Overcoming challenges and advancing (bio)degradation guidelines: OECD TG309 revisited

Workshop Report No. 41

Overcoming challenges and advancing (bio)degradation guidelines: OECD TG309 revisited

Workshop Report No. 41

Brussels, January 2025
ISSN-2078-7219-041 (online)

ECETOC Workshop Report No. 41

European Centre for Ecotoxicology and Toxicology of Chemicals
Rue Belliard 40, B-1040 Brussels, Belgium.

DISCLAIMER:

Representatives of the European Commission, ECHA and Member State Competent Authorities expressed no official view or position on the scientific matters discussed during the breakout sessions and their participation cannot be taken as agreement/disagreement to the views and/or positions expressed during the breakout sessions.

This workshop report is intended to accurately reflect the workshop discussions and conclusions, but it is noted that not all workshop participants provided written input to the workshop write-up.

CONTENTS

SUMMARY

The Organisation of Economic Cooperation and Development (OECD) test guideline No. 309 (Aerobic Mineralisation in Surface Water – Simulation Biodegradation Test) is designed to assess the degradation of substances in aerobic natural water. Information derived from this test is used in a regulatory context to evaluate the persistence of substances, as well as to inform exposure assessments. Experience with the test guideline has grown in recent years due to its increasing use within regulatory frameworks, and the guideline was recently included in the work plan for revision under the OECD Test Guidelines Programme.

In response to these developments, an ECETOC workshop was held in January 2025 to discuss the latest developments and challenges concerning the test guideline and to identify the key updates and further research that were needed. The workshop was attended by 36 in-person participants, with a further 76 participants attending online. The workshop was structured around three pillars (robustness, implementation and relevance) in order to provide framing for discussions and to ensure that these different aspects received adequate focus. The workshop began with scene-setting presentations, followed by presentations in plenary for each of the three pillars. Each presentation block was followed by questions and answers from the audience. After the plenary sessions, in-person participants were divided into three breakout groups, each assigned to one of the three workshop pillars. Each breakout group was tasked with identifying and discussing challenges and corresponding solutions, relevant research, and proposed next steps for their respective pillar. There was a final plenary session during which an overview of each of the breakout sessions was presented before closing the workshop.

Scene-setting presentations included experiences from the initial development and ring testing of the guideline, and a recent expert survey of update needs and regulator's viewpoint. In the robustness pillar, a number of challenges were identified relating to inter- and intra-test variability that affect the robustness of OECD Test Guideline (TG) 309 as a tool for regulatory prioritisation and assessment. In the implementation pillar, practical challenges with the routine use of the test guideline and potential solutions were highlighted, particularly in relation to dealing with substances with difficult-to-test properties and complex compositions. In the relevance pillar, it was discussed that the test guideline in its present form and current regulatory implementation did not reflect well the degradation of substances in the environment, and therefore that the environmental relevance of the test was poor.

The issues identified in the plenary presentations were further discussed in the breakout sessions, with possible solutions and next steps for OECD implementation identified. It was clear that many issues were cross-cutting and relevant to more than one of the three workshop pillars (robustness, implementation and relevance). Key themes identified were:

- Microbial biomass concentration and characterisation as a key source of variability in test results
- Address co-metabolic vs growth-linked biotransformation kinetics within the test and its relationship to test substance-biomass concentration ratios
- Inclusion of suspended sediment as an option for improving test robustness, with corresponding concerns of an impact on relevance

- Impacts of inoculum sampling and storage on test outcomes
- Permitted variations in experimental setups/conditions impacting test outcomes
- Refine the applicability domain (AD) of the test to address substances with challenging properties and complex compositions
- More stringent requirements on study reporting
- Clearly define aerobic conditions and address issues with closed test systems
- Improve guidance on pre-tests, sampling strategy, and sterile and solvent controls
- Choice of surface water for the inoculum
- Establish more clearly the test guideline title and objectives, potentially splitting up the guideline according to different objectives
- Define clear parameters for the inclusion of diffuse light

1. INTRODUCTION

1.1 Background

Understanding the fate and degradation of substances in the environment is a key component of chemical hazard and risk assessment. The OECD test guideline (TG) No. 309 (Aerobic Mineralisation in Surface Water – Simulation Biodegradation Test) is designed to assess the degradation of substances in aerobic natural water (OECD, 2004). Information derived from OECD TG 309 tests is used in a regulatory context to evaluate the persistence of substances, as well as to inform exposure assessments.

The guideline was originally published in 2004, although there was limited experience with it until it was later introduced as a standard information requirement in the EU regulation of plant protection products (EC, 2009, Hand and Moreland, 2013, Solomon et al., 2013). More recently, OECD TG 309 has been identified as the preferred first choice of test guideline for assessing the persistence of industrial chemicals, pharmaceuticals and biocides, according to European Chemicals Agency (ECHA) guidance, and there has been an increase in the number of these tests requested for substances registered under the EU Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation (ECHA, 2017).

As a result of these developments, the OECD TG 309 test guideline has taken on increased significance in the assessment and management of substances, particularly within the EU. There has been a corresponding increase in the amount of research publications related to this guideline. One notable project funded by the German Environment Agency (UBA) asked experts for their recommendations for improvements of a suite of OECD test guidelines, in which the OECD 309 received the highest number of suggested changes (Polcher et al., 2023). Another project, Cefic-LRI ECO55 “Assessing the Impact of Sample Collection on Microbial Population and Validity Criteria in the OECD 309 Surface Water Mineralisation Test”, led to a proposal for revision of the test guideline being submitted to the OECD Working Group of National Co-ordinators of the Test Guidelines programme (WNT). This proposal has subsequently been accepted and was introduced into the 2024 work plan for the OECD Test Guidelines Programme (OECD, 2024).

Given the recent developments concerning OECD TG 309 and its growing importance within chemicals assessment and management, ECETOC considered it timely to host a workshop to bring together experts from across the scientific community to discuss the latest research and experiences with this test guideline and develop proposals for its improvement.

1.2 Workshop structure and aims

The workshop was organised as a 1.5-day event that was attended by 36 in-person participants from across industry, academia, regulatory bodies, contract research organisations (CRO), and consultancies. The workshop was structured to include presentations in plenary format (with audience Q&A) and breakout discussions in smaller groups. Plenary sessions were also open to virtual participants, with a further 76 participants attending online.

The workshop's stated aims were:

- Discussing and addressing current challenges and limitations in conducting and interpreting degradation simulation studies.
- Defining and aligning on the key updates needed for the OECD TG 309 guideline.
- Establishing a roadmap to ensure that prioritized updates are addressed in time for the OECD TG 309 update.
- Identifying research needs and potential methodological advancements to enhance the relevance and applicability of this and other guidelines.

It was decided by the workshop organising committee (Appendix C) to structure the workshop according to three pillars: robustness, implementation and relevance. This was done to provide a framing for discussions and to ensure that the different aspects of OECD TG 309 requiring consideration received adequate focus. The three pillars were defined as follows:

- **Robustness** – Reproducibility of the test guideline and impact of permitted experimental conditions
- **Implementation** – Experiences with running the test guideline for a range of substances and/or physicochemical properties
- **Relevance** – Extent to which the test guideline reflects what occurs in the environment

The event began after a welcome lunch with an introduction by ECETOC, followed by two scene-setting keynote presentations. Following this there were three further plenary sessions, one for each of the three pillars. The first two pillars (Robustness and Implementation) completed the first day of the workshop, with the third pillar (Relevance) being conducted in the morning of the second day. Each of these sessions included a keynote presentation followed by several shorter presentations of related research from workshop participants. The presentations for each of the pillars were followed by a 20-minute Q&A session with the audience (both in-person and online).

Following the plenary sessions, the in-person workshop participants were divided into pre-assigned breakout groups corresponding to each of the pillars. Breakout discussions were held over two sessions with a lunch break in-between. Each breakout group was tasked with identifying and discussing challenges and corresponding solutions, relevant research, and proposed next steps for their respective pillar, and the time needed for the relevant update. Following the conclusion of the breakout discussions there was a final plenary session during which an overview of each of the breakout sessions was presented, with audience Q&A, followed by the close of the workshop.

2. PRESENTATION SUMMARIES

All workshop presentations have been made publicly available to download from the ECETOC website (link).

2.1 Scene-setting

2.1.1 Welcome, introduction and workshop objectives

Dr. Blanca Serrano Ramòn, Secretary General at ECETOC, opened the workshop by welcoming workshop participants, introducing the organisation of ECETOC, and describing the objectives of the workshop. She then introduced Megan D'Souza who took on the role of introducing the speakers and chairing the workshop.

2.1.2 Lessons learned in ISO ring test in 1998-99 for OECD TG309

Dr. Lars Toräng, Bjørn Thorsen A/S, Denmark

Presentation: link

The first scene-setting presentation was given by Dr. Lars Toräng. Lars had been involved in the development of the original OECD 309 test guideline as part of his PhD project and had kindly agreed to join the workshop to provide this important perspective on the origins of the test guideline.

Lars first described the aim of his PhD project, which was to increase knowledge and process understanding of biodegradation in batch simulation tests. His PhD was laid out in four objectives:

- Study biodegradation mechanisms and phenomenon in batch simulation tests in surface and groundwater
- Effect of the test concentrations
- Other factors affecting the microbial adaptation
- Investigate whether the standardized simulation tests are adequate for estimation of real-world biodegradation rates

The above objectives signify that microbial community adaptation was an important consideration from the outset.

It was indicated that the need for a batch simulation test was originally identified for the purposes of chemical risk assessment (i.e. determination of Predicted Environmental Concentration, PEC). These tests would be conducted in laboratory systems with conditions that are realistic for the particular environmental compartment (e.g. surface water, sediment, soil, groundwater). They should mimic the actual environmental

conditions such as redox potential, pH, temperature, microbial community, concentration of test substance, suspended solids/sediment, and occurrence and concentration of other substrates.

The OECD 309 test guideline was preceded by a ring-test report prepared and organised by Niels Nyholm and Lars Toräng (Technical University of Denmark) for the International Standards Organisation (ISO, 1999). The drafting of the method started in 1991, and the ring-test was run in 1998/99, with seven laboratories participating (UBA, BASF, Henkel, NIVA, Shell, Finnish Environment Institute and Technical University of Denmark).

The ring-test used two model compounds: aniline and 4-chloroaniline, and a concentration range of 0.5 – 500 µg/L. All laboratories used the ^{14}C-radiotracer technique. The ISO 14592-1 ring-test report has been made available for download on the ECETOC website (link). The OECD 309 test guideline was subsequently published in 2004.

In the ring-test experiments, concentrations of the test substance and degradation rate constants were calculated based on measurements of the remaining radioactivity in test vessels. Degradation of both aniline and 4-chloroaniline were characterised by an initial slow phase, followed by a second, faster phase of degradation. This biphasic pattern was attributed to the adaptation of the microbial community. In the case of aniline, adaptation was always observed (apart from in two cases) and the time to adaptation was faster than for 4-chloroaniline. 4-chloroaniline degradation rate constants for both non-adapted and adapted inocula were determined, with non-adapted degradation rates corresponding to half-lives of 54-340 days. High variation in degradation rates of aniline was observed depending on the test concentration. Separate data showing variable lag phases (time to adaptation) for the substance 2,4-D was also presented.

A separate field study of aniline degradation in the river Rhine was conducted. The degradation rate of aniline was determined by measuring the concentration at various distances from a known discharge site and considering the river flow rate. Dilution of the plume was accounted for by also measuring conductivity and concentrations of sulfate and 1,4-dioxane. The degradation rate of aniline (half-life of 9 hours) was observed to be in the same range as the adapted degradation rates observed in the laboratory batch experiments.

A semi-continuous pre-exposure procedure (SCEP) was also developed to investigate adaptation of biomass under “natural” conditions. This followed the same setup as the batch tests, but with one third of the test volume being replaced with freshly collected water at regular intervals. This was demonstrated with the substance 2,4-D where, at test concentrations of 100 µg/L, the lag phase was gradually reduced from 25 days initially to virtually zero after pre-exposure of the inoculum for 4-5 weeks. By contrast, when tested at 0.5 µg/L, no adaptation of 2,4-D was observed, and degradation rates after pre-exposure were comparable to the freshly collected river water. The population density of specific degraders could also be traced by a ^{14}C-MPN (most probable number) assay. The SCEP was concluded to be an environmentally realistic procedure that eliminates or reduces lag phases without increasing degradation rates. It produces adapted microbial populations from which well-defined adapted rate constants could be obtained (similar to those obtained in batch systems after longer lag periods). However, the method was very labour intensive.

Some methodological and technical problems with the batch test system were highlighted, which included limited experiences with the method, getting enough data points in the “degradation window”, addressing primary vs ultimate degradation, the high variation observed between tests, and the challenges of testing

difficult substances and substances with complex compositions. It was concluded that biodegradation rates obtained in studies at high concentrations (> 10-100 µg/L) may grossly overestimate the actual degradation rates at environmentally relevant concentrations. For direct interpretation of measured rates, it is important to keep test concentrations sufficiently low to ensure the same kinetics as in the environment. Biodegradation rates in adapted environments receiving continuous exposures of substances can be achieved by pre-adaptation of the batch tests using a semi-continuous pre-exposure procedure.

2.1.3 Advancing Environmental Assessment: Revising OECD Test Guidelines with Insights from the German Environment Agency (UBA)

Dr. Ulrich Jöhncke, German Environment Agency, Germany

Presentation: link

In this presentation, the results of an UBA-sponsored project "Review of the OECD Test Guidelines relevant to environmental assessment with regard to the state of the art in science and technology" were presented, along with UBA's experience with OECD 309 and some specifications on test conditions given in the ECHA REACH guidance.

The OECD test guidelines (TG) review project was conducted between 2020 and 2023 by Ramboll, Fraunhofer IME and Fraunhofer ITEM. This was intended to address the fact that, despite their importance for assessment under all legislative frameworks, there is no regular update of OECD TGs. The goal was therefore to provide a list of potential update needs across Phys-chem, environmental fate and ecotoxicology test guidelines.

The project sought input from the international community on update needs by a questionnaire via an open online survey. In total there were 257 completed questionnaires, approximately 500 individual suggestions and 59 additional comments via email. Responses were pre-evaluated by ranking and weighting the suggestions. The pre-evaluation results were then discussed in a series of workshops, before recommendations for the revisions of the TGs were prioritised and presented to the OECD Working Group of National Co-ordinators of the TGs programme (WNT). The OECD TG 309 received the highest number of suggestions (>40) and was rated as one of the most relevant TGs for revision.

The second part of the presentation discussed the relevance of OECD TG 309 under different legal frameworks in the EU. This considered both the hazard assessment and the risk assessment, and different regulatory frameworks, such as for REACH, pharmaceuticals, biocides and plant protection products. In terms of the assessor's view on the need for modifications, the need for adequate controls was emphasised, including reference substances and sterile controls, and more information on the analytical methods, e.g. extraction efficiency. It was also recommended to update the title of the TG to 'transformation in surface water' as it currently only mentions mineralisation, which is misleading. It was also recommended to take new scientific findings into account, including using kinetic information from the lag phase, low-level spiking to measure environmentally relevant co-metabolism, considering the use of non-labelled studies, and avoiding any form of (pre)adaptation or measures that change community composition, enzyme expression, or matrix parameters.

It was reminded that aspects related to temperature and suspended solids concentrations were specified in REACH guidance. According to this guidance, new kinetic simulation studies should be conducted at temperature of 12 °C (9 °C for marine environment) to correspond with average temperatures of European surface waters. Representative suspended particulate matter (SPM) concentrations in the EU are considered to be 10-20 mg dw/L or 5 mg dw/L for freshwater and marine water, respectively. Finally, it was emphasised that the OECD TG 309 plays a very important role in hazard assessment under REACH, and for this purpose it needs to have a non-adapted micro-biocenosis, as little adsorption to surfaces as possible, and a clear picture of what happens in the test system (i.e. relevant controls and a high recovery rate).

2.2 Pillar 1: Robustness

2.2.1 Keynote: Enhancing Standardisation and Reproducibility in OECD 309 Biotransformation Studies for Chemical Persistence Assessment

Dr. Carolin Seller, Eawag – Swiss Federal Institute for Aquatic Science and Technology, Switzerland

Presentation: link

This presentation began by defining what robustness means in terms of biotransformation. The robustness of an experimental procedure can be defined as its capacity to maintain its intended function invariant despite possible variations in its inputs. In the case of OECD 309, the intended function is to determine the co-metabolic biodegradation of a test substance in aerobic surface water. Co-metabolic biodegradation describes the situation where the microorganisms are not using the substance as a main substrate for growth (metabolic biodegradation) but instead transform it incidentally as they grow on other substrates. While degradation rates may vary in the field, and between different tests conducted on the same substance using different surface waters, a robust test should demonstrate reproducibility between replicates, and this robustness should be reflected across a range of chemicals.

A second question to be addressed was “does the OECD 309 guideline allow for robust experiments?”. It was first recognised that various experimental variations were permitted within the OECD 309 TG. Data were presented for cold OECD 309 tests in three different surface waters for a mixture of 43 test chemicals, each at a concentration of 1 µg/L. These included pelagic tests and suspended sediment tests, at two different concentrations (1 and 10 g solids/L). In the pelagic tests, only 15 compounds showed any biotransformation after 50 days, and biotransformation only started after extensive lag phases. Significant inter-replicate variabilities were observed for seven compounds. In the experiments with suspended sediment, inter-replicate variability was reduced, with the 10 g solids/L tests showing lower variability than at 1 g solids/L. The specificity of biotransformation pathways and their associated enzymes influences degradation kinetics and is important to consider. Cell type compositions of the different surface waters had been investigated using flow cytometry. This indicated that sediment-borne biomass was generally more stable than in water only. In general, the OECD 309 TG was considered to show significant inter-study and inter-replicate variabilities. It was impossible to disentangle the “natural” and “artificial” variabilities, and current margins allowed within OECD 309 did not allow co-metabolic biotransformation of a wide range of chemicals to be robustly studied.

A third question of the presentation was whether it was generally possible to robustly study biotransformation in the laboratory. Sludge biotransformation experiments were presented where a mixture of 21 pharmaceuticals and pesticides were tested. These types of experiments appeared to show good agreement between experiments, and also when compared to soil experiments. Considering different reaction types, robust biotransformation kinetics were observed for hydrolysis and oxidation reactions, but not for reduction reactions. These data suggested that it was indeed possible to robustly measure biotransformation rates of substances in certain experimental systems.

The switch between co-metabolic and metabolic biotransformation was also investigated by assessing biotransformation of 26 pharmaceuticals, pesticides and food additives at different concentrations (0.5, 5 and 50 µg/L). It was shown that test concentration had a clear influence on degradation kinetics, and that concentrations of ~ 5 µg/L could trigger the onset of metabolic biotransformation (growth kinetics) in sludge. It was therefore considered that both the concentration of the chemical and the amount of biomass present was important in the design of robust co-metabolic biotransformation experiments, and that higher biomass-to-chemical concentration ratios were needed to achieve this. It was concluded that the current biomass-to-chemical ratio in OECD 309 studies is too low to allow for a robust setup. It was therefore recommended that suspended sediment concentrations should be increased to > 10 g solids/L, and that test chemical concentrations should be below 1 µg/L.

2.2.2 Inter-laboratory ring test and proposal to revise OECD 309: Outcome of LRI-ECO55 project

Chris Hughes, Embark Chemical Consulting, United Kingdom

Presentation: link

This presentation summarised the results of an inter-laboratory ring test and a proposal to revise the OECD 309 tests guideline that was part of the work of the Cefic-LRI ECO55 project. Candidate reference substances caffeine and 2,4-D had been identified following a stepwise process that included: literature search, persistence categorisation and physicochemical property screening, OECD 301D biodegradation screening, and validation under OECD 309 conditions. A ring test was then conducted with nine participating laboratories, who ran caffeine and 2,4-D as candidate reference substances alongside either aniline or sodium benzoate. Experiments were run in parallel to ongoing tests, and thus reflected experimental setups and conditions currently being employed by each of the participating laboratories. All experiments were conducted as pelagic tests using 10 µg/L test substance concentrations and dark conditions, corresponding to current guideline requirements for reference substances. Results of the ring test demonstrated high variability in results for both aniline and caffeine, with extents of degradation ranging from negligible to complete (and rapid). 2,4-D generally showed no biodegradation apart from in one lab, which showed appreciable degradation at two temperatures with no discernible lag phase.

A proposal for a revision of the guideline, supported by the United Kingdom, was prepared and submitted to OECD in November 2023. The Standard Project Submission Form (SPSF) proposed revisions across the following areas:

1. Conditions for sampling and storage of the inoculum prior to test initiation

2. Refinement of validity criteria for the degradation of existing reference substances and reflecting current practices of carrying out testing at lower temperatures

3. Inclusion of additional new reference substances to better assess inoculum viability and facilitate benchmarking of test substance degradation.

4. Recognition of the influence that permitted variations in experimental setup can have on the degradation of test substances.

The proposal was unanimously accepted onto the workplan of OECD, and various additional comments were received from member states. The nature of key comments was presented for consideration and discussion at the workshop.

2.2.3 Robustness of the OECD 309 – Reproducible holistic results

Dr. Marie Collard, dsm-firmenich, Belgium

Presentation: link

This presentation reiterated that decisions made based on the OECD 309 TG need to be robust and lead to scientific and evidence-based regulatory actions. It is therefore of critical importance that the test is designed such that results are reproducible. Evidence gathered as part of an ECETOC Task Force looking at OECD 309 suggests that the results can be highly variable according to the experimental design applied (but within the TG boundaries). It was recommended that the TG be redesigned to allow application of conditions that reflect all processes involved in degradation in the natural environment. Substances susceptible to indirect photolysis, volatilisation and hydrolysis were highlighted as situations where relevant processes can potentially lead to variability in outcomes.

2.2.4 Does changes in microbial community lead to divergent degradation patterns in the OECD 309 test?

Dr. Fola Ogungbemi, Currenta, Germany

Presentation: link

This presentation described experiences with OECD 309 testing of an unspecified test substance. This included pre-tests that showed high inter-replicate variability in tests conducted at 100 µg/L. In the main test, significantly lower degradation was observed at 10 µg/L than at 100 µg/L. In a repeat experiment, similar observations were observed at 10 µg/L, whereas high inter-replicate variability was observed at the 100 µg/L test concentration (similar to the pre-tests). It was suggested that variability in the microbial community development was the cause of variability between replicates, and that the higher degradation observed at the 100 µg/L test concentration was due to the increased amount of substrate available for the microbial

degraders. Proposed solutions were enhanced microbial monitoring (density and diversity), alternative reference substances, and higher numbers of replicates.

2.3 Pillar 2: Implementation

2.3.1 Keynote: Challenges in practical routine testing with OECD 309

Dr. Dieter Hennecke, Fraunhofer IME, Germany

Presentation: link

This presentation discussed practical experiences and issues with the OECD 309 as a routine standard test. The requirements for a standard test were listed as 1) guideline compatibility (acceptance, validity), 2) straightforward (producing clear results, conducted according to GLP), and 3) able to be run on a (limited) budget. Importantly, standard tests could not be considered as research projects. As consequence of these requirements, most labs have a specific setup of OECD 309 studies, with a limited number of options for modifications.

Regarding the validity criteria of the test concerning the inoculum, it was acknowledged that practically any surface water can be used, but it is a responsibility of the laboratory to get "the right one". The determination of the activity of the inoculum using a reference substance was very critical for the test validity, but the validity criteria themselves were unclear (degradation within "usually less than two weeks"). The actual results of reference compound degradation depended on several factors, with limited further guidance. Results are also only available after the test has been started, and the laboratory is responsible to bear the costs of repeating the test if the inoculum fails the validity criteria. Reference compound data were presented showing very different degradation kinetics depending on the setup (flow-through vs closed) and test temperature. CFU measurements were reported to sometimes be inconsistent with reference compound degradation, suggesting these may lead to uncertainties with evaluation of study validity. Incomplete recoveries and mass balance was also reported to be a problem with sodium benzoate.

The inclusion of suspended matter (SM) in OECD 309 tests was also an area of confusion. The OECD 309 test is divided between a pelagic (water only) test and a suspended sediment test, where 10 – 1000 mg/L additional sediment is added. ECHA guidance requires the pelagic test for persistence assessment and typically requires SM content of 10-20 mg/L. This is often greater than the SM content of surface waters, raising the question of whether additional SM should be added. Further challenges included accurately measuring the SM content and keeping SM in suspension, with magnetic stirring leading to grinding of the sediment (and stirrer bars!).

Test substance properties are an important consideration for the performance of the test. Poorly water-soluble substances present significant challenges for analytical quantification and test substance dosing. Use of a co-solvent is not straightforward, and poorly soluble substances are often observed to precipitate out of solution during application, or to not form stable solutions. The test substance may also interact with test vessel materials. Pre-tests are required and these can take significant time and cause issues with planning in

the laboratory. For test materials with a complex composition, such as substances of unknown or variable composition, complex reaction products, or biological materials (UVCBs), there are further significant challenges and laboratories often need to find individual solutions for specific cases (not consistent with the idea of "standard tests").

A further challenge is with the stability of the OECD 309 test system itself. The system sounds rather simple (e.g. water with very low matrix components, vessel setup, easy to control), but the reality can be very different. An example is with sterile controls. Some test substances have been observed to undergo greater transformation in sterile samples, or to degrade into different transformation products than in non-sterile samples. One potential cause is that autoclaving can lead to significant pH changes, with no buffer capacity in the test system. Overall limited guidance in the TG is available on sample sterilisation. Biological growth can also be observed in the test systems, with a gelatinous organic matrix being formed over time on the surface or suspended in the water. The extent of this growth can vary with test temperature and with use of co-solvents. The formation of biofilms presents challenges for extraction procedures and may contribute to non-extractable residue (NER) formation. It can also lead to oxygen depletion, particularly in closed systems. There is presently no guidance on how to handle NER in OECD 309. The amount of NER can be significant, but characterisation is technically not possible due to the very low amounts of solid phase (mg-range) in the system.

Maintaining mass balance for isotope labelled test substances is a further challenge. Highly sorptive, volatile and/or poorly soluble substances, in particular, are often difficult to recover. Small changes in the experimental system can have a large effect on mass balance. The reasons for this are not always clear but the responsibility remains with the laboratory for maintaining a mass balance. It is important to recognise that the required range for mass balance of 90 – 110% is not a validity criterion but is important for process understanding.

Finally, it was emphasised that the test guideline is referred to as a mineralisation test with an optional secondary objective to obtain information on primary degradation and the formation of transformation products. However, there had been no instance in the presenter's experience when a test for mineralisation only had been requested. In recent years the regulatory emphasis (e.g. REACH) has shifted from identifying readily or rapidly degradable chemicals toward potentially persistent chemicals. It is important to recognise that the OECD 309 test was originally designed for "degradation" rather than "persistence" determination. The presentation concluded that the OECD 309 guideline was being predominantly used outside of its original scope, with the exception becoming the rule in terms of the inclusion of specific analysis in the test. Most of the challenges described in this presentation can be traced back to this aspect.

2.3.2 Experiences in Assessing Persistence in Aquatic Environments

Prof. Jason Snape, University of York, United Kingdom

Presentation: link

In this presentation, the presenter's experiences with assessment of persistence using standardised and non-standard experimental systems were discussed. The range of standardised biodegradability test systems were

first presented with commentary on their microbial and ecological relevance and the relevance of measured degradation rates. The presenter considered that OECD 301 and 310 ready biodegradability tests were of poor relevance, whereas inherent biodegradability tests provided better screens for activated sludge. In terms of simulation tests, the relevance of wastewater treatment (OECD 303 and 314 A-E) and soil (OECD 307) tests were generally considered to be good, whereas the water-sediment (OECD 308) and surface water (OECD 309) simulation tests were considered to be of low relevance. It was reminded that carbon dioxide is actually an inefficient by-product of microbial metabolism, and is higher in carbon-rich, growth-rich environments.

Non-standard systems were described, including systems with added sediment, extended test durations, purging into liquid or headspace, nutrient addition, varying vessel sizes, and semi-continuous operation. Enhanced biomass studies from the Cefic-LRI ECO11 project were also described. Various inoculum concentration techniques have been investigated including centrifugation, tangential flow filtration, membrane filtration and column colonisation. Data were presented indicating that the probability of biodegradation was directly related to cell concentrations across a range of inoculum sources.

According to the presenter, the OECD 309 TG was regarded as just another ready biodegradation test. It is not to be considered a 'higher tier' simulation study, has little ecological relevance and a high false negative rate. It was argued that the test suffers from a loss of microbial viability/diversity and that the duration is far too short. Proposed improvements were the inclusion of sediment, concentration of biomass to improve predictability and reproducibility, or semi-continuous operations, which were able to demonstrate adaptation and maintain microbial diversity. Microbiologically relevant test systems were identified as a critical research need. Validation of semi-continuous approaches and field-based measurements were suggested as paths forward. It was suggested that there is a need for new laboratory tools that are truly predictive of persistence, rather than tinkering with the existing OECD standard test methods (and quantitative structure-activity relationships (QSARs) that are built on the data from these tests).

2.3.3 Hydrophobic and volatile chemicals in OECD 309 tests – challenges and required test modifications

Dr. Heidi Birch, Technical University of Denmark, Denmark

Presentation: link

This presentation discussed a modified test system for measuring primary degradation of hydrophobic and volatile test substances in surface water. There are various modifications to the test design to address the challenge of evaporative and sorptive losses of these substances during test setup and incubation. The modifications included a closed system with sufficient headspace, no plastic parts in contact with the test substances, and aqueous solutions handled with gas-tight syringes. For the analytical method, close alignment between testing and analysis is employed to minimise losses during sample treatment. Sampling, enrichment and analysis is conducted directly on unopened vials using automated solid phase microextraction (SPME) coupled to gas chromatography mass spectrometry (GC-MS). No manual extraction or evaporative enrichment is performed. Primary biodegradation kinetics are determined from peak area ratios between biotic and abiotic systems, accounting for losses of hydrophobic and volatile test substances. Test system dosing by either microvolume spiking or passive dosing can be utilised to avoid solubility limitations and toxicity. Microvolume

spiking allows quantitative transfer of test substances with minimised solvent volumes (limiting oxygen depletion). Passive dosing is applicable for more hydrophobic test substances (log K_{ow} > 3) and controls for composition and ensures dosing below solubility. The details of the modified test system are published in the journal MethodsX (Birch et al., 2023).

2.3.4 How to determine the persistence of UVCBs – whole UVCB testing combined with constituent specific analysis

Prof. Philipp Mayer, Technical University of Denmark, Denmark

Presentation: link

The compositional complexity (i.e. the large number of known and unknown constituents) of UVCB substances was identified as a source of testing challenges for these substances. This leads to issues with screening testing, where non-specific parameters (CO_2, O_2, DOC) cannot discriminate constituents, and with simulation testing, where ^{14}C labelling of UVCBs is often not possible and also cannot discriminate constituents (due to sensitivity limitations). Most UVCBs also contain hydrophobic and/or volatile constituents which adds further challenges. Simulation testing of UVCBs typically involves the testing of a representative constituent, but this leaves uncertainty around the degradation of all other constituents. A proposed new approach was presented which combined whole UVCB testing with constituent-specific analysis.

An experimental system similar to the one described in the previous presentation was utilised for the UVCB substance testing. This approach was able to test, analyse and determine the biodegradation kinetics of both known and unknown UVCB constituents. Fast biodegradation of the main constituents of lavender oil in stream water was observed using the approach. The abiotic degradation of linalyl acetate (38.9% of lavender oil composition) was also quantified. Biodegradation kinetics of unknown constituents in lavender oil and black pepper were also determined, with all constituents being found to be degraded. The approach can be used to direct the identification and assessment of the most persistent constituents of a UVCB substance. The approach has also been used for determining biodegradation kinetics of chemicals in produced water discharged from offshore oil platforms into the marine environment.

2.4 Pillar 3: Relevance

2.4.1 Keynote: Environmental relevance of the OECD 309 test

Prof. Michael McLachlan, Stockholm University, Sweden

Presentation: link

This presentation discussed the environmental relevance of OECD 309 by working through a series of decision points. The presenter began by outlining the scientific ambition of the research group, which was to determine rates of primary degradation of substances occurring *in the environment*. Their interest was in the environmental relevance of OECD 309 and the focus was on hydrophilic substances. It was postulated that the

semi-logarithmic biodegradation kinetics described in the OECD 309 test guideline, and often observed in experiments conducted at higher test concentrations, were not environmentally relevant. Decision Point #1 suggested that assessments should rather be based on the initial attenuation kinetics (i.e. the kinetics observed before any adaptation of the microbial community), which were potentially more environmentally relevant. The adapted biodegradation kinetics were potentially still useful, but should not be regarded as a simulation test.

The sensitivity of the pelagic test was then discussed. In experiments assessing the biodegradation of 64 compounds (pharmaceuticals, pesticides, etc) using water from a Swedish river downstream of a wastewater treatment plant (WWTP), rate constants were measurable for only seven out of 64 compounds. Of these, only four had half-lives below the persistence threshold of 40 days. The influence of adding suspended sediment was also explored, with data presented suggesting a linear correlation between the degradation rate and the concentration of sediment added. This led to Decision Point #2 or whether to add sediment in OECD 309 tests. It was suggested that OECD 309 tests without the addition of sediment would find most chemicals to be persistent and would have no discriminatory power. They would therefore not be useful in prioritising chemicals and would likely be redundant vis-à-vis the ready biodegradability test. As an aside, it was suggested that most chemicals would likely be found persistent in the pelagic test because most chemicals *are* persistent in water alone, since the microbial density is simply too low. This would imply that most chemicals would eventually be subject to Persistent, bioaccumulative and toxic/Persistent, mobile and toxic (PBT/PMT) regulation. The question was posed whether this was really what was desired from regulatory policy, and that perhaps persistence should rather be regulated in a different way, such as at the level of aquatic systems. Decision Point #2 concluded that to include additional sediment would result in better discriminatory power but would lead to potential regulatory misalignment since persistence in water was no-longer being measured. A question was posed whether a water-sediment mixture could instead be considered a surrogate for aquatic systems.

Decision Point #3 was then how much sediment to use in such a water-sediment mixture. A fixed amount would support standardisation and comparability of microbial activity across sites, but not comparability of half-lives. Alternatively, site specific amounts could be used, which would be more environmentally relevant and support comparability of half-lives. However, it was unclear how to choose the amount of sediment and whether this would be standardisable.

Data from various studies were presented. An investigation into the effect of spiking surface waters with chemicals on the degradation rates of other chemicals suggested that testing multiple chemicals in the same experiment was possible. An investigation into the influence of the storage time of added sediment on the degradation rates of chemicals suggested only small differences between storage times of 5 and 24 hours. In an investigation comparing degradation rates upstream and downstream of WWTPs, intended to elucidate the influence of adaptation, no systematic influence of exposure to contaminants on k was found. An investigation into spatial-temporal variability of degradation rates was conducted using 129 mostly highly water-soluble chemicals and a total of 38 experiments across Europe and Australia. Degradation rates were found to be highly variable across sites, with 95% of data falling into the range of four times the standard deviation (for comparison, 0.5 log units of standard deviation is equivalent to a range of two orders of magnitude). Hierarchical clustering of Pearson correlation coefficients of log k was conducted to identify the potential to use benchmarking for subsets of chemicals. In exploring the possibility to reduce variability in k

using normalisation or benchmarking, it was observed that the standard deviation of log k was about four times larger than the uncertainty in measurements. Normalisation to total organic carbon (TOC) was found to increase, rather than reduce, variability. Benchmarking to a best-performing universal benchmark was found to have a negligible effect. Optimised group-based benchmarking using best-performing benchmarks in each group was found to reduce the variability for 87% of compounds, but the variability remained large.

Decision Point #4 concerned which sediment to use in OECD 309 testing. A fixed sediment may not be practical for storage and transport, and would still be subject to high seasonal variability. There would also be the question of which site would be most relevant. A flexible sediment would likely lead to large variability in test outcomes, with no methods yet available to correct for this variability.

It was concluded that OECD 309 is currently not an environmentally relevant test and that the use of initial kinetics and the addition of sediment are required to make OECD 309 potentially relevant and useful. It was argued that the persistence criterion for half-life in water (only) is of limited use for prioritising chemicals for regulation, and that the very large variability of biodegradation rates in the environment is a fundamental obstacle for using a simulation test to assess a regulatory threshold.

2.4.2 Relevance of the OECD 309 – Consistency with screening data

Dr. Marie Collard, dsm-firmenich, Belgium

Presentation: link

In this presentation, preliminary results of an ongoing ECETOC Task Force were presented (title: Building knowledge from available degradation simulation studies: analysis of the most influential factors & development of a grouping approach). It was stated that OECD 309 is a higher-tier biodegradation study used to refine the persistence assessment, e.g. when a substance does not pass a ready biodegradation test. This implies that ready biodegradation tests are more stringent than OECD 309 and that there should possibly be a relationship between amounts of biodegradation observed in both tests.

Analysis of preliminary data comparing biodegradation (%) in OECD 301 with the percentage of applied radioactivity as CO_2 in OECD 309 found that, for substances showing low degradation, degradation was mostly very limited in OECD 309 compared to OECD 301. For only a limited number of cases was the degradation higher in OECD 309. It was suggested that the results of OECD 309 do not represent the potential for degradation observed in a screening test, and therefore that the design of OECD 309 is inadequate to further characterise the biodegradation kinetics. When comparing biodegradation in screening tests with dissipation (primary degradation) half-lives in OECD 309, the results of OECD 301 had only limited predictivity of the OECD 309 result. It was suggested that other processes than biodegradation may need to be taken into consideration to improve the predictivity, and the question was posed whether the current OECD 309 test was too stringent.

2.4.3 Definition, Impact, and Relevance of Diffuse Light for OECD 309 Tests

Dr. Martin Brüggemann, Bayer, Germany

Presentation: link

This presentation detailed an investigation into the influence and relevance of light in OECD 309 testing. In parallel OECD 309 experiments using pelagic (PE), suspended sediment (SE), suspended sediment with diffuse light (LI) setups to assess the metabolism of two substances, it was observed through visual inspection that the SE and LI setups showed clear growth of microorganisms over 62 days, whereas the PE setup remained clear. Both sediment and light were observed to enhance the metabolism of the two substances. The presence of light was particularly important for degradation of the observed metabolites. The observations were regarded to be in agreement with other published studies.

It was reported that, although the application of diffuse light was an option in OECD 309, there was a lack of agreement and guidance regarding the intensity. Hand and Moreland (2014) proposed to use lower limit and midpoint of the euphotic zone water bodies, i.e. 1% and 10% of the incident Photosynthetically Active Radiation (PAR). Long-term turbidity data for lakes in EU were presented, suggesting that the majority of euphotic zone depths were between ~ 5 and 25 m.

Average PAR intensities were obtained from remote sensing data for EU FOCUS water locations. Based on these, a PAR intensity of ~6.6 W/m^2 (10 % of the geomean PAR, corresponding to the midpoint of the euphotic zone) was suggested as a representative worst-case scenario for diffuse light conditions in OECD 309 tests. For comparison purposes, ~7 W/m^2 corresponds to the light intensity (PAR) on a winter day (end of December) at 9:00 am in Brussels. The proposed PAR intensity may need to be adjusted to account for light:dark cycles.

UV-vis absorption coefficients for relevant wavelengths were determined for several water sources across Germany. Results were extrapolated to estimate light extinction for the upper epilimnion layer (3 m depth). Geomean light intensity at this depth across all sites was calculated to be 27 W/m^2.

3. BREAKOUT SESSIONS

3.1 Introduction to the breakout sessions

The in-person workshop participants were divided into three pre-assigned breakout groups to discuss the three pillars: reliability, implementation and relevance. The breakout groups were given the task, for their respective pillars, of identifying challenges with the current OECD 309 test guideline, and discussing the corresponding solutions, relevant research (previous and ongoing), and proposed next steps for OECD implementation.

In preparation for the workshop, registered participants were invited to provide input to pre-populate a template that would be use as a basis for facilitating and recording the discussions of the breakout groups (see Appendix D). The template was a table with the following headers:

- Challenge
- Impact of challenge
- Solution
- Relevant past research
- Relevant ongoing research
- Next steps for OECD Implementation
- Timeline (immediate, 1-3 yrs, >3 yrs)
- Priority

The breakout discussions were run in two 90-minute sessions. Each breakout group was assigned a moderator and a rapporteur. Members of each of the breakout groups are listed in Appendix A. The breakout groups reported back on their discussions to the rest of the workshop participants in a plenary session.

A summary of the discussions within each of the breakout groups is presented below, along with the main findings as reported back in the plenary session. Further details are also provided in Appendix D.

3.2 Breakout session 1: Robustness pillar

Robustness: Reproducibility of the test guideline and impact of permitted experimental conditions

Moderator Chris Hughes, rapporteur: Katherine Santizo

The robustness pillar of the workshop was principally concerned with the reproducibility of the OECD 309 test guideline and the extent to which permitted variations in the experimental setup and other aspects of execution of the guideline could lead to differences in results. An important consideration for the robustness of a test guideline, and its reliability for the testing and assessment of chemicals, is the degree of variability of results. Inter-replicate variability within individual tests should be minimal in order to produce statistically reliable results. Similarly, inter-test variability should be as low as possible so that results from individual tests are able to be compared, and the sources of any such variability should be understood and mitigated, to the extent possible.

In addition to the fact that OECD 309 utilises natural surface water as the inoculum, which represents an important potential source of inter-test variability, the guideline also contains a number of elements of flexibility concerning its execution. These permitted variations may introduce variability in the test outcomes. The discussions within this breakout group centered on these aspects and considered the information submitted in advance of the workshop in addition to that which had been presented and discussed during the workshop.

The breakout group followed a structured approach to the discussions, whereby the information submitted in advance of the workshop was first briefly reviewed. Then, the group brainstormed additional challenges and issues to capture based on the information presented at the workshop, as well as from their own experience. Once the full list of ideas had been collated, the group then discussed each of these individually in more detail. Following this, the complete list was again reviewed to determine whether any further additions were needed. Consideration was given to potential overlap of the issues with the other pillars of the workshop. Finally, the list was prioritized to identify the most pressing issues concerning the robustness of the guideline.

A constructive discussion was held between breakout group participants. It was generally recognized that the current version of the OECD 309 test guideline has several issues concerning its robustness. In order to be considered robust, the test needs to be reproducible. Reproducibility could be defined in multiple ways, e.g. within the test (inter-replicate variability), and between tests (inter-test variability), with different tests having various ways that they could be similar, or differ, from one another in terms of their experimental methods, conditions, inoculum sources, etc. The variability in the test results for OECD 309 is believed to arise principally due to inoculum/matrix differences and the experimental setup and conditions used (provided the substance doesn't have properties that make it difficult to test – see the Implementation Pillar).

Addressing inter-replicate variability was considered crucial in order for the guideline to deliver results that could be considered reliable. Evidence presented during the workshop suggested that the currently used pelagic test did not deliver this due to the low biomass concentrations in these tests, which led to inconsistent degradation observations between replicates. Evidence had been presented that seemed to suggest this was improved by increasing biomass concentrations in the test through the addition of suspended sediment. The use of tangential flow filtration methods to increase cell concentrations has also shown promise. Although

there had been limited experience of its application within OECD 309, there has been experience with this for biodegradation testing in seawater and a new test guideline was being developed under the OECD Test Guidelines Programme.

Regarding inter-test variability, it was considered that some of this variability was expected to arise due to natural differences between the inoculum, and that efforts to increase test reproducibility should not compromise relevance. In other words, care should be taken to not modify the nature of the test, that is to simulate the degradation of the test substance within natural surface waters. This did, however, raise questions as to the utility of the test as a prioritization tool for assessing (for example) the persistence of substances in the regulatory context, if there was so much variability arising as a result of the inoculum.

It was also recognized that individual permitted variations of the test setup were likely to lead to differences in results, and that this reduced comparability of the results between tests and was generally undesirable. It was therefore important to better understand how the permitted variations in performance of the test contributed to variation in study outcomes and to reduce these to the extent possible. One suggestion was to consider constraining some of the experimental conditions currently permitted, such as defining a default or standard set of experimental conditions, procedures and apparatus. Test conditions to be considered here could include:

- Test substance concentration
- Test volume and sampling method (sacrificial vs sequential)
- Open (flow-through)/closed systems
- Shaken/stirred systems
- Darkness/diffuse light
- Concentration of suspended sediment (if included)

Test conditions could potentially be modified to meet the needs of different test substance properties, such that different variants of OECD 309 could correspond with different test substance profiles, similar to the approach in OECD 301. It was also suggested that it was an option for these modifications to be specified in separate guidance documents, rather than within the test guideline itself.

Whilst increasing the biomass concentrations employed in OECD 309 tests was considered an imperative to address challenges with robustness by some breakout group participants, there was a concern voiced that such an approach could lead to the degradation capacity of the inoculum to be increased beyond that which would be observed in the environment. It was suggested that this ‘boosting’ of the test would impact the relevance of the test and render results not directly comparable to regulatory persistence criteria. The use of high amounts of suspended sediment could also potentially introduce additional challenges for interpretation, such as high NER formation, which is difficult to deal with in OECD 309 tests. It was therefore considered that no agreement could be reached on the need to increase biomass concentrations to improve the robustness of the OECD 309 test, as it is currently performed.

Another issue that was discussed was the verification of inoculum activity in the OECD 309 test. At present, the guideline utilizes reference substances to demonstrate this, with mineralisation of sodium benzoate or aniline being measured in parallel to the degradation of the test substance. However, it was questioned whether this approach made sense in the context of OECD 309, or whether other techniques were available that could provide more useful information on inoculum viability. Further, the work of the Cefic-LRI ECO55 project suggested that there was wide variation in the performance of the current reference compound aniline between test laboratories. It was considered that direct measurements of inoculum cell count, or other indications of activity/functional capacity, could be considered instead. Suggestions of potentially suitable techniques included microbial community characterization (DNA and RNA analysis), Adenosine Triphosphate (ATP) measurements, flow cytometry and plate counts. It was generally considered that such techniques would provide more useful and relevant information within OECD 309 tests. It was felt that a relevant biomass range could be specified that could improve standardization of the test. However, further investigation would be needed into the performance, robustness, practicalities and relative costs of these techniques before they could be introduced as alternatives to the current reference compounds. It was also felt that removal of the validity criteria surrounding the degradation of reference compounds would introduce challenges for the assessment of test validity and complicate the interpretation of results.

Although principally concerned with relevance, the test substance concentration was discussed as an important potential source of variability. Evidence presented at the workshop suggested that test concentrations should be as low as possible in order to observe co-metabolic transformation of test substances and avoid inoculum adaptation leading to metabolic biotransformation (growth-linked kinetics). It was discussed that this desire needs to be balanced with what is practically achievable, especially when using ^{14}C-labelled test substances, which is generally a requirement for regulatory purposes. It was also recognized that higher test concentrations may be needed to measure the formation of transformation products.

The sampling and storage of the inoculum in OECD 309 was discussed and recognized as an important potential source of variability in test results. The work of the Cefic-LRI ECO55 project and elsewhere has demonstrated that microbial communities will shift after having been sampled, and that this can impact the performance of the inoculum in biodegradation testing. It was agreed that it was necessary to minimize the time between inoculum sampling and the initiation of testing under OECD 309, and that the recommendation of a maximum 7-day storage period from the ECO55 project should be considered for the update to the guideline. It was also mentioned that including suspended sediments with water samples could be a means of maintaining microbial communities stable and keeping systems ‘alive’ for longer periods.

In the case of abiotic controls used in OECD 309, it was discussed that several sterilisation techniques are available and that recent experience should be captured to ensure that the guideline makes appropriate recommendations in this area (e.g. to ensure effective sterilization and to avoid artefacts). It was discussed that reporting requirements should be reviewed to ensure that relevant details captured and reported are sufficient to adequately evaluated and interpret OECD 309 studies for regulatory use. The importance of pre-tests was also discussed for establishing aspects such as test concentrations, sampling time intervals and an appropriate analytical method. These can support the effective design and execution of OECD 309 tests, which can ultimately improve the reliability and robustness of the resulting data.

Updated Table from Breakout presentation

CHALLENGES	SOLUTIONS	NEXT STEPS FOR OECD IMPLEMENTATION
Microbial biomass as a source of variability: Low biomass concentrations lead to high inter-replicate variability, particularly in pelagic tests Microbial inoculum poorly standardized Validity criteria provide limited verification or useful information	Include microbial cell counts and/or activity measurements in the guideline and provide an acceptable range (min. threshold) Consider including suspended sediments to increase biomass concentrations Revisit the need for reference substances for validity criteria	Investigate techniques for microbial biomass/activity measurement Suspended sediments require additional research and consideration on concentration ranges Analyze LRI ECO55 results on reference substances and provide recommendations
Current biomass-to-chemical ratios do not support co-metabolic biotransformation: Growth-linked degradation kinetics common	Use a lower test concentration. As low as practically possible (consideration for analytics) Higher concentrations may be required for metabolite quantification Consider increasing biomass concentration (e.g. using suspended sediment)	To discuss in OECD Expert Group

CHALLENGES	SOLUTIONS	NEXT STEPS FOR OECD IMPLEMENTATION
Inoculum sampling and storage impacts test outcomes: Microbial community changes over time Activity and biodegradation performance impacted	Provide clearer instructions and sampling and storage Limit storage duration to 7 days (ECO55 recommendation)	Discussions for how storage conditions and stringent language would improve robustness
Variation in experimental setups/conditions impact test outcomes Various permitted setups Difficult substance properties require modifications	Constrain experimental conditions that can be used Identify a standard set and recommendations on where to apply amendments based on substance chemistries	Compile studies that address this aspect to build on standardization and sources of variability

3.3 Breakout session 2: Implementation pillar

Implementation: Experiences with running the test guideline for a range of substances and/or physicochemical properties

Moderator: Dieter Hennecke, rapporteur: Louise Camenzuli

The second breakout group dealt with the aspects of the practical application of OECD 309 in the test routine, also referred to as implementation pillar. The basis for the discussions was a pre-defined list of different practical challenges when conducting the OECD 309 test with the current test substances requiring regulatory testing. The aim was to determine whether and, if so, how the challenges can best be addressed and what effort(s), both in terms of time and scientific knowledge, will be required to implement the potential solution in a revised OECD 309. The fundamental prerequisite for the suitability of the developed solutions was the applicability in routine testing. However, some of the challenges listed were also considered by the group to be irrelevant or already solved. The composition of this breakout group consisted of scientists from both regulation and CROs, complemented by scientists from industry, research institutes and academia, ensured that the challenges considered were examined from various relevant perspectives.

Challenges related to the implementation focussed largely on the fact that the OECD 309 TG is currently being used on a large variety of test substances, with a range of physico-chemical properties, very different from the test chemicals on which the guideline was developed. This was apparent in the discussions around applicability domain, dosing of difficult-to-test substances, and the need to establish definitions of aerobic conditions. For example, it seems that the closed test set up that is described in the guideline, and the advice associated with it, were not part of the testing during the establishment of the guideline. Another example is the testing of UVCBs, which has various challenges even if the guideline would describe a straightforward testing procedure due to the general challenges of testing UVCB-substances. In the following section, the most important discussions are summarized.

Applicability domain: this was by far the most discussed point, even though opinions on it did not differ greatly in the group. Many of the reported difficulties in 309 tests are based on conducting studies with substances that are out of the Applicability Domain of the TG 309. Section 7 of the TG defines the applicability of the TG, however, does not define specific limits beyond the Henry's law constant (H). The TG does not define clear physico-chemical limits for which the guideline is applicable, leading to the test being regularly used on substances which require significant method development. Results received under these "newly developed" conditions are no longer necessarily comparable to results under the conditions for which the TG was developed, and such studies can certainly not be concluded as "standard" studies. This could be addressed through the definition of an AD for which the OECD 309 has been developed and is hence suitable for. It would be of value to ask for ranges of Log K_{aw} (replacing H), water solubility, pKa, Log P_{ow} and k_d, because this helps establish validity criteria for analysis and nominal concentrations at environmentally relevant pH. It was recognised that modifications/testing options necessary for substances that go beyond these ranges should also be established.

It was acknowledged that the TG has already been used to assess substances that are likely to be considered outside of a future AD. It is expected that knowledge relating to a working AD, for example accounting for substance physico-chemical parameters that require practical modifications for testing, is already applied in practice although unpublished in Contract Research Organisations (CROs). Furthermore, additional modifications are likely to be available in scientific literature. Thus, beside a targeted literature review to collate available information on AD of the TG - e.g. the OECD Guidance Document 23 on "Aquatic Toxicity Testing of Difficult Substances and Mixtures" (OECD, 2019) – it was recommended to send out questionnaire or interviews to CROs to collate available first-hand knowledge in identifying potential strategies for testing substances with properties beyond the AD.

As a sub-aspect the applicability of OECD 309 to UVCBs, multi-constituent substances (MCS) and mixtures. It was recognised that TG 309 is not relevant to evaluate the biodegradation of these substances, and hence guidance on how to test these substances would be valuable. There is a significant body of recent research that has focused on developing testing methods for UVCBs, MCS, and mixtures, particularly petroleum hydrocarbons and essential oils. These methods are designed to analytically follow the primary degradation of the whole substances' constituents. So, alternative methods exist and have been published to evaluate the biodegradation of UVCBs and MCSs. However, they have not been evaluated within the limits of the OECD TG 309. Hence, the group recommend that the newer methods are validated and, for better acceptance, are included in TG 309.

The use of ^{14}C-radiolabelled test substance was discussed as well briefly for the UVCBs. It was concluded that ^{14}C-labelling of UVCB will in most cases be impossible and even if, it would not solve the major issues of testing UVCBs.

Going along with AD definition is also dosing methods of the test substances. Par. 22 of the OECD 309 TG would need to be refined, in line with the applicability domain challenge, to clarify the applicability of currently listed dosing methods and future options. Best practices could be identified to ensure consistency across testing laboratories.

The next steps identified and recommended by the group under the topic Applicability Domain are:

1. A literature review could be performed to collate available information on dosing methods for difficult-to-test substances. The OECD 23 Guidance document was also discussed as a source of information that could support definitions of applicability domain.
2. A questionnaire or interviews could be sent out to CROs to collate available first-hand knowledge on methods used for dosing difficult-to-test substances.
3. It was recognised that regulators and academics would also have knowledge to this capacity and should be included in above questionnaire / interviews.
4. A validation of these methods for suitability of testing within an OECD 309 TG would be required.

Based on the information gathered in the steps listed above, a guidance document similar to OECD Guidance Document 23 can be developed, to allow for e.g. a standardisation of dosing methods used across different tests. It was discussed that the inclusion of any new technologies would need to be sufficiently described and validated.

Data reporting: a challenge that has been flagged is that there is a high variation of the reporting detail from TG309. More detailed instructions and requirements on measuring and reporting e.g., the mass balance, are needed. The solution that the group identified was the development of an OECD harmonised template, which could be included as an annex to the TG. The reporting, currently in the form of a list, could be developed into a table format, allowing for easier utility. The table could then identify parameters which must be reported. The group was informed that The OECD Working Party for Hazard Assessment (WPHA) are currently developing a guidance document on the regulatory use of research data. The first draft of the OECD GUIDANCE DOCUMENT ON THE REGULATORY USE OF RESEARCH DATA was reviewed in August 2024. This is not published yet but should be coming through the WPHA.

Aerobic conditions: only a challenge for testing in a closed system, since the open system can be assumed to be sufficiently aerated. The solution that the group identified was that the term “aerobic” should be defined, through the establishment of relevant oxygen concentrations or redox potentials.
Hence, as a first step before defining any thresholds, the group recommends that a discussion needs to be had, on which is the most relevant metric to be used. CRO involvement is recommended, since it is expected that they would have historical start and end data across a variety of test substances. It would also be important to ask the CROs how the different measurements are made and identify the different instrumentation.

However, there was quite some discussion, with no agreed upon conclusion, whether such criteria would be included as validation criteria. The group also recommended that this solution would require recommendations on how to measure dissolved oxygen concentrations or redox potential, to ensure consistency.

Number of samplings: Par. 30 of the OECD 309 TG states that "usually a minimum of five sampling points is required during the degradation phase". However, especially for slowly degrading substances, the selection of five sampling points is difficult to establish. Additionally, it is recognised that 5 sampling points can be a limitation in the statistical analysis. For slowly degradation substances, this challenge was concluded to have high impact.
The solution identified was a re-write of the relevant paragraph in the TG, with the recommendation for the inclusion of preliminary work. As such, data from pre-testing could be used to establish the sampling points, and the number of replicates needed to optimise the derivation of the kinetics. Inclusion of specific kinetic models in the guideline was refused since these models develop further independently from the OECD TGs. The group also discussed the possibility to evaluate different sampling options based on the purpose of the test. E.g. if the goal is to evaluate the disappearance of the parent, or to identify metabolites, the sampling options may vary.

Sterile and solvent controls: the OECD 309 TG suggests the inclusion of sterile controls, however, provides no information as to how to prepare a sterile control. For the sake of consistency, it was recommended that the guideline should mandate the inclusion of sterile controls, with a clear rationale/guidance around the utility of such controls. A new section should be included in the guidance which would include methods available for setting up sterile controls. The group discussed that information on sterile controls likely is available with CROs, hence the necessary step to develop the required guidance is through discussions with CROs.

Collection of surface water: Par 18 of the TG 309 provides guidance around the collection and transport of surface water. The guidance specifically states that *"if it is known that the aquatic environment has been contaminated with the test substance or its structural analogues within the previous four years, it should not be used."*. It is recognised that this text is difficult to justify for industrial chemicals in surface water. The group discussed that the text highlighted under this challenge is likely legacy text from plant protect product assessment. The relevance of the text in the OECD 309 TG vs the OECD 307 TG needs to be examined and should potentially be revised to align with the overall goal of the TG.

The estimated time needed to develop respective solutions as described above was discussed in the group. It was concluded that not more than 3 years should be needed for any of the solutions and for most even 1 year should be a realistic estimate.

Updated Table from Breakout presentation

CHALLENGES	SOLUTIONS	NEXT STEPS FOR OECD IMPLEMENTATION
Limited Applicability domain (AD): • Lack of physico-chemical limits • Applicability to UVCBs, MCS • Dosing methods	1. Define clear physico-chemical limits for which the OECD TG 309 was developed 2. Update OECD TG 309 with new AD limits 3. Develop Guidance document (GD) for chemicals outside of AD to improve consistence across testing laboratories	1. Questionnaire / interviews with CROs, regulators, academics, potentially through the OECD to leverage knowledge and methods available 2. Validation of methods outside of domain 3. Develop guidance
Data reporting: • High variation in reporting of tests • Key parameters not consistently reported	1. Development of an OECD harmonised template for test reporting 2. Identifying parameters that must be reported for study evaluation	1. Discussion with relevant OECD WG to understand process 2. Develop harmonised template 3. Decide on accessibility of template
Aerobic conditions definition: • No clear definition (DOC / redox) • Issues with closed system	1. Define relevant values for closed systems 2. Include recommendations on measurements	1. Discussion on relevant metric: mg/L or % saturation, or if equivalent exists 2. Discussions with CRO for relevant thresholds

CHALLENGES	SOLUTIONS	NEXT STEPS FOR OECD IMPLEMENTATION
Number of sampling intervals: • Minimum of 5 sampling points during deg phase • Can be difficult to establish	1. Re-write relevant paragraphs in TG 2. Recommendation: inclusion of prelim work 3. Potential to evaluate different sampling options based on purpose of test	1. Check existing guidelines for information 2. Discussions with CROs to capture learnings
Lack of information on controls (sterile and solvent controls): • No guidance on how to set up • Limited guidance on when to use • Limited guidance purpose of controls	1. Mandate when controls are needed in TG 2. Develop guidance to assess validity of controls (specifically for sterile controls) 3. Describe suitable methods in TG	1. Discussion with CROs on methods and best practices 2. Compilation of state of knowledge 3. Develop guidance
Collection of surface water: • No contamination in previous 4 years • Potentially difficult for industrial chemicals • Balance between "representative" and "conservative" water sources	1. Text likely legacy text from pesticide use, and could be revised to align with goal 2. Potential for representative inoculum	Not discussed

3.4 Breakout session 3: Relevance pillar

Relevance – Extent to which the test guideline reflects what occurs in the environment

Moderator: Miriam Leon Paumen, rapporteur: Delina Lyon

The third breakout group was Pillar 3 – Relevance, which focused on the extent to which the OECD 309 test reflects the potential biodegradability of test substances in European surface waters. The breakout group reviewed the list of challenges which were collated via survey before and during the workshop (see Appendix D) and ranked them according to impact of the challenge on relevance. Many of the Relevance challenges overlapped with issues identified in the other pillars, for example, on the test substance concentration, inclusion of sediment, and handling of the inoculum. In theory, the OECD 309 simulation test is relevant if it mimics what might happen in the specific surface water from which the inoculum is taken. The biodegradation half-lives or the degradation products observed would be specific for that surface water. As such, some participants suggested that no modifications should be made that would alter the test system to improve or reduce the inherent biodegradation capacity of that surface water, as these would lower the relevance of the test while other participants and presenters questioned the relevance of the test in the first place. The presentations did not support the hypothesis that the OECD 309-derived half-lives/rate constants would be similar to what occurs in the surface water since the sampled water changes over the testing period. The presenters showed clear evidence that the test is not directly relevant (does not reflect how the chemical was being degraded in the surface water body when it was sampled). It was also discussed that the OECD 309 test guideline was not written with the specific regulatory applications for which it is currently being used. A number of the presenters and attendees expressed frustration that the guideline was not environmentally relevant; other attendees felt that the test outcome provided relevant data for regulatory purposes. While no consensus was obtained as to how to remedy this incongruity, there were still many suggestions as detailed below on how to improve the guidance.

The OECD 309 test guideline offers multiple possible endpoints, i.e., mineralisation, primary biodegradation half-life, or identification of degradation products. The selection of the endpoint of interest greatly impacts how the test would be set up. For example, a radiolabelled test substance is crucial for determining mineralisation but may not be as important to measure the primary transformation half-life. Consequently, for determining primary transformation half-life, one could consider running the OECD 309 with a cold test substance, with expected analytical recovery between 70-110%, rather than use a radiolabelled substance, with an expected recovery of 90-110%. Additionally, the test substance concentration may need to be increased if one is looking for degradation products. One sweeping suggestion from the breakout group would be to divide the OECD 309 test guideline according to the endpoint of interest so more appropriate test set-up recommendations could be made.

There were also a number of overall challenges for the relevance of the test, which were rated as high/medium; they have been grouped according to topic and discussed below.

Inoculum handling artefacts: Since the microbial community of the inoculum drifts quickly from its initial composition during storage (Hakvåg et al., 2024), it could be recommended to use the inoculum within 24 h (OECD 301 gives a max of 5-7 days). The temperature at which the inoculum is stored, and the test is conducted

may also reduce the relevance. The guideline does not require sampling the inoculum at a temperature close to the test temperature, while it may be that bacteria colonies adapted at 12°C are not the same as at 20°C (psychrophile vs mesophile).

Test substance concentrations: As mentioned above, the choice of test substance concentration will need to be adjusted depending on the objective of the test. However, more generally, since environmental concentrations of most test substances would be expected to be low, any biodegradation processes would likely occur due to co-metabolism rather than active microbial processes. Thus, the test substance concentration in a simulation test should remain low to allow for the expected types of transformation processes. Presently, the concentration range given in OECD TG 309 is 100 µg/L and preferably in the range of <1-10 µg/L and thus is not sufficiently restrictive to ensure environmentally relevant concentrations. However, testing at concentrations as low as possible will depend on the analytical feasibility and may restrict the ability to use radiolabelled substances. There may also be increased variability of the outcome. As discussed above, identification of degradation products would likely not be feasible if the test substance concentrations are too low. It should also be noted that the degradation products of co-metabolic processes may differ from those at higher test concentrations, which trigger specific biodegradation pathways.

Choice of surface water for the inoculum: The OECD 309 test guideline only mentions testing one surface water at two different test substance concentrations. If the objective of the test is to derive biodegradation half-lives/rate constants for European surface waters which will then be used to conclude on the persistence of the test substance, the choice of the surface water takes on a greater significance. However, during the presentations, it was demonstrated that site-to-site variability in half-life/rate constants is very high in tests with sediment/water mixtures, and reproducibility of the pelagic test outcomes is poor. The issue of possible site-to-site variability in biodegradation rate constants was discussed, with suggestions to add more surface water samples or to have better guidance on how to select more representative samples. In the end, the group did not have sufficient information to formulate recommendations for testing strategies.

Inclusion of sediment: While there were presentations showing that the inclusion of sediment into the OECD309 test system improved the reproducibility of testing outcomes, some of the breakout group members clarified that such a test would no longer be considered a simulation of surface water.

Presence of light: It was suggested that clearer instruction regarding the lighting conditions during the testing should be included. In the presentations, it was shown for one substance that inclusion of even diffuse light increased the rate of degradation. There were suggestions to use the euphotic zone definition for the surface water sampling location to define the illumination regime for the test OR to perform the test completely in the dark to avoid any influence from the light. There was no conclusion within the breakout group as to what suggestion should be included.

Updated Table from Breakout presentation

CHALLENGES	SOLUTIONS	NEXT STEPS FOR OECD IMPLEMENTATION
<u>Clear Testing Objectives:</u> OECD 309 includes different objectives: mineralisation half-life as a main objective with primary biodegradation half-life and identification of degradation products as secondary objectives. Appropriate test system setups (e.g., test substance concentration, radiolabelled test substance, etc.) and analytical (e.g., mass balance) should be tailored for each objective.	Be clear about the objective of the test – it is difficult to look at mineralisation and identify degradation products in one test. Separate the different goals, as they will give 2 different rates (does maximum rate give environmental potential but not environmental relevance?).	Split up the 309 guideline into 3 sections: could have description for mineralisation, for co-metabolic simulation, and for identification of transformation products. Have discussions on the utility of different types of degradation rates. Draft text for the guideline.
<u>Inoculum handling artefacts:</u> The guideline is unclear on the limits for the storage of surface water after it has been sampled. The guideline does not recommend sampling the inoculum at a temperature close to the test temperature, while it may be that bacteria colonies adapted at 12°C are not the same as at 20°C (psychrophile vs mesophile).	Since the inoculum changes during storage, use the inoculum as quickly as possible (OECD 301 gives a maximum of 5-7 days). Avoid changing the testing temperature from the environmental temperature, as this has a high impact on the microbial community.	Draft text for the guideline.

Choice of surface water for the inoculum There is a requirement to test only one surface water at two test substance concentrations.	Provide further instruction in the guideline to define the sampled sites depending on the use of the test result.	Have discussions to define parameters to choose sites.
Inclusion of sediment The definition of suspended solids concentrations for a pelagic test vs suspended sediment test (0.01 to 1 g/L dry weight) is not clear.	Provide a clearer definition of what "pelagic" means and indicate whether it differs for different types of surface water. Explain the utility of adding sediment based on the goal of the test, e.g., co-metabolism (addition of sediment may increase precision and sensitivity, but not environmental relevance).	Draft text for the guideline.
Test substance concentrations The test substance concentrations should relate to the objective of the test but also be environmentally relevant.	Include text to clarify testing as low a concentration as possible depending on the substance, the goal, and analytical feasibility.	Draft text for the guideline (different for each objective).
Presence of light The guideline does not define criteria or parameters for diffuse light.	Use euphotic zone definition of that sampling point; timing of the light.	Draft definitions for the guideline.

4. CLOSE OF THE WORKSHOP

In the closing session of the workshop, the results of each breakout discussion were communicated in plenary to the rest of the workshop participants (including online participants), followed by a brief Q&A. The workshop chair then concluded the workshop by thanking all participants for their active engagement and contributions to the discussions. It was commented that the discussions held at the workshop were valuable and timely given the ongoing implementation of the OECD 309 test in a regulatory context. Participants were advised to look out for future updates and communications on the outcomes of the workshop.

ABBREVIATIONS

AD	Applicability domain
ATP	Adenosine triphosphate
CRO	Contract research Organisation
GC-MS	Gas chromatography mass spectrometry
MPN	Most probable number
NER	Non-extractable residues
OECD	Organisation of Economic Cooperation and Development
PAR	Photosynthetically Active Radiation
PBT	Persistent, bioaccumulative and toxic
PEC	Predicted Environmental Concentration
PMT	Persistent, mobile and toxic
PPP	Plant Protection Products
Q&A	Questions and answers
QSAR	Quantitative structure-activity relationships
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
SCEP	Semi-continuous pre-exposure procedure
SM	Suspended matter
SPM	Suspended particulate matter
SPME	Solid phase microextraction
SPSF	Standard project submission form
SVHC	Substance of very high concern
TG	Test guideline
TOC	Total organic carbon
UBA	German Environment Agency
UVCB	Substance of unknown or variable composition, complex reaction products or biological materials
WNT	Working Group of National Co-ordinators of the TGs programme
WWTP	Wastewater treatment plant

BIBLIOGRAPHY

BIRCH, H., ANDERSEN, H. R., COMBER, M. & MAYER, P. 2017. Biodegradation testing of chemicals with high Henry's constants – Separating mass and effective concentration reveals higher rate constants. *Chemosphere,* 174**,** 716-721.

BIRCH, H., HAMMERSHØJ, R. & MAYER, P. 2018. Determining Biodegradation Kinetics of Hydrocarbons at Low Concentrations: Covering 5 and 9 Orders of Magnitude ofKowandKaw. *Environmental Science & Technology,* 52**,** 2143-2151.

BIRCH, H., HAMMERSHØJ, R., MØLLER, M. T. & MAYER, P. 2023. Technical guidance on biodegradation testing of difficult substances and mixtures in surface water. *MethodsX,* 10**,** 102138.

EC 2009. Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/ 414/EEC.

ECHA 2017. Guidance on Information Requirements and Chemical Safety Assessment - Chapter R.11: PBT/vPvB assessment. Version 3.0.

ECHA 2022a. Medium chain chlorinated paraffins ('MCCP') and other substances that contain chloroalkanes with carbon chain lengths within the range from C14 to C17. *Annex XV Restriction Report.* Helsinki, Finland.

ECHA 2022b. MEMBER STATE COMMITTEE SUPPORT DOCUMENT FOR IDENTIFICATION OF BIS(2-ETHYLHEXYL) TETRABROMOPHTHALATE COVERING ANY OF THE INDIVIDUAL ISOMERS AND/OR COMBINATIONS THEREOF AS A SUBSTANCE OF VERY HIGH CONCERN BECAUSE OF ITS VPVB (ARTICLE 57E) PROPERTIES.

ENVIRONMENT AGENCY 2019. Substance Evaluation Conclusion and Evaluation Report for Medium-chain chlorinated paraffins/alkanes, $C_{14\text{-}17}$, chloro. Helsinki, Finland.

ENVIRONMENT AGENCY 2022. Persistent, Bioaccumulative and Toxic (PBT) properties of Long Chain Chlorinated Paraffins (LCCPs). Bristol.

GOULIARMOU, V., SMITH, K. E. C., DE JONGE, L. W. & MAYER, P. 2012. Measuring Binding and Speciation of Hydrophobic Organic Chemicals at Controlled Freely Dissolved Concentrations and without Phase Separation. *Analytical Chemistry,* 84**,** 1601-1608.

HAKVÅG, S., GIRAY, K., HUGHES, C., RIBICIC, D., BRAKSTAD, O. G., BROWN, D. M., SCHOFIELD, K., MAEß, C., SCHEERBAUM, D. & FIEBIG, S. Impacts of Sample Storage and Reference Compounds in the OECD 309 Surface Water Mineralisation Test. SETAC Europe 34th Annual Meeting, 2024 Seville, Spain.

HAND, L. H. & MORELAND, H. J. 2013. Surface water mineralization of isopyrazam according to OECD 309: Observations on implementation of the new data requirement within agrochemical regulation. *Environmental Toxicology and Chemistry,* 33**,** 516-524.

HAND, L. H. & MORELAND, H. J. 2014. Surface water mineralization of isopyrazam according to OECD 309: Observations on implementation of the new data requirement within agrochemical regulation. *Environmental Toxicology and Chemistry,* 33**,** 516-524.

HOFFMANN, T. O., BAULIG, Y., VOLLMER, S., BLÖTHE, J. H., AUERSWALD, K. & FIENER, P. 2023. Pristine levels of suspended sediment in large German river channels during the Anthropocene? *Earth Surf. Dynam.,* 11**,** 287-303.

HOLZMANN, H., CLAßEN, D., ACKERMANN, J. & SCHÄFFER, A. 2022. Fate of 14C-labelled ionic organic chemicals in a water-sediment system and surface water. *Chemosphere,* 303**,** 134885.

HUGHES CB, GRIFFITHS M, COOK S, VAN DE MEENT D, PARSONS J, LYON D, et al. Developing a weight-of-evidence methodology for persistence assessment of substances in the environment. ChemRxiv. 2025; doi:10.26434/chemrxiv-2025-g1jpg This content is a preprint and has not been peer-reviewed.

ISO 1999. Water Quality - Evaluation of the aerobic biodegradability of organic compounds at low concentrations. Ring Test Report - ISO 14592 Part 1.

MENDO DIAZ, O., PATINY, L., TELL, A., HUTTER, J., KNOBLOCH, M., STALDER, U., KERN, S., BIGLER, L., HEEB, N. & BLEINER, D. 2024. A Quasi Real-Time Evaluation of High-Resolution Mass Spectra of Complex Chlorinated Paraffin Mixtures and Their Transformation Products. *Anal Chem,* 96**,** 12378-86.

OECD 2004. Test No. 309: Aerobic Mineralisation in Surface Water – Simulation Biodegradation Test. *OECD Guidelines for the Testing of Chemicals, Section 3.* Paris: OECD.

OECD 2024. Work plan for the OECD Test Guidelines Programme (TGP).

POLCHER, A., SCHLECHTRIEM, C. & HAHN, S. 2023. Review of the OECD Test Guidelines relevant to environmental assessment with regard to the state of the art in science and technology. German Environment Agency.

SELLER, C., HONTI, M., SINGER, H. & FENNER, K. 2020. Biotransformation of Chemicals in Water–Sediment Suspensions: Influencing Factors and Implications for Persistence Assessment. *Environmental Science & Technology Letters,* 7**,** 854-860.

SMITH, K. E. C., OOSTINGH, G. J. & MAYER, P. 2010. Passive Dosing for Producing Defined and Constant Exposure of Hydrophobic Organic Compounds during in Vitro Toxicity Tests. *Chemical Research in Toxicology,* 23**,** 55-65.

SOLOMON, K., MATTHIES, M. & VIGHI, M. 2013. Assessment of PBTs in the European Union: a critical assessment of the proposed evaluation scheme with reference to plant protection products. *Environmental Sciences Europe,* 25**,** 10.

APPENDIX A: WORKSHOP PROGRAMME

Day 1: 21st January 2025		
12:30 – 13:00	Arrival and registration	
13:00 – 14:00	***Welcome Lunch***	
14:00 – 14:15	Welcome, introduction and workshop objectives	**Dr. Blanca Serrano Ramòn**
14:15 – 15:00	**Scene setting keynotes**: i. Lessons learned in ISO ring test in 1998-99 for OECD TG309 – *(14:35)* *ii.* Advancing Environmental Assessment: Revising OECD Test Guidelines with Insights from the German Environment Agency – *(15:00)*	**Dr. Lars Toräng [1]** **Dr. Ulrich Jöhncke [2]**
15:00 – 15:30	***Coffee break***	
15:30 – 16:30	**Robustness pillar** *(1 hour duration)*: i. Main presentation on key challenges – *(15:50)* ii. Short follow-on presentations on selected issues *a) Chris Hughes (16:00)* *b) Marie Collard (16:05)* *c) Fola Ogungbemi (16:10)* *iii.* Questions, research needs, discussion points, and additional topics raised by the audience. *(16:30)*	**Dr. Carolin Seller [3]** **[3a]** **[3b]** **[3c]**
16:30 – 17:30	**Implementation pillar** *(1 hour duration)*: i. Main presentation on key challenges – *(16:50)* ii. Short follow-on presentations on selected issues *(20 min)* *a) Jason Snape* *b) Heidi Birch* *c) Philipp Mayer (17:10)* *iii.* Questions, research needs, discussion points, and additional topics raised by the audience. *(17:30)*	**Dr. Dieter Hennecke [4]** **[4a]** **[4b]** **[4c]**
19:00	***Evening event***	

Day 2: 22nd January 2025		
9:30 – 9:45	Welcome and setting the day	
9:45 – 10:45	**Relevance pillar** *(1 hour duration)*: i. Main presentation on key challenges – (10:05*)* ii. Short follow-on presentations on selected issues *(20 min)* *a) Marie Collard* b) *Martin Brüggemann (10:25)* *iii.* Questions, research needs, discussion points, and additional topics raised by the audience. *(10:45)*	**Dr. Michael McLachlan [5]** **[5a]** **[5b]**

10:45 – 11:00	***Coffee break***	
11:00 – 12:30	Breakout Sessions Part 1 *(Only In-person participants)*	
12:30 – 13:30	***Lunch***	
13:30 – 15:00	Breakout Sessions Part 2 *(Only In-person participants)*	
15:00 – 15:30	***Coffee Break***	
15:30 – 16:15	**PLENARY**: Overview of Breakout Sessions (15 *minutes/ pillar*)	
16:15 – 16:30	Conclusion & Close	

Abstracts and Speakers' bios

Lessons learned in ISO ring test in 1998-99 for OECD TG309 **Lars Toräng, Ph.D.**
First batch simulation test of surface water was drafted by Niels Nyholm, Technical University of Denmark and Udo Pagga, BASF in "*ISO/DIS 14592 Part 1 - Water quality - Evaluation of the aerobic biodegradability of organic compounds at low concentrations — Part 1: Shake flask batch test with surface water or surface water/sediment suspensions*" under ISO/TC 147/ SC5/ WG4 Biodegradability. Batch simulation tests aim at estimating degradation rates, conducted in a laboratory system with conditions that are realistic for the particular environmental compartment. Simulation tests should mimic the actual environmental conditions such as redox potential, pH, temperature, microbial community, concentration of test substance, and occurrence and concentration of other substrates. Seven laboratories participated and the main results and important considerations behind the method will be presented. Selected model compounds were aniline and 4-chloroaniline in the concentration range 0.5-500 µg/L. All laboratories used ^{14}C-radiotracer technique with ^{14}C supplied from recognized sources, and none used specific analysis. In addition to the mandatory pelagic tests with water, two included tests with sediment spiking. Only one laboratory carried out optional biodegradation studies with aniline and 4-chloroaniline after a pre-adaptation step with semi-continuous operation.
Regulatory Affairs manager at Bjørn Thorsen A/S (2022-) and Agilent Technologies (2017-2022), REACH expert at Sun Chemical (2007-2017), Post Doc. about Equilibrium Sampling Through Membranes (2004-2007) at National Environmental Research Institute (2004-2007), and Ph.D. (1998-2003) at Technical University of Denmark. *Analysed data and author of the first batch simulation ring test back in 1998-99. My Ph.D. was about understanding adaptation and biodegradation of organic compounds in batch simulation tests. For direct interpretation of measured rates, it is important to keep the concentrations of simulation studies as low as the actual concentrations, or sufficiently low, to ensure the same kinetics as in the environment. Biodegradation rates obtained in degradation studies at high concentrations (> 100 µg/L) may grossly overestimate the actual rates at environmental relevant concentrations, even when investigated in the same environment.*

Advancing environmental Assessment: Revising OECD Test Guidelines with Insights from the German Environment Agency
Ulrich Jöhncke, German Environment Agency

The talk will give information on the Research and Development project "Review of the OECD Test Guidelines relevant to environmental assessment with regard to the state of the art in science and technology" done in behalf of the German Environment Agency. This project investigated the revision needs of OECD test systems from 2020 to 2023. Background and procedure chosen by the consultants which led to identification of the OECD TG 309 identified as the TG in most need of revision will be described. A compilation of the suggestions received will be given. In addition, experiences that the German Environment Agency made with and views on the test depending on the assessment concerned will be presented from the perspectives of sections which operate under different legal frameworks. This will be accompanied by background information on the specifications given in the REACH guidance on the test procedure of OECD TG 309.

1990: Graduated with a degree in ecotrophology at the University of Applied Sciences at campus Niederrhein, Mönchengladbach
1991 to 2007: section IV 2.3 Biodegradation and Bioaccumulation at the Federal Environment Agency, specialising in the evaluation of test protocols for chemicals, but also for plant protection products or pharmaceuticals.
Since 2007: section IV 2.3 Chemicals, evaluating of or preparing dossiers as part of the REACH processes, e.g. SVHC identification. In charge of the degradation team, member of UBA's coordinating group on degradation and exposition and supporting expert for the German member of MSC and the Agency's expert in ECHA's PBT expert group. Initiation and supervision of scientific research projects, e.g. on NER.

PILLAR 1: ROBUSTNESS: Enhancing Standardization and Reproducibility in OECD 309 Biotransformation Studies for Chemical Persistence Assessment
Dr. Carolin Seller, Eawag – Swiss Federal Institute for Aquatic Science and Technology

For chemical persistence assessment, regulatory frameworks following the REACH regulation rely on chemical half-lives (DT_{50}) derived from OECD biotransformation simulation studies, including the OED 309 guideline, which aims to assess biodegradation in surface waters. REACH pursues a cautious approach, meaning that the longest DT_{50} available from at least two different OECD 309 tests is compared against persistence cut-off criteria to depict a "worst-case" scenario. This implies the assumption that variable DT_{50} values, to large extents, represent variability of a chemical's biotransformation behavior in natural surface water bodies. Unfortunately, differing OECD 309 DT_{50} values also stem from laboratory artefacts resulting from insufficiently standardized conditions allowing for differences in test setups that can significantly impact study outcomes. As just a few examples, the OECD 309 guideline allows for (i) batch or semi-continuous operation, (ii) systems with or without suspended sediment, (iii) a range of initial chemical concentrations, (iv) largely undefined inoculum sources, (v) various test temperatures and many more, making the comparison between study outcomes highly uncertain. Our research suggests that proper standardization of those test conditions, e.g., addition of suspended sediment and reduction of chemical concentrations, holds the potential to significantly reduce laboratory artifacts and therewith increase reproducibility of test results. Finally, the OECD 309 guideline does not require any analysis and reporting of the composition or diversity of the microbial community present in the test systems, even though both parameters strongly influence a compound's biotransformation behavior in aquatic environments and should be included to allow for a proper data analysis.

.	*Carolin's research is focused on exploring biotransformation behavior of anthropogenic chemicals, i.e., pharmaceuticals, pesticides or industrial chemicals, in aquatic environments. The main objective is to understand which chemical properties in combination with which environmental parameters would favor a rather rapid substance degradation and avoid its release and widespread transport in the environment. Thereby, she has been gathering and working with field and laboratory data describing biotransformation in water, water-sediment, or activated sludge systems.*

PILLAR 2: IMPLEMENTATION
Dr. Dieter Hennecke, Fraunhofer Institute

From theory to testing practice: to conduct practical e-fate testing under GLP for regulatory purposes is in general a challenging task. It might become a nightmare if the testing guideline is unspecific in relevant sections or if requests by a respective regulatory framework are unclear (e.g. suspended sediment amount in OECD 309 pelagic test for REACH). Though on the other hand some flexibility is necessary to be able to test chemicals with a certain range of Phys-chem properties within the frame given by a guideline.

Examples will be given of challenges, questions and experiences in practical OECD 309 standard tests conducted at a GLP testing facility under the required standardised conditions with typical substances. It will demonstrate that a guideline cannot have an answer to all – foreseeable and unforeseen – issues that may arise during a real GLP-test, and that the line between a standard test and a research project can be very thin. This is important to understand for registrants and regulators as expectations for executing laboratories for a "standard test" are usually quite different.

.	*Dieter Hennecke is a chemist with more than 30 years of experience in applied research on the fate of chemicals in the environment. In his department at Fraunhofer IME, guideline-driven standard studies are carried out for regulatory purposes under GLP, mostly with substances that are difficult to test. He is further active in the development of new and improvement of existing guidelines to support chemical regulation. His particular expertise lies in ^{14}C-radiolabeling technology and its application in novel research fields (e.g. fate of polymers).*

PILLAR 3: RELEVANCE: Environmental relevance of the OECD 309 test
Dr. Michael McLachlan, University of Stockholm

Recent research exploring the application of OECD 309 type tests to study the biodegradation rates of organic contaminants in rivers will be presented. The underlying premise of the work is that the initial biodegradation rate in the test is the best measure of the biodegradation rate in the environment; all changes in biodegradation rate during the test reflect departures of the test conditions from environmental conditions. The research shows that almost all of the chemicals studied did not degrade in a pelagic test. Large quantities of sediment (50 g/L) were required in the incubation to increase the microbial community density to a level sufficient to biodegrade most test chemicals to a measurable extent. While the repeatability of the test for a specific water/sediment mixture was good, when the test was conducted with water/sediment mixtures from different rivers then the biodegradation rate constant for a given chemical varied by up to several orders of magnitude. These results will be

discussed in the context of the environmental relevance of the OECD 309 test and the information that the test can provide in chemical regulation.	
·	*Michael is a professor of environmental chemistry in the Department of Environmental Science at Stockholm University. Michael's research interest is the fate of organic contaminants in the environment. During the fast years his work has focused on biodegradation in aquatic systems, and he is particularly interested in quantifying biodegradation rates in the natural environment. He has led research investigating the use of OECD 309 for this purpose.*

Breakout groups

ROBUSTNESS	IMPLEMENTATION	RELEVANCE
Marie Collard	Jason Snape	Sandrine Sourisseau
David Saunders	Philipp Mayer	Amine Tchéré Kakia
Marco Kraas	Diedrik Schowanek	Roy Geerts
Aina Charlotte Wennberg	Pippa Curtis-Jackson	Harriet Moreland
Des Kelly	Adrian Beard	Anu Kapanen
Silvia Berkner	Emma Danby	Martin Brüggemann
Fola Ogungbemi	Sami Belkhiria	
Katie Endersby	Daniela Classen	
Heidi Birch		
Chris Hughes **(Moderator)**	Dieter Hennecke **(Moderator)**	Miriam Leon Paumen **(Moderator)**
Carolin Seller	Lars Toräng	Michael McLachlan
Katherine Santizo **(Rapporteur)**	Louise Camenzuli **(Rapporteur)**	Ulrich Johncke
		Delina Lyon **(Rapporteur)**

APPENDIX B: WORKSHOP PARTICIPANTS

In-person participants

First Name	Last Name	Affiliation
Martin	Brüggemann	Bayer AG
Lars	Toräng	Bjørn Thorsen A/S
Katherine	Santizo	Cefic
Adrian	Beard	Clariant Corp.
Delina	Lyon	Concawe
Fola	Ogungbemi	Currenta GmbH
Marie	Collard	dsm-firmenich
Carolin	Seller	Eawag
Andrea	Salvadori	ECETOC
Blanca	Serrano Ramon	ECETOC
Megan	D'souza	ECETOC
Anu	Kapanen	ECHA
Chris	Hughes	Embark Chemical Consulting
Pippa	Curtis-Jackson	Environment Agency, UK
Marco	Kraas	Evonik Operations GmbH
Louise	Camenzuli	Exxonmobil
Miriam	Leon Paumen	Exxonmobil
Christian	Schlechtriem	Fraunhofer Institute
Dieter	Hennecke	Fraunhofer Institute
Daniela	Classen	German Environment Agency
Silvia	Berkner	German Environment Agency
Ulrich	Johncke	German Environment Agency
Bruno	Hubesch	Hubesch Consult BV, Scientific services for industry
Aina Charlotte	Wennberg	NIVA - Norwegian Institute for Water Research
Roy	Geerts	Nouryon
Diederik	Schowanek	P&G
Emma	Danby	Scymaris
David	Saunders	Shell
Des	Kelly	Smithers Ltd
Michael	McLachlan	Stockholm University
Harriet	Moreland	Syngenta
Heidi	Birch	Technical University of Denmark
Philipp	Mayer	Technical University of Denmark
Sandrine	Sourisseau	TotalEnergies
Katie	Endersby	Unilever
Jason	Snape	University of York

Online participants

First Name	Last Name	Affiliation
Emma	Jack	3rd Party Environmental
Leon	Rockett	Afton Chemical
Andrew	Barrick	Auburn University
Daniel	Tschentscher	BASF
Patrizia Elinor	Thoma	BASF
Bernhard	Jene	BASF SE
Björn	Hidding	BASF SE
Patricia	Golombek	BASF SE
Simon	Schroeter	BASF SE
Sven	Peter	BASF SE
Doris	Ebert	BASF SE, Agricultural Center Limburgerhof
Kat	Colvin	BP
Angelica	Candido	Cefic
Ece	Koc Martin	Cefic
Francesca	Trivelloni	Cefic
Sara	Lozano García	CEHTRA
Marieke	de Rooij	Charles River Labs
Jens	Bietz	Clariant
Nicole	Brun	CP Kelco
Philippa	Kearney	DEFRA (UK)
Na	Guan	Dow
Nathalie	Vallotton	Dow
Lucy	Wilmot	ECETOC
Sergio	Perez	ECETOC
Anna-Lena	Rehrl	Environment Agency Austria
Elina	Uotila	ERA-Consult
Milagros	Vega	ERA-Consult
Florian	Möllmann	Evonik Operations GmbH
Zahra	Farmani	Evonik Oxeno
Huiting	Chen	ExxonMobil
Biyao	Han	ExxonMobil (China) Investment Co., Ltd
Phil	Lamond	Fera
Philip	Rooney	Fera
Victoria	Pratt	Fera
Juha	Einola	Finnish Safety and Chemicals Agency
Tomi	Kaartinen	Finnish Safety and Chemicals Agency
Boris	Meisterjahn	Fraunhofer IME
Eva-Maria	Teggers	Fraunhofer IME
Prasit	Shrestha	Fraunhofer IME-AE
Stefan	Hahn	Fraunhofer ITEM
Gordon	Sanders	Givaudan International SA

Bettina	Seiwert	Helmholtz-Centre for Environmental Research (UFZ)
Jakob	Vogel	Helmholtz-Centre for Environmental Research (UFZ)
Milena	Voll	Helmholtz-Centre for Environmental Research (UFZ)
Marlies	Bergheim	Henkel AG & Co. KGaA
Thuy	Bui	Heubach Colorants Germany GmbH
Nathalie	Schraepen	Huntsman
Anwen	Clementson	Infineum UK Ltd
Mahdis	Ghasemi	Interested Attendee
Ioannis	Liagkouridis	IVL Swedish Environmental Research Institute
Xiaohua	Dr. He	Knoell Germany GmbH
Robert	Unsworth	Labcorp
Michael	Essers	LANXESS Deutschland GmbH
Pamela	Abdallah	L'Oréal
Zhi-Yi	Rong	Lubrizol Corporation
Aminé	Tchéré Kakia	Noack Laboratorien GmbH
Kristina	Miller	Noack Laboratorien GmbH
Jennifer	Menzies	Procter and Gamble
Megan	Griffiths	Ricardo
Emiel	Rorije	RIVM
Eric	Verbruggen	RIVM
Joop	De Knecht	RIVM
Aymeric	Bellemain	Rovaltain Research Staphyt
Chiara	Barnaba	Staphyt
Daria	Clemm	Stockholm University
Malte	Posselt	Stockholm University
Yijing	Li	Stockholm University
Eszter	Simon	Swiss Federal Office for the Environment (FOEN)
Peter	Takacs	SWS-Weiterbildung
Anne-Lise	Mandrillon	Syensqo
Neil	Wang	Syensqo
Rebecca	Southwell	Syngenta
Emma	France	Syngenta Ltd
Rune	Hjorth	Technical University of Denmark
Vurtice	Albright	The Dow Chemical Company
Silke	Fiebig	Vitis Regulatory GmbH

APPENDIX C: ORGANISING COMMITTEE

Michael McLachlan
Stockholm University
Sweden

Dieter Hennecke
Fraunhofer IME
Germany

Louise Camenzuli
Exxonmobil
Belgium

Chris Hughes
Embark Chemical Consulting
United Kingdom

Delina Lyon
Concawe
Belgium

Katherine Santizo
Cefic
Belgium

Christian Schlechtriem
Fraunhofer IME
Germany

Blanca Serrano
ECETOC
Belgium

Megan D'souza
ECETOC
Belgium

Andrea Salvadori
ECETOC
Belgium

APPENDIX D: BREAKOUT TABLES

Robustness pillar

Challenge	Solution	Relevant research	Next steps for OECD Implementation	Timeline <1, 1-3, >3 yrs
The guideline does not define any standard method for the preparation of sterile controls. Use of poisons (Azide and Hg salts) can result in undesirable analytical matrix effects / damage MS and may modify constitution of matrix where suspended sediment is added back in to the test system.	Develop a reliable and relevant standardised method; include check of sterility and taking caution of pH changes that may occur Suggest consulting CROs' experience Review the relevant ECHA guidance *R.7b (2023) (https://echa.europa.eu/documents/10162/17224/information_requirements_r7b_en.pdf/1a551efc-bd6a-4d1f-b719-16e0d3a01919) and ECHA note on sterile controls (https://echa.europa.eu/fi/technical-scientific-reports)*			
The range of suspended sediment concentration defined in the guideline is too low (0.01 – 1g/L). Comment: Relates to comment on increased suspended sediment concentration: The differences	Suggested change to 0.1 – 10 g/L would lead to a more standardized test condition. Comment: There should be further discussions and ensure the overlap on relevance. What are the cut-offs that	Seller et al 2020 ECHA guidance R11, R7b Decreasing suspended sediment concentrations in river Elbe and Rhine, measured		1-3

between replicates observed by Seller et al. are very high and, to our knowledge, are not observed in studies from CRO laboratories. Increasing suspended sediment concentration enhances transformation, which leads to issues with comparability. The purpose of simulation-type studies is to use environmentally relevant, unchanged test matrices, i.e. surface waters with environmentally relevant suspended sediment concentrations. According to REACH guidance document R.11, this range is typically 10–20 mg/L for rivers in the EU. Comment: In addition, increasing suspended sediment concentrations in OECD 309 tests leads to quantities of NER which are comparable to OECD 308 tests as shown by Holzmann et al. (2022). Furthermore, as only low amounts of sediment (<1g/L) are added to the surface water in the suspended sediment variant of the test, total NER quantification is already challenging, and NER characterization is not possible anyway with the available methods due to low sediment amounts (Chemosphere 303 (2022) 134885).	are needed for robustness and relevance? SS has shown impact on variabilities in test system. Should the range be increased? If so to what extent? Comment: Suggest separating the SS range for pelagic test and suspended sediment test. Amount of SS in pelagic test should represent average surface water conditions. Comment: The 10 g /L seems very high to me, and the proposed range remains in the 2 orders of magnitude range. Would 0.1 to 1 g/L be more realistic? Agree that this can be a very influential variable which needs standardising. Comment: Keep pelagic variant Comment: Additionally, different suspended sediment concentrations could be tested in a ring test. An environmentally relevant concentration should be included into the ring test to allow for comparability with regulatory acceptable tests. For a more standardised TG, however, suspended sediment concentration should then be defined clearly and not given as a range.	values might even be below 10 mg/L Hoffmann et al., 2023, Earth Surf. Dynam., 11, 287–303, 2023 https://doi.org/10.5194/esurf-11-287-2023 Holzmann et al (2022). (Chemosphere 303 (2022) 134885)		

The guideline does not define microbial activity limits beyond the degradation of the reference substance within 2 weeks. The guideline utilises the degradation of the reference substance as a proxy for microbial activity, however, it is well established that this is a poor descriptor, hence alternative methods for microbial activity estimation should be defined.	Limits to microbial activity need to be defined in a way that better reflects the method used. Include flow cytometry for potential quantification / ATP/ respiration as feasible options. Should not only be reference substances. Comment: More information on the reasoning why this is considered important could be helpful for discussion. Comment: Microbial activity? I would suggest that there are at least two components here to look at: Microbial heterogeneity (diversity) and Microbial numbers/abundance. RNA analyses at sampling, Day 0 and test end to assess changes. Analysis and comparison of population heterogeneity at Day 0d Sampling, Day 0 and End of incubation. What other methods exist? Can be address via other challenges	See ECO55 and interim report on Sample storage (05 Dec. 2023). Dieter H. mentioned past work done in Fraunhofer.		
The guideline recommends the use of sodium benzoate or aniline "*to ensure microbial activity of the test water is within certain limits"*. However, it is unclear what these limits are, and it is well established that these two reference substances do not differentiate between surface water samples.	Different reference substances need to be identified that capture the variability in inoculum concentration. Lower degradation should be taking place at lower inoculum concentration. *(Comment: How would this information be used for the*	ECO 55 Comment: Na Benzoate was the only Ref. Subst. In ECO55 that "functioned", But it will likely function anytime, anywhere and under most any conditions. ANILINE, on the other hand, a known Reference		

Comment: The TG does not stipulate any specific Ref. Substance pass levels and uses the term "degradation" when this probably should read "mineralisation".	*assessment?* *Also, as a remark, OECD TG 309 has optional measurements of* *- microbial biomass (e.g. acridine orange direct count or colony forming units);* *- chlorophyll-a concentration as a specific estimate for algal biomass.)*	Substance and READY Biodegradable chemical, demonstrated huge variability in the ECO55 Ring-test with a reported median "DT50" of 117.5 days! This really begs the question – Just how relevant the OECD 309 is. Comment: I am not familiar with this project but there is a large difference to the ring test mentioned in the OECD TG 309. Any information how this difference can be explained? (OECD TG 309: In an ISO ring-test of the method where seven laboratories located around Europe participated, adapted degradation rate constants for aniline ranged from 0.3 to 1.7 day-1 with an average of 0.8 d-1 at 20oC and a standard error of 0.4 d-1 (t½ = 0.9 days). Typical lag times were 1 to 7 days.))		
The guideline states that it is possible to test slightly volatile substances using a closed flask. However, tests in closed systems have been shown to be inadequate and of questionable reliability due to e.g. reduced degradation of the reference substance.	Current work exploring, look at ongoing research. Potential of GD for specific subgroups Move to a set-up where air is constantly and slowly drawn through (under vacuum) the test system. The air exiting the individual vessels would then go through a series of volatile			

Comment: I don't think this applies to all tests with closed bottles. Is it known what explains the reduced degradation of the reference substance? Comment: Overlap with implementation	scrubbers.			
Testing is allowed in the dark or diffuse light. This may cause diverging results for light-sensitive test substances. Comment: Overlap with relevance but part of the experimental design options that impact robustness may be specific to plant protection products and therefore can be a specific point to address in GD that develops.	It would be helpful to increase reproducibility to unambiguously require testing in the dark	Regulatory experience (data cannot be shared)		
In addition, increasing suspended sediment concentrations in OECD 309 test led to quantities of NER which are comparable to OECD 308 tests as shown by Holzmann et al (2022). Furthermore, as only low amounts of Sediment (<1g/L) is added to the surface water in suspended sediment variant of the test, already total NER quantification is challenging, and NER characterization is not possible anyway with the available methods due to mentioned low sediment amounts (Chemosphere 303 (2022)		Hoffmann et al., 2023, Earth Surf. Dynam., 11, 287–303, 2023 https://doi.org/10.5194/esurf-11-287-2023 Holzmann et al (2022) (Chemosphere 303 (2022) 134885)		

134885)				
The guideline proposes sequential sampling, i.e. one flask for all sampling days or harvesting of whole flask at each sampling day.	Delete the option for sequential sampling. Comment: To be discussed and further explored whether there are substances where this is preferred (potential impact on pelagic vs SS). When to use and when not to. Part of guidance document on recommendations.	Research done by Holzmann et al. (2021) shows reduced recoveries and mineralisation during sequential sampling compared with sacrificial sampling (Science of the Total Environment 753 (2021) 142101)		
The guideline is unclear on the limits for the storage of surface water after it has been sampled. Since the inoculum composition drifts quickly from its initial composition during storage, it could be recommended to use the inoculum within 24 h (OECD 301 gives a max of 5-7 days).	Language correction to recommend more stringent language on storage and general language corrections to clarify storage including transport	ECO55 (Comment: ECO55 findings help with this, where they suggest storing at +4°C if the water sample cannot be used immediately on the same day of sampling. It also recommends a maximum storage period of 7 days, which is significantly shorter than the 4 Weeks indicated in the current version of the TG. Furthermore, it may also be advisable to store and even run the 309 study at the actual temperature that the sample was taken.)		Immediate
Challenge the necessity to run 2	Refine the goal on why the two			

different concentration levels (states between 5 and 10 x for the 2nd test concentrations).	concentrations are used and what each concentration achieves. Align with existing TGs that state the high concentration purpose. Re-emphasize Lower range => cometabolism and have more robust reporting			
More detailed reporting? Impact on how data details are shared.	There is high variation of reporting detail. More detailed instructions and requirements on measuring and reporting e.g. mass balance components is needed.			
Requirement to test only one surface water with two concentrations.	To consider increasing the number of surface waters to be tested to increase the relevance. OECD 308 and 307 include requirement for multiple test environments.			
Aim of the potential modification to increase relevance.	Any modification proposed should not make the test more favorable for degradation but increase robustness and repeatability. Increase in degradation potential of the test setup is not equal to increasing its relevance.			
Pretests are undefined. Analytics criteria but concentration ranges could be improved.	Pre-test done with highest concentration maybe should consider others and this could be aided by adding recommendations on ranges to ensure low values. Use similar wording as in other guidelines e.g., 307,308			
High inter-replicate variability				
High inter-test variability				
Permitted experimental setups				

Implementation pillar

Challenge	Solution	Relevant research	Next steps for OECD Implementation	Timeline <1, 1-3, >3 yrs
Par. 7: Defines the **applicability** of the test guideline (TG), however, does not define specific limits beyond the Henry's law constant (H). The TG does not define **clear physico-chemical limits** for which the guideline is applicable, leading to the test regularly being used on substances which require significant method development. Results received under these "developed" conditions are no longer necessarily comparable to results produced under the conditions for which the TG was developed. "Method development" should be no go in standard testing that aims to get valid and reproducible results under identical test conditions.	This can be addressed through the definition of an **Applicability Domain (AD)** for which the OECD 309 has been developed and is suitable for. Ranges of Log K_{aw} (replacing H), water solubility and Low K_{ow} could be established. Additional physico-chemical properties such as pKa, log P and log D should be also evaluated. It was recognised that modifications/testing options necessary for substances that go beyond these ranges should also be established.	It was acknowledged that the TG has already been used to assess substances that are likely to be considered outside of a future AD. It is expected that knowledge relating to a working AD, for example accounting for substance physico-chemical parameters that require practical modifications for testing, is already applied in practice although unpublished in Contract Research Organisations (CROs). Furthermore, additional modifications are likely to be available in scientific literature.	1. A targeted literature review could be performed to collate available information on AD of the OECD 309 TG. The OECD Guidance Document (23) on Aquatic Toxicity Testing of Difficult Substances and Mixtures (OECD, 2019) was also discussed as a source of supporting information. 2. A questionnaire, or interviews, could be sent out to CROs to collate available first-hand knowledge in identifying substance properties within and beyond the AD. 3. It was recognised that regulators and academic researchers would also have knowledge to this capacity and should be included in above questionnaire / interviews.	<1

The group concluded that the impact to this challenge is **high,** since there is currently no clear definition of what can be tested with an OECD 309, and that clarity is **urgently** needed for a reliable assessment.			4. Theoretical calculations could also be performed (multi-media fate modelling) as a foundation for deriving thresholds. The group proposed that one option for coordinating interviews and questionnaires could be via the OECD, since precedent for such an exercise already exists (ED).	
Applicability of OECD TG 309 to **UVCBs, multi-constituent substances (MCSs) and mixtures**. It was recognised that TG 309 is not relevant to evaluate the persistence of these substances, and hence guidance on how to test these substances would be valuable. The additional issue of testing UVCBs is toxicity to relevant microorganisms. Due the potential large number of constituents being tested in parallel, there is a potential of constituent toxicity. The group discussed that the	Whole UVCBs, MCSs and mixtures cannot be interpreted reliably using radiolabelled test substances, and so, this aspect of the guideline was discussed. Please note that 'individual' components of UVCBs, MCSs, and mixtures can be assessed with a radio-isotopically labelled substance. One solution was to add an option of assessment using unlabelled test substance. This could be added as an appendix to the TG or a new test guideline could be developed. This could then be used as the recommended option for whole UVCBs, MCSs and mixtures,	There is a significant body of recent research that has focused on developing testing methods for UVCBs, MCSs, and mixtures, particularly petroleum hydrocarbons and essential oils. These methods are designed to analytically follow the primary degradation of the whole substances' constituents (Birch et al. 2017, 2018, 2023). Additionally, several UVCBs / MCSs have been assessed under REACH, for which methods have been developed. For example, chlorinated paraffins (ECHA, 2022a, Environment Agency, 2019, Environment Agency, 2022, Mendo Diaz et al., 2024), bis(2-ethylhexyl)	Sufficient methods exist to evaluate the persistence of UVCBs and MCSs. However, they have not been evaluated within the limits of the OECD TG 309. Hence, the group recommends that these newer methods are validated for inclusion in the OECD TG 309. Development of a guidance document at the OECD requires a working group of experts to be established with a leading country. It should also be acknowledged that methods that are in development should also be considered for future inclusion. The guidance document, Birch et	1-3

impact of this challenge is **high.** Particularly because interpreting data from UVCBs, MCS and mixtures is challenging for regulatory purposes in the OECD TG 301 screening methods. It was discussed that establishing an OECD 309-like test guideline for UVCBs and MCS would probably be easier than the OECD 301 and would likely have higher regulatory acceptability.	providing clarifications of the pros & cons of the two aspects to the test. The limitations of the methodology could be clearly presented and clarified as part of the updated or new TG, e.g., that this would only allow for the assessment of primary biodegradation. Given that the radio-isotopic labelling is not impossible for UVCBs, MCSs, and mixtures, the group also discussed the development of an OECD 23-like guidance for screening and higher tier degradation studies. A guidance document would allow for more consistency across laboratories and would clearly lay out accepted testing options for these substances.	tetrabromophthalate (ECHA, 2022b).	al. (2023), was nominated as a starting point, to help establish further research needs. Additional research was recommended for the assessment of transformation products in studies using unlabelled test material. While this is not specifically targeted towards the assessment of UVCBs, it would support the establishment of the alternative methods. Non-target analysis (NTA) was discussed as a potential analytical technique to address metabolites assessment. Stable isotope labelling of a test substance can aid in NTA.	
Par. 22 defines the **methods for addition of the test substance**, specifically for substances with poor water solubility, and for low to non-volatile substances. The methods defined in this paragraph cannot be considered comprehensive or encompassing of dosing options for the numerous physico-	Par. 22 of the OECD 309 TG would need to be refined, in line with the applicability domain challenge, to clarify the applicability of currently listed dosing methods and future options. Best practices could be identified to ensure consistency across testing laboratories. The group recognised that this	Not discussed – however, relevant passive dosing is well developed. E.g. Birch et al. 2017, 2018, Gouliarmou et al. 2012, Smith et al. 2010. OECD GD 23 might give further suggestions on how to proceed with substances with poor water solubility.	The next steps identified here are closely related to those under the challenge listed under the applicability domain. 1. A literature review could be performed to collate available information on dosing methods for difficult-to-test substances. The OECD 23 Guidance document was also	1-3

chemical variables that make substances difficult to introduce into a test system. Consequently, this leaves too much room for interlaboratory variability. This challenge was discussed to have a **high** impact, particularly on the short term.	challenge is a parallel activity/solution listed under the applicability domain definition. Such work needed to better understand the different dosing methods could be well tied into any work needed for a better definition of an applicability domain. Passive dosing was suggested as a potential solution for a certain range of substances, however, feasibility across a broad range of laboratories would need to be established. General guidance is also needed here and could be rolled into the OECD23-like guidance document needed for UVCBs and MCSs. Respective chapters could include alternative dosing methods that could help for certain difficult-to-test substances. The inclusion of any new technologies would need to be well described and be broadly applicable.		discussed as a source of information that could support definitions of applicability domain. 2. A questionnaire or interviews could be sent out to CROs to collate available first-hand knowledge on methods used for dosing difficult-to-test substances. 3. It was recognised that regulators and academics would also have knowledge to this capacity and should be included in the above questionnaire / interviews. 4. A validation of these methods for suitability of testing within an OECD 309 TG would be required. Based on the information gathered in the steps listed above, a guidance document can be developed to allow for a standardisation of dosing methods used across different tests. It was discussed that the inclusion of any new technologies would need to be sufficiently described and broadly applicable,	

			e.g., microdosing.	
Par. 53 of the OECD 309 TG lists the minimum information that should be available in the **test report**. However, a challenge that has been flagged is that there is a high variation of the reporting detail. More detailed instructions and requirements on measuring and reporting e.g., the mass balance components are needed. Lack of reporting information leads to issues around assessment of the test method. This challenge was discussed as having a **high** impact, since improved data reporting would allow for better quality assessment.	The solution that the group identified was the development of an **OECD harmonised template (**Environmental Fate and Behaviour (Degradation and accumulation) - Harmonised templates | OECD**),** which could be included as an annex to the TG. The reporting, currently in the form of a list, could be developed into a table format, allowing for easier utility. The table could then identify parameters which must be reported.	The OECD Working Party for Hazard Assessment (WPHA) is currently developing a guidance document on the regulatory use of research data (OECD WPHA Project: OHTs for Research Data (OHTR) | Risk Assessment Portal | US EPA). The Persistence Assessment Tool (PAT) (Hughes et al., submitted; preprint available), an excel based tool, has recently developed and captures key meta data needed to evaluate the validity of studies.	A discussion with the OECD secretariat of the Working Party for Hazard Assessment and/or Test Guideline Coordinators was suggested to better understand the path needed for the development of an OECD harmonised template. Such a template could then be included as part of a TG update, either through inclusion in the annex, or hosted on the OECD website.	<1
The OECD TG 309 is a simulation test for **aerobic** mineralisation in surface water. However, the guideline does not define criteria for dissolved oxygen concentrations and/or redox values to ensure that the test systems are truly aerobic. It was	The solution that the group identified was that the term "aerobic" should be defined, through the establishment of relevant oxygen concentrations or redox potentials. However, there was quite some discussion, with no agreed upon conclusion, whether	Not discussed	The group discussed the different metrics available for dissolved oxygen concentrations e.g., mg/L or % saturation, and which they would consider the most relevant, however, no consensus could be reached.	<1

recognised that this challenge is only a challenge for testing in a closed system, since the open system can be assumed to be sufficiently aerated, that aerobic conditions could be assumed. This challenge was concluded as having a **high** impact for closed systems.	such criteria would be included as validation criteria. The group also recommended that this solution would require recommendations on how to measure dissolved oxygen concentrations or redox potential, to ensure consistency.		Hence, as a first step before defining any thresholds, the group recommends that a discussion needs to be had, on which is the most relevant metric to be used. CRO involvement is recommended, since it is expected that they would have historical start and end data across a variety of test substances. It would also be important to ask the CROs how the different measurements are made and identify the different instrumentation.	
The OECD 309 TG suggests the inclusion of **sterile controls**, however, provides no information as to how to prepare a sterile control; for example consideration of whether additional consumables should be incorporated in-line 0.2 µm filters before an exposure vessels in a flow-through system; the methods that would confirm that sterilisation measures established a 'non-viable system' which was	A detailed discussion took place around the definition of sterile control, or whether it should be more accurately described as an abiotic control. For the sake of consistency, it was recommended that the guideline should mandate when sterile controls are needed, with a clear rationale/guidance around the utility of such controls. A new section should be included in the guidance which would include	Not discussed	The group discussed that information on sterile controls likely is available with CROs, hence the necessary step to develop the required guidance is through discussions with CROs. A state of the knowledge could be complied, which would include the key points identified under the solution. Efficacy of the different methods could be identified, and prioritised.	<1

maintained throughout the exposure period. The guideline also does not state when sterile controls are required, nor does it indicate the purpose of the sterile controls. The group concluded that this challenge is of **high** impact, especially for systems in which suspended sediment is introduced.	1. Methods available for setting up sterile controls. 2. Clarity on whether methods are suitable for both the suspended sediment test **and** the pelagic test. 3. Guidance on how to assess validity of sterile controls (e.g., maintaining pH). 4. Number of flasks (at least duplicates). 5. Relevant time points when flasks should be sacrificed.			
Par. 30 of the OECD 309 TG states that "*usually a minimum of five sampling points are required during the degradation phase*". However, especially for slowly degrading substances, the selection of five sampling points is difficult to establish. Additionally, it is recognised that five sampling points can be a limitation in the statistical analysis. For slowly degradation substances, this challenge was concluded to have **high** impact.	The solution identified was simply a re-write of the relevant paragraph in the TG, with the recommendation for the inclusion of preliminary work. As such, data from pre-testing could be used to establish the sampling points, and the number of replicates needed to optimise the derivation of the kinetics. The group also discussed the possibility to evaluate different sampling options based on the purpose of the test. E.g. if the goal is to evaluate the disappearance of the parent, or to identify metabolites,	Not discussed	In development of the new paragraph, the group suggested that it would be relevant to check other existing guidelines for information. Also, it was recommended to include CROs, as they would have relevant insights as to the number of sampling points needed, and how these are set out.	<1

	the sampling options may vary.			
Par. 18 of the OECD 309 TG provides guidance around the collection and transport of surface water. The guidance specifically states that *"if it is known that the aquatic environment has been contaminated with the test substance or its structural analogues within the previous four years, it should not be used .....*". It is recognised that this text is difficult to justify for industrial chemicals in surface water. The key challenge highlighted was that this text leads to a conservative assessment that is potentially out of scope of the TG, particularly for substances that have diffuse emissions. A balance between a "representative" water source and a "conservative" water source needs to be established in the guideline. The group concluded that this challenge is of **medium** impact.	The group discussed that the text highlighted under this challenge is likely legacy text from plant protection product assessment. The relevance of the text in the OECD 309 TG vs the OECD 307 TG needs to be examined and should potentially be revised to align with the overall goal of the TG. Another solution that was discussed, is the potential to have a representative / 'standardised' inoculum, with a more robust microbiology. This would require a better description of the inoculum, potentially through a better description of relevant locations or inoculum characterisation.	The plenary presentation given by Michael McLachlan provide potential insights into the relevant surface water, since they assessed differences in biodegradation from upstream and downstream sources.	Not discussed	<1

Par. 23 of the OECD 309 TG states that poorly water-soluble substances can be dosed with a volatile organic solvent. The paragraph also states that when an organic solvent is used, solvent controls should be *"treated similarly to active test vessels amended with test substance in solvent carrier."* The TG, however, does not state that solvent controls must be included, and hence are not always included in the test set-up. This can lead to difficulties when interpreting the results, due to a lack of knowledge around the influence of the solvent. The group concluded that this challenge is of **medium** impact.	The solution to this challenge is an update to the TG text to state that solvent controls should be mandated whenever a solvent is used for dosing. This should provide sufficient clarity.	Not discussed	Not discussed	<1
Par. 5 of the OECD 309 TG states that *"The concentrations should differ from each other by a factor of 5 to 10 and should represent the expected range of concentrations in the environment. The maximum concentration of the test*	The group discussed that while the concentrations were generally suitable, the wording in this paragraph could be rephrased to state: "low environmentally relevant concentrations" was suggested as the solution. Options to work test lower concentrations should be	Not discussed	Not discussed	

substance should not exceed 100 µg/L, but maximum test concentrations below 10 µg/L or less are preferred to ensure that the biodegradation follows first order kinetics." Challenges identified with the wording in this paragraph include that the wording is too specific, and at times limiting. The group concluded that the impact of this challenge to be **low.**	provided, if analytical techniques of sufficient sensitivity are established.			
Par. 5 of the OECD 309 TG states that "*At least two different concentrations of test substance should be used in order to determine the degradation kinetics."* It was unclear to the group why this was listed as a challenge since, in general, it was discussed that testing two concentrations was not an issue. The group concluded the impact of this challenge to be **low,** and agreed that this text was still	The group agreed that there was indeed value in testing a chemical at two different concentrations. An alternative option of potentially performed the test with two different inocula sources may be more beneficial than testing two different concentrations, was also discussed. The group recommended that text could be included in the TG to allow for the omission of a second concentration, provided that a justification is given.	Not discussed	Not discussed	

valid within the guideline, and recommended keeping it in.				
Testing should always be performed below the limit of solubility. This is a challenge for substances with low water solubility, as often there would not be sufficient radioactivity in the test system. An assumed solution would be to choose sufficiently high specific activity; however, it should be noted that this is not always technically feasible. **Impact not discussed.**	Provided that it is not possible to increase the specific activity of the test substance, a solution could be solely high-resolution liquid chromatography mass spectrometry which offers far higher sensitivity than radio-detection. Stable and radio-isotope patterns differ between radiolabelled and non-labelled compounds. In that case stable isotope labelling such as ^{13}C or ^{15}N could be used, but analysis can be more complex due to the higher natural abundance of the ^{13}C-isotope. It was flagged that this is an issue linked to the applicability domain issue, and a better definition of the applicability domain could help identify when an alternative solution may be needed.	Not discussed	Not discussed	
The OECD 309 TG does not specify which models to use to estimate the kinetics and indicates that the biodegradation rate constant should be determined from the	The group disagreed that this was a challenge of the test and agreed that the TG should not specify which models to use. The group agreed that the TG should remain unchanged and state that the most		No next steps needed	

period showing the fastest degradation. This has **no impact** on the outcome of the test.	appropriate models should be used. While it was recognised that the revision of the OECD TG 309 could be an opportunity to introduce new models to derive kinetic information, it was highlighted that models (e.g. FOCUS) are updated at a higher frequency than the TG, leading to a potential situation where the TG recommends outdated models. Additionally, the group agreed that the OECD TG should not dictate which aspect of the results should be used, but that this should be part of implementation guidelines by the relevant regulatory authority.			

Relevance pillar

Challenge	Solution	Impact of challenge on relevance (high/medium/low)	Relevant research	Next steps for OECD Implementation	Timeline (immediate, 1-3 yrs, >3 yrs)
The guideline is unclear on the limits for the storage of surface water after it has been sampled.	Since the inoculum composition drifts quickly from its initial composition during storage, it could be recommended to use the	high/medium	Eco 55	Suggestion to maintain a layer of sediment with the stored water sample but take only the water fraction for the test	0

	inoculum within 24 h (OECD 301 gives a max of 5-7 days).				
Changing the testing temperature from the environmental temperature has a high impact on the test. The guideline does not recommend sampling the inoculum at a temperature close to the test temperature, while it may be that bacteria colonies that adapted at 12°C are not the same as at 20°C (psychrophile vs mesophile).	Sample the inoculum at a temperature close to the test temperature,	High		Decide on the storage/acclimation temperature. Depending on the storage duration, it may be preferable to store the sample at 4° C or at the test temperature or at the source temperature. Determine whether testing at the source temperature would avoid microbial adaptation. This may be difficult to implement in a laboratory (having many temperature options). One could then correct the half-life with Arrhenius equation. This assumes temperature is the driver of the biodegradation rate. Or one could use the half-life value at that particular site and temperature.	0
The microbial diversity may be restricted in smaller test vessels.	Propose larger rather than smaller vessels - greater volume will theoretically lead to a broader microbial diversity and will also reduce wall-/apparatus-effects. One could also prolong the test or use sequential batches, although competent degraders may be lost over	Low (mainly reflects growth-linked biodegradation and not cometabolic conditions)	Study on 301F showed that vessel changed biodiversity -	One could use larger vessels, although this may not be practical in the laboratory and may need more research to identify the proper metric for microbial diversity.	

	time.				
The “relevance” of the test system may change over the duration of the study. Specifically, the microbial heterogeneity and abundance may change from the Day of sampling; Day 0 of in-life phase; Intermediate and End of in-life phase.	We currently use reference substances (although these might be different microbes than the target substance). It is possible to see this via the kinetics of the test (the most relevant indicator of the integrity/stability of the microbial community is something that we already measure, namely the slope of the ln concentration vs. time plot (i.e., rate constant). If that changes during the experiment, then the key function of the microbial community has changed).	Low		If the concern is the change in the test system over time, perhaps it is more "relevant” to have a sequential batch reactor (adding fresh inoculum periodically) to "simulate” an environmental system.	
Do not use exaggerated test concentrations to develop metabolites for identification – must try to stay at as low concentrations as possible, within analytical constraints (to avoid moving beyond co-metabolic range).	Be clear about the objective of the test – cannot do simulation and transformation in one test.	High		Make the test objectives clear. Split up the guidance into 3 sections: mineralisation, co-metabolic simulation, and identification of transformation products. Then link to the reporting requirements and label the type of half-life that is generated	0
There is no guidance on measurement and	Oxygen concentrations are usually maintained with the	Low			

maintenance of oxygen concentrations in closed bottle tests.	relatively low endogenous respiration. It is advised to saturate oxygen of the water phase at the beginning. This may impact the pH of the test.				
There is currently no characterization of the microbial diversity to ensure that the inoculum represents the breadth of metabolic pathways relevant to the test substance.	Possible to have reference substances more appropriate for the test substance.	Low		For understanding variability around half-life (taking Michael's presentation) this type of information could inform on how to interpret the half-life in a global view Can make this a research project; this is for pathway/transformation information. We use cfu/mL but could use other methods. However, is it useful to get this data? Appropriate indicators of biodiversity are not needed for the validity of the result but for interpretation. Implementation may be difficult – there is more discussion about surface water characterization in the guideline which can be applied.	
The guideline does not define criteria or parameters for diffuse light.	A clearer definition or description of diffuse light OR delete the option for illumination altogether	medium	Hand, L. H., & Moreland, H. (2014). ET&C, 33, 516–524. Have Canadian guideline for type of light (Canadian Guidelines for	What type of light (non-UV)? Could use more research to see impact on different tpes of chemicals?	0

			Determinate Environmental Chemistry and Fate of Pesticides) Use euphotic zone definition of that sampling point; timing of the light – make this a more defined part of the guidelines		
The TG has mineralisation as primary objective and primary degradation and identification of degradation products as secondary objective	Emphasize the relevance of the determination of the primary degradation (+ transformation products) with accompanied detailed test setup and reporting obligation including mass balance.	high		See comment above	0
While the test results are being applied to a broad range of environmental conditions, there is only a requirement to test one surface water at two concentrations.	To consider increasing the number of surface waters to be tested to increase the relevance. OECD 308 and 307 include requirement for multiple test environments.	medium		Guidance already says that samples can be selected according to use of the data. But no practical definition of what sites to sample. To capture multiple potential receiving waters, provide further instruction in the guideline to define the sampled sites depending on the use of the test result. (could use other pre-tests to help define sample sites).	1-3
There is little guidance on what is a pelagic test and Suspended sediment test:	Specify test conditions separately for pelagic and sediment test e.g. Amount	high		Provide clearer definition on what is the permissible suspended solids concentration allowed in a pelagic	0-1

(0.01 to 1 g/L dry weigh)	of suspended solids.			test (in suspended sediment addition 0.01-1 g/L, so would pelagic be ???? g/L, different for lake/river/sea?). Explain the utility of adding sediment based on the goal of the test, e.g., co-metabolism (addition of sediment may increase precision and sensitivity and not environmental relevance)	
Aim of the potential modification to increase relevance	Any modification proposed should not make the test more favorable for degradation but increase robustness and repeatability. Increase in degradation potential of the test setup is not equal to increasing its relevance.	NA		(This was for our discussion, not for guideline)	
Presently, the concentration range given in OECD TG 309 is 100 µg/L and preferably in the range of <1-10 µg/L and thus above environmentally relevant concentrations.	If technically feasible include lower concentrations into ring test - Test as low as you can (regardless of environmental relevance, depending on the substance, depending on the goal) depending on the analytical feasibility (not limited to radiolabelled).	high		Use environmentally relevant test substance concentrations. This will be more specific to each separate objective of the guideline (co-metabolism should be as low as possible, transformation should be much higher)	0
What is the comparability		Low, not relevant for this			

of results to degradation observed in the field?		guideline			
It is not clear if the test misses relevant environmental processes, e.g. impact of flow, water-sediment interactions, influence of flocs, biofilms. Is a more holistic approach needed?	The introduction makes clear what you can/cannot measure with this test.	Low, not relevant for this guideline, since this is only a pelagic test.			
The relevance of the fixed water: sediment ratio is not clear.	(see above discussions)	Low, not relevant for this guideline (what is real environment)			
Looking at only non-adapted phase (linear degradation at beginning) may not give transformation products.	Solution: separate the different goals	Low (addressed elsewhere)			
There should be some kind of analytical method (not just standard method) in the test system	There is some text in the guideline already and laboratories can certified as being GLP	Low			
A measure of the maximum or initial degradation kinetics should be provided, perhaps linked to test goal, or if the goal is for co-metabolic vs. metabolic transformation.	Separate the different goals; will have 2 different rates (does maximum rate give environmental potential but not environmental relevance?).	High		Much more discussion needed to understand application of the different rates.	1-3
Testing of mixtures of chemicals is not covered by the guideline		Low, typically test guidelines are intended for testing of individual chemicals and not of			

		composed mixtures. Mixture testing could potentially be addressed in a separate guidance document. The testing of mixtures carries important potential benefits in terms of efficiency but may impact relevance if the presence of other chemicals affects degradation.			
The current guidance states that “Ideally, the radiolabelled mass balance should range from 90% to 110%, whereas the analytical accuracy should lead to an initial recovery of between 70% and 110% for non-labelled test substances.” However, it is not clear why the mass balance needs to be provided.	Should be as complete as possible – include more detail - tie to objective	high		The mass balance requirements should be tied to the specific objective of the test being performed.	0

APPENDIX E: QUESTIONS AND COMMENTS FROM ONLINE PARTICIPANTS

DAY 1:

A) Besides the reduction of lag phase after pre-adapt, will the kinetic rate increase?

B) Will the kinetic rate be over-estimated when using high exposure concentration?

C) Some requirements for the test system mentioned are contradictory and therefore a challenge to the study's performance. E.g. microorganisms should not be adapted (so coming from pristine waters), but suspended solids should reflect large rivers (to help survive the microorganisms); suspended solids are important, but particle surfaces should be as little as possible to avoid adsorption...

D) Question to Lars Toräng: According to the conclusion of your presentation, overestimation of biodeg rates at high concentrations may grossly overestimate the rates at environmentally relevant concentrations. Is this conclusion only based on the examples in the presentation or rather a general finding? Sometimes the contrary is found in OECD 309 studies, at least when comparing 5-10 µg/L versus 50-100 µg/L.

E) Lars's Answer: This conclusion is related to a general finding in my Ph.D. However, I agree that this is not always the case if adaption at high concentration (100 µg/L) is absent. In these cases, it can be very difficult to detect any degradation at all at test concentration of 100 µg/L while a more rapid and first order degradation kinetic can be observed at lower concentrations e.g. 1 µg/L. Variable lag phases (1-180 days have been observed in surface simulation tests) followed by very rapid degradation is however the most frequent scenario that I have observed for the 20-25 different test substances (industrial chemicals, pesticides, pharmaceuticals) that I have investigated.

F) Question to Fola: Is the test substance sorptive?

G) Question to Carolin Seller: Were the tests performed with radiolabeled test materials?

H) Question to Chris Hughes and All: The very high variability in mineralisation of Aniline (a known Readily Biodegradable material ..... and even reference substance) in the ECO55 project. This is concerning to observe - a known Ready Biodeg substance failing 50% of the 309 studies in the ring test. Any Comments?

I) Is the availability and change over time of other sources of carbon for energy and growth in the 309 samples not determining the abundance and viability of bacteria? This would not be very reflective of natural conditions.

J) Comment to Jason Snape: 11 years ago, the Japanese Authorities discovered that their "Golden Standard" for Ready Biodeg was, alas, no longer Readily Biodegradable in the 301C (already quite targeted and poor in microbial distribution). I agree with you - we can probably put this down to changes in environmental "status". This then begs questions about a necessary "renewal" period for relevant Reference Substances!?

K) Question to Dieter Hennecke: The case of the very poorly soluble and probably volatile substance hopefully would warrant an expert statement that testing the substance in the OECD 309 test system is not feasible.

L) Question to Jason Snape: Could you say that environmental inoculum is still environmentally realistic if you concentrate on increasing the microbial biomass? Or could you justify it by saying a chemical could move around in the real environment and so would mix with a much wider variety of microbes than you would get in an unconcentrated sample?

M) Comment: Good morning, all, and many thanks for the very interesting presentations and discussions yesterday! I work at a member state competent authority. My tasks are related to REACH and CLP. There are many issues from yesterday's meeting that I would like to have more time to consider. However, just to mention a couple of first impressions/comments I find relevant for the discussions:

   I) Some indication of viability of inoculum can be obtained from comparison between sterile control and non-sterile test (using whatever relevant parameters that are measured, e.g., mineralisation, decrease in parent concentration, transformation products). Often at least some difference is seen, maybe challenging the claims about non-representative biodegradation potential in the OECD TG 309 and at least opposing the claim that it would be just a hydrolysis test. No degradation (and no difference between sterile and non-sterile) can mean that the substance is non-degradable, it does not necessarily mean that the test is non-reliable.

   II) I disagree with the statement that 309 would be just one more screening test. One key difference in OECD 309 and screening test is test concentration. 309 aims at determination of degradation half-life and therefore uses low concentration compared to what is generally used in screening tests.

   III) The addition of suspended sediment may increase biodegradation potential; therefore, I find it difficult using such tests to support a P or not vP conclusion.

   IV) In one of the presentations, it was mentioned (if I remember correctly) that it is difficult to know what amount of suspended sediment is allowed in the pelagic test. To my understanding, for the pelagic test, it is not allowed to add suspended sediment. For REACH purpose the water for the pelagic test should be selected so that its original suspended matter content is in the range of the ECHA guidance.

   V) As mentioned in one of the comments at the end of yesterday's meeting, the relevance of the methodology used to reflect the biodegradation capacity in the environment should be taken into account. We should not use methods which increase the biodegradation capacity in the test so that it is not anymore representative of that occurring in the environment.

DAY 2:

A) Comment on Michael's presentation: [addition of sediment] "does not measure persistence in water - regulatory misalignment". That's not necessarily the case. If I'm not mistaken, neither REACH nor CLP legal texts tell you not to add sediment to your water, and neither REACH nor CLP tell you that you cannot directly compare the resulting half-lives with the P/vP criteria. Indeed, REACH and CLP tell you to compare the half-

life in SURFACE water, not PELAGIC water. According to OECD 309, adding up to 1 g/L sediment is still a surface water test. In addition, REACH and CLP tell you to perform the tests under relevant conditions, i.e. those conditions that allow for an OBJECTIVE assessment of the intrinsic property of the substance (ECHA Board of Appeal decision). If the pelagic test is too variable and too stringent in comparison to what really happens in the environment, can we say that the pelagic test allows for an objective P/vP assessment? If modifications of the test allow for a more OBJECTIVE comparison and it is still within the limits of the TG, then changes in the practices can already be accepted today without any modification of any text. But I understand more work on the sediment addition may be needed to better understand the variability of the results. Alternatively, I wonder whether 308 and 307 could be sufficient for the P/vP assessment.

B) Question to Martin Bruggeman: when does the formation of biofilms become an operational problem within the test set-up?

C) Question to Michael McLachlan: Thank you for the excellent presentation. Multiple substances have the potential to sorption to solids. Could you comment on a likely upper boundary of sorption applicable to the presented approach?

D) Question to Michael McLachlan: When you are talking about the pelagic test, are you talking about surface water without any suspended solids? For REACH purposes it is recommended to work with SPM around 10-20 mg/L, representative of the level of suspended solids in EU surface water. In the context of the OECD TG 309, this would make it a suspended sediment test and not a pelagic test.

E) Question to Chris Hughes: You proposed to analyse LRI ECO55 results on reference substances – You collected a selection of historical ref. substance data during that initiative. With this new discussion, do you think it might be able to obtain more historical data for ref. substance (+ sample parameters) from OECD 309 studies from a broader community to allow a more powerful analysis of this data? The data can be anonymised.

F) Question to Delina Lyon: Perhaps I missed it in your RELEVANCE review …… but the use of the traditional Ref. substances (Na Benzoate, Aniline) and what they may (or may not) tell you as far as the vague validity criteria are concerned wrt to the output results for the test item ……. Are the results relevant or not?

G) Does “co-metabolism” here mean that degradation is not by specific degraders but generally by enzymes produced by microbes primarily utilizing other substrates than the test chemical? If so, degradation of reference substance might be more relevant (e.g. as validity criterion) for “co-metabolic” degradation than “metabolic” degradation.

H) General Question: Almost all scientific presentations showed that the classical OECD 309 test is not suitable for deriving persistence endpoints because it is not discriminating between compounds (all are persistent), it has a lot of practical challenges, the variability of results can even not be improved by adding sediment, the relevance for the environment is questionable, etc. - how should more reporting (in tabular form), more samplings, more replicates, more oxygen controls, more detailed reporting of test parameters help to overcome the issue that the test is just not suitable for the evaluation it is used for in the regulatory context?

I) Are we condemned to performing full in-field mesocosm studies?

J) Measures Required Revisions:

- Measures to reduce and control variability.
- Strengthen and clarify the validity criteria.
- Clarify the use and interpretations of the outcome, such as a cut-off mineralization rate of the overall test system without specification of test substance, co-metabolites and transformation product; primary degradation half-life of test substance =/= DegT 50 of test substance. In case no half-life of major transformation products is determined, the obtained DT50 is not sufficient making conclusion of persistence of test substance.

ECETOC PUBLISHED REPORTS

The full catalogue of ECETOC publications can be found on the ECETOC website:
http://www.ecetoc.org/publications

Responsible Editor:
Blanca Serrano Ramon
ECETOC AISBL
Rue Belliard 40
B-1040 Brussels, Belgium
VAT: BE 0418344469
www.ecetoc.org
D-2018-3001-

ECETOC, is a collaborative space for leading scientists from industry, academia and governments to develop and promote practical, trusted and sustainable solutions to scientific challenges which are valuable to industry, as well as to the regulatory community and society in general.