or non-guidelines studies. | | | 3 (?)/ 1 | 4 |
| Summaries of assessment of adequacy of available non-testing information including uncertainty of the predictions | | 1 | 3 (?)/ 4 | |

Comment: Input to the work or output? (input data. Needs to be reliable and adequate; Output was meant. Distinguish between "summaries" on tools and on chemicals (related to guidance documents). Important: Transparency about the selection of tools and the adequacy of information about chemicals.

| **Key Topic 2: Tools and instruments for testing strategies of OSIRIS** | |
|---|---|
| **Existing tools in OSIRIS:** What building blocks should be developed preferentially for inclusion in ITS? | |
| **Building blocks of ITS** | **1 = hightest priority**<br>**9 = lowest priority** |
| Animal data generated according to accepted guidelines | 3,1/ 2,1/ 4,9/ 1,7: difference: a) it is needed in general b) it is not the focus of Osiris c) generating primary data is not purpose of Osiris |
| Non-guideline animal data | 3,2/ 2,2/ 4,7/ 1,5 |
| *In vitro* methods generated according to validated methodologies | 3,4/ 2,7/ 4,6/ 1,4 |
| *In vitro* methods generated according to 'suitable' methodologies | 3,4/ 2,8/ 4,5/ 1,3 depending on endpoints, if suitable, it is very important |
| Hazard data from structurally related chemicals (read-across) | 3,3/ 2,3/ 4,1/ 1,1: if high quality data is there then high priority for all groups |
| (Q)SARs | 3,6/ 2,6/ 4,2/ 1,1: again contingent on high quality data |
| Threshold of Toxicological Concern | 3,7/ 2,5/ 4,4/ 1,2: contingent on high quality data and perspective on Osiris contribution |
| Human data | 3,8/ 2,9/ 4,8/ 1,6: low in the context of Osiris; yet generally, of course, very important |

| | |
|---|---|
| Tool to estimate low level exposure | 3,5/ 2,4/ 4,3/ 1,2: contingent on the ability to develop acceptable concepts for low exposure: if so, then important |
| Comment: yes (4) on the condition of having high quality data, dissent is disappearing<br>Two dimensions: (1) high priority in general (2) priority for specific Osiris contribution<br>Inclusion of a "environmental threshold of no concern" level | |

**Note**: First number indicates group, second number indicates priority, for example 3,1 means that group number 3 gave first priority to the ITS building block.

Two questions:

1) How much emphasis should Osiris place on getting high quality data?

- High emphasis on best data because it adds an increase in precision
- Quality data is not necessarily connected with known databases
- However, if that data is not available this should not a reason for abandon the respective activity

2) How should Osiris deal with knowledge gaps?

- REACH explicitly asks to include all relevant information even if they are of lower quality
- One of the objectives is to develop methods to process lower quality data (need to characterize uncertainty and variability)
- Degree of accuracy of data needed depends on purpose, context and application (for example labelling)

**New tools:** Should new tools be developed for the REACH process? And if so which ones do you have in mind?

*Tools:* Exposure assessment tools including exposure scenarios with updated default values; TTC for non-food chemicals and for non-oral routes. (2)

*Comments:* No need for new building blocks. Weight of evidence approach/ decision theory should be further developed. Guidance is rather needed than mandatory decision theory. Intelligent databases/ knowledge bases (for example: hyperlinks, data mining, intelligent routing, relational datasets, etc.). (4)

| **Key Topic 3: Databases of Chemicals** | |
|---|---|
| Which databases will be more important in the future for (eco)toxicity and exposure assessment than today? | |
| **Kind of databases** | **1 = highest priority**<br>**7 = lowest priority** |
| Exposure and exposure categories | 3,2/ 2,1/ 4,1 |
| Descriptions of categories of chemicals related to mode of toxic action | 3,1/ 2,3/ 4,2/ 1,2 |
| Physicochemical properties | 3,5/ 2,4/ 4,3 |
| Toxicity data | 3,6/ 2,6/ 4,6 |
| Ecotoxicity data | 3,7/ 2,6/ 4,5 |
| Information of modes of toxic action (e.g. chemical reactivity) | 3,3/ 2,2/ 1,2 |

| | |
|---|---|
| Estimates of fate and (eco)toxicity data including an estimate of their uncertainty | 3,4/ 2,6/ 4,4/ 1,1 |
| Comment: yes (4), yes (1). Comment: Uncertainty and Variability are very important, last item has two different targets (ecotoxicity) and uncertainty<br>Comment: if we have uncertainty characterisation there is no priority | |
| Which ongoing international efforts should be considered by OSIRIS? | |
| **International activities** | **1 = highest priority**<br>**N = lowest priority** |
| **Tools** | |
| OECD QSAR toolbox | 3,1/ 2,2/ 4,3 |
| RIP's | 3,2/ 2,1/ 4,1/ 1,1 |
| Test guidelines OECD | 3,6/ 2,4/ 4,7 |
| EU CAESAR | 3,4/ 2,3/ 4,5 |
| Health Canada | 3,3/ 2,5/ 4,4 |
| Fobig/ECVAM data quality | 3,5/ 2,7/ 4,6 |
| *US-PMN-Activities*<br>*It is wise for OSIRIS to use the existing data bases and this one is worth considering (Group 4): Response; this has been done already (1992); interesting to revisit this database after OSIRIS is completed; also check how much "real" data is available* | 3,7/ 2,6/ 4,3 |
| **Data (distinction is not quite consistent)** | |
| Predictomics: it lines up with the EU-FP: should mine them | Low (3) 2,3/ 4,2 |

| | |
|---|---|
| EU-FP: reprotect, Acute-tox, Sensitiv….. | Low (3) 2,2/ 4,2 |
| EPAA | Low (3) 2,4 |
| Others: SIDS Data (2,1), yes (1) | |

Should OSIRIS incorporate all existing tools and databases or should OSIRIS ensure simple communication with them (=full compatibility)?

3. Ensure simple communication, no full incorporation

1. same opinion

2. same opinion

4. same opinion (based more on feasibility, and cost-efficiency not desirability)

| On what databases/endpoints should OSIRIS focus in order to reduce or replace vertebrate testing? | |
|---|---|
| **Databases needed** | **1 = highest priority**<br>**7 = lowest priority** |
| Local toxicity (skin, eye) | 3,7/ 2,8/ 4,7 |
| Reproductive toxicity | 3,1/ 2,1/ 4,1/ 1,1 |
| Carcinogenicity | 3,5/ 2,5/ 4,2/ 1,5 |
| Repeated dose toxicity | 3,6/ 2,4/ 4,3/ 1,5 |
| *Sensitisation* | 3,3/ 2,2/ 4,6/ 1,3 |
| *Mutagenicity* | 3,2/ 2,6/ 4,4/ 1,2 |
| *Aquatic tox (BCF) longterm (3)* | 3,4/ 2,7/ 4,5/ 1,5 |
| *Others: Chronic aquatic toxicity (2,3)Carcinogenicity: related on number of chemicals = low priority; related animals per chemical higher priority, if multiplied – medium to low priority* | |

| **Key Topic 4: Quality of data** | | | |
|---|---|---|---|
| Risk assessors use scoring systems to assess the quality of the available information. Which aspect do you consider important with respect to a scoring system? | | | |
| **Quality aspect** | **Should be included** | **Should not be included** | **Relevant, but not integratable** |
| **Single studies** | | | |
| Data uncertainty | 3/ 2/ 4/ 1 (?) | | |
| Model uncertainty | 3/ 2/ 4/ 1 (?) | | |
| Distinction between model uncertainty for non-testing data and data uncertainty for test data (Interspecies extrapolation not relevant in this context) | | | |
| **Data sets** | | | |
| Representativeness of results (generalisability) | 2/ 4 | | 1 |
| Stochastic effects<br>a) correlation versus causation<br>b) is covered by outlayers (percentiles)<br>c) emphasis on robust systems that are resilient against outlayers | | 2/ 1 | (4) |

| | | | |
|---|---|---|---|
| Outlayers and surprises | 4/ 1 | 2 | (4) |
| Comment: yes (4) Different meanings of generalisability: across chemical domains, space, populations, ecosystems, machine learning capability<br><br>• Outlayers and surprises cannot be scored yet it may detect other endpoints<br>• Oulayers and surprises depend on the limitations of the research framing<br>• Leverage points (these are crucial points that exert influence in a regression): dominate the slope<br><br>Identify surprises: (intelligent) speculation, trial and error, | | | |
| What kind of information on uncertainty needs to be communicated to the regulators?<br><br>Ask the regulators (1) | | | |

| **Key Topic 6: Support for industry and regulation** | | | | |
|---|---|---|---|---|
| How can we support industry and regulators in providing effective and efficient testing methods and procedures in a timely manner? Please be as specific as possible. | | | | |
| **Effort** | **Very important** | **Important** | **Partly important** | **Not important at all** |
| Demo version of the OSIRIS webtool to get feedback and enhance | 3/ 2 | 4/ 1 | | |

| | | | | |
|---|---|---|---|---|
| implementation | | | | |
| Training | 3/ 2/ 4/ 1 | | | |
| Guidance documents | 3/ 2 | 4/ 1 | | |
| Case examples | 3 | 2/ 4/ 1 | | |

Comment: yes (1):

4: Two-way communication rather than just training; case examples could be part of the demo

1: Beta version should be out to get feedback and acceptance; to be in accordance with REACH time lines (aggressive time line: so a real challenge; needs to have it ready in 2010-2011)

How should the results of OSIRIS be evaluated after the project is completed?

Criteria for success will be if the tool will be used by industry and regulators! (3)

- Have a test case early on: Start early as possible with one. (1)

What is the goal: how many replacements? How many accepted? How many used?

- Usual evaluation processes are included (publication, peer review).
- Consistent and reproducible outcome.
- International harmonisation on a global scale (starting with the EU).
- Three goals: adequacy, acceptance and global implementation.

BE REALISTIC!! AND HELP US TO MAKE THE GOALS BECOME TRUE!

**Annex 3: Agenda**

**Wednesday, 28th of November 2007**

| | |
|---|---|
| 13.00 | **On-site Registration** |
| 14.00 | **Welcome and Introduction**<br>*Dr. J.J. M. (Han) van de Sandt, TNO Quality of Life, Zeist, the Netherlands*<br>*Prof. Dr. Ortwin Renn, DIALOGIK & University of Stuttgart, Germany* |
| 14.15 | **OSIRIS Pillar 4: Envisioned products and procedure**<br>*State of the art; focus and output of OSIRIS*<br>*Dr. Han van de Sandt, TNO Quality of Life, Zeist, the Netherlands*<br>*Questionnaire topics 2-6*<br>*Dr. Dinant Kroese, TNO Quality of Life, the Netherlands*<br>*Results of the questionnaire*<br>*Prof. Dr. C. J. (Kees) van Leeuwen, TNO Quality of Life, the Netherlands* |
| 15.15 | **Delphi method: aim and procedure**<br>*Introduction to method*<br>*Prof. Dr. Ortwin Renn, DIALOGIK & University of Stuttgart, Germany* |
| 15.30 | **First Round: Group Experts Delphi: break out in smaller groups** |
| 17.00 | **Coffee break** |

| | |
|---|---|
| 17.30 | **Plenary discussion: Justification of Group Results** |
| 19.00 | **Adjourn, invitation for a joined dinner** |

**Thursday, 29th of November 2007**

| | |
|---|---|
| 9.15 | **Feedback from day 1**<br>**Second Round: Group Experts Delphi: break out in smaller groups** |
| 10.45 | **Coffee break** |
| 11.15 | **Plenary Discussion: Justification of Group Results** |
| 12.30 | **Lunch** |
| 13.30 | **Third Round: Group Experts Delphi: break out in smaller groups** |
| 14.30 | **Coffee break** |
| 15.00 | **Plenary Discussion: Justification of Group Results** |
| 15.30 | **Concluding Session: General Feedback** |
| 16.00 | **End of the workshop** |

# 2 Second Stakeholder Workshop – Opponents and World Café

*Christina Benighaus, Dr. Michael Ruddat, Prof. Dr. Dr. h.c. Ortwin Renn*

## 2.1 Summary

The 2nd Stakeholder Workshop in November 2008 in Brussels gave the participants an overview on the preliminary results of OSIRIS and the contributions to the hazard assessment and the risk assessment process. Experts from industry, academia and government presented the available testing methods and explained how they can be used in REACH. Invited critical commentators and the audience discussed the pros and cons of the approaches taken by OISRIS to fit the REACH testing requirements and took stock of the merits and problems of ITS approaches.

Integrated Testing Strategies (ITS) will give the opportunity to accelerate the use of non-testing information for regulatory decisions making of chemicals without reducing the required level of safety. OSIRIS will develop approaches for the many-to-one replacements of animal tests, i.e. Integrated Testing Strategies (ITS). But using ITS will not entail one-by-one replacements, but several different approaches will be combined and integrated. A systemic combination of the testing strategies like in vitro testing, QSAR, read-across or TTC will help to develop innovative non-animal

approaches. One major requirement is to use contextual information with category data, read across and Mode of Action information. It was suggested to combine endpoints with specific tests, for example RDT and in vivo Mutagenicity.

An important limiting factor in implementation of the ITS will be the level of uncertainty that one is willing to accept when applying the modified testing strategies. The new testing strategies demand a new concept dealing with uncertainty. The open question remains of how much uncertainty one is willing to accept. The acceptance of uncertainty needs an integrated concept that links:

- risk assessment with risk perception and socio-cultural processing of risk
- physical risk analysis with financial, economic and social risk
- risk theory with organizational capacity building and management competency.

OSIRIS should take into account the extent of coverage of "chemical space" rather than considering chemicals individually. Looking at chemical spaces can facilitate the selection of testing priorities as a basis to advance testing methodologies. Exposure considerations and in particular use categories are also influential factors for ranking chemicals.

The participants recommended to define QA/QC (quality criteria) for new testing methods, old none GLP-data etc., to decide which data are available and which data can be used for which purpose. The criterion of data "quality" in the context of development of databases should inform predictive tool development. The weight of evidence should be assimilated across broader data sources, taking into account factors such as consistency, specificity, biological plausibility, etc.

OSIRIS efforts should focus on 'in time delivery' of its tools and approaches targeted at substances with a registration deadline of December 2013. In order to be considered as an information

source for the 2013 substances OSIRIS should be functional already by end 2010. Deliverance end 2010 does allow the consortia to assess applicability & remaining uncertainty of the non-test information, and either take the decision whether to accept the non-test information as is, or start generation of Annex VII & VIII information prior to registration.

Data from industry is needed both to develop the ITS and for meeting the regulatory requirements of REACH. Sharing data will be one of the important factors in reducing animal test and costs.

The framework has to be easily accessible and user friendly. That means among others that the input information should be clearly captured, that the algorithms used are transparent, that the results are reproducible, and the outputs formatted in such a way that they are ready for use in REACH. Finally, OSIRIS should develop a vision how to deal with end-users when questions arise, or when bugs are discovered, and also how it will ensure sustained development, support and maintenance for the tool.

OSIRIS should consider the political dimension with respect to the acceptance of ITS by, for example, ECHA, Member States or the EC. A new integrated concept interrelated with the different levels of uncertainty have to be accepted not only from the user side (mainly industry) but also from European (ECHA, Member States, ECVAM) and Non-European authorities.

A business plan is needed to deal with many of the above challenges. The development and assessment process of ITS needs time. In order to gain confidence and continuous feedback for alternative testing methods from stakeholders, an open and transparent process is absolutely essential.

## 2.2 Concept and methods of the workshop

**Target group of second workshop**

The second workshop was designed to represent a broad audience (minimum 50 up to 100 participants), to disseminate results as well as to collect feedback from the participants. As participants DIALOGIK invited representatives from:

- European Chemicals Agency (ECHA), European Chemicals Bureau (ECB), National Competent Authorities
- EU industry (individual companies and sector groups),
- NGOs (environmental groups, public health groups, consumer groups)
- Experts from universities and research institutes
- Key internal and external OSIRIS partners of the consortium members
- Advisory board members
- Experts from related activities worldwide (OECD, US-EPA, Health Canada).

**Objectives and strategies of the workshop**

The overall objectives of the second workshop were:

- to communicate and disseminate the preliminary results of the first one and a half years of research to key stakeholders
- to discuss issues of handling application, uncertainty and limitations of the ITS.

Secondary goals of the second workshop were:

- to ensure early input of OSIRIS results into the ongoing REACH process

- to initiate a dialogue between the project members and the EU-stakeholders
- to increase the acceptance of the proposed models, non-testing methods, web-based tools and ITS.

Subjects of the second workshop and working questions

As explained above, the workshop addressed the topics of human and environmental toxicology and the exposure of the biological domain. It included the framework of the OSIRIS project and envisioned application in the REACH process. The subtopics of the second stakeholder workshop covered:

- approaches of integrated testing strategies and their potential for REACH, such in vitro testing, QSAR's, TTCs and read across,
- REACH requirements and dealing with uncertainty,
- benefits of OSIRIS for industry, NGOS and regulators in the European Union, and
- replacement, refinement and reduction of animal testing

The working groups were asked to deal with the following three questions:

- Under which conditions are the proposed integrated testing strategies operational for being used under REACH?
- Is the pool of existing information sufficient to conduct integrated testing strategies and if not, which additional information needs to be accumulated?
- Can you reduce the amount of testing, especially animal testing (reducing costs and time) without sacrificing accuracy, validity and reliability of the results?

### 2.2.1 Key Methods of the workshop

The workshop was divided in three main parts, an introductory part to explain the basic approach of the team to risk assessment and ITS, a lecturing part with presenters and opponents in which the audience received detailed information about the project and ITS. This information was the main input to a general discussion and a question-and-answer period using the world café or carousel method as a means to facilitate the exchange of arguments, comments and ideas.

**Welcome and introduction in the workshop and OSIRIS**

In the first part of the project the coordinator of the project, Prof. Dr. Gerrit Schüürmann and the coordinator of the EU-research programme, Dr. Georges Deschamps, introduced the OSIRIS project and explained the risk assessment process in the framework of REACH. A representative of the European Chemicals Agency, Evelin Fabjan, listed the requirements of REACH as a reminder for the discussion to follow.

**Presenters and opponents: results and critical comments**

The second part of the workshop contained the lectures of the presenter and the opponent with respect to each major topic followed by a plenary discussion. The Pillar leaders or his/her representative (Mark Cronin, Dr. Dinant Kroese and Dr. Theo Vermeire) presented the preliminary results of the consortium after one and a half year research. The main topics of the agenda were:

- "Human and environmental toxicity and exposure: results and critical points of the OSIRIS framework" and
- "Integration of the components in the OSIRIS framework and use in the REACH process: results and critical points".

After each topical presentation, an opponent (Bette Meek and Dr. Watze de Wolf) pointed out critical issues and posed open questions. Both opponents focused on a handful of critical points, to which the presenters responded. The opponents were invited by the OSIRIS team in advance to stimulate the discussion.

### 2.2.2 World Café

DIALOGIK selected a special communication method called the carousel technique. This technique is a modification of the World Cafe Method[1] and has been proven very effective in similar situations. It is well suited for involving large groups with more than 20 people. It can be easily practiced and is flexible with respect to varying group compositions. It can be applied to solution-oriented as well as evaluation-oriented topics.

**World Café Ambiance**

For informal and personal working atmosphere it is essential to create an environment that evokes the informal feeling of a café house. Therefore DIALOGIK tried to make the workshop rooms look like a Café Ambiance, with small tables designed to host four or five people. Less than four people at a table may not provide enough diversity of perspectives, more than five limits the amount of personal interaction.

The Café tables were arranged in a staggered, random fashion rather than in neat rows. They looked like tables in a sidewalk café after it has been opened for a few hours. DIALOGIK placed

[1] "Café Conversations are an easy-to-use method for creating a living network of collaborative dialogue around questions that matter in service of the real work.[1"] For a detailed description of the method, please have a look at the webpage of "the World Café" (www.theworldcafe.com)."

at least two large sheets of paper over each table cloth along with a mug filled with markers. Paper and pens encouraged scribbling, drawing, and connecting ideas. To honour the tradition of community and hospitality associated with a Cafe, DIALOGIK provided beverages and snacks, because a Café is not complete without food and refreshments.[2]

**Host and travellers**

DIALOGIK invited four or five participants each to gather at the small Café-style tables and let them discuss three rounds of approximately 20-30 minutes. They worked on the three working questions mentioned above. DIALOGIK encouraged both table hosts and members to write key ideas on their tablecloths or to note key ideas on large index cards in the centre of the group. After the initial round of conversation, DIALOGIK asked one person to remain at the table as the "host" while the others served as travellers. The travellers carried key ideas, themes and questions to their new conversations tables.

DIALOGIK asked the table host to welcome the new guests and briefly share the main ideas, themes and questions of the initial conversation. DIALOGIK encouraged guests to link and connect ideas coming from their previous table conversations - listening carefully and building on each other's contributions.

By providing opportunities for the participants to move from one table to the next, they were able to link ideas, questions, and themes. At the end of the second round, all tables in the room were cross-pollinated with insights from prior conversations. In the third round people returned to their home (original) tables to synthesize their previous discoveries or they continued travelling

[2] In according 2008 The World Café. Free to copy and distribute with acknowledgement & a link to: http://www.theworldcafe.com

to new tables, leaving the same host at the table.[3]" After three rounds of conversation, we asked the hosts of each table to share their impressions of the three rounds and draft some conclusions about the results for each of the three questions- These conclusions were summarized on a flip-chart. One host per question presented then the shared insights to the audience at the end of the meeting. The audience were invited to comment on the results. However since all participants have been exposed to almost all conversations during the carousel methods, only a few amendments were made.

## 2.3 Presentations[4]

### 2.3.1 Introduction to OSIRIS, risk assessment in REACH

The OSIRIS Co-ordinator **Gerrit Schüürmann (UFZ, Germany)** opened the meeting by providing an outline of OSIRIS and explaining the context of the project in terms of the 3Rs-concept of Russell and Burch, i.e. Replacement, Reduction and Refinement of animal testing and the need to develop approaches for the many-to-one replacement of animal tests, i.e. Integrated Testing Strategies (ITS). Using ITS means therefore that there is no one-by-one

[3] In according 2008 The World Café. Free to copy and distribute with acknowledgement & a link to: http://www.theworldcafe.com

[4] Parts of this chapter are written by Andrew Worth who summarized the presentations of the workshop.

replacement, but several different approaches are combined and considered instead.

Then **Georges Deschamps (European Commission, DG Science and Research)**, emphasised the global dimension of the risk assessment and the need for international dialogue and communication. It was noted that the engagement of stakeholders is an important and integral part of OSIRIS.

Evelin Fabjan (European Chemicals Agency, ECHA, Finland) gave an overview of the information requirements under REACH, on the basic principles of ITSs and emphasised the safety aspects and the importance of ITSs in REACH. She also indicated a number of outstanding needs for scientific development, including the need to:

- integrate different methods/information
- study the applicability of the Threshold of Toxicological Concern (TTC) concept
- gain more experience of quantitative read-across for human health endpoints and
- have a readily accessible and reliable source of information on QSAR validity

### 2.3.2 REACH: alternative testing - a new practical approach

*By Evelin Fabjan*
*(European Chemicals Agency, ECHA, Finland)*

The REACH Regulation entered into force on 1st June 2007 with an aim to streamline and improve the former legislative framework on chemicals of the European Union (EU). It places greater

responsibility on industry to manage the risks that chemicals may pose to the health and the environment. It requires manufacturers and importers of chemical substances (≥1 tonne/year) to obtain information on the physicochemical, health and environmental properties of their substances and to use this information to determine and document how these substances can be used safely.

In order to achieve a high level of protection of human health and the environment while limiting the need for additional testing, all available data on the intrinsic properties of a substance, including testing data (in vivo, in vitro) as well as non-testing data (obtained with (Q)SAR models, grouping of substances, weight of evidence etc.) must be evaluated first. Annexes VI to X of the REACH Regulation specify the minimum data requirements for registration purposes according to the tonnage. The standard information set may be adapted according to the specific rules in column 2 of the above-mentioned Annexes and general rules described in Annex XI of the REACH Regulation (e.g. in cases where testing is not technically possible, or testing does not appear scientifically necessary, or based on exposure considerations). Where available data are not adequate to meet the requirements of the REACH Regulation, additional testing may be needed.

Whereas the legislation provides the legal framework that registrants need to follow when deciding if, when and what type of information needs to be submitted, to facilitate this, extensive guidance on integrated testing strategies was developed in close collaboration with experts from Member States, industry and NGO's[5].

The presentation briefly outlined the information requirements under the REACH Regulation, the elements of integrated testing strategies (ITSs), their current applicability for human health end-

[5] Guidance on information requirements & Chemical Safety Assessment.

http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_en.htm?time=1233748148

points (based on the guidance documents), and summarised the main areas where further development is needed

### 2.3.3 Human and environmental toxicity and exposure

**Human and environmental toxicity and exposure: results and critical points of the OSIRIS framework**

*Presenter: Prof. Dr. Mark Cronin, School of Pharmacy and Chemistry, Liverpool John Moores University, UK*

*Second Presenter: Dr. Theo Vermeire, National Institute of Public Health and the Environment - RIVM, Bilthoven, The Netherlands*

*Opponent: Dr. Bette Meek, Chemical Risk Assessment, McLaughlin Centre for Population Health Risk Assessment, University of Ottawa, Canada*

**Prof. Dr. Mark Cronin (Liverpool John Moores University, LJMU)** gave an overview of progress made in pillar 2 (biological domain), e.g. the collection and structuring of toxicological databases, evaluation of data quality, the application of mode and mechanism of action information in ITS, formation of categories for read-across and the optimisation of proposals for in vivo testing. He emphasised the importance of, and difficulty, in establishing the quality and adequacy of the test and non-test toxicological data. It was noted that adequacy is highly context and policy-dependent. He also described ongoing work aimed at developing a better understanding of the role of mechanistic information in ITS.

**Dr. Theo Vermeire (RIVM, NL)** gave an overview of progress made in pillar 3 (exposure-informed testing), including both exposure-based waiving (EBW) and exposure-based testing (EBT). He indicated that according to the legal text of REACH, the possibilities for EBW are quite limited and the burden of proof is very high. EBW should be justified by a thorough exposure assessment. However, he described that there are opportunities to explore the possible application of the TTC concept, as well as the Environmental Threshold of No Concern (ETNC) concept. He referred to both of these as instances of a more generic No Further Action Level (NFAL). He also described ongoing work aimed at the development of probabilistic modelling approaches for assessing the relationship between exposure levels and NFALs.

## 2.3.4 Exposure informed testing under REACH

*By Dr. Theo Vermeire[1], Dr. Marja van de Bovenkamp[1], Dr. Hans Marquart[2]*

[1] National Institute of Public Health and the Environment (RIVM), Bilthoven, The Netherlands

[2] TNO Quality of Life, Zeist, The Netherlands,
Email contact: theo.vermeire@rivm.nl

**Introduction**

Within the REACH framework, but also within OECD, there is understanding that for reasons of animal welfare, costs and logistics, it is important to limit the number of tests to be conducted. Integrated Testing Strategies (ITS) will make it possible to increase the use of non-testing information for regulatory decision making of chemicals, and to effectively reduce animal testing without increasing the overall uncertainty. Exposure is one of the

decision elements in ITS. Testing can be waived triggered on the basis of exposure considerations. This presentation aims to describe criteria for exposure informed testing as foreseen in the REACH regulation and to give more detail to the REACH requirements for exposure-based waiving. General guidance for Exposure Based Waiving is given in the REACH TGD Chapter R.5 (Adaptations on information requirements). Besides, this presentation is further based on research done within the EU Sixth Framework project OSIRIS (Optimized Strategies for Risk Assessment of Industrial Chemicals through Integration of Non-Test and Test Information).

Exposure informed testing includes both Exposure Based Waiving (EBW) and Exposure Based Triggering (EBT). The principle behind any EBW is that there are situations when human or environmental exposures are so low that there is a very low probability that the acquisition of additional effect information may lead to an improvement in the ability to manage risk. In contrast, EBT refers to situations where human or environmental exposures are considered high enough to justify testing above the regulatory requirements.

In the Annexes VII-X of REACH, specific rules are presented when standard toxicity testing, as specified in Annex VI, may be omitted, triggered, replaced or adapted. No possibilities for EBW exist below a tonnage of 10 tonnes per annum. Therefore, so-called 'column 2' adaptations for EBW/EBT only come into play from Annex VIII. In addition, Annex XI, section 3, presents the possibility of the waiving of certain toxicity studies in Annex VIII (repeated dose toxicity, sub-chronic toxicity, reproductive toxicity) and the tests in Annexes IX and X based on 'the exposure scenario(s) developed in the Chemical Safety Report ('substance-tailored exposure driven testing').

This presentation will discuss the criteria for the justification for EBW and EBT, including (eco)toxicological reference values. Examples will be given for both human and environmental exposure

assessment. The consequences for exposure assessment methodology will also be presented.

**Results and discussion**

EBW and EBT can best be considered within the context of risk-based decision making. Extensive and detailed knowledge of exposure throughout the life cycle for human and environmental exposure is essential for exposure informed testing. Human exposure includes occupational exposure, consumer exposure and human exposure via the environment. For humans, both external and internal exposure should be considered. All stages in the life-cycle of a chemical should be taken into account for a valid justification of waiving: production, formulation, industrial or professional or private use, service life and disposal.

The justification for EBW/EBT can be based on either a qualitative argumentation or a quantitative argumentation. Qualitative justification for EBW could be based on specific use or limited emissions, on specific operational use or use conditions and on substance properties. Examples are:

- Substances reacting away or binding covalently to a matrix
- Use in strictly controlled, closed systems with extensive personal protective equipment (PPE)
- Infrequent use
- Substances with low volatility, fugacity.
- Absorption is unlikely.

If absence of exposure cannot be argumented in a qualitative sense, a quantitative exposure assessment and risk characterization based on hazard and exposure may be needed, considering the exposure scenario developed in the Chemical Safety Report. Quantitative justification for EBW needs a assessment that exposure is below a 'no further action level' such as PNECs (Predicted No-Effect Concentrations), DNELs (Derived No-Effect Levels),

DMELs (Derived Minimal-Effect Levels) or TTCs (Thresholds of Toxicological Concern). The 'no further action level' should be applicable even when little toxicological information is available for a substance and exposure via different routes and in different compartments should be taken into account. TTCs will be discussed separately in this symposium. The kinetics of the compound (especially bioavailability) can refine the exposure estimate to justify EBW.

Quantitative justification will further be based on exposure scenarios. An exposure scenario describes what a substance is used for, how it is used and under which operational conditions, and what risk management measures are taken to control the exposure of man and the environment. The REACH Guidance details how an exposure scenario is built and how it is used for the exposure assessment. The quantitative exposure estimate, obtained either by modelling or by measuring, and relevant to the test that is to be waived, will be compared to the 'no further action level'. EBT requires the outcome of a Chemical Safety Assessment showing risk levels that indicate the need for further research based on testing strategies such as in the REACH Guidance.

Both the exposure estimate and the 'no further action level' are uncertain because of uncertainties and variability in scenarios, models, and parameters, leading theoretically to a distribution of risk characterization ratios (RCRs) like PEC/PNEC, Estimated Intake/DNEL, PEC/TTC, Estimated Intake/TTC. Therefore the real question is what the probability is that the estimated RCR is exceeding the trigger value of 1 and what probability of exceeding is acceptable to warrant the conclusion that EBW is justified. For instance, if the distribution is such that only the far right end of the exposure distribution is exceeding the trigger value, EBW may be acceptable. Also, a tier 1 realistic worst case assessment can be performed the result of which can be considered to be equivalent to a 'far right end' estimate. If a significant part of the distribution

exceeds the trigger value, EBW should be declined. Distributions far above one would trigger testing (EBT).

The 'no further action level' can be very low, below levels for which methods have been developed and validity can be assumed to be reasonable. Therefore, it needs to be determined whether available methods and models can make a valid estimate of (very) low exposure, while incorporating the relevant parameters of the exposure scenarios with sufficient sensitivity. For measured and modeled data this means that the exposure situation used to derive the exposure estimate should be comparable to the situation under study with respect to potential determinants of exposure. For modeling, additional criteria are that the model estimates exposure accurately given the exposure situation and that the model parameters can be estimated accurately. A selection of available models will be discussed in the light of these requirements: EUSES for the environment, Stoffenmanager, RISKOFDERM and ECETOC TRA for workers and CONSEXPO for consumers.

**Conclusions**

In the justification for EBW a number of conditions should be met. First, it should be determined whether current exposure models and measurement data are suitable to accurately estimate exposure in the lower exposure range. When valid exposure estimates or measurements are obtained, they should be compared to a relevant toxicological threshold to determine whether exposure is below the 'no further action level'. Although some thresholds are available it is as yet unclear to what extent they meet the criteria stated above. This needs to be evaluated. In addition, it needs to be determined whether it is valid to assume that exposure to substances in REACH at levels below the given thresholds do not pose any risk. Further evaluation of the identified exposure scenarios that may give reason to EBW and EBT, using the model outcomes and measurements and the available toxicological

thresholds, should give insight in the necessary improvements and criteria to make the EBW and EBT concept feasible.

**References**

[1] Vermeire TG, Bakker J, Bessems JGM, Van de Bovenkamp M, Dang Z, Van Engelen JGM, Gunnarsdottir S, Hagens WI, Links I, Marquart H, Mikkers J, Van Zijverden M. 2008. Exposure informed testing under REACH. RIVM report 601017001/2008, TNO-report V 8125. RIVM, Bilthoven and TNO, Zeist, The Netherlands.

Acknowledgement: The authors thank the sponsors of this project: the European Commission, the Dutch Ministry of Housing, Spatial Planning and the Environment, the Dutch Ministry of Social Affairs and Employment, and RIVM. The authors also thank all contributing coleagues of TNO and RIVM.

## 2.3.5 Integrated testing strategies in the OSIRIS framework and use in the REACH process

**Integration of the components in the OSIRIS framework and use in the REACH process: results and critical points**

*Presenter: Dr. Dinant Kroese, Chemical Safety, TNO Quality of Life, Zeist,The Netherlands*

*Opponent: Dr. Watze de Wolf, Environmental Sciences Europe, DuPont, Belgium*

**Dr. Dinant Kroese (TNO Quality of Life, The Netherlands)** gave an overview of progress made in pillar 4 on integration of test and non-testing information: e.g. how to add (Q)SAR data, and *in vitro* data. He showed the need to develop a formal weight of evidence (WoE) framework for evaluating and documenting the integration of these different types of information that may be asked for in an endpoint ITS, and illustrated this for human health, but indicating that the same concept holds as well for environmental health. This should be happened in a transparent and objective manner to quantify uncertainties and resolve conflicting values.

He also described ongoing investigations of the applicability of decision analysis (DA) and cost-effectiveness analysis (CEA) in the design and analysis of ITS. Ideally, in case endpoint-specific information is not yet considered sufficient (by the WoE approach), one should upfront be able to choose the optimal way – in terms of duration, cost, animal usage etc - of achieving the situation of sufficient information.

Finally, he presented the OSIRIS webtool which is an important and challenging development in the project: this tool is to integrate the ITS, WoE and to take account of DA and CEA considerations. It should advise the user on the adequacy of information within ITS, and on whether provided information is sufficient or not. Though not fully crystallised yet, the idea is that this webtool should have access to publicly available databases, and be able to consider and import information from various sources, including those provided by the end user in a confidential way.

## 2.4 Results of the workshop

### 2.4.1 Critical Comments of the opponents

**Bette Meek (University of Ottawa, Canada),** a member of the OSIRIS advisory board, offered some insights and suggestions based on her substantial experience of the development of priority setting methods for the Canadian Domestic Substances List (DSL). According to the Canadian experience, exposure considerations and in particular use categories had been very influential in ranking chemicals according to their concern. She also emphasised the importance of obtaining information on early effects and modes of action in the risk assessment process, and referred to a conceptual framework developed by the International Life Sciences Institute (ILSI). Dr. Meek also noted the importance of characterising the chemical space of regulatory inventories (especially the REACH inventory) and comparing this with the applicability domains of potentially useful QSAR models.

**Watze de Wolf (ECETOC, Belgium)** suggested a number of success criteria for judging the successful uptake of ITS, including the need to gain acceptance by all parties involved in the risk assessment process and the sustainability of ITS tools, such as those developed within OSIRIS, beyond the end of the project in 2011. It was acknowledged that for ITS to gain widespread acceptance, all parties will need to embrace a change of mind-set, and transparent software tools will need to be openly accessible to all.

## 2.4.2 Comments to human and environmental toxicity and exposure

*By Bette Meek, Associate Director, Chemical Risk Assessment, McLaughlin Centre for Population Health Risk Assessment, University of Ottawa, (on Interchange from Health Canada),*
*E-mail bmeek@uottawa.ca*

Comments offered here are based on experience acquired in meeting the time limited legislated mandate in Canada to set priorities for health risk assessment and management from amongst the 23, 000 compounds on the Domestic Substances List (DSL) under the Canadian Environmental Protection Act (i.e., "categorization"). This exercise involved development and testing of predictive methodologies for both exposure and effect.

**Exposure**

One of the important observations from the categorization exercise was the limited influence of quantity of production on potential for exposure, based on relatively simple exposure profiling conducted for all of the entries on the DSL. In fact, the nature and pattern of use of the chemical was far more influential, with a significant number of high volume production substances considered to present "lowest potential for human exposure".

This observation likely has implications for exposure based waiving for the high production volume chemicals in Europe. In addition, the methodology which was developed to relatively rank potential for exposure for all 23, 000 chemicals on the Canadian DSL based on their production volume and use profile may be additionally helpful in this context. For example, there is potential to quantify exposure based on this profiling (in addition to physical/chemical properties) through comparison with quantita-

tive estimates for well characterized Priority Chemicals with similar use patterns and properties.

I wished also to comment on some aspects related to the threshold of toxicological concern (TTC). While it offers potential in priority setting (including exposure based waiving), I believe that there are significant barriers to its widespread adoption, currently, the most important of which relates to transparency of the supporting underlying database on toxicity. In addition, there seems to be limited understanding that the TTC represents essentially "negligible exposure", based on consideration of relevant data in a manner similar to that which serves as the basis for quantitative structure activity relationship (QSAR) models. I particularly liked Theo's characterization of the TTC in the context of a "no further action level". However, as per a number of commercially available (Q)SAR models, there is limited transparency concerning the nature of the relevant original data which support the TTC; development of a software tool to enable users access to the relevant underlying primary toxicological data would likely contribute considerably to increasing understanding and its potential application. Certainly, the TTC may offer promise for consideration in the context of industrial chemicals though it was developed originally for application in relation to food additives, based on recent comparison of the "chemical space" of the underlying databases with that for the Canadian DSL.

**Biological Domain**

There is also potential to fairly efficiently identify chemicals which are relatively "non-toxic" based on hierarchical consideration of available data on hazard and relatively conservative criteria for dose-response for relevant endpoints. A tool of this nature developed for DSL categorization permitted efficient identification of approximately 20% of the chemicals examined as not requiring additional consideration with very limited investment of resources.

It will also be important to make optimum use of the available toxicological data since it is the limiting determinant of the potential contribution of (Q)SAR modelling and read across (including categories and analogues). The likely contribution of these interdependent lines of evidence, in a predictive context particularly from a human health perspective is limited considerably by the extent of the existing dataset on their toxicity and its mining in a structure activity context. The limited information captured in databases that underlie some commercially available (Q)SAR models for complex endpoints such as developmental toxicity is simply inadequate to consider, for example, relationships between various endpoints and potential patterns of effects associated with specific modes of action. This issue has been considered recently in a project of the International Life Sciences Research Foundation funded by Health Canada and the U.S. Environmental Protection Agency which brought together endpoint specialists, (Q)SAR model developers and risk assessors, as a basis to design and populate a database to better inform (Q)SAR modelling for this endpoint.

It's also critically important to reconsider the criterion of data "quality" in the context of development of databases as a basis to inform predictive tool development. Objectives are necessarily considerably different than that for which rather narrow reliability criteria (e.g., Klimisch) have been applied in the past in the consideration of individual toxicity studies. Rather, what is critically relevant in this context is assimilation of the weight of evidence across broader data sources, taking into account factors such as consistency, specificity, biological plausibility, etc.

The need for early consideration of mode (mechanism) of action in the development of efficient and integrated test strategies is also critical. Indeed, the lack of same in previous traditional testing strategies for hazard for human health endpoints has severely limited the potential value of available data on hazard in the development of predictive tools. The sole possible exception is can-

cer/genotoxicity (in particular for DNA-reactive carcinogens), for which there is at least crude consideration of how the chemical may be inducing the effect. In fact, it is envisaged that chemicals may be more meaningfully grouped for further consideration in future based on their genomic profiles. These profiles can be further linked to early key events for particular modes of action for critical effects; focus on early key events in a mode of action continuum (versus measures of overt toxicity) should in future obviate the need for longer term studies.

Also, rather than considering chemicals individually, there is a need to take into account the extent of coverage of "chemical space" in determining testing priorities, as a basis to advance predictive methodologies.

### 2.4.3 Comments to integration of components in the OSIRIS framework & use in the REACH process

*By Dr. Watze de Wolf, ECETOC, Av. E. Van Nieuwenhuyse 4, B-1160 Brussels*

*Member ECETOC Scientific Committee, Director Health & Environmental Sciences, DuPont*

Comments offered are based on experiences acquired in preparing an industrial chemicals company for REACH, as well as experiences in the use of non-testing information in research and development activities.

**OSIRIS Elements for Success**

Several points are critical to successful application of the OSIRIS framework in the context of the new EU chemicals legislation REACH.

First and foremost the endpoints addressed in OSIRIS need to match the information requirements as stipulated in the different annexes of REACH. The challenge lies not with the development of non-test approaches for environmental endpoints, or local and acute toxicity endpoints. However, repeat dose toxicity is where most animals are used. Non-test information is expected to have the most significant animal-use reduction potential for reproductive toxicity assessment.

OSIRIS efforts should focus on 'in time delivery' of its tools and approaches targeted at substances with a registration deadline of December 2013. It is unrealistic to expect an impact for substances with a registration deadline of end 2010. For these substances the information requirements need to be fulfilled already mid 2009 to allow the Consortia to finish their hazard assessment part of the Chemical Safety Report end 2009, thus allowing just enough time for the exposure assessment and subsequent registration dossier submission by the Lead Registrant by mid 2010. In order to be considered as an information source for the 2013 substances OSIRIS should be functional already by end 2010.

Deliverance end 2010 does allow the consortia to assess applicability & remaining uncertainty of the non-test information, and either take the decision whether to accept the non-test information as is, or start generation of Annex VII & VIII information prior to registration. Within industry these decisions are not only made by scientists, who can assess the technical merits of the non-test information, but also by business decision makers (risk managers). The latter group will have to balance the risk of non-acceptance by the authorities in the context of their overall business planning. Are they willing and able to accept the residual uncertainties and the potential that their scientists will have to spend (extended)

time and resources in interpretation discussions with ECHA representatives?

Acceptability considerations are not restricted to industry and ECHA as the sole actors. In a growing global world hazard information has no regulatory or geographical boundaries. Hence, other authorities such as for instance EFSA, FDA, US EPA, Health & Environment Canada make use of the same hazard information. Hence, OECD activities on Mutual Acceptance of Data, and the development (Q)SAR Toolbox Phase II are important elements that will have a significant impact on the use of OSIRIS Framework outputs.

The Framework has to be easily accessible, and user friendly. That means among others that the input information should be clearly captured, that the algorithms used are transparent, that the results are reproducible, and the outputs formatted in such a way that they are ready for use in REACH. Finally, OSIRIS should develop a vision how to deal with end-users when questions arise, or when bugs are discovered, and also how it will ensure sustained development, support and maintenance for the tool.

A business plan is needed to deal with many, if not all, of the above challenges. Without such a plan I expect that OSIRIS will deliver scientific developments for an R&D environment, not a regulated one.

### 2.4.4 Results of the plenary discussion

The plenary discussion after the presentations focussed on the following main points and open questions:

**Uncertainty of testing strategies**

- Industry takes a special view on uncertainty. If a test is legally accepted, than it is regarded as reliable.

- Uncertainty relates to a social construct: certainty or safety are both social constructs. This means: these are mental instruments to explain variability of results without knowing the exact cause for each variation. There is always uncertainty involved in every testing (false negative/ false positive).
- Science-based risk assessments are not sufficient for evaluating and managing risks. It's a question of how much uncertainty one is willing to accept. There is a need of an integrated concept that links:
  - risk assessment with risk perception and socio-cultural processing of risk
  - physical risk analysis with financial, economic and social risk
  - risk theory with organizational capacity building and management competency.
- Is uncertainty greater when using animal testing or ITS?

**Available and sharing data**

- Data from industry is needed both to develop the ITS in OSIRIS and meet the requirements of REACH. Sharing data will be one of the important factors in reducing animal test and costs. The problem is that some partners, mostly the industry, must see a benefit if they agree to share data with others. They have to provide data continuously for research and OSIRIS will rely on continuous data flows for their webtool. Therefore data transfer and sharing should be harmonized and be obligatory for all actors. This is in the best interest of the public. However, on should respect that some sort of sensible data is proprietary and will not leave companies.

**Selecting endpoints**

- It was discussed why sensitisation and mutagenicity were selected as endpoints instead of reproductive toxicity. On the first Expert Workshop, reproductive toxicity got the highest ranking, too, because of the number of animals and costs involved. But it was argued that a lot of animal testing is also necessary in the case of sensitisation. In addition, there is more data available for sensitisation than for reproductive toxicity. This is certainly an important point in developing ITS. Mutagenicity was selected because a great amount of in vitro data are already available.
- The participants raised the question whether two generation testing is much more valid than one generation testing. Is there a great loss of information when performing a one generation test only? A result of one study does not confirm this hypothesis, but this must be more validated.

**Setting priorities of chemicals and endpoints**

- OSIRIS should take into account the extent of coverage of "chemical space" rather than considering chemicals individually. This could be important to determine testing priorities as a basis to advance testing methodologies.
- There was support for the idea of including as many lists and endpoints as possible. However, resources (e.g. budget, time) are limited. The question might be: Which endpoints should be considered?[6] OSIRIS researchers pointed out that, at the beginning, the focus was on a narrow selection of lists and endpoints. This choice will be broadened further as the project proceeds.

[6] This is a point that was already being discussed at the first Expert Workshop (see results in the report on the OSIRIS homepage, http://www.osiris.ufz.de/).

**Learning from other projects and using their routines**

- It was advised that OSIRIS should avoid doing research that has been done before in other projects but instead go beyond that. OSIRIS researchers pointed out that research findings of other relevant projects are considered in a routinized manner and all relevant studies will be taken into account.
- OSIRIS can perhaps learn from other similar projects. For example, the question was raised of how many cases of the 23.000 analyzed substances of the Canadian research program QSARs turned out to be relevant. Although exact figures are not available, one can assume that it was quit a great amount. QSARs could become a promising perspective for OSIRIS, too. Actually, it was emphasized that, in the first two years of the project, QSARs will be developed and made ready for easy access. Additionally approaches of ITS will be incorporated for the ongoing work in OSIRIS.

**Open question, which should be taken into account by OSIRIS:**

- OSIRIS should consider the political dimension in the question of acceptance of ITS by for example ECHA, Member States or the EC. How can acceptance by these European and governmental institutions be best achieved?
- The goal of OSIRIS is not to write deliverables but also to circulate testing methods and make them acceptable and usable by different stakeholders. What can be delivered by OSIRIS and in what time? How do the timelines of OSIRIS and REACH match?
- What will come after 2011 when the funding from the EC will stop at end of the project? Will OSIRIS simply end? What will come after the OSIRIS project and who will support and take care of the web tool? Will it be an open source product which everybody is allowed to use and update?

### 2.4.5 Results of the World Café

According to the World Café, the participants discussed three main questions at different tables. After three rounds the table hosts summarized the results on three flip charts separately for each question. These three flip-overs are presented below.

**Question 1: Under which conditions are the proposed integrated**

**Flip-over 1**

**Basics:**

- Models must be available
- Comparable substances (data) are crucial
- Basic knowledge about specific substance.

**Other conditions:**

- model transparency
- transparency of weight of evidence (incl. waiving)
- scientific sound basis
- interaction between stakeholders
- dissemination, communication and training
- easy to use, costs
- for all stakeholders: confidence in the ITS
- ITS with classification DNEL, DMEL, PNEC
- Data base needs to include human data (epidemiological, etc.)
- ITS tools must be available
- substance should fit into a domain of applicability of the model.

**Question 2: Is the pool of existing information sufficient to conduct integrated testing strategies and if not, which additional information needs to be accumulated?**

**Flip-over 2**

Is it true that the majority of chemicals no data exists?

One has to differentiate three options:

- Volumes: Chemicals are produced in very different amounts of tonnes. If considering only the chemicals with high volume, there is data available. For 75% of the chemicals volumes data do exist.
- Number of substances: If considering the absolute number of substances, there is no data for a lot of single substances.
- Endpoints: Only if we know all possible endpoints we can say, that there is enough data or not. Do we know them all?

There is a contradiction: On the one hand, data gaps are minimal considering volumes (active groups). On the other hand, basic data is missing for a lot of substances (rest of substances). Besides ITS are already being used with the data available (Testing Strategy Steer).

Some elements of ITS aren't as new, superior or different from well known practices as is being assumed. For example, a QSAR is a formalized expert judgement.

**Recommendation:** Definition of QA/QC (quality criteria) for new test methods, old none GLP-data, etc.

Additionally, there are some individual statements by members of the group studying this question:

- It is not sufficient. There are data that are not public. Companies have to share these data but they need to be confident on the "downstream users" of data.
- It is too early to say if adequate data exists for ITS.
- The suitability of ITS depends on the substance class and the selected endpoint.
- As technologies continue developing, the existing data will never be sufficient.
- How to assess complex mixtures (e.g. natural oils)?
- Easier for local rather than systemic effects.
- In case of lack/insufficient information, more information needs to be generated on the Mode of Action.
- Evaluation framework can be highly subjective, e.g. we need more transparency about current Risk Assessment methods.
- There should be a common sense about the endpoints companies and scientists are working with.
- Access to alternative test methods should be possible (in addition to OECD etc. as the validation takes years).
- Read-across – grouping of chemicals by SA/ QSAR, Mode of action, Pharmakokinetics

**Question 3: Can you reduce the amount of testing, especially animal testing (reducing costs and time) without sacrificing accuracy, validity and reliability of the results?**

**Flip-over 3**

Accuracy, Validity and Reliability of the Golden Standard?

➔ Surrogate for human test, uncertainty is there, so you can adopt ITS as well.

Can we reduce the amount of testing strategies?

Yes, we can under three conditions:

- use contextual information
    - category data
    - read across
    - Mode of Action information
- combine endpoint with specific test, for example RDT and in vivo mutagenicity
- use early indicators instead of "late" indicators, shorten exposure of animals to chemicals.

➔ needs category – ITS

## 2.5 Conclusion

The implemented guidelines of REACH require a new strategy to minimise the use of animals in testing methods. Gerrit Schüürmann explained in the context of the project the principle of Humane Experimental Technique from Russell and Burch (1959)[7] "3Rs" (reduce, replace and refine animal testing) which is internationally accepted and promoted in the partnership between the European Commission and industry (EPAA). The industry understands that, for reasons of animal welfare, costs and logistics, it is important to limit the number of tests to be conducted. Integrated Testing Strategies (ITS) will give the opportunity to accelerate the use of non-testing information for regulatory decisions making of chemicals without reducing the required level of safety. OSIRIS will develop approaches for the many-to-one replacements of animal tests, i.e. Integrated Testing Strategies (ITS). But using ITS will not entail one-by-one replacements, but several different approaches will be combined and integrated, introduced Gerrit Schüürmann. A systemic combination of the testing strategies like in vitro testing, QSAR, read-across or TTC will help to develop innovative non-animal approaches. One major requirement is to use contextual information with category data, read across and Mode of Action information. It was suggested to combine endpoints with specific tests, for example RDT and in vivo Mutagenicity.

[7] William .M.S. Russell and Rex. L. Burch (1959): The Principles of Humane Experimental Technique.
http://altweb.jhsph.edu/publications/humane_exp/het-toc.htm,
downloaded 9.12.2008

**Uncertainty of ITS**

An important limiting factor in implementation of the ITS will be the level of uncertainty that one is willing to accept when applying the modified testing strategies. The participants asked if the uncertainty boundaries will be higher with the new testing strategies compared to conventional animal tests. But uncertainty is involved in all testing methods and cannot be reduced to zero. The boundaries of uncertainty associated with traditional testing are also not well known in quantitative terms too.

Industry has a special view on uncertainty: if a test is legally accepted, than it is regarded as reliable. However, there is no 100% safety or reliability with any test method. If something is regarded as safe it means that the remaining uncertainties are judged acceptable to society. It is a judgement rather than a scientific fact.

The new testing strategies demand a new concept dealing with uncertainty. The open question remains of how much uncertainty one is willing to accept. The acceptance of uncertainty needs an integrated concept that links:

- risk assessment with risk perception and socio-cultural processing of risk
- physical risk analysis with financial, economic and social risk
- risk theory with organizational capacity building and management competency.

**Priorities in OSIRIS and in time delivery**

OSIRIS should take into account the extent of coverage of "chemical space" rather than considering chemicals individually. Looking at chemical spaces can facilitate the selection of testing priorities as a basis to advance testing methodologies. Exposure considerations and in particular use categories are also influential factors for ranking chemicals. In fact, the nature and pattern of the chemicals' usage have proven out to be more influential than the

volume of the respective substances. It is important to obtain information on early effects and "Modes of Action" in the risk assessment process. This information should be linked to a conceptual framework like the one developed by the International Life Sciences Institute (ILSI). TTC offers potential for priority setting (including exposure based waiving) if the process is transparent and open for viewing the underlying database on toxicity. There is also a potential to identify fairly efficiently those chemicals that are relatively "non-toxic" based on hierarchical considerations of available data on hazard and relatively conservative criteria for dose-response for relevant endpoints. A tool developed for DSL categorization permits efficient identification of approximately 20% of the chemicals examined as not requiring additional consideration with very limited need of resources.

OSIRIS efforts should focus on 'in time delivery' of its tools and approaches targeted at substances with a registration deadline of December 2013. In order to be considered as an information source for the 2013 substances OSIRIS should be functional already by end 2010. Deliverance end 2010 does allow the consortia to assess applicability & remaining uncertainty of the non-test information, and either take the decision whether to accept the non-test information as is, or start generation of Annex VII & VIII information prior to registration. Within industry these decisions are not only made by scientists, who can assess the technical merits of the non-test information, but also by business decision makers (risk managers). The latter group will have to balance the risk of non-acceptance by the authorities in the context of their overall business planning. Are they willing and able to accept the residual uncertainties and the potential that their scientists will have to spend (extended) time and resources in interpretation discussions with ECHA representatives?

**Data existence for ITS and REACH**

On the one hand, data gaps are minimal considering volumes (active groups). On the other hand, basic data is missing for a lot of substances (rest of substances). The adequacy of data is highly context- and policy-dependent. ITS are already being used with the data available (Testing Strategy Steer). The participants recommended to define QA/QC (quality criteria) for new testing methods, old none GLP-data etc., to decide which data are available and which data can be used for which purpose. The criterion of data "quality" in the context of development of databases should inform predictive tool development. The weight of evidence should be assimilated across broader data sources, taking into account factors such as consistency, specificity, biological plausibility, etc.

**Sharing of Data**

Data from industry is needed both to develop the ITS and for meeting the regulatory requirements of REACH. Sharing data will be one of the important factors in reducing animal test and costs. The problem is that some partners, mostly the industry, must see a benefit if they agree to share data with others. In addition, they have to provide data continuously for research and OSIRIS will rely on continuous data flows for their webtool. Therefore data transfer and sharing should be harmonized and be obligatory for all actors. This is in the best interest of the public. However, on should respect that some sort of sensible data is proprietary and will not leave companies. The industry must rely on the confidentiality of potential "downstream users" of data. As technologies continue developing, the existing data will never be sufficient.

The framework has to be easily accessible, and user friendly. That means among others that the input information should be clearly captured, that the algorithms used are transparent, that the results are reproducible, and the outputs formatted in such a way that

they are ready for use in REACH. Finally, OSIRIS should develop a vision how to deal with end-users when questions arise, or when bugs are discovered, and also how it will ensure sustained development, support and maintenance for the tool.

**ITS under REACH conditions**

First and foremost the endpoints addressed in OSIRIS need to match the information requirements as stipulated in the different annexes of REACH. The challenge lies not with the development of non-test approaches for environmental endpoints, or local and acute toxicity endpoints. However, repeat dose toxicity is where most animals are used. Non-test information is expected to have the most significant animal-use reduction potential for reproductive toxicity assessment.

For the use of ITS under REACH the transparency of models and ITS, especially weight of evidence, is absolutely necessary; otherwise the stakeholder will have no confidence in the methods. OSIRIS has to disseminate and communicate the new methods to all interested parties and train the stakeholders to use them properly.

Basic knowledge about specific substance and comparable substances (data) are crucial. The testing strategies should be easy to use, have low costs and contain the classification DNEL, DMEL, PNEC. Data bases needs to include human data (epidemiological, etc.).

Common sense should be employed to choose the most sensible endpoints for academic and industrial research. The evaluation framework can be highly subjective, e.g. OSIRIS needs more transparency about current Risk Assessment methods. Access to alternative test methods should be granted to all interested parties (in addition to OECD etc.), as the validation takes years.

OISRIS should develop a formal weight of evidence (WoE) framework for evaluating and documenting ITS and integrating

the different types of information. This should be done in a transparent and objective manner. This refers particularly to the quantification of uncertainties and the resolution of conflicting values.

The two-generation study required by REACH could be replaced by an extended one-generation study. As an additional opportunity the use of early indicators instead of "late" indicators may shorten the exposure of animals to chemicals.

**Acceptance and sustainability of using ITS**

One goal of OISRIS is to develop ITS and a webtool that will offer a wide range of applications beyond the end of the project of 2011. To accomplish continuous service and availability of the results of OSIRIS it is necessary to gain acceptance by all parties involved in the risk assessment process. It was acknowledged that all parties will need to embrace a change of mind-set, and transparent software tools will need to be openly accessible to all. OSIRIS should conceive a practical solution of how the webtool and research results could be made available to all interested parties and further sustained after the end of the project. For practical reasons, the timeline of OISIRIS and REACH should be aligned.

OSIRIS should consider the political dimension in with respect to the acceptance of ITS by, for example, ECHA, Member States or the EC. A new integrated concept interrelated with the different levels of uncertainty have to be accepted not only from the user side (mainly industry) but also from European (ECHA, Member States, ECVAM) and Non-European authorities. Hence, OECD activities on Mutual Acceptance of Data, and the development (Q)SAR Toolbox Phase II are important elements that will have a significant impact on the use of OSIRIS Framework outputs.

**Gaps in the acquisition of exposure**

According to the legal text of REACH, the possibilities for exposure based waiving are quite limited and the burden of proof is very high. Exposure based waiving should be justified by a thorough exposure assessment. It should be determined whether current exposure models and measurement data are suitable to estimate exposure accurately in the lower exposure range. When valid exposure estimates or measurements are obtained, they should be compared to a relevant toxicological threshold to determine whether exposure is below the 'no further action level'. Although some thresholds are available it is still unclear to what extent they meet the criteria stated above. This needs to be evaluated. In addition, it needs to be determined whether it is valid to assume that exposure to substances in REACH at levels below the given thresholds do not pose any (substantial) risk.

A business plan is needed to deal with many of the above challenges. The development and assessment process of IST needs time. In order to gain confidence and continuous feedback for alternative testing methods from stakeholders, an open and transparent process is absolutely essential.

## 2.6 Annex: Agenda

**Monday, 17th of November 2008**

| | |
|---|---|
| 8.30 | On-site Registration, Coffee and refreshments |
| 9.30 | **Welcome and introduction to the OSIRIS framework**<br>*Prof. Dr. Gerrit Schüürmann, Helmholtz Centre for Environmental Research - UFZ, Germany*<br>*Dr. Georges Deschamps, European Commission, Brussels, Belgium*<br>*Moderation: Prof. Dr. Ortwin Renn & Christina Benighaus, DIALOGIK and University of Stuttgart, Germany, Frederic Bouder, King's Centre for Risk Management, King's College London, UK* |
| 9.45 | **REACH: alternative testing – a new practical approach**<br>*Evelin Fabjan, European Chemicals Agency - ECHA, Finland* |
| 10.15 | **Human and environmental toxicity and exposure: results and critical points of the OSIRIS framework**<br>*Defend: Mark Cronin, School of Pharmacy and Chemistry, Liverpool John Moores University, UK*<br>*Second Defend: Dr. Theo Vermeire, National Institute of Public Health and the Environment - RIVM, Bilthoven, The Netherlands*<br>*Opponent: Dr. Bette Meek, Chemical Risk Assessment, McLaughlin Centre for Population Health Risk Assessment, University of Ottawa, Canada* |
| 11.30 | **Coffee break** |

| | |
|---|---|
| 12.00 | **Integration of the components in the OSIRIS framework and use in the REACH process: results and critical points**<br>*Defend: Dr. Dinant Kroese, Chemical Safety, TNO Quality of Life, Zeist, The Netherlands*<br>*Opponent: Dr. Watze de Wolf, Environmental Sciences Europe, DuPont, Belgium* |
| 13.15 | **Lunch** |
| 14.00 | **World Café/Carousel method: process addressing the three leading questions (see below) and the three issues addressed in the papers by the opponents:**<br>*- Under which conditions are the proposed integrated testing strategies operational for being used under REACH?*<br>*- Is the pool of existing information sufficient to conduct integrated testing strategies and if not, which additional information needs to be accumulated?*<br>*- Can you reduce the amount of testing, especially animal testing (reducing costs and time) without sacrificing accuracy, validity and reliability of the results?*<br>*Categorisation of the proposed methods and procedures according to relevance and implementability* |
| 15.30 | **Coffee break** |
| 16.00 | **Presentation of the Group results** |
| 16.30 | **Plenary discussion and summary of the results**<br>*Dr. Andrew Worth, European Chemicals Bureau, Ispra, Italy*<br>*Prof. Dr. Gerrit Schüürmann, Helmholtz Centre for Environmental Research - UFZ, Germany* |

| | |
|---|---|
| 17.00 | **End of the workshop, closing and farewell address**<br>*Prof. Dr. Gerrit Schüürmann, Helmholtz Centre for Environmental Research - UFZ, Germany* |