Executive summary:

Authorization and restriction of chemicals (REACH) requires demonstration of the safe manufacture of chemicals and their safe use throughout the supply chain. REACH is based on the precautionary principle, but aims to achieve a proper balance between societal, economic and environmental objectives. Both new and existing chemicals will be evaluated within REACH, amongst others aiming to efficiently use the scarce and scattered information available on environmental fate and effects of chemicals. REACH thus aims at closing huge gaps of knowledge on physicochemical properties and adverse effects of large numbers of chemicals. Thereupon REACH aims to reduce animal testing by optimized use of qualitative and quantitative information on related compounds. The REACH proposals advocate the use of non-animal testing methods, but guidance is needed on how these methods should be used.

It is within this context that CADASTER aims at providing the practical guidance to integrated risk assessment within REACH by performing a full hazard assessment for chemicals belonging to four compound classes. The main goal is to exemplify the integration of information, models and strategies for carrying out safety-, hazard- and risk assessments for large numbers of substances. Real risk estimates are delivered according to the basic philosophy of REACH of minimizing animal testing, costs, and time. CADASTER thus shows how to increase the use of non-testing information for regulatory decision whilst meeting the main challenge of quantifying and reducing uncertainty. On the basis of fusing the research findings with other ongoing research and regulatory developments, recommendations on a viable management strategy for optimized testing and in-silico modeling of hazardous organic chemicals, are provided. The focus of the activities was on assessing and quantifying uncertainty and variability in probabilistic risk assessment, as introduced by the use of non-testing information.

To achieve the main goals set for CADASTER, an intensive screening of the available experimental data and predictive (QSAR) models for endpoints relevant for risk assessment of the four chemical classes that were the topic of study within CADASTER, was initially performed. A subsequent gap analysis identified the most essential data and models that are lacking for performing future fate and effect assessment. Subsequent integration of the predictive tools thus obtained allowed for establishing a quantitative framework for integration of predictive tools within risk assessment. Case studies on a variety of endpoints relevant for chemical hazard and risk assessment for a blend of chemicals form within the four chemicals classes that were the topic of study within CADASTER, exemplified the application of testing alternatives and the subsequent uncertainties and variability.

The case studies included:
- Assessing uncertainty and propagation of uncertainty in environmental fate modelling;
- ranking of compounds on the basis of their environmental hazard;
- ranking of compounds on the basis of their environmental risk, as calculated by combining effect assessment with fate assessment, whilst taking uncertainty and variability in input data into account;
- assessing uncertainties in risk assessment based on QSARs;
- applying read across approaches to chemical classes of a heterogeneous nature without much structural resemblance;

- prioritization based on evaluation of persistence, bioaccumulation potential and toxicity of chemicals and associate uncertainty analysis, and based on hazard assessment;
- prioritization of chemicals based on risk assessment;

All data, models, fate and effect assessments, risk assessments, and additional tools and methods developed within the project are operationalized in the CADASTER website http://www.cadaster.eu. The website provides the user with the opportunity to develop his/her new models and perform risk assessment on chemicals that were not within the domain of CADASTER.

Project Context and Objectives:

Authorization and restriction of chemicals (REACH) requires demonstration of the safe manufacture of chemicals and their safe use throughout the supply chain. REACH is based on the precautionary principle, but aims to achieve a proper balance between societal, economic and environmental objectives. Both new and existing chemicals will be evaluated within REACH, amongst others aiming to efficiently use the scarce and scattered information available on environmental fate and effects of chemicals. REACH thus aims at closing huge gaps of knowledge on physicochemical properties and adverse effects of large numbers of chemicals. Thereupon REACH aims to reduce animal testing by optimized use of qualitative and quantitative information on related compounds.

The REACH proposals advocate the use of non-animal testing methods, but guidance is needed on how these methods should be used. Operational procedures are to be developed, tested, and disseminated that guide a transparent and scientifically sound evaluation of chemical substances in a risk-driven, context-specific and substance-tailored manner. Various alternatives are available to supplement existing data or even to substitute for lacking toxicity data, as advocated within REACH in order to reduce unnecessary animal testing. In line with the paradigm shift that has taken place when establishing REACH of performing risk management instead of risk assessment, the concept of Intelligent (or: Integrated) Testing Strategies (ITS) was developed to optimize the integration of available experimental data and alternative means of assessing adverse effects, whilst adhering to one of the main objectives of REACH of minimizing the use of test animals. ITS are the most efficient way to obtain the necessary information to carry out hazard and risk assessments of large numbers of chemicals, while reducing costs to industry and minimising animal testing. Intelligent testing strategies are integrated approaches comprising of multiple elements aimed at speeding up the risk assessment process while reducing costs and animal tests.

Within an ITS, all alternatives to experimental testing are integrated and experimental testing is carried out only as a last resort, i.e.when no information at all or when no reliable information can be obtained by means of any of the following alternatives:
- Quantitative Structure Activity Relationships or Quantitative Property Activity Relationships (QSAR or QSPR);
- Read across, commonly performed by interpolating information on related compounds;
- In-vitro testing;
- Exposure based waiving, providing evidence that biota are not exposed to the chemical of interest, or at concentrations well below the no observed effect level.

Concerted action and intensive efforts are needed to operationalize all possible alternatives into a workable, consensually acceptable, and scientifically sound strategy for hazard and risk assessment of large numbers of chemicals. The production of guidance and (web-based) tools is essential in this respect. So far, the use of non-testing methods in the European regulatory context is quite limited and fragmented. Reasons include the lack of distinct application criteria and guidance, and the fact that uncertainty has not been addressed rigorously. Industry is primarily made responsible for carrying out the risk assessments, and

practical guidance is therefore needed on how to apply the elements of newly derived testing strategies in a consistent manner.

It is within this context that the CADASTER project was designed. CADASTER aims at providing practical guidance to integrated hazard and risk assessment procedures by exemplifying a hazard and risk assessment for chemicals belonging to four specific compound classes by integrating the various tools that are made available within the project for each of the four compound classes. The tools and the underlying data and models are made available via the project website http://www.cadaster.eu as an on-line and standalone tool for development, publishing and use of QSAR models for REACH. The CADASTER tools predict physico-chemical properties and toxicities for four analysed classes. The predictions provided are compatible with the OECD QSAR Application Toolbox and the EPI SuiteTM Toolbox developed by the EPA's Office of Pollution Prevention Toxics and the Syracuse Research Corporation (SRC). Operational procedures were developed that explicitly take account of variability and uncertainty in data and in models. The objectives of CADASTER are in line with the basic idea of REACH to obtain the information needed for carrying out hazard and risk assessments for large numbers of substances by integrating multiple methods and approaches with the aim to minimize testing, costs, and time.

CADASTER facilitates the selection of the relevant fate and effect parameters as it supplements the existing database on fate and effect properties of the following compound classes that were selected as the chemical classes of choice for CADASTER:
1- Polybrominated diphenylethers (PBDE), typically being a class of hydrophobic chemicals, some of which have been used as flame retardants, that pose a threat to man and the environment.
2- Perfluoroalkylated substances and their transformation products, like perfluoroalkylated sulfon¬amides, alkanoic acids, sulfonates. Fluorinated compounds are typically a class of persistent, relatively hydrophilic compounds that may be toxic for man and environment.
3- Substituted musks/fragrances, being a heterogenic group of chemicals of varying composition. Examples include substituted benzophenones, polycyclic musks, terpene derivatives. In view of their typical use pattern, the chemicals have a common emission pattern in the environment.
4- Triazoles/benzotriazoles, a class of chemicals that are increasingly used as pesticides and anti-corrosives.

The main goal of CADASTER is to exemplify the integration of information, models and strategies for carrying out safety-, hazard- and risk assessments for large numbers of substances to the new categories of risk assessors within REACH. Real risk estimates are delivered according to the basic philosophy of REACH of minimizing animal testing, costs, and time. CADASTER thus shows how to increase the use of non-testing information for regulatory decision whilst meeting the main challenge of quantifying and reducing the level of uncertainty. By fusing the research findings with other ongoing research and regulatory developments, recommendations on a viable management strategy for optimized testing and in-silico modeling of hazardous organic chemicals, are provided. The focus of the activities was on assessing and quantifying uncertainty and variability in probabilistic risk assessment, as introduced by the use of non-testing information.

To achieve the main goals set for CADASTER, four objectives were identified and operationalized within four workpackages:

Objective 1 (workpackage 2): Collection of data and models

Sub-activities within this objective included:
1) Collection of existing experimental data on the most common regulatory endpoints considered in the Screening Initial Data Set Dossier (SIDS - internationally agreed data on the intrinsic hazards of a chemical) for the four classes of chemicals selected.
2) Collection of existing (Q)SARs for the endpoints considered in the SIDS for the four classes of chemicals selected.
3) Generation of new data on endpoints and chemicals for which, as identified in workpackage 3, insufficient data are available for model validation and proper hazard/risk assessment.
4) Development of a database on experimental data and (Q)SAR models for dissemination of the results of the activities 1 - 3 to all project partners and to other interested bodies.

Objective 2 (workpackage 3): Development and validation of QSAR models

Sub-activities within this objective include:
1) Evaluation of the performance of existing QSARs for the chemical classes studied.
2) Similarity analysis and multivariate ranking methods for identification of priority chemicals in the selected classes to orient the experimental tests in WP2.
3) Development of new QSARs where gaps are identified due to lack of existing models or due to models of insufficient quality.
4) Documentation of the performance of the (final) models selected and developed.
Sub-activities 1 and 3 are directly linked with the activities in WP4, as the evaluation of existing QSARs as well as the identification of models of insufficient quality will be carried out within the requirements for reliable application in chemicals legislation.

Objective 3 (workpackage 4): Integration of QSARs within hazard and risk assessment

The objectives of this WP are:
1) To propose how QSAR models can be applied and integrated into a probabilistic risk assessment framework,
2) To explore the feasibility of economic valuation of chemicals impacts,
3) To evaluate how QSAR models can meet the legal requirements of current and upcoming chemicals legislation (REACH),
4) To synthesise research findings and recommendations from the other work-packages in the programme and prioritize chemicals.

Objective 4 (workpackage 5): Outreach via development of website, newsletters/workshop(s) and standalone tools for dissemination of project results

Sub-activities within this objective include:
1) Development of an on-line and standalone Decision Support System (DSS) for development, publishing and use of QSAR/QSPR models for REACH.
2) Integration of the developed models with the QSAR Application Toolbox developed by the OECD and establishing the compatibility of the models with the (Q)SAR Model Reporting Format (QMRF) format.
3) Provision of a sustainable dissemination of project results by the http://www and as standalone tools.

4) Development of communication mechanisms including newsletters and workshops.

Overall, the core of the project is constituted by objective 3 (workpackage 4): Integration of QSARs within hazard and risk assessment. It should be noted that this title is misleading as the integration is not restricted to QSAR, but also includes the additional non-testing approaches that are available for implementation in REACH, like read across and exposure based waiving. Data and predictive models (Workpackages 2 and 3) are the core ingredients needed for future hazard assessment for chemicals within the four classes of compounds identified, and were supplemented in Workpackage 4 with additional tools potentially available within REACH (like read across, in vivo-in vitro extrapolation, etc.).

An intensive screening of the available experimental data and predictive (QSAR) models for endpoints relevant for risk assessment of the four chemical classes that were the topic of study within CADASTER, was initially performed. A subsequent gap analysis identified the most essential data and models that are lacking for performing future fate and effect assessment. Subsequent integration of the predictive tools thus obtained allowed for establishing a quantitative framework for integration of predictive tools within risk assessment. Case studies on a variety of endpoints relevant for chemical hazard and risk assessment for a blend of chemicals from within the four chemicals classes that were the topic of study within CADASTER, exemplified the application of testing alternatives and the subsequent uncertainties and variability.

The case studies included:
- Assessing uncertainty and propagation of uncertainty in environmental fate modelling;
- ranking of compounds on the basis of their environmental hazard;
- ranking of compounds on the basis of their environmental risk, as calculated by combining effect assessment with fate assessment, whilst taking uncertainty and variability in input data into account;
- assessing uncertainties in risk assessment based on QSARs;
- applying read across approaches to chemical classes of a heterogeneous nature without much structural resemblance;
- prioritization based on evaluation of persistence, bioaccumulation potential and toxicity of chemicals and associate uncertainty analysis, and based on hazard assessment;
- prioritization of chemicals based on risk assessment;

All data, models, fate and effect assessments, risk assessments, and additional tools and methods developed within the project are operationalized in the CADASTER website http://www.cadaster.eu (Workpackage 5). The website provides the user with the opportunity to develop his/her new models and perform risk assessment on chemicals that were not within the domain of CADASTER.

Based on the case studies and the additional integrating activities with regard to alternatives to testing in a probabilistic framework, the following recommendations were made:

General recommendations
- Support the use of QSAR predictions and other alternatives in risk assessment under REACH but continue to acknowledge the weaknesses/uncertainties of the alternatives to avoid any misuse.

- Support the use of QSAR predictions by stressing their advantages in relation to other in-Silico techniques.

Recommendations with respect to the development of future QSARs to facilitate the integration in risk assessment
- Build QSARs that are probabilistic, such as Bayesian modeling.
- Build QSARs that model sources of variation in hierarchical levels to open up for the consideration of variability in experimental data, varying quality in experimental data.
- Build QSAR integrated assessment models that in a hierarchical fashion integrate QSAR data, available experimental data, available data from other testing alternatives, and sources of variability and uncertainty to properly assess the hazard and risk endpoints.

Recommendations with respect to reporting of information of testing alternatives
- Extend/modify information requirements and reporting formats such that uncertainty becomes naturally associated to a prediction.
- Consider uncertainty to be quantified by probabilities (e.g. it should be possible to attach a random sample from a predictive distribution to open up for Monte Carlo simulation).
- Include the assessment of quantitative uncertainty including evaluation of the approach taken (its theoretical bases and if possible by evaluation on QSAR data).
- Include a QSAR-specific recommended approach to judge the confidence in QSAR predictions (based on some metric to evaluate predictive reliability and reference cut-offs to aid judgment).

Recommendations with respect to the practical integration of testing alternatives into risk assessment
- There is a need for methods to propagate both qualitative and quantitative uncertainty associated to input parameters to assessment supporting decision making.
- There is a need to find simple rules of thumb that can be used to facilitate the reporting of uncertainty in QSAR predictions.

Recommendations for future activities
- Support workshops on the assessment and consideration of uncertainty in testing alternatives with practical training.
- Support the development of more case-studies to show the impact and usefulness of considering uncertainty in data generated by means of testing alternatives, in the regulatory decision context under REACH.
- Support more research on the QSAR integrated assessments and development of user-friendly tools for uncertainty analysis and evaluation of quality assessment output with respect to quality in available background knowledge.

Project Results:
General introduction
The REACH proposals advocate the use of non-animal testing methods, but guidance is needed on how these methods should be used. As an example: the REACH system requires that non-animal methods should be used for the majority of tests in the 1-10 tonne band, even though such methods are not yet available for most of the endpoints relevant at this tonnage. In an attempt to resolve the issue of lack of guidance, the European Commission made suggestions on how reduction, refinement and replacement strategies could be applied to animal use in the REACH system:

1 – Encouragement of the use of validated in silico techniques such as (Q)SAR models.
2 – Encouragement of the development of new in vitro test methods.
3 – Minimization of the actual numbers of animals used in the required tests, and replacement of animal tests wherever possible by alternative methods.
4 – Formation of Substance Information Exchange Forums (SIEFs) for the obligatory provision of data and cost sharing.
5 - Requirement of official sanctioning of proposals for tests for compounds with production volumes of above 100 tonnes to minimize animal testing.

Within REACH, operational procedures are to be developed, tested, and disseminated that guide a transparent and scientifically sound evaluation of chemical substances in a risk-driven, context-specific and substance-tailored manner. The procedures include alternative methods such as chemical and biological read-across, in vitro results, in vivo information on analogues, qualitative and quantitative structure-activity relationships (SARs and QSARs, respectively), thresholds of toxicological concern, and exposure-based waiving. Concerted action and intensive efforts are needed to operationalize all possible alternatives into a workable, consensually acceptable, and scientifically sound strategy for hazard and risk assessment of large numbers of chemicals. The production of guidance and (web-based) tools is essential in this respect. So far, the use of non-testing methods in the European regulatory context is quite limited and fragmented. Reasons include the lack of distinct application criteria and guidance, and the fact that uncertainty has not been addressed rigorously. Industry is primarily made responsible for carrying out the risk assessments, and practical guidance is therefore needed on how to apply the elements of newly derived testing strategies in a consistent manner.

CADASTER aims at providing practical guidance to integrated hazard and risk assessment procedures by exemplifying a hazard and risk assessment for chemicals belonging to four specific compound classes by integrating the various tools that are made available within the project for each of the four compound classes. The tools and the underlying data and models are made available via the project website http://www.cadaster.eu as an on-line and standalone tool for development, publishing and use of QSAR models for REACH, compatible with the OECD QSAR Application Toolbox and the EPI SuiteTM Toolbox developed by the EPA's Office of Pollution Prevention Toxics and the Syracuse Research Corporation (SRC). Operational procedures were developed that explicitly take account of variability and uncertainty in data and in models. The objectives of CADASTER are in line with the basic idea of REACH to obtain the information needed for carrying out hazard and risk assessments for large numbers of substances by integrating multiple methods and approaches with the aim to minimize testing, costs, and time.

CADASTER facilitates the selection of the relevant fate and effect parameters as it supplements the existing database on fate and effect properties of the following compound classes that were selected as the chemical classes of choice for CADASTER:
1 – Polybrominated biphenylethers (PBDE), typically being a class of hydrophobic chemicals that pose a threat to man and the environment.
2 - Perfluoroalkylated substances and their transformation products, like perfluoroalkylated sulfon¬amides, alkanoic acids, sulfonates. Fluorinated

compounds are typically a class of persistent, relatively hydrophilic compounds that may be toxic for man and environment.
3 – Substituted musks/fragrances, being a heterogenic group of chemicals of varying composition. Examples include substituted benzophenones, polycyclic musks, terpene derivatives. In view of their typical use pattern, the chemicals have a common emission pattern in the environment.
4 - Triazoles/benzotriazoles, a class of chemicals that are increasingly used as pesticides and anti-corrosives.

The main goal of CADASTER is to exemplify the integration of information, models and strategies for carrying out safety-, hazard- and risk assessments for large numbers of substances to the new categories of risk assessors within REACH. Real risk estimates are delivered according to the basic philosophy of REACH of minimizing animal testing, costs, and time. CADASTER thus shows how to increase the use of non-testing information for regulatory decision whilst meeting the main challenge of quantifying and reducing the level of uncertainty.

General overview of activities within CADASTER

The current regulatory developments within REACH have provided important considerations in the initial design of the CADASTER project. CADASTER is aimed at exemplifying the integration of the various alternatives to experimental testing and the consequences for environmental risk assessment, explicitly considering uncertainties associated with the use of alternatives to replace experimental data. In line with its acronym, the project is designed to provide case studies on the development and application of in-silico techniques for environmental hazard and risk assessment, and to use the project results as an illustration of how to deal with the major limitations and uncertainties related to the implementation of alternatives to testing in hazard and risk assessment.

To this end, the following activities were amongst others performed:

- Collection of existing data and predictive models on the endpoints that are essential for performing hazard and risk assessment of chemicals within REACH (Aim of this activity: filling the research gap of lack of data for endpoints relevant for Environmental Risk Assessment (ERA) and for in silico model building);
- Exploring and making available of (eco)toxicological data, amongst others for the purpose of in silico model development (Aim of this activity: filling the research gap of making data available for in silico model building);
- Assessing the quality of available toxicity data (Aim of this activity: filling the research gap of low or difficult to ascertain quality of available toxicity data, and need of characterization and validation of existing toxicity data);
- Supplementing existing experimental data to allow for in silico model development for endpoints essential for risk assessment, and to allow for validation of existing and newly developed models (Aim of this activity: filling the research gap of lack of data and models for endpoints relevant for ERA and for in silico model building);
- Collection of existing alternatives to experimental testing, with a focus on QSAR models on the endpoints that are essential for performing hazard and risk assessment of chemicals within REACH (Aim of this activity: filling the research gap of lack of models for endpoints relevant for ERA);
- Development of new (QSAR) models and validation of existing and newly developed (QSAR) models, including development of consensus models (Aim of this activity: filling the research gap of lack of models for endpoints relevant for ERA);
- Implementation of tools to estimate the applicability domain of models and to optimize experimental design. This activity includes characterisation of variability and uncertainty of models and underlying, and sensitivity analysis of individual models (Aim of this activity: filling the research gap of lack of definition of the applicability domain of models);
- Development of a computational framework for QSAR based probabilistic risk assessment, including uncertainty analysis of the risk characterisation ratios (Aim of this activity: filling the research gap of need of probabilistic risk assessment);
- Development and public release of the QSPR-THESAURUS Website and associated databases containing all data and models made collected and generated within the CADASTER project (Aim of this activity: filling the research gap of lack of robustness of different database entries for toxicity, consequently leading to different QSAR models and results);

- To improve and validate individual QSAR models, and prepare standardized reporting formats for the models (like QMRF – the QSAR Model Reported Format developed by JRC and implemented in the OECD QSAR Toolbox). Aim of this activity: filling the research gap of lack of validation of individual QSAR models which are incorporated in complex existing tools such as ECOSAR);
- Perform training to risk assessors, national chemicals authorities (particular from Eastern European countries), industry and SMEs on the use of alternative tools for risks assessment in REACH, amongst others demonstrating how the tools developed within CADASTER as well as the models available in the OECD QSAR toolbox can be used to estimate REACH end-points for chemical compounds and thus decrease the number of animal tests. This activity included training on how to develop new models for the assessment of REACH-end points (in particular for new scaffolds of compounds for which there are no reliable QSAR models) and how to use the software developed by the CADASTER project participants (Aim of this activity: filling the gap of need of appropriate training and understanding of personnel, particularly of those dealing with registration dossiers);

The core of the activities was directed towards the following topics:
- Collection and generation of fate and effect data essential for risk assessment;
- Collection and development of predictive models for endpoints essential for risk assessment;
- Development of methodologies for assessment of the applicability domain of models;
- Making data and models available to any outside user via the project websites http://www.cadaster.eu and http://www.qspr-thesaurus.eu.
- Characterisation of uncertainty, variability, model sensitivity;
- Training of (future) risk assessors and outreach of project results.

In line with the four objectives that were identified, the work was organised within four workpackages, supplemented with a separate workpackage (workpackage 1) on project management. The main Science And Technology (S&T)results are provided below, organised per scientific workpackage.

WORKPACKAGE 2 – COLLECTION OF DATA AND MODELS
Within WP2, various milestones and deliverables have been identified to warrant proper monitoring and timely execution of the project:

Deliverables WP2
D2.1 (Month 12) Overview of data on physicochemical properties, fate and environmental effects of chemicals within the four classes of chemicals selected (report)
D2.2 (Month 12) Overview of (Q)SAR models and their specific features for assessing fate and effects (report)
D2.3 (Month 18) Overview of non-testing approaches available for implementation in REACH (report)
D2.4 (Month 36) Establishment of a database on properties and fate/effect parameters of chemicals within the four classes of chemicals selected (report).
D2.5 (Month 38) Overview of new data generated (report).

Milestones WP2
M2.1 Prototype of a user-friendly database on properties and fate/effect parameters operational.

The work within WP2 was subdivided along 4 tasks:
Task 2.1 Collection of existing experimental data.
Task 2.2 Collection of (Q)SAR models and non-testing approaches.
Task 2.3 Generation of new data.
Task 2.4 Establishment of database.

Task 2.1 Collection of existing experimental data
D2.1 (Month 12) Overview of data on physicochemical properties, fate and environmental effects of chemicals within the four classes of chemicals selected

Overview

A data search on all endpoints of relevance was performed for the environmental risk and hazard assessment of the groups of chemicals included in the case studies. Physicochemical properties, environmental fate parameters, and aquatic and terrestrial ecological effect parameters are included, among other available toxicity data. This task was carried out by means of a literature search, supplemented with searches of existing databases on risk and hazard assessment parameters, like IUCLID, AQUIRE, etc. Thereupon, additional data were collected from industry sources and regulatory agencies (Dupont, RIFM).
Activities performed

Existing experimental data on polybrominated diphenylethers (PBDE), perfluoroalkylated substances and their transformation products, substituted musks/fragrances, and triazoles/benzotriazoles were collected from the literature and from existing databases on physico-chemical properties, environmental fate parameters, and aquatic and terrestrial ecological effects parameters. Thereupon, tools for automatic querying of on-line databases were applied, and US_EPA dockets databases with information about four related classes of chemicals (greater than 5,000 documents) were uploaded and made searchable on the web at http://www.cadaster.eu/DocSearch/ using natural language search tools. This tool should help to search relevant information in dockets without a need to open all of them.

Additional data were requested from industry sources, amongst others via a request to the Research Institute for Fragrance Materials (RIFM) for substituted musks/fragrances and via a request to the Dupont Company for polybrominated diphenylethers (PBDE) and polyfluorinated alkylated compounds.

The search was extended to additional biological activities of the chemicals of interest. Data on chemicals containing molecular structures which were considered relevant for the QSAR modeling were also included in the database (for instance: data on flame retardants that are structurally different from the PBDEs that are the topic of interest within CADASTER, polyfluorocompounds in addition to perfluorochemicals, PFCs, etc.). The extension is essential to allow for application of read-across techniques in future stages of the project. All CADASTER project participants involved in data collection have already uploaded the corresponding experimental data in the CADASTER online database:
- Partner 1, RIVM, coordinated the data collection for substituted musks/fragrances
- Partner 2, PHI, coordinated the data collection for triazoles/benzotriazoles (TAZ/BTAZ)

- Partner 3, UI, collected and shared data for all four classes of chemicals
- Partner 4, IVL, coordinated the data collection for the polybrominated diphenylethers (PBDE)
- Partner 6, HMGU, collected and shared data from on-line databases for all four classes of compounds.

The existing experimental data within the four classes of chemicals selected are thus available for use by the project participants.

There are two sets of molecules. The core datasets were provided by UI group. In these datasets all molecules were manually verified and annotated to the 4 CADASTER classes of molecules. There are in total 939 molecules in this set with 4994 records. These data are the core data for model development.

The extended database was created by using simple fragmental definitions to extend the basic set. The selected molecules do not necessarily belong to our 4 classes, but could be used to extend the datasets with molecules having structural similarity. These additional molecules were found using fragment search, e.g. triazole and benzotriazole rings for classes of triazoles/benzotriazoles class, bromobenzene derivatives for polybrominated diphenyl ethers, polyfluorinated alkyl chains for PFC compounds. No search has been done for musks/fragrances due to the intrinsic high diversity of molecules within these classes. While a total number of 939 chemicals were selected for the primary CADASTER classes by UI, the number of extended chemicals was 3682 molecules. These data can be used to validate model for predictions outside of applicability domain as well as for read-across studies.

The extended database includes the previous one and contains 6838 experimental data within the four classes of chemicals:
-Heterogeneous brominated compounds including 209 PBDEs and some other flame retardants: 243 structures and 1093 experimental data.
-Perfluoroalkylated substances (PFC), including their transformation products: 691 structures and 1725 experimental data.
-Substituted musks/fragrances: 533 structures and 1138 experimental data.
-Triazoles and benzotriazoles (TAZ/BTAZ): 447 structures and 2726 experimental data.

All details are available from the CADASTER website.

One of the main conclusions that may be drawn from the detailed information on the data made available in the CADASTER database, is that in general the number of experimental data is quite large.
However, in general only limited data are available for the endpoints of primary interest for environmental risk assessment within REACH, i.e. the SIDS endpoints. Especially data on adverse effects for relevant species are lacking and most toxicity are available for mammals (rats/mice). This again stresses the need for application of read across approaches on top of additional generation of data on SIDS endpoints already foreseen within the CADASTER project.

In line with the Description of Work of CADASTER, the set of experimental data were collected and uploaded in the database during the whole project. This included extension of information on chemicals newly added to the database whenever this was required. Amongst others, a multivariate characterization approach was applied by HIK for

polybrominated diphenyl ethers and the use of mixtures to simultaneously determine BCFs was applied for this purpose.

Task 2.2 Collection of (Q)SAR models and non-testing approaches

D2.2 (Month 12) Overview of (Q)SAR models and their specific features for assessing fate and effects

Overview

A survey of the existing QSAR/QSPR models for the four CADASTER classes of chemicals was completed. The analysis of these models according to the requirements of the 'OECD principles for QSAR validation' for regulatory applicability was the topic of Deliverable 3.2. Publicly available EPI Suite models were also taken into consideration. In this case, it has been assessed, which of these models are reliably applicable to the four classes of chemicals.

At this stage of knowledge, QSPR models are available only for some SIDS physico-chemical properties of BFRs (Henry's low constant, vapor pressure, water solubility, LogKOW, photodegradation rate), while for the other three classes of chemicals EPI Suite models are the only tools available to predict SIDS physico-chemical properties.
The same applies for existing QSAR models, which are predominantly developed for non-SIDS endpoints, such as endocrine disruption (for BFRs and PFCs) or skin sensitization (for fragrances). There is only one QSAR model based on acute toxicity to fish which is developed for a large data set containing a few substituted triazoles. In the absence of ad hoc QSAR models for the four classes of chemicals selected in this project, the ECOSAR estimation program included in EPI Suite could be used to predict acute and chronic toxicity to fish, aquatic invertebrates and algae (SIDS endpoint for ecotoxicity). The problems linked to the applicability of ECOSAR models to BFRs, fragrances, PFCs, TAZs and BTAZs are better explained in the following paragraph.

Activities performed
Below are summarized the existing QSARs collected for individual classes. Tables containing all the useful information and details on the models are attached in Appendix 2. The same models have been verified for their application of the OECD Principles of QSAR in regulation (Deliverable 3.2).
BFRs
QSPRs models have been published for the following physico-chemical and degradation properties:
- Henry's Law Constant, H (SIDS) (Xu et al., 2007);
- Subcooled Liquid Vapor Pressure, PL (SIDS) (Xu et al., 2007; Wania and Dugani, 2003; Wang et al., 2008; Öberg, 2002);
- Water Solubility, WS (SIDS) (Wania and Dugani, 2003);
- Octanol-Air Partition Coefficient, LogKOA (Xu et al., 2007; Wania and Dugani, 2003 ; Wang et al., 2008; Chen et al., 2003 ; Zhao et al., 2005) ;
- Octanol-Water Partition Coefficient, LogKOW (SIDS) (Wania and Dugani, 2003; Braekevelt et al., 2003; Li et al., 2008);
- Photodegradation rate, Kp, measured in different solvents (methanol/water, methanol, hexane) (SIDS) (Niu et al.,2006; Chen et al., 2007 ; Fang et al., 2009);

- Quantum yield, Fp, measured in different solvents (methanol/water, methanol, hexane) (Niu et al.,2006; Chen et al., 2007 ; Fang et al., 2009).

To our knowledge, no ad hoc QSAR models have already been published for ecotoxicity endpoints (acute toxicity to algae, daphnids and fish), probably due to the lack of sufficient amounts of experimental data. However different QSARs have been found in literature for endpoints related to endocrine disrupting activity of BFR, that is:
- Aryl hydrocarbon Receptor Relative Binding Affnity, RBA (Wang et al, 2005; Wang et al., 2006; Zheng et al., 2007);
- Luciferase activity (Harju et al., 2002);
- Androgen Receptor Antagonism, ARANT (Harju et al., 2007; Yang et al., 2009).

In addition to the above mentioned ad hoc QSARs/QSPRs, different EPI Suite estimation programs can be used to predict SIDS physicochemical properties, such as Henry's Law Constant, Melting Point (MP), Boiling Point (BP), Vapor Pressure (VP), Water Solubility (WS), LogKOA and LogKOW.

For most of EPI Suite models no experimental data of brominated flame retardants, and in particular of PBDEs, are included in the training set used to build the models. Comparing available experimental physicochemical data of BFRs with those estimated by the EPI Suite programs, the latter ones tend to either overestimate or underestimate the experimental values, especially for highly brominated compounds. This evidence suggests that BFRs are not included in the structural applicability domain of EPI Suite models.

Another important tool included in EPI Suite is the ECOSAR estimation program, which predicts toxicity to fish, aquatic invertebrates and algae using an extensive set of structure-activity relationship models (SARs). The majority of BFRs (PBDEs, CH3O-BDE metabolites, hexabromocyclododecane, hexabromobenzene, 2,4,6-tribromophenol) are classified by ECOSAR as 'Neutral Organics'. For these chemicals baseline toxicity is estimated by logKow.

Since acute toxicity models (e.g. Fish LC50 96-h, Daphnid LC50 48-h, Green Algae EC50 96-h) can be used only for compounds with LogKow values less than 5.0 (6.4 for Green Algae), these models are not applicable to BFRs, which are characterized by LogKow values higher than 5.

For chronic toxicity endpoints (i.e. Fish ChV 30-d, Daphnid ChV 16-d, Green Algae ChV) ECOSAR models can be used for compounds with LogKow values less than 8.0. Thus, these models could be applicable for most of BFR, with the exception of higher brominated PBDE.

In any case, ECOSAR highlights that, according to water solubilities estimated by the program WSKOWWIN, BFRs “may not be soluble enough to measure the predicted effects”.

Fragrances

Till now, to our knowledge, no ad hoc QSAR/QSPR models have been developed for the prediction of physico-chemical properties and environmental toxicity of fragrances. Nevertheless different QSARs exist

for skin sensitization, an endpoint related to human toxicity but not included in SIDS.

Hence, the only existing models for the prediction of SIDS endpoints for fragrances are EPI Suite models. EPI Suite programs can be used to estimate physico-chemical properties such as Melting Point, Boiling Point, Vapor Pressure, Water Solubility, LogKOA and LogKOW.

In general, EPI Suite estimations are quite in agreement with experimental data, although they are less accurate than those obtained using ad hoc models recently developed by UI within the CADASTER Project. In addition, most of the models used by the ECOSAR program to estimate ecotoxicity for fragrances are not validated QSARs, and based on very few experimental data.

Perfluorinated chemicals (PFCs)

For QSPRs on SIDS physico-chemical properties, data on boiling point, Fluorophilicity - Fluorous partition coefficient are modeled (Rucker et al., 2005; Kiss et al., 2001). In addition, commercial softwares were used to derive 'polyparameter linear free energy relationships' for various end points. EPI Suite models were also considered. Their performances have been compared with those of some preliminary models developed by UI on MP, BP and VP (presented in Conferentia Chemometrica 2009, Siofok, Hungary).

QSAR models on T4-TTR binding are published using the PLS approach (Weiss et al., 2009). In addition, ECOSAR predictions can be considered as toxicity models for PFCs but as in above cases (eg. Fragrances) they are derived from a fragment based approach, with either few or with predicted data, and the baseline toxicity value which is used belongs to different chemical classes and not specifically for the PFCs. In some cases, the differences between the predictions from different chemical classes are 10 fold.

Triazoles and Benzotriazoles (TAZs and BTAZs)

QSPR models specifically on TAZs and BTAZs have not been found in literature. Only logP data are modeled where few TAZs are part of the larger dataset. Regarding EPI Suite models, their predictions for MP, VP, LogKOW and WS do not show lare deviations from available experimental data. However, preliminary ad hoc QSPRs developed by UI for triazoles and benzotriazoles have RMSE values always lower than those calculated for EPI Suite models, the main exception being the LogKow model.

QSAR models on the following end-points are found for TAZs/BTAZs (Trohalaki et al., 2002; Klink, 2003; Wei et al., 2006): EC25MTT (mM), LSCROS (mM), EC25LDH (mM), EC50GSH (mM), Sensitization and/or irritancy potential, Fungicide (wheat head blight) (FA) (50 ug/ml). Although these models are not specifically for SIDS endpoints, they are related to human and to plant toxicity. In addition, models based on 'LC50-96h acute toxicity of rainbow trout' are developed for large datasets which contain only few TAZs (Benfenati, 2006; Benfenati, 2008). With regard to the commonly available ECOSAR models, it was found that in general, most ECOSAR models available are based on very few data and thus they are of limited utility for the specific classes of compounds studied within CADASTER. From this analysis the need of the development of specific QSAR models in the CADASTER Project (WP3) is highly evident. The main problem

in this respect is the lack of SIDS data in sufficient amounts to develop new QSARs. For this reason, all the available data that have been collected will be modeled.

Task 2.2 – Collection of (Q)SAR models and non-testing approaches

D2.3 (Month 18) Overview of non-testing approaches available for implementation in REACH

Overview

In this task, an overview is provided of the non-testing options given under REACH to either replace experimental testing, or to strengthen confidence in experimental results. The latter is needed as the (in general scarcely available) experimental data for specific (SIDS) endpoints and for specific chemicals, might on their own not be sufficiently convincing as a proper reflection of the actual value of specific endpoints. The non-testing options available under REACH are: Quantitative Structure Activity Relationships (QSARs), read-across, category approaches, and exposure based waiving.
In other CADASTER reports attention has already been given in detail on the use of QSAR techniques to generate data for chemical risk assessment (Deliverable 2.2: Overview of (Q)SAR models and their specific features for assessing fate and effects – December 2009). Therefore, in the report covering deliverable 2.3, the focus was on the possibilities to apply read-across and category approaches to the CADASTER selection of substances, and on providing an overview of tools as well as guidance for the application of read across and category approaches.

Task 2.3 Generation of new data

Delivery 2.5 (Month 38) Overview of new data generated

Overview

This task was carried out by means of experimental testing of chemicals and the report provides an overview of the CADASTER testing of chemicals. New data were generated on endpoints and chemicals for which, as identified in WP3, insufficient data were available for model validation and proper hazard/risk assessment. Following evaluation of available experimental data and available (Q)SAR models within WP3, new experimental data on the most essential endpoints of assessment (i.e. new data for the endpoints that are responsible for most of the variance in the safety, hazard and risk assessment) were generated.

The following testing of toxicity and fate and behaviour was performed:
1 – Polybrominated diphenylethers (PBDE)
28-day sediment testing of PBDEs was performed on bioaccumulation with aquatic oligochaeta Tubifex tubifex by PHI.
2 - Perfluoroalkylated substances and their transformation products
Toxicity testing of fluorinated compounds was performed with lettuce (Lactuca sativa) and green algae (Pseudokirchneriella subcapitata) at the RIVM. Thereupon, testing was performed with two cladoceran species (Daphnia magna and Chydorus sphaericus), as well as with embryos of the zebrafish (Danio rerio), also at the RIVM.
3 – Substituted musks/fragrances
Toxicity testing of fragrances was performed with green algae (Pseudokirchneriella subcapitata) and with Daphnia magna at the PHI.

Substituted musks/fragrances were tested also on ready biodegradability, at the PHI.

4 - Triazoles/benzotriazoles

Toxicity testing of substituted (benzo)triazoles was performed with Daphnia magna and with embryos of the zebrafish (Danio rerio) at the RIVM. Toxicity testing of substituted (benzo)triazoles was performed with green algae (Pseudokirchneriella subcapitata) at the PHI. Substituted (benzo)triazoles were tested also on ready biodegradability, at the PHI.

Activities performed

The chemicals were selected for the testing on the basis of existing toxicity data for both vertebrate and invertebrate species, as well as on principal component analysis aimed at selecting a minimum number of compounds to be tested in order to maximize the spanning of the chemical domain of both compound classes. In addition to the selection of the optimal set of chemicals to be tested, the criteria of availability of the chemicals in terms of possibilities of being able to purchase the chemicals was essential in the final selection of compounds to be tested.

Toxicity and fate and behaviour testing by PHI was performed on a pre-selected number of polybrominated diphenylethers, substituted musks/fragrances and substituted (benzo)triazoles. The testing performed includes testing of polybrominated diphenylethers on bioaccumulation with Tubifex tubifex. Ecotoxicity testing on substituted musks/fragrances includes testing with algae and daphnids. Fragrances were tested also on ready biodegradability. Substituted (benzo)triazoles were tested with algae and on ready biodegradability.

Toxicity testing by RIVM was performed on a pre-selected number of perfluorinated compounds and substituted (benzo)triazoles. The ecotoxicity testing performed includes testing of perfluorinated compounds with lettuce, algae, daphnids and zebrafish embryos. Substituted (benzo)triazoles were tested with daphnids and zebrafish embryos.

In the report prepared by PHI and RIVM “Delivery 2.5 (Month 38) Overview of new data generated” an overview is given of the chemicals tested within each class of compounds, the materials and methods that were employed, the test setup, as well as the results obtained.

RIVM prepared a paper “Toxicity of poly- and perfluorinated compounds to lettuce (Lactuca sativa) and green algae (Pseudokirchneriella subcapitata)”, published in Archives of Environmental Contamination and Toxicology.

RIVM prepared a paper “Acute toxicity of poly- and perfluorinated compounds to two cladocerans, Daphnia magna and Chydorus sphaericus” and “Physicochemical Properties and Aquatic Toxicity of Poly- and Perfluorinated Compounds”, published in Critical Reviews in Environmental Science and Technology.

LU and PHI prepared a paper “Assessing bioaccumulation of polybrominated diphenyl ethers for aquatic species by QSAR modelling”, published in Chemosphere.

PHI and RIVM collaborated for the preparation of a paper “Bioaccumulation of different PBDEs by Tubifex tubifex”, submitted to Environmental Toxicology and Chemistry.

PHI, RIVM and UI collaborated for the preparation of a paper “Experimental assessment of environmental fate and effects of triazoles and benzotriazoles”, submitted to ATLA (as proceedings of the 2nd CADASTER Workshop in Munich).

UI, IVL, IDEA, HMGU, PHI and RIVM collaborated for the preparation of a paper “Evaluation of CADASTER QSAR models for aquatic toxicity of (benzo-)triazoles and prioritization by consensus”, submitted to ATLA (as proceedings of the 2nd CADASTER Workshop in Munich).

Task 2.4 Development of a database on experimental data and (Q)SAR models

Delivery 2.4 (Month 36) Establishment of a database on properties and fate/effect parameters of chemicals within the four classes of chemicals selected

The CADASTER QSPR-THESAURUS database has been developed within the framework of the CADASTER project. The database is based on the On-line Chemical Modeling Environment (QSPR THESAURUS) http://www.qspr-thesaurus.eu, which has been developed by Dr Tetko's group at HMGU1 and is currently being offered as a commercial software from eADMET GmbH http://www.eadmet.com. The database was further developed according to the request of the CADASTER project partners and database users. The database provides the main repository to store and handle endpoint data collected and measured during the CADASTER project. Below, we describe the main features of the database.

Overview of the database structure

The front page of the QSPR THESAURUS database provides an access to the four classes of chemicals which are the focus of the CADASTER project. After selection of one of the classes, the user accesses the database of experimental and calculated properties. The database contains experimentally measured biological and physicochemical properties of molecules belonging to the four classes, together with the conditions under which the experiments have been conducted and references to the sources where the data were published. These data were collected or measured by CADASTER partners during the project.

The experimental measurements are the central entities of the database. They combine all the information related to the experiment, in particular the result of the measurement, which can be either numeric or qualitative depending on the measured property. The central system component, where the experimental measurements can be introduced, searched and manipulated, is the compound property browser.

The experimental measurement includes information about the property that was measured and the associated chemical compound. The compounds and the properties can be marked with particular keywords, also known as tags that allow convenient filtering and grouping of the data. The CADASTER database uses five tags to differentiate compounds from each of four analyzed classes as well as all compounds.

For each measurement stored in QSPR THESAURUS, it is obligatory to specify the source of the data. The source is usually a publication in a scientific journal or a book. The strict policy of QSPR THESAURUS is to accept only those experimental records that have their source of information specified. This improves the quality of the data and allows it to be verified by checking the original publication. Although a user can also introduce an unpublished article and link the data to it, records from such sources should be treated with caution. The ways to browse, introduce and automatically fetch the publications from the PubMed database are described below in the "Sources of information" section.

Every numeric property has a corresponding category of units, for example, the category of units for Inhibition Concentration 50% (IC50) is

'Concentration'. By default the QSPR THESAURUS database keeps experimental endpoints in the original format (i.e., in units as reported in the publication). For this purpose all units are grouped into strictly defined unit categories, for example Kelvin, Celsius and Fahrenheit degrees belong to the 'Temperature' category. For the purpose of compatibility and for modeling of the combined sets from different publications, the system provides on the fly conversion between different units.

An important feature of our database, which is also unique among other chemical databases, is the possibility to store the conditions of experiments. This information is crucial for modeling: in many cases, the result of an experimental measurement is senseless without knowing the conditions under which the experiment has been conducted. For example, it does not make sense to specify the boiling point for a compound without specifying the air pressure. Such conditions should be introduced as obligatory conditions, i.e., a new record will be rejected by the system if there is no information about these conditions provided. Conditional values stored in the database can be numerical (with units of measurement), qualitative or descriptive (textual). Moreover, in the 'conditions' section it is possible to note additional information related to the experiment, even if it is not a 'condition' in the classical sense. Examples of such additional information are assay descriptions, a target of the ligand (the receptor) or species on which the biological activity has been tested. For simplicity, we further universally refer to all this information as 'conditions'.

Features overview
In brief, the distinguishing features of the QSPR THESAURUS database are as follows.
- The wiki principle: most of the data can be accessed, introduced and modified by users
- Different access levels: guests, registered users, verified users, administrators
- Tracking of all the changes
- Obligatory indications of the source of the data
- Possibility to indicate conditions of the experiment, which can be later used for QSAR modeling
- Search by substructure, molecule names, by publication where the measurements were referenced, by conditions of experiments, etc.
- Control of duplicated records
- Batch upload and batch modification of large amounts of data
- Different units of measurements and utilities to interconvert between units
- Organizing the records in re-usable sets (“baskets”)
- Hidden and public records to allow collaborative development on the web

The database contains convenient tools for upload, search, editing, curation, tracking and download of experimental data. It is integrated with models that were developed during the CADASTER project. It provides an easy and professional way to store and share information about the environmental toxicity of chemical compounds. The estimation of the accuracy of predictions and applicability domain of models allows external user to make a proper choice on whether the calculated values of a particular models can be used to avoid experimental measurements.

WORKPACKAGE 3 – DEVELOPMENT AND VALIDATION OF QSARs

Paola Gramatica1, Ester Papa1, Simona Kovarich1, Stefano Cassani1, Barun Bhhatarai1, Magnus Rahmberg2, Sara Nilsson2, Tomas Öberg 3, Ullrika Sahlin3, Igor Tetko4, Stefan Brandmaier4, Nina Jeliazkova5, Nikolay Kochev 6, Ognyan Pukalov6

1. QSAR Research Unit in Environmental Chemistry and Ecotoxicology, University of Insubria, Via J.H. Dunant 3 - 21100 Varese, Italy
2. IVL Swedish Environmental Research Institute, Box 210 60, SE- 100 31 Stockholm, Sweden
3. Linnaeus University, School of Natural Sciences, 391 82 Kalmar, Sweden.
4. Helmholtz Zentrum Muenchen - German Research Center for Environmental Health, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany
5. IdeaConsult Ltd., 4 A.Kanchev str. Sofia 1000, Bulgaria
6. University of Plovdiv, Department of Analytical Chemistry and Computer Chemistry Plovdiv, Bulgaria

Major outcomes for WP3

WP3 deals with the development of individual and consensus QSAR models for different endpoints regarding the physico-chemical properties and toxic activities of the compounds of interest for CADASTER, i.e. BFRs, PFCs, Fragrances, TAZs/BTAZs (here referred to as "CADASTER chemicals"). This activity raised different issues related to data availability, harmonization of structural information available for the modeling, validation of the quality and of the applicability domains of models, which gave the opportunity to improve the quality of the proposed approaches and to highlight the importance of good collaboration among the partners with different scientific background and experiences.

The major outcomes and recommendations related to future risk assessment within the REACH framework, for WP3 activities can be summarized in the following points:
- A strong effort was made to fill in the QSPR-THESAURUS web-database, created by HMGU, chemical structures, molecular descriptors, experimental data, and to upload QSAR models developed by WP3 partners. This work was useful to improve the web-database and to adapt it to the needs of QSAR developers and public users. This work also highlighted that the lack of experimental data for the SIDS endpoints, of interest for studying the CADASTER chemicals in the environment, represents the biggest problem to develop new, ad hoc, QSARs specific for these chemicals. For this reason, the available experimental data of physico-chemical properties and toxicity (i.e. beyond the SIDS endpoints) were collected and modelled, in order to support the prioritization exercises with additional information. These models developed on non SIDS endpoints, such as rodent toxicity of PFCs, endocrine activity of BFRs and cytotoxicity of fragrances are also relevant to support risk assessment on human and ecosystems. The non-SIDS information could be also considered to perform exposure assessment through prediction of chemical behaviour in the environment, and to define appropriate measures for risk management.
- The evaluation of existing QSARs according to the "OECD principles for QSAR validation and application in regulation" demonstrated that only a small number of models for SIDS endpoints are currently available for the CADASTER chemicals. Additionally, the majority of the existing QSARs were shown to have no external validation, nor a definition of the applicability domain. Thus they do not fulfill the OECD principles for use of QSAR models in regulatory assessment of chemicals. Therefore these

models are of limited utility for the specific classes of compounds studied under the project CADASTER.
- The upload of new QSAR and QSPR models in the QSPR-THESAURUS web-database was particularly challenging since their reproducibility in the database was dependent on the calculation of molecular descriptors which were derived by different WP3 partners, from the molecular structures, by different tools. The application of this procedure raised a number of issues related to the need of harmonization for structural design and SMILES writing, as well as for the calculation of molecular descriptors starting from different structural inputs and using different software. These problems have been addressed in a tutorial paper (Gramatica et al. Mol.Inform. 2012). However the calculation of different molecular descriptors using different methodologies gave also the possibility to improve the quality of the online CADASTER Thesaurus tool, and to explore the availability and the applicability of alternative software for the calculation of molecular descriptors (i.e. commercial software DRAGON, freely available software PADEL-Descriptors, QSPR-Thesaurus descriptors, etc…)
- The limited availability of experimental data was challenging for the procedures of external validation and evaluation of the applicability domain of the models. Different techniques were applied to grant the external predictivity of the models and to evaluate the applicability domain of the models. This led to the in depth study of the behavior of the most used statistical validation parameters and to the proposal of the Concordance Correlation Coefficient (CCC) as an additional index of model quality. These parameters were implemented in the QSAR-INS software developed by UI. It was extensively demonstrated that models based on limited training sets have limited applicability domains and their application for predictive purposes should be carefully evaluated. This is particularly relevant for those models which were not externally validated due to the insufficient number of experimental data points available for this analysis.
- Similarity analysis and multivariate ranking were applied for the identification of priority chemicals in the four chemical classes of interest. Various priority lists were generated for all the four CADASTER classes and used to focus the experimental tests, performed in WP2, on the prioritized chemicals. The prioritization was performed:
A) on the basis of toxicological and chemical-physical profiles defined for the studied classes of chemicals obtained from the experimental data available in the CADASATER database and also considering data predicted by the new "ad hoc" QSAR models developed in WP3. Experimental and predicted data available for non SIDS endpoints were also included in these analyses, in particular for the identification of potentially harmful compounds on the basis of their toxicity.
B) on the basis of chemical structure, in order to select representative chemicals in the structural space of each studied chemical class.
- The predictions obtained from the individual QSAR models, developed by different modeling approaches and different molecular descriptors, were averaged in a consensus approach to propose predicted data that are not biased by any specific model. These exercises, performed for physico-chemical properties of PFCs, aquatic toxicities of BTAZs and ready biodegradability of fragrances, highlighted again the superior quality of consensus modeling in comparison to the majority of the individual approaches.
- A comparison of prediction accuracy of the QSAR/QSPR models developed ad hoc for the CADASTER classes with some freely available tools was performed. In particular,

1) QSPR models developed for the prediction of physico-chemical properties of BFRs, PFCs and TAZs/BTAZs were compared with EPI Suite estimation program,
2) QSAR models developed for aquatic toxicity of TAZs/BTAZs were compared with the ECOSAR toll,
3) classification models for the prediction of ready biodegradability of fragrances were compared with BioWIN software (implemented in Epi Suite v 4.1) and START estimation tool (implemented in Toxtree v. 2.5.1).
This study highlighted the importance of ad hoc models, which resulted more accurate than general models, to predict specific classes of chemicals, like those studied in the CADASTER project.

The present report provides a summary of all the activities performed from January 2009 to December 2012, within the CADASTER Project by WP3 project partners, i.e University of Insubria (UI), Swedish Environmental Research Institute (IVL), Linnaeus University (LnU), Helmholtz Zentrum Muenchen (HMGU), and Ideaconsult Ltd (IDEA).
Activities are summarized in the following paragraphs according to the timeline of the CADASTER Project, and organized in 7 tasks:

- Task 3.1- Chemical Structures and molecular descriptors database
- Task 3.2 Evaluation of existing QSARs according to OECD principles
- Task 3.3 Gap analysis and plan for QSAR development work
- Task 3.4 Prioritization through similarity analysis and ranking methods
- Task 3.5 Development of new QSARs
- Task 3.6 Development of multi-model approaches
- Task 3.7 External Validation of QSAR models

Task 3.1- Chemical Structures and molecular descriptors database

The design of the database was completed in June 2009 (i.e. within the first 6 months of the project). The database was made available via the project website CADASTER.eu, and filled with the relevant information on molecular descriptors for chemicals belonging to the four classes of interest for the CADASTER project. Over 900 compounds for which at least one experimental data was available or which were structurally related to the CADASTER project for their specific usage were initially included in the database.

The basic database published in 2009 included:

- Flame retardants, including PBDEs: 240 structures and 1403 molecular descriptors
- PolyFluorinated chemicals, including their transformation products: 366 structures and 1862 molecular descriptors
- Fragrances: 79 structures and 1429 molecular descriptors
- Triazoles and Benzotriazoles: 279 structures and 1879 molecular descriptors

Additional chemicals were introduced in the database during the course of the project.

DRAGON descriptors (ver. 5.5 for Windows, Talete srl., 2007) were calculated starting for all the chemicals reported above starting from the x,y,z coordinates of the chemical structure. Chemical structures and DRAGON Descriptors were reported in Deliverable 3.1 - Appendix 1.

Related Documents and Papers:

1) Deliverable 3.1 Chemical Structures and molecular descriptors database
2) Igor Tetko, Pantelis Sopasakis, Prakash Kunwar, Stefan Brandmaier, Sergii Novotarskyi, Larisa Charochkina, Volodymyr Prokopenko, Willie Peijnenberg. Prioritization of PolyBrominated Diphenyl Ethers (PBDEs) using the QSPR-THESAURUS webtool. ATLA submitted (2013)

Task 3.2 Evaluation of existing QSARs according to OECD principles

The existing QSARs (collected in WP2, Deliverable 2.2) were evaluated according to the OECD principles for QSAR models validation for regulatory application:1) a defined endpoint; 2) an unambiguous algorithm; 3) a defined domain of applicability; 4) appropriate measures of goodness-of-fit, robustness and predictivity; 5) a mechanistic interpretation, if possible.

Some models for BFRs and PFCs were reported in Deliverable 3.2, and a complete analysis of the existing QSARs, also for TAZ/BTAZs, was reported in Deliverable 3.2 - Appendix 3.2.

In general, the majority of the existing QSARs available for the prediction of physico-chemical and toxicological properties of BFRs, PFCs, TAZs/BTAZs and Fragrances had external validation, nor applicability domain definition, thus they did not fulfil the 'OECD principles for QSAR validation' for regulatory applicability. Due to these limitations these models were considered of limited utility for the specific classes of compounds studied under the project CADASTER.

This analysis highlighted the need for the development of specific QSAR models developed according to the OECD principles in the CADASTER Project (WP3). The lack of SIDS data in sufficient amount to develop new QSAR models, highlighted by WP2 (deliverable 2.2) was bypassed by using all the available experimental measures of toxicity and physico-chemical properties collected in the database from literature, to develop QSAR models.

Related Documents and Papers:
1) Deliverable 3.2 Evaluation of existing QSARs according to OECD principles
2) Deliverable 2.2 Collection of (Q)SAR models and non-testing approaches. Overview of (Q)SAR models and their specific features for assessing fate and effects

Task 3.3 Gap analysis and plan for QSAR development work

Under this Task, in CADASTER Project, it has been written:
"Identification of gaps in the model library with input from task 3.2. Gaps can be either lack of models for end-points or models with insufficient predictive per¬formance. Both ecotoxicological and fate related end-points are considered, and existing QSARs will be applied as much as possible to fill data gaps."
In Tasks 2.2 (Deliverable 2.2) and 3.2 it was verified that only a small number of models for SIDS end points were available and with some problems of applicability to CADASTER molecules.

This justified the subsequent development of specific QSAR models in CADASTER Project (WP3), according to the OECD principles and based on both SIDS and not-SIDS endpoints.

All the Partners involved in WP3 (UI, HIK, IVL, Idea, HMGU and RIVM) participated to the development and validation of individual models (Deliverable 3.5), which were finally used to create consensus models at the end of the project (Deliverable 3.6).

During the first year of the project, UI developed some QSAR models, mainly for BFRs, PFCs and for fragrances. These models were developed with MLR (OLS method), based on DRAGON descriptors, selected by Genetic Algorithm. Also HIK and IDEA modeled physico-chemical properties (mp, bp, vp) of PFCs on data selected by UI and a joint publication was published in 2011 (Bhhatarai et al. 2011) All the models were developed taking into account the OECD principles for acceptability of QSARs for regulation purposes.

Additionally IDEA performed an investigation of the coverage of the 4 classes of compounds of interest for CADASTER by the ECHA preregistration list by structural alerts. A large number of chemical structures have been identified. This information was used by partners to facilitate selection of compounds for testing and in the model development to meet the requirements set forward within the OECD principles for QSAR models validation for regulatory application.

Related Documents and Papers:
1) Deliverable 3.2 - Evaluation of existing QSARs according to OECD principles
2) Deliverable 2.2 - Collection of (Q)SAR models and non-testing approaches. Overview of (Q)SAR models and their specific features for assessing fate and effects
3) Deliverable 3.3 - Gap analysis and plan for QSAR development work
4) Deliverable 3.5 - Development of new QSARs
5) Barun Bhhatarai, Wolfram Teetz, Tao Liu, Tomas Oberg, Nina Jeliazkova, Nikolay Kochev, Ognyan Pukalov, Igor Tetko, Simona Kovarich, Ester Papa, Paola Gramatica. CADASTER Models for Predictions of Melting and Boiling Points of Perfluorinated Chemicals" Molecular informatics (proceedings EuroQSAR2010), 2011, 30 (2-3), 189-204.

Task 3.4 Prioritization through similarity analysis and ranking methods

The activity on prioritization to optimize the experimental testing (WP2) was performed during the first 24 months of the Project. Similarity analysis and multivariate ranking methods were applied for experimental design and the identification of priority chemicals in the four chemical classes of interest.

The prioritization was performed:
- on the basis of toxicological and chemical-physical profiles defined for the studied classes of chemicals. The basic idea is that the priority chemicals are those more hazardous sensu latu, thus those that have any demonstrated kind of toxicity. Owing to the general paucity of experimental data, in particular for SIDS endpoints, for chemicals belonging to the four CADASTER classes, the available experimental data on toxicity end points (also mammalian toxicity and endocrine disruption activity) were used for the characterization of toxicological profile. In addition to the available experimental data also data predicted by the new “ad hoc” QSAR models developed by Insubria University (UI) in the Project (WP3, task 3.5) were used.

- on the basis of chemical structure, in order to select representative chemicals in the structural space of each studied chemical class.

1) Brominated Flame Retardants (BFRs) A priority list of BFRs was provided by UI on the basis of their toxicological profile obtained from some preliminary UI-QSAR models developed for several endpoints related to dioxin-like activity (AhR RBA, EROD induction, AhR agonism) and endocrine disruption (ED) potency (ER agonism, PR antagonism, T4-TTR competition, E2SULT inhibition). The PBDE congeners predicted with higher activity for both dioxin-like activity and endocrine disruption potency were listed and were suggested as priority compounds for the experimental tests. In addition to PBDEs, the most active OH-PBDE congeners, found to be even more active than parental compounds by both experimental evidences and QSAR predictions, were proposed for testing. Multidimensional scaling (MDS), based on the molecular descriptors selected in the QSAR models for ED property, was applied to explore the structural similarity/diversity of the suggested priority compounds, in the space of the 243 BFRs of interest for CADASTER. This analysis showed that the priority BFRs were representative of the BFRs chemical domain. A prioritization based on multivariate structural characterization was performed by LNU (Linnaeus University). Principal component analysis (PCA) was applied to theoretical molecular descriptors calculated for 209 PBDEs. After these preliminary works of experimental design, based on the prioritization performed by UI (on the toxicity profile), LNU (on structural diversity) and on the chemical availability, the partner PHI (Public Health Institute, Maribor), which was responsible for the experimental testing of PBDEs, selected a final set of 12 PBDE congeners. These compounds were subsequently tested by PHI in the specie Tubifex tubifex (56 days, sediment) for the measurement of BCF.
2) Fragrances A priority list of fragrances was provided by UI on the basis of their toxicological profile. QSAR models developed by UI for mouse acute toxicity (LD50 oral) and rat cytotoxicity (EC50 NADH oxidase and EC50) were applied to 79 fragrances and, among these, 30 compounds were identified as always active ("AA") for all the modeled endpoints (Papa et al. 2009). Multidimensional scaling (MDS), based only on molecular descriptors selected in the developed QSARs, was then applied to the 79 studied fragrances in order to select, among the 30 AA compounds, the most structurally representative for each functional class. Finally, 10 fragrances were selected and suggested for experimental testing. The unavailability of four of the suggested fragrances for the experimental testing implied a further analysis by UI, in order to identify some alternative compounds. A Factorial design on 4 principal components of molecular descriptors was performed, identifying 16 groups of structural similar compounds. Hence, for each of the four fragrances several possible replacements were provided. In addition, from the results of the Factorial Analysis, other compounds were suggested for testing, in order to cover a wider part of the structural domain. On the basis of the prioritization performed by UI and the availability of the prioritized chemicals, the partner PHI selected a final set of 11 fragrances.
3) Perfluorinated compounds (PFCs) A priority list of PFCs was provided by UI on the basis of their toxicological profile. QSAR models developed for mouse and rat oral toxicity (Bhhatarai et al. 2011) and inhalation toxicity (Bhhatarai et al. 2010) were applied to predict the activity of 376 PFCs (including some in ECHA list). A principal component analysis (PCA) was then performed on the experimental and predicted data (taking into account the 204 compounds within the AD of the models and 25 commonly known PFCs with experimental data). 22 long chain PFCs were

identified as the most toxic for all the four considered endpoints and thus suggested as priority perfluorinated compounds for the experimental testing. A multivariate characterization, by principal component analysis based on theoretical descriptors (Dragon), was performed by LNU (Linnaeus University). The strategy for the compound selection suggested by LNU was to select compounds always taking into consideration the availability of test compounds and any additional information (e.g. toxicity data). Finally, on the basis of the prioritization performed by UI (on toxicity profile), by LNU (on structural representativity), and compound's availability, Partner RIVM (Rijksinstituut voor Volksgezondheid en Milieu) selected a final set of 13 perfluorinated compounds for the experimental testing.
4) Triazoles and Benzotriazoles (TAZs and BTAZs) A priority list of (B)TAZs was provided by UI on the basis of their toxicological profile and structural similarity. Available experimental data for several eco-toxicity and environmental behavior endpoints (i.e. EC50 Algae, EC50 Daphnia, LC50 fish, LogP, BCF) collected from the Footprint Database were analysed in order to identify the most active compounds. A Principal Component Analysis was thus performed using only those compounds for which experimental data for EC50 algae, EC50 daphnia, LC50 fish, LogP and BCF were available (21 (B)TAZs). The most active (benzo)triazoles for the considered endpoints were identified, which were characterized by high toxicity in both plants (algae) and animals (daphnia and fish) and bioaccumulation potential in the food chain (high LogP and BCF). The subsequent step was to identify, and suggest for further testing, (benzo)triazoles structurally similar to those highlighted in the aforementioned PCA, for which experimental data were missing. Multidimensional Scaling (MDS) and Kohonen maps were applied for this purpose. Results obtained using the two different approaches were compared and a list of common priority (B)TAZs for experimental testing was suggested. A multivariate characterization, by principal component analysis based on theoretical descriptors, was performed by LNU (Linnaeus University). In particular, two different PCA-models based on 2D-Dragon descriptors (calculated directly from the SMILES of 421 molecules given by RIVM) were provided for the final selection of the priority (B)TAZs.

Development of active learning approaches
The investigation of the active learning approaches (i.e., methods which create datasets by selection of compounds which will increase accuracy of models to be developed, so called optimal experimental design) was performed as part of planned activities in WP 3.4. In collaboration of LNU and HMGU a new approach for experimental design was developed (Brandmaier et. Al 2012a, Brandmaier et. Al 2012b), The proposed approach consisted in an extension of standard approaches like Full or Fractional Factorial design or D-Optimal design, to a stepwise procedure that utilizes the D-Optimal design and combines it with partial least squares techniques (PLS) to iteratively refine the descriptor space for the compound selection. This refinement was realized by the usage of the PLS latent instead of principal components. The focus within the development was on several aspects. Apart from a good performance, the approach should not just be a theoretical solution, but it should be close as possible to the typical data analysis performed in laboratories. Furthermore, it should take into consideration cost efficiency and enable a maximum of flexibility, to customize the approach to specific needs. On the web version of the CADASTER database (see http://www.cadaster.eu/database online) both a traditional D-Optimal approach and the stepwise approach were implemented to prioritize chemicals (Tetko et al. 2013).

Related Documents and Papers:

1) Deliverable 3.4 - Prioritization through similarity analysis and ranking methods
2) Deliverable 3.5 - Development of new QSARs
3) Ester Papa, Mara Luini, Paola Gramatica. QSAR modelling of oral acute toxicity and cytotoxic activity
of fragrance materials in rodents SAR QSAR Environ Res., 2009, 20 (7-8), 767-779.
4) Barun Bhhatarai and Paola Gramatica. Per- and Poly-fluoro Toxicity (LC50 inhalation) Study in Rat and Mouse using QSAR Modeling. Chem. Res. Toxicol., 2010, 23(3), 528-539.
5) Barun Bhhatarai and Paola Gramatica. Oral LD50 Toxicity Modeling and Prediction of Per- and Polyfluorinated Chemicals on Rat and Mouse, Molecular Diversity, 2011, 15 (2), 467-476.
6) Stefan Brandmaier, Igor Tetko, Tomas Öberg. An evaluation of experimental design in QSAR modelling utilizing the k-medoid clustering. Journal of Chemometrics, 2012, 26, 509-517.
7) Stefan Brandmaier, Ullrika Sahlin, Igor Tetko, Tomas Oberg. PLS-Optimal: A Stepwise D-Optimal Design Based on Latent Variables. J Chem Inf Model, 2012, 52, 975-983.
8) Igor Tetko, Pantelis Sopasakis, Prakash Kunwar, Stefan Brandmaier, Sergii Novotarskyi, Larisa Charochkina, Volodymyr Prokopenko, Willie Peijnenberg. Prioritization of PolyBrominated Diphenyl Ethers (PBDEs) using the QSPR-THESAURUS webtool. ATLA submitted (2013)

Task 3.5 Development of new QSARs

The various Partners, involved in WP3, have developed models for the endpoints/classes by applying different modelling approaches. The models were developed taking into account the OECD principles for validation and acceptability of QSARs for regulation purposes, in particular external validation and check of applicability domain. The models, developed in the Project, have been documented in publications on international journals, peer reviewed (ISI), and in meeting presentations, listed below, and also in the CADASTER database (qspr-thesaurus) and CADASTER website (see http://www.cadaster.eu online).

1) Polybrominated diphenyl ethers (PBDEs and Brominated Flame retardants (BFRs)).

Local QSAR/QSPR models have been developed for several physico-chemical properties, environmental fate and toxicity endpoints. The selection of the endpoints for QSAR modelling was based on the few experimental data available.

The modelled endpoints and the developers of the respective models are listed below:

- melting point (MP) (Developers: UI, HMGU)
- Boiling Point (BP) (Developer: HMGU)
- vapour pressure (log VP) (Developers: UI, LnU)
- water solubility (log WS) (Developer: UI)
- Henry's low constant (log H) (Developer: UI)
- octanol-air partition coefficient (log KOA) (Developer: UI)
- octanol-water partition coefficient (log KOW) (Developer: UI)
- photolysis rate constant (log KP) (Developer: UI)
- photolysis half-life (log HLP) (Developer: UI)
- Biovoncentration (Developer: LnU)

- endocrine disrupting (ED) potency (different endpoints, i.e. aryl hydrocarbon receptor binding, agonism and antagonism, EROD induction, estrogen receptor agonism and antgonism, androgen receptor antagonism, progesterone receptor antagonism, T4-TTR competition and E2SULT inhibition) (Developers: UI, HMGU)

Some of the QSPRs (MP, VP, WS, H, logKoa, logKow) were compared with the EPI Suite Estimation programs. As expected, prediction accuracy of the local models developed under CADASTER Project was higher than prediction accuracy obtained by applying the general EPI Suite models.

Equations, statistical performances, predictions and information on applicability domain (interpolated/extrapolated predictions) for the models developed for the properties reported above are provided in the following documents and publications:

1) Deliverable 3.5 - Development of new QSARs
2) Ester Papa, Simona Kovarich, Paola Gramatica. Development, Validation and Inspection of the Applicability Domain of QSPR Models for physico-chemical properties of Polybrominated DiphenylEthers QSAR and Combinatorial Science, 2009, Volume: 28 Issue: 8, 790-796.
3) Ester Papa, Simona Kovarich, Paola Gramatica. QSAR modeling and prediction of the endocrine disrupting potencies of brominated flame retardants, Chem Res. Toxicol, 2010, 23 (5), 946-954.
4) Ester Papa, Simona Kovarich, Paola Gramatica. On the use of local and global QSARs for the prediction of Physico-Chemical Properties of Polybrominated Diphenyl Ethers Molecular informatics (proceedings EuroQSAR2010), 2011, 30 (2-3), 232-240.
5) Simona Kovarich, Ester Papa, Paola Gramatica. QSAR classification models for the prediction of endocrine disrupting activity of brominated flame retardants, J.Hazardous Materials, 2011,190 (1-3), 106-112.
6) Alexander Safanyaev, Vladimir Palyulin, Eugene Radchenko, Igor Tetko. Quantitative Structure-Activity Relationships of AhR Ligands. Journal of Chemical Information and Modelling submitted (2013)

2) Perfluoroalkylated substances (PFCs) (and Poly-fluorinated compounds)
PFCs studied under the CADASTER Project include 382 chemicals, both linear and aromatic chemicals, with different carbon chain length, fluorination degree (per- and polyfluorinated compounds) and functional groups (carboxylates, sulfonates, sulfonamides, alcohols, etc.). The majority of these compounds are included in the ECHA pre-registration list. Local QSAR/QSPR models have been developed for several physico-chemical properties, environmental fate and toxicity endpoints. The selection of the endpoints for QSAR modelling was based on the few experimental data available.
The modelled endpoints and the developers of the respective models are listed below:

- Melting point (MP) (Developers: UI, LnU, HMGU)
- Boiling point (BP) (Developers: UI, LnU, HMGU)
- Vapour pressure (log VP) (Developers: UI, LnU)
- Water solubility (log WS) (Developers: UI, HMGU)
- Critical micelle concentration (log CMC) (Developers: UI, HMGU)
- Mammalian (rat/mause) acute toxicity (oral/inhalation) (log 1/ LD50) (Developers: UI, HMGU)
- T4-TTR competing potency (T4-TTRcomp) (Developers: UI, HMGU)
- Organic carbon-water partition coefficient (log KOC) (Developer: UI – reported in Deliverable 4.6)

- Inhibition of root elongation in Lettuce (Lactuca sativa): (Developer: RIVM)
- Inhibition of photosynthesis of green algae (Pseudokirchneriella subcapitata) (Developer: RIVM)
- Immobilisation in Cladoceran (waterflea) (Daphnia magna) (Developer: RIVM)
- Immobilisation in Benthic cladoceran species (Chydorus sphaericus) (Developer: RIVM)
- Interspecies relationships (Daphnia Magna to Chydorus sphaericus; Lactuca sativa to Pseudokirchneriella subcapitata) (Developer: RIVM)

QSPRs models developed for MP, BP, VP and WS were compared with the EPI Suite Estimation programs. As expected, prediction accuracy of the local models developed under CADASTER Project was higher than prediction accuracy obtained by applying the general EPI Suite models.

Equations, statistical performances, predictions and information on applicability domain (interpolated/extrapolated predictions) for the models developed for the properties reported above are provided in the following documents and publications

1) Deliverable 3.5 - Development of new QSARs
2) Deliverable 4.6: Synthesis of research findings and recommendations for prioritization
3) Barun Bhhatarai, Wolfram Teetz, Tao Liu, Tomas Oberg, Nina Jeliazkova, Nikolay Kochev, Ognyan Pukalov, Igor Tetko, Simona Kovarich, Ester Papa, Paola Gramatica. CADASTER Models for Predictions of Melting and Boiling Points of Perfluorinated Chemicals" Molecular informatics (proceedings EuroQSAR2010), 2011, 30 (2-3), 189-204.
4) Barun Bhhatarai and Paola Gramatica. Prediction of Aqueous Solubility, Vapor Pressure and Critical Micelle Concentration for Aquatic Partitioning of Perfluorinated Chemicals Environ. Sci. Technol., 2011, 45(19), 8120-8128.
5) Barun Bhhatarai and Paola Gramatica. Oral LD50 Toxicity Modeling and Prediction of Per- and Polyfluorinated Chemicals on Rat and Mouse, Molecular Diversity, 2011, 15 (2) 467-476.
6) Barun Bhhatarai and Paola Gramatica. Per- and Poly-fluoro Toxicity (LC50 inhalation) Study in Rat and Mouse using QSAR Modeling. Chem. Res. Toxicol., 2010, 23(3), 528-539.
7) Tomas Öberg and Tao Liu. Global and local PLS regression models to predict vapor pressure QSAR and Combinatorial Science, 2008, 27, 273-279.
8) Guanghui Ding, Marja Wouterse, Rob Baerselman, Willie Peijnenburg. Toxicity of poly- and perfluorinated compounds to lettuce (Lactuca sativa) and green algae (Pseudokirchneriella subcapitata). Arch. Environ. Sci. Technol., accepted for publication (2011).
9) Guanghui Ding , Evert-Jan van den Brandhof, Rob Baerselman, Willie Peijnenburg. Acute toxicity of poly- and perfluorinated compounds to two cladocerans, Daphnia magna and Chydorus sphaericus. Environ. Toxicol. Chem., accepted for publication (2011).
10) Simona Kovarich, Ester Papa, Jiazhong Li, Paola Gramatica. QSAR classification models for the screening of the Endocrine Disrupting activity of perfluorinated compounds, SAR QSAR Environ Res., 2012, 23, 207-220.
11) Ester Papa, Simona Kovarich, Paola Gramatica. QSAR prediction of the competitive interaction of emerging halogenated contaminants with human transthyretin, SAR QSAR Environ Res., in press (2013)

3) Fragrances. Fragrances studied and modelled under the CADASTER Project over 140 chemicals, selected in the literature and uploaded in the web by Partner 3 (UI) or provided by RIFM, belonging to different chemical classes. Local and global QSAR/QSPR models have been developed for the following physico-chemical properties and toxicity endpoints:

- Boiling Point (BP) (Developer: HMGU)
- Melting Point (MP) (Developer: HMGU)
- vapour pressure (log VP) (Developer: UI)
- water solubility (log WS) (Developer: UI)
- octanol-water partition coefficient (log KOW) (Developer: UI)
- acute toxicity in mouse (log 1/LD50 oral) (Developer: UI)
- cyto-toxicity in rat (Log EC50 NADH-Ox, Log EC50 D) (Developer: UI)
- Biodegradation (Developers: UI, IDEA, HMGU)

The selection of the endpoints for QSAR modelling was based on the few experimental data available.

Predictions obtained by the local QSPR models developed for VP, WS, and logKow were compared with predictions calculated by EPI Suite models. VP and logKow local models are characterized by higher prediction accuracy than the general EPI Suite models.

Equations, statistical performances, predictions and information on applicability domain (interpolated/extrapolated predictions) for the models developed for the properties reported above are provided in the following documents and publications

1) Ester Papa, Mara Luini, Paola Gramatica. QSAR modelling of oral acute toxicity and cytotoxic activity
of fragrance materials in rodents SAR QSAR Environ Res., 2009, 20 (7-8), 767-779
2) Deliverable 3.7 External validation of QSAR models
3) Simona Kovarich, QSAR models for the (eco) toxicological characterization and prioritization of emerging pollutants: case studies and potential applications within REACH, PhD thesis, 2013.
4) Susann Vorberg and Igor Tetko, Biodegradability Data Modeling. Molecular Informatics submitted (2013)

4) Triazoles and Benzotriazoles (B)-TAZs)) (B)TAZs studied under the CADASTER Project include 386 compounds, included also in the ECHA pre-registration list, structurally highly heterogeneous, and characterized by different using pattern and mechanism of actions. Local QSAR/QSPR models have been developed for several physico-chemical properties and eco-toxicity endpoints. The selection of the endpoints for QSAR modelling was based on the available experimental data, which were mainly collected from freely available databases (e.g. SRC PhysProp database, PPDB Footprint Database).

The modelled endpoints and the developers of the respective models are listed below:

- melting point (MP) (Developer: UI)
- vapour pressure (log VP) (Developer: UI)
- water solubility (log WS) (Developer: UI)
- octanol-water partition coefficient (log KOW) (Developer: UI)
- acute toxicity in fish (LC50 96h, Oncorhynchus mykiss) (Developer: UI, IVL, LnU, HMGU, IDEA)

- acute toxicity in aquatic invertebrates (EC50 48h, Daphnia magna) (Developer: UI, IVL, LnU, HMGU, IDEA)
- acute toxicity in algae (EC50 72h, Pseudokirchneriella subcapitata) (Developer: UI, IVL, LnU, HMGU)
- terrestrial toxicity (LD50 earthworms, honeybees, birds) (Developer: UI)

All the models, when compared with the respective models available in EPI Suite, showed a better accuracy in prediction (comparable prediction accuracy was found only for the logKow model).

Equations, statistical performances, predictions and information on applicability domain (interpolated/extrapolated predictions) for the models developed for the properties reported above are provided in the following documents and publications

1) Stefano Cassani, Simona Kovarich, Ester Papa, Partha Pratim Roy, Leon Van der Wal, Paola Gramatica, Daphnia and fish toxicity of (benzo)triazoles: validated QSAR models, and interspecies quantitative activity-activity modeling. Submitted to J. Hazardous Materials, 2013
2) Stefano Cassani, Simona Kovarich, Ester Papa, Partha Pratim Roy, Magnus Rahmberg, Sara Nilsson, Ullrika Sahlin, Nina Jeliazkova, Nikolay Kochev, Ognyan Pukalov, Igor Tetko, Stefan Brandmaier, Mojca Kos Durjava, Boris Kolar, Willie Peijnenburg, Paola Gramatica, Evaluation of CADASTER QSAR models for aquatic toxicity of (benzo-)triazoles and prioritization by consensus. Submitted to ATLA, 2013.
3) Paola Gramatica, Stefano Cassani, Partha Pratim Roy, Simona Kovarich, Chun Wei Yap, Ester Papa, QSAR Modelling is not «Push a button and find a correlation » : a case study of toxicity of (benzo-)triazoles on algae. Molecular Informatics, 2012, 31, 817 – 835
4) Barun Bhhatarai and Paola Gramatica. Predicting physico-chemical properties of emerging pollutants: QSPR modeling of Benzo(triazoles), Water Research, 2011, 45 (3) 1463-1471.

Other Publications into the CADASTER project and references:

1) Nicola Chirico and Paola Gramatica. Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient J. Chem. Inf. Model., 2011, 51 (9), pp 2320–2335.
2) Nicola Chirico and Paola Gramatica. Real External Predictivity of QSAR Models. Part 2. New inter-comparable thresholds for different validation criteria and the need for scatter plot inspection J. Chem. Inf. Model., 2012, 52 (8), pp 2044–2058.
3) Partha Pratim Roy, Simona Kovarich, Paola Gramatica. QSAR model reproducibility and applicability: a case study of rate constants of hydroxy radical reaction models applied to Polybrominated Diphenyl Ethers and (Benzo-)Triazoles Journal of Computational Chemistry, 2011, 32, 2386-2396.
4) Iurii Sushko, Sergii Novotarskyi, Robert Körner, Anil Kumar Pandey, Vasily Kovalishyn, Volodymyr Prokopenko, Igor Tetko. Applicability domain for in silico models to achieve accuracy of experimental measurements. Journal Of Chemometrics, 2010, 24, 202-208.
5) Iurii Sushko, Sergii Novotarskyi, Robert Korner, Anil Kumar Pandey, Artem Cherkasov, Jiazhong Li, Paola Gramatica, Katja Hansen, Timon Schroeter, Klaus-Robert Muller, Lili Xi, Huanxiang Liu, Xiaojun Yao, Tomas Oberg, Farhad Hormozdiari, Phuong Dao, Cenk Sahinalp, Roberto Todeschini, Pavel Polishchuk, Anatoliy Artemenko, Victor Kuz'min, Todd

Martin, Douglas Young, Denis Fourches, Eugene Muratov, Alexander Tropsha, Igor Baskin, Dragos Horvath, Gilles Marcou, Christophe Muller, Alexander Varnek, Volodymyr Prokopenko, Igor Tetko. Applicability Domains for Classification Problems: Benchmarking of Distance to Models for Ames Mutagenicity Set. Journal of chemical information and modeling, 2010, 50, 2094-2111.
6) Sergii Novotarskyi, Iurii Sushko, Robert Korner, Anil Kumar Pandey, Igor Tetko. A comparison of different QSAR approaches to modeling CYP450 1A2 inhibition. J Chem Inf Model, 2011, 51, 1271-1280.
7) Iurii Sushko, Elena Salmina, Vladimir Potemkin, Gennadiy Poda, Igor Tetko. ToxAlerts: A Web Server of Structural Alerts for Toxic Chemicals and Compounds with Potential Adverse Reactions. Journal of chemical information and modeling, 2012, 52, 2310-2316.
8) Stefan Brandmaier, Sergii Novotarskyi, Iurii Sushko, Igor Tetko. From descriptors to predicted properties: Experimental design using the applicability domain estimation. ATLA submitted (2013)
9) Stefan Brandmaier and Igor Tetko. Robustness in experimental design: A study on the reliability of selection approaches. Computational and Structural Biotechnology Journal submitted (2013)
10) U.S. Environmental Protection Agency - Estimations Programs Interface for Windows (EPI Suite) – 2009-2012.

Task 3.6 Development of multi-model approaches

The present task provides a summary of all the consensus models developed by WP3 partners for the melting and the boiling point of PFCs, the aquatic toxicity of (B)TAZs and the Biodegradation of fragrances. These models were chosen, among those reported in task 3.5, for the consensus analysis because they are the only QSA(P)Rs developed on SIDS endpoints among the models created within the CADASTER project by all the WP3 Partners. In brief, local models for the endpoints of interest were individually developed by different WP3 Partners for the four chemical classes studied within CADASTER (Task 3.5) by different approaches and methodologies. Finally, the consensus approach was used to combine the individual models developed for the same endpoint and chemical class. The combination of different modeling approaches into a consensus approach helps to complement the deficiencies of one model with the support of another. The application of this procedure to CADASTER models allowed reaching better and, potentially, more reliable QSAR predictions.

1) Consensus models for Perfluorinated compounds (PFCs):

Within the CADASTER Project, WP3 Partners developed local QSPR models for the prediction of Melting Point (MP) and Boiling Point (BP) of per- and poly-fluorinated chemicals (PFCs). MP and BP are important physicochemical properties that indirectly affect the solubility, and hence the transport, distribution and environmental fate of these compounds. All these models are statistically robust, internally and externally validated, and with a verified applicability domain.

Consensus model for MP
Consensus predictions were derived by averaging the predictions obtained by individual models.The highest accuracy in prediction was obtained when the consensus approach was applied. Indeed, the Consensus Model was characterized by the highest R2 values and the lowest value of RMSE. WP3 individual models and the consensus model were compared with the EPI Suite model for the prediction of MP, by calculating the value of the RMSE (Root Mean Square of Errors). As expected, the accuracy in

prediction of the local models (WP3 models) and the consensus model, which are specific for PFCs, was higher than the accuracy in prediction obtained by applying the general EPI Suite model. The results obtained by the application of the consensus approach for the prediction of MP of PFCs shows that in the reported study the consensus by predictions calculated by independently developed models using different descriptors and using different algorithms, delivers the best prediction results. In this specific case study, simple statistical algorithms (like MLR and PLS) applied to complex descriptors perform about as good as complex algorithms (like NN) applied to simple descriptors. In fact, the combination of different modeling approaches helps to increase the prediction power as the different approaches complement each other. Model equations, statistical performances, predictions and information on applicability domain (interpolated/extrapolated predictions) are provided in the publication of Bhhatarai et al. 2011.

Consensus model for BP

As it was done for the MP models, consensus predictions were derived by averaging the predictions obtained by individual models. The comparison of WP3 local models (Individual Full Models and Consensus Model) with EPI Suite estimations of BP data highlighted again the higher prediction accuracy obtained by applying local models specifically developed for PFCs than general models.
Model equations, statistical performances, predictions and information on applicability domain (interpolated/extrapolated predictions) are provided in the publication of Bhhatarai et al. 2011.

Related Documents and papers:
1) Deliverable 3.6 Development of multi-model approaches
2) Barun Bhhatarai, Paola Gramatica. Predicting physico-chemical properties of emerging pollutants: QSPR modeling of Benzo(triazoles), Water Research, 2011, 45 (3) 1463-1471.

2) Consensus models for Triazoles and Benzotriazoles (BTAZs):

Within the CADASTER project, several QSAR models have been developed by WP3 Partners UI, LnU, IVL, IDEA and HMGU for the prediction of aquatic toxicity of (B)TAZs. The QSARs were realized by different modeling approaches (e.g. MLR-OLS, PLSR, Kohonen Neural Network) starting from theoretical molecular descriptors calculated by commercial and freely available software (DRAGON, PaDEL-Descriptor, QSPR-THESAURUS web). The considered end-points were:

- EC50 (72h) in Pseudokirchneriella subcapitata
- EC50 (48h) in Daphnia magna
- LC50 (96h) in Onchorhynchus mykiss

Experimental data for algae, zooplankton and fish are among the key data required for risk assessment of chemicals.
Predictions by consensus were obtained by combining predictions from different models and approaches, taking into account statistical performances and applicability domains of individual models. Models were also applied for the screening of more than 300 (B)TAZs without experimental data, many of them included in the European Chemicals Agency (ECHA) pre-registration list. This screening allowed for the identification of the compounds potentially most problematic in the aquatic environment.

Model equations, statistical performances, predictions and information on applicability domain (interpolated/extrapolated predictions) are provided in the publication Cassani et al 2013.

Consensus models for acute toxicity in algae (Pseudokirchneriella subcapitata), aquatic invertebrates (Daphnia magna), and fish (Oncorhynchus mykiss)

Consensus predictions were derived by averaging (arithmetic mean) the predictions obtained by individual models (Cassani et al, 2013). Individual predictions were evaluated in comparison to predictions by consensus (individual model predictions deviation, and median deviation). Individual WP3 predictions and applicability domains, predictions by consensus, and model deviations are reported in Deliverable 3.6 - Appendix II, III and IV, and in Cassani et al, 2013.

General considerations on Consensus models for aquatic toxicity

Consensus predictions were calculated for more than 386 (B)TAZs, many of them included in the ECHA pre-registration list. Since the experimental data were not available for the majority of the studied chemicals, particular attention was given to the analysis of the applicability domain of the models used to calculate predictions by consensus. As expected, higher disagreement in prediction among different models was observed for compounds falling outside the domains of individual QSARs. Predicted values for these chemicals are considered as less reliable since model extrapolations, and should be used carefully.
Similar individual predictions were obtained when only (B)TAZs within in the applicability domain of all the WP3 models, were included in the calculations. The fact that different models, based on different descriptors and/or modeling approaches, lead to similar predictions adds confidence and reliability to QSAR predictions obtained by the consensus approach.

Finally predictions by consensus calculated to predict acute toxicity in algae, daphnia and fish for (B)TAZs included in the AD of the models were analyzed by PCA (Principal Component Analysis). This analysis was done in order to characterize the toxicological profile of BTAZs and to identify the most active compounds in the studied aquatic environment.
The globally “more toxic” chemicals (B)TAZs (i.e. with an overall higher toxicity for the three key organisms in the analyzed aquatic scenario) were discriminated from less hazardous ones by assigning an arbitrary cut-off along the PC1. For these chemicals, the following ranges of toxicity were predicted for the three organisms:

- Pseudokirchneriella subcapitata: EC50(72h) 0.18 - 4.62 mg/L
- Daphnia magna: EC50(48h) 0.19 – 5.37 mg/L
- Onchorhynchus mykiss: LC50(96h) 0.12 – 4.65 mg/L

According to EU classification criteria, as described in Annex VI of Directive 67/548/EEC (EC (1991)), these prioritized (B)TAZs can be all classified as “very toxic” (EC(LC)50 = 1 mg/L) or “toxic” (EC(LC)50 = 10 mg/L), and are threfore highlighted for the necessary experimental tests.

Related Documents and papers:

1) Deliverable 3.6 Development of multi-model approaches
2) Stefano Cassani, Simona Kovarich, Ester Papa, Partha Pratim Roy, Magnus Rahmberg, Sara Nilsson, Ullrika Sahlin, Nina Jeliazkova, Nikolay Kochev, Ognyan Pukalov, Igor Tetko, Stefan Brandmaier, Mojca Kos Durjava, Boris Kolar, Willie Peijnenburg, Paola Gramatica, Evaluation of CADASTER

QSAR models for aquatic toxicity of (benzo-)triazoles and prioritization by consensus. Submitted to ATLA, 2013.

3) Consensus models for Fragrances

Within the CADASTER project, several classification QSAR models have been developed by WP3 Partners UI, IDEA and HMGU for the prediction of biodegradation (Deliverable 3.7, Kovarich 2013 (PhD thesis)). The models were realized by different modeling approaches and were based on different dataset (specific for fragrances (UI) or general (IDEA, HMGU). Among the models developed by different WP3 partners, only those characterized by the best classification performances and based on the most different approaches (modelling methods and molecular descriptors) were selected to derive consensus predictions for the 45 fragrances included in the validation set.

The application of the consensus approach allowed enhancing predictive power by reducing misclassification errors of individual models.
As a general comment it was observed a fair agreement among the classes predicted by different WP3 models for Ready Biodegradable and Not Ready Biodegradable chemicals. The fact that different QSAR models, based on different descriptors and/or modeling approaches, lead to similar predictions adds confidence and reliability to QSAR predictions obtained by the consensus approach.

Related Documents and papers:
1) Deliverable 3.7 External validation of QSAR models
2) Simona Kovarich, QSAR models for the (eco) toxicological characterization and prioritization of emerging pollutants: case studies and potential applications within REACH, PhD thesis, 2013.

Task 3.7 External validation of QSAR models

The aim of this task was to validate the QSAR models developed within the CADASTER project, using the new experimental data obtained from WP2. This was done by comparing predicted and actually measured fate and effect endpoints for (B)TAZs and Fragrances.

1) External validation of QSARs developed for aquatic toxicity of Triazoles and Benzotriazoles

The QSAR regression models of toxicity of triazoles and benzo-triazoles to algae (EC50 72h in Pseudokirchneriella subcapitata), EC50 48h in Daphnia magna and fish (LC50 96h in Onchorhynchus mykiss) were developed by five partners in WP3 (UI , LnU, IVL, IDEA and HMGU) and reported in Deliverable 3.5 and 3.6. The models were developed by different methods (MLR-Ordinary Least Squares (OLS), PLSR, Bayesian Lasso on PLS components (BLASSO/PLS), and ASsociative Neural Network (ASNN)), using various molecular descriptors (DRAGON, PaDEL- Descriptors and QSPR-THESAURUS web), and different procedures for variable selection, validation and applicability domain inspection. The predictions of the developed models, as well as those obtained in a consensus approach by averaging the data predicted from each model, were compared with the results of experimental tests that were performed by two CADASTER partners (PHI and RIVM in WP2). The individual and the consensus models are able to correctly predict the chemicals tested in CADASTER according to their toxicity class, confirming the utility of the QSAR approach. The models were also applied to predict the aquatic toxicity for over 300 (B)TAZs, many of which are

included in the REACH pre-registration list, and are without experimental data. This highlights the importance of QSAR for screening and prioritization of untested chemicals, in order to reduce and focus experimental testing.

A summary of the modeLling approaches and molecular descriptors used by different partners was provided in Appendix 1 of Deliverable 3.7.

Acute toxicity in algae: External validation of Consensus models for EC50 72h in Pseudokirchneriella subcapitata by a new tested Evaluation set.

The predictive ability of individual and consensus models for the prediction of EC50 72h in Pseudokirchneriella subcapitata was checked on experimental data newly generated for 13 (B)TAZs which were prioritized and tested by PHI within the CADASTER project. This external dataset is here referred as "evaluation set" (EV set). A comparison of experimental data (training set and PHI EV set) and values predicted by the consensus model was performed. All the algae models, even if based on different descriptors and methods, are in reasonable agreement also in the prediction of (B)TAZs in the EV set. Additionally, RMSE calculated for EV set chemicals are comparable with RMSE calculated for the training set (with the exception of the IVL model), which confirms the accuracy of the QSAR predictions for new chemicals. Only four chemicals have errors in prediction larger than 0.5 and lower than 0.7 log units. Among these myclobutanil, epoxiconazole, and triazophos include in their structures particular groups and atoms such as nitrile group, phosphate group, epoxy group and halogen atoms (F, Cl) which are limitedly represented (or nor present) in the training set and therefore in the modeling descriptors. Additionally, the agreement in prediction across the different modelling methods was performed by converting experimental and predicted EC50 data into four classes of toxicity: (1) very toxic (EC50 = 1 mg/L), (2) toxic (EC50 = 10 mg/L), (3) harmful (EC50 = 100 mg/L), and (4) not harmful (EC50 greater than 100 mg/L). Experimental (PHI) and predicted (individual and consensus) classes of the B-TAZs included in the EV set were reported in Appendix 1 of Deliverable 3.7. All the models predict the 13 PHI compounds in Class 2 (Toxic) with only two exceptions: triazophos is predicted in class 1 (very Toxic) only by the UI-DRAGON model, and difeconazole, which has the highest pEC50 value and is the most toxic among the 13 tested compounds, is predicted as very toxic (in class 1 instead of experimental class 2) by all the models. Furthermore the toxicity of paclobutrazol and myclobutanil, which, according to experimental data meausured by PHI, should belong to Class 3 (Harmful), was overestimated to Class 2 (Toxic) by all the models. It should also be noted that the toxicity value reported in literature for paclobutazol [9] (7.2 mg/l; Class2 (Toxic)) is in agreement with the here calculated QSAR predictions.

Acute toxicity in daphnids External validation of Consensus models for EC50 48h in Daphnia magna by a new tested Evaluation set
The predictive ability of individual and consensus models for the prediction of EC50 48h in Daphnia magna was checked on experimental data newly generated for 12 (B)TAZs which were prioritized and tested by RIVM within the CADASTER project.

The comparison of RMSE values shows that all the models are able to predict the RIVM chemicals with a satisfactory accuracy. The consensus model has the best RMSE in prediction after the HMGU model.

Differently from the PHI experimental data for algae toxicity, the RIVM set of data has a larger range of EC50 values, but it is important to note that one chemical (guanazole) is a strong outlier, which is underestimated by all the QSAR models (residual by consensus 1.1 log units), justifying the increasing of the RMSEEV values for the majority of the models.

On the basis of the comparison performed on classes of toxicity, all the QSAR models are in general agreement and, with the only exception of guanazole, they overestimate (from class 3 to class 2) the toxicity of 4 compounds. Benzotriazole and diclobutrazol are overestimated by all the QSAR models. Paclobutrazol and myclobutanil are classified as "Toxic" by the consensus model, but not by all the QSARs, which are, in these two cases, in disagreement.

Acute toxicity in fish External validation of Consensus models for LC50 96h in Oncorhynchus mykiss by an external Evaluation set. An independent test set of LC50 values for Onchorynchus mykiss became available for 18 (B)TAZs, after the development of QSARs. This set was used as evaluation set for the external validation of the models. The external validation of the models shows that the QSARs proposed for the consensus have good external predictive ability, in the case of UI/IDEA-OLS and IVL-PLS models (Q2ext greater than 0.84), and satisfying predictive ability calculated for the ASNN model (Q2ext = 0.7). LnU model is the only QSAR with Q2ext less than 0.7. The RMSE value calculated for the consensus model is 0.37, lower than RMSE calculated for the training set (RMSETR=0.44). This demonstrates that despite the differences in variables composition and fitting performances across the individual models, their combination provides results in the prediction of external compounds which are better than the fitting of training chemicals, and in general better than the external predictive ability of each model taken individually.

A publication on consensus models of (B)TAZ aquatic toxicity and their external validation has been submitted in ATLA as proceedings of the "CADASTER Workshop on the development and application of QSAR models in REACH", held in Munich (7-9 October 2012).

Related Documents and papers:

1) Deliverable 3.7 External validation of QSAR models
2) Stefano Cassani, Simona Kovarich, Ester Papa, Partha Pratim Roy, Magnus Rahmberg, Sara Nilsson, Ullrika Sahlin, Nina Jeliazkova, Nikolay Kochev, Ognyan Pukalov, Igor Tetko, Stefan Brandmaier, Mojca Kos Durjava, Boris Kolar, Willie Peijnenburg, Paola Gramatica, Evaluation of CADASTER QSAR models for aquatic toxicity of (benzo-)triazoles and prioritization by consensus. Submitted to ATLA, 2013.
3) Simona Kovarich, QSAR models for the (eco) toxicological characterization and prioritization of emerging pollutants: case studies and potential applications within REACH, PhD thesis, 2013.

2) Consensus models and External validation of QSARs developed for ready biodegradability of fragrance materials

The consensus model developed within the CADASTER project for the prediction of ready biodegradability of fragrance materials was externally validated on the dataset composed of 45 fragrances, 11 of them were tested by the WP2 Partner PHI. These 11 chemicals were previously selected by UI as priority compounds on the basis of the available

information on potential toxicity (cyto-toxicity and mammalian toxicity) and structural representativeness (Task 3.4 – report Deliverable 3.4). The tested fragrances were separated into ready and not ready biodegradable according to the biodegradation percentage, i.e. RB if biodegradation greater than 60% (pass level for OECD 301D test) and NRB if biodegradation less than 60%.

Consensus predictions for the 45 fragrances included in the validation set were obtained by assigning the class predicted by the majority of the six models considered. In case of conflicting predictions (i.e., 3 models classify the compound as RB and 3 models classify as NRB), the NRB class was assigned, as the most precautionary.

Predictions obtained by individual and consensus models for 45 fragrances were reported in Appendix II of Deliverable 3.7.

As a general comment it was observed a fair agreement among the classes predicted by different WP3 models for RB and NRB chemicals. The fact that different QSAR models, based on different descriptors and/or modeling approaches, lead to similar predictions adds confidence and reliability to QSAR predictions obtained by the consensus approach.

To further on verify the predictive ability of the models toward new chemicals, the individual and consensus models were tested on a final "RIFM dataset" consisted of 34 chemicals belonging to different chemical classes i.e. Aldehydes/Cyclic, Esters/Salicylates, Heterocycles/Oxygen Containing/Pyrans, Ketals/Cyclic and Phenols/Alkoxy, with satisfying results.

Predictions obtained by the consensus model were also compared with prediction generated by applying two freely available tools, i.e. the START program implemented in ToxTree (v. 2.5.1) and the widely used BioWIN, which is implemented in the EPI Suite estimation program (v. 4.1).

The estimation programs used for the prediction of ready biodegradability of fragrance materials showed poor classification performances in comparison to the consensus model developed within the CADASTER Project, whose overall classification accuracy exceed of 20-30% the OA% of BioWIN and ToxTree models.

This highlights the importance to develop ad hoc QSAR models when dealing with specific classes of chemicals.

Related Documents and papers:

1) Deliverable 3.7 External validation of QSAR models
2) Simona Kovarich, QSAR models for the (eco) toxicological characterization and prioritization of emerging pollutants: case studies and potential applications within REACH, PhD thesis, 2013.

WORKPACKAGE 4 – INTEGRATION OF QSARs WITHIN HAZARD AND RISK ASSESSMENT

The objectives of this WP, as specified in the Description of Work, are:
1) To propose how QSAR models can be applied and integrated into a probabilistic risk assessment framework,
2) To explore the feasibility of economic valuation of chemicals impacts,
3) To evaluate how QSAR models can meet the legal requirements of current and upcoming chemicals legislation (REACH),

4) To synthesise research findings and recommendations from the other work-packages in the programme and prioritize chemicals.

These objectives have been addressed by completing the five tasks distinguished in the DoW. These tasks are presented below together with a description of the work carried out and major conclusions.

WP4 - task 4.1 QSAR models in a probabilistic risk assessment framework
Several case-studies have been made to demonstrate the integration of QSAR models into a probabilistic risk assessment framework. These case-studies have been or are in a process of being published in peer-reviewed journals.

Variability and uncertainty of QSARs e.g. identified in WP2 and developed in WP3 have been characterized both qualitatively (e.g. as the confidence in individual predictions) and quantitatively (by a probability distribution for the error compared to experimental obtained estimate). The characterization of uncertainty has been made possible by suggesting a conceptual framework to define and understand uncertainty in a QSAR prediction in a decision context and to provide a framework for the theoretical understanding of approaches to assess uncertainty in predictions covering probability models and principles for statistical inference (submitted to CADASTER second workshop proceedings in ALTA). Focus has been to derive a conceptual framework for predictions for QSAR regressions (i.e. with a continuous as oppose to categorical response) which is common within environmental risk assessments and for which uncertainty is less often reported probabilistically. The characterization of predictive uncertainty would benefit from a probabilistic formulation of QSAR models and models with point estimates as output may be turned into a probabilistic framework without any loss of validity from a chemical point of view. However, a QSAR model (or a supervised learning algorithm in combination with a QSAR data set and analogy statement) for use in probabilistic risk assessment needs to be validated for its ability to make reliable predictions and to quantify associated uncertainty.

Some of the case-studies include an extension of classical uncertainty analysis, where the treatment of uncertainty in QSAR predictions also considers qualitative uncertainty. Based on a measure of similarity or distance to the QSAR models domain of applicability, the confidence in a QSAR prediction is evaluated. When confidence (i.e. predictive reliability) is low, this is allowed to be reflected in the uncertainty analysis by enlarging uncertainty for that input parameter. We have evaluated two ways to do this, by widening the predictive probability distribution or by creating a probability box for the input parameters. The latter case results in a need to propagate non-probabilistic measures of uncertainty into an assessment model. Therefore a classical Mackay level I model was implemented in the Risk Calc computation environment and two approaches to treat systematic uncertainty in experimental data, fuzzy arithmetic and probability bounds analysis, were evaluated. Propagation of uncertainty by discrete probability calculus or interval arithmetic can be done at a low computational cost and gives maximum flexibility in applying different approaches. Propagation of uncertainty using single probability distributions was made with Monte Carlo simulation techniques. The extended uncertainty analysis is a novel approach to consider qualitative uncertainty in quantitative assessments, with a potential of wide use for any kind of applications.

The contribution of individual QSAR predictions to the overall risk assessment framework has been evaluated by sensitivity analyses in a set of QSAR integrated fate and effect assessments (Golsteijn, Iqbal et al. in review; Iqbal, Golsteijn et al. in review; Golsteijn, Papa et al. manuscript). QSAR modeling of variability have been made with regard to their application in hazard assessments based on the approaches of Assessment Factor (AF) approach and Species Sensitivity Distribution (SSD). Current guidance for QSAR integrated probabilistic risk assessment together with actions taken within CADASTER to address identified gaps in guidance are described in Deliverable 4.2.

A modeling framework to assess QSAR uncertainty in a comparative toxicity assessment of triazoles based on QSARs has been developed. The modeling framework can also be used for life cycle impact assessments and risk assessments. It was implemented on a case-study on triazoles. We found that a uncertainty in a measure of Comparative Toxicity Potential (CTP) was mainly determined by uncertainty in biodegradation rates and soil sorption, together with the small number of species sampled in the hazardous concentration. By contrast, the contribution of uncertainty in species-specific toxicity predictions was relatively small. Therefore, the reliability of CTP predictions for triazoles can particularly be improved by including experimental data for biodegradation and soil sorption, and developing toxicity QSARs for more species.

WP4 - task 4.2 Testing ECETOC's TRA tool

The feasibility of using QSARs in risk assessment has been evaluated from a user's perspective (Deliverable 4.3). The ECETOC TRA screening risk assessment tool was evaluated establish the ability of the tool for identifying chemicals of concern, at varying levels of data/information. The potential and current use of QSARs in the legal framework was described in Deliverable 4.5. Methods and decision points for the establishment of scientific validity and applicability domains for QSAR models have been evaluated both in the regulatory context and in practical assessment (Deliverable 4.3 and 4.5). The possibilities for validation to judge the confidence in individual QSAR predictions in the context of decision making has been explored by definition of suitable terminology and method development (Deliverable 4.2).

Within this task, the ECETOC Targeted Risk Assessment (TRA) tool for screening chemicals within a risk assessment were evaluated to assess its potential for SMEs as a screening level alert tool. It is essential under REACH that chemicals that could be of high concern are quickly identified and assessed. SMEs are, by their nature, unlikely to have the resources to enable them to do this, and the availability of such a tool would be helpful to them in allowing a quick assessment, prioritisation and allocation of resources where it is needed.

The aim was to understand the model in the TRA tool in order to perform a complete evaluation. Originally some of the QSAR models developed in work package 3 were to be used as input to the model. However, due to the lack of data, no QSAR models for ecotoxicological effects have so far been developed under this task duration. This has limited the evaluation to include only predicted environmental concentration without any relation to estimated effect reference values. We collaborated with the Chemitecs project to retrieve more data and also to have a case study to compare with. The model evaluation was performed following mainly two different strategies. A comparison with risk assessments made in the project Chemitecs was done and the results evaluated. And secondly, the case

molecules involved in the CADASTER project were all assessed in the tool and the results evaluated. In the evaluation of the case molecules a Design of Experiment (DoE) for each chemical group (BDE, PFC, (B)TAZ) was also included.

The focus of the evaluation of ECETOC's TRA tool was to assess its usage potential especially for SMEs. The user needs to understand how his choices will affect the output of the risk assessment. Therefore, gathering an understanding of how the model works is important from a non-developer point of view. If the tool is supposed to be used by an SME and perhaps personnel, which are not greatly experienced with risk assessments, the explanations of how the tool is working needs to be greatly improved. At the moment the documentation is mainly concerned with handling the tool.

The data analysis has shown that octanol-water partitioning constant and water solubility has a large effect on the output of the model. When compared to other tools this is not surprising. In the analysis, due to lack of data, the biodegradability was not included. This is unfortunate since the biodegradability is an important factor when performing risk assessments. It was harder to establish with certainty how influential the vapour pressure is on the model and how much effort needs to be spent at assessing this parameter. The evaluation shows that the molecular weight has very little or no effect on the output. It is surprising that this parameter is a compulsory parameter anyway. A surprising aspect of the ECETOC TRA tool is that the release area is not explicitly specified. In the tool the area is specified through the SpERCs, which seems like a very uncertain estimation.

When using different Specific release classes in the tool generally it seems as the only difference is a scaling. This has not been fully evaluated though but should be considered in the future. There could be other methods which might be more efficient for this target. Data availability is a key factor to get accurate results. The greatest limitation when working with the tool is data. Especially when dealing with SMEs, this need to be attended to and the tool should be designed in a way, which keeps the input data at each different scope at the minimum.

To automatically identify chemicals of concern there is a lot of input data needed. Not only the physico-chemical data but also tonnage, Environmental Release Class and at least one PNEC (Predicted No Effect Concentration) -value. There would be much to be gain from cutting down the information needed for the first scope and simplify the use. For instance just by looking at Kow (Partitioning coefficient of octanol-water), water solubility and vapour pressure it can be established whether a compound is likely to reach high concentrations or not.

Related Documents and Papers:
Deliverable 4.3. Evaluation of ECETOC TRA tool. Available at http://www.cadaster.eu/node/110
WP4 - task 4.3 Economic valuation of chemical impacts on ecosystems
One of the aims of REACH is to obtain a proper balance between societal, economic and environmental objectives. Analyzing the trade-offs between the benefits and the environmental impacts of chemical substances is, however, not straightforward. Benefits of the production and application of chemicals can relatively easily be expressed in monetary units, but ecosystem properties and processes are not or only partly captured in commercial markets.

A literature review revealed that only very few studies have translated chemical impacts on the environment into cost estimates. In these studies, the costs of unwanted chemical impacts were based on surveys of people's willingness-to-pay (WTP) for a reduction of the impact. Yet, such surveys provide no univocal monetary estimates, as WTP depends on characteristics of the respondents, like their origin and socio-economic status, as well as characteristics of the survey design, like survey mode, payment frequency, response rate, and survey year. To obtain more objective cost estimates, we developed a model designed to quantify the direct costs that are required to keep the number of animals in a wildlife population on the level corresponding with uncontaminated conditions.

We applied the model to assess the costs of the impacts of polybrominated diphenyl ethers (PBDEs), which is one of the chemical classes of focus in CADASTER, on a population of peregrine falcons (Falco peregrinus). To that end, a model was developed to quantify the direct costs that are required to maintain a wildlife population at a user-defined size at a given level of contamination. The population model as described in the deliverable has been improved in order to account for differences in survival and fecundity between the different life stages of the species of concern. Moreover, the model has been made fit for assessing the joint impacts of multiple chemicals. The model has been applied to assess the impacts and costs of polybrominated diphenyl ethers (PBDEs) as well as dichlorodiphenyldichloroethylene (DDE) on a population of peregrine falcons.

The population modeling study revealed non-linear responses to PBDEs and DDE of the density-dependent peregrine falcon population in California. Initially, decreases of fecundity due to increasing toxicant exposure were largely compensated for by an increasing probability of territory acquisition. However, once the probability of territory acquisition approached a maximum, the toxicant impacts were no longer buffered and a steep decline was observed in the equilibrium population size (carrying capacity). Cost estimates showed non-linear increases with the contamination levels as well as the user-defined target equilibrium population size. Our findings imply that density-dependence in a population may create pollution thresholds that can influence population persistence: the carrying capacity of a population may be unaffected over certain ranges of environmental pollution and show a sudden, strong decline when the pollution approaches critical levels.

Two scientific papers are being prepared based on the results of this task. The first paper deals with the effects of PBDEs and DDE on the peregrine falcon population and has been submitted in July 2012 (Journal of Applied Ecology). A second manuscript, focused on the corresponding economic valuation, is currently in preparation.

Related Documents and Papers:

Deliverable 4.4. Evaluation of options for economic valuation of chemical impacts. Available at http://www.cadaster.eu/node/110
A.M. Schipper, H. Hendriks, M.J. Kauffman, A.J. Hendriks and M.A.J. Huijbregts (submitted). Density-dependent population responses to multiple toxicants: implications for wildlife management.

WP4 - task 4.4 QSAR models in the legal framework

The objectives of this task were:
1 - To evaluate the methods and decision points for the establishment of scientific validity and applicability domains for QSAR models (in cooperation. with WP 3).
2 - To evaluate the need for documentation (of models) with regard to current progress in the OECD and REACH implementation (cooperation with WP 5).
Since the start of the project and the actual time frame for this task, the first registration round in REACH for chemicals produced and used at the highest tonnages, has been finalized. This means that information now is available on the outcome of the use of QSAR in REACH. In this task we reviewed this information and summarized the status from different other sources and projects about the use of QSAR in REACH.

The use of QSAR is in the REACH framework treated in Article 13, 25 and in Annex XI. Article 25 declares that tests on animals should be a last resort when all other options, such as QSAR, are considered. In Article 13 the rules are given for how to generate necessary information by actual testing or by QSAR models.

In Annex XI it is stated that a result of QSARs may be used if certain criteria are fulfilled:
- Results may be derived only from a (Q)SAR model whose scientific validity has been established
- The substance must fall within the applicability domain of the (Q)SAR model
- Results must be adequate for the purpose of classification and labelling and/or risk assessment
- Adequate and reliable documentation of the applied method needs to be provided

Considering the use of QSAR in REACH there are a few guidance documents and a number of practical guides published by ECHA:
- Regulation (EC) No 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)
- Practical guide 5: How to report (Q)SARs
- Practical guide 10: How to avoid unnecessary testing on animals
- Guidance on information requirements and chemical safety assessment - Chapter R.6: QSARs and grouping of chemicals

In a report from JRC-ICHP (Joint Research Centre - Institute for Health and Consumer Protection) the frame work for documenting QSARs and their predictions are reviewed and it is documented in this report how to develop the frame work for documenting QSARs for better guidance in regulatory purposes. A check list is suggested of 10 key questions that the risk assessor should go through when evaluating a QSAR model in a regulatory purpose. Not all questions should be answered if they are not needed in the actual regulatory context, and in the same way additional questions could be added. These questions will assess the practical applicability of the QSAR and the adequacy of the predictions.

The deadline for the first registration round in REACH was the 30th of November 2010. The outcome of this registration round has now been evaluated and reported by ECHA. There are 33 test types to be reported in the REACH registration process. This includes different routes of the chemical into the test object and also the duration of the test. But to compromise the result, the different routes and the test duration are merged together in one resulting test type for each category. This

resulted in 20 test types presented. In the reporting of the substances seven different categories can be chosen for how the information of the current test type is derived.

They are:
- Experimental studies.
- Testing proposal.
- Read-across.
- IUCLID flags to omit the study. These options are to be used to indicate when testing does not appear to be: scientifically necessary; technically not possible; or not necessary based on low exposure considerations (abbreviation FO).
- Weight of Evidence.
- QSAR.
- Miscellaneous. None of the above fit the information inserted in the registration process.

Registrants in the first registration round in REACH have almost in every test type given experimental studies as the main source. The dominating alternative method reported is read-across. For the use of QSAR, which is our focus, there are very little results reported. In twelve of the test types QSAR has been reported but the number is very small, less than 1 percentage. These test types and endpoints include: Acute toxicity, Skin irritation (in vivo), Skin sensitisation (in vivo), Genetic toxicity (in vitro), Toxicity to reproduction, Developmental toxicity, Carcinogenicity, RDT all routes and duration, and "Additional ecotoxicological information". The figures are a little bit higher for the test types concerning the environmental compartments: Bioaccumulation (Fish), Short term toxicity to fish, Long term toxicity to fish. For these types the use of QSAR is about 2-3 percent. This corresponds to respectively 25, 267 and 153 actual number of QSAR models used. In total 433 QSAR models were reported.

Read-across and Weight of Evidence (WoE) approaches as alternatives methods are more applied in the first registration round, than QSAR models. However, a QSAR model could nevertheless have been applied in a Read-across approach to fill data gaps. When it comes to WoE, QSARs are probably included in this approach but it is categorized as a WoE without stating that QSARs have been used. The outcome of a WoE approach includes different types of information collected, for a specific end-point; if a QSAR model is used the results are evaluated together with the additional information that is available. The QSAR model is not alone in providing the information needed in the REACH requirements. This will probably be one reason for the registrant to choose the WoE approach instead of "just" one or more QSAR models since the justification and evaluation are based on more information than just the QSAR model, thus probably reducing the uncertainty on how ECHA will treat the reported QSAR model. Therefore, the registrants put all information available in the WoE approach to be on the safe side in the evaluation process. In the statistics from ECHA, the QSAR models applied in Read-across and WoE are not included. It is hard at this stage to give any number of QSAR models inside of these approaches but these numbers certainly will be high in WoE approaches since in the guidelines there is a recommendation to incorporate all possible information on individual end-points. It will be interesting to see if ECHA will further analyse the QSARs used in these two approaches and how these models are reported.

Related Documents and Papers:

Deliverable 4.5. Evaluation of QSAR models in the legal framework. Available at http://www.cadaster.eu/node/110

WP4 - task 4.5 Policy and management

The activities within this objective focus on the integration of the core findings of the CADASTER project and on the generation of actual risk assessment for chemicals belonging to the four classes of chemicals selected within the project. The activities in this respect have been the integration of predictive tools and the assessment of the uncertainties in the probabilistically-based risk assessments of the chemicals selected.

The recommendations on a viable management strategy for optimized testing and in-silico modeling of hazardous organic substances have been based on merged research findings from CADASTER with other ongoing research and regulatory developments (Deliverable 4.6). The synthesis of research findings is backed up with quantitative assessments in co-operation with the other work-packages. Such co-operation has for example resulted in case-studies on QSAR integrated hazard, fate and risk assessments using data or models identified in WP2 and QSARs developed in WP3. There has been a close co-operation between WP4 and WP5, and a flow of suggestions, code and assessment frameworks to support the development of tools such as the CADASTER web and the Ambit software. The selection of compounds to test in WP2 has been done in co-operation with WP4 and WP3.

The aim of REACH of achieving a proper balance between societal, economic and environmental objectives have been addressed by case-studies exploring the possibilities by economic evaluation of the impacts from hazardous chemicals (Deliverable 4.4 and Appendix PBDE.2) and by discussing and demonstrating the impact and usefulness of considering uncertainty in QSAR predictions in decision making (Sahlin in review). The consideration of uncertainty opens up for more precise characterization of uncertainty allowing for different attitudes towards risk, and support both safer risk management as well as more precise risk assessments.

In general it is to be concluded that the integration of alternatives into probabilistic risk assessment is possible given proper assessments of (predictive) uncertainty and (predictive) reliability to inform the characterization of parameter and model uncertainty. A framework has been demonstrated that allows for quantifying the predictive distribution of parameter estimates, as for the case of QSAR predictions denoted in the workflow for QSAR-based risk assessment. The case-studies demonstrate the need of a general conceptual framework for probabilistic risk assessment. This conceptual framework is needed to define and understand uncertainty in a QSAR prediction in a decision context and to provide a framework for the theoretical understanding of approaches to assess uncertainty in predictions covering probability models and principles for statistical inference.

Related Documents and Papers:
In addition to the papers listed under task 4.1. there are:
Deliverable 4.6. Synthesis of research findings and recommendations for prioritization. Available at http://www.cadaster.eu/node/110
Igor Tetko, Pantelis Sopasakis, Prakash Kunwar, Stefan Brandmaier, Sergii Novotarskyi, Larisa Charochkina, Volodymyr Prokopenko, Willie

Peijnenberg. Prioritization of PolyBrominated Diphenyl Ethers (PBDEs) using the QSPR-THESAURUS webtool. ATLA submitted (2013)
Rorije, E., Aldenberg, T and W. Peijnenburg (in review). Read-Across estimates of aquatic toxicity for selected fragrances. Submitted to ATLA.

WORKPACKAGE 5 – OUTREACH VIA DEVELOPMENT OF WEBSITE, NEWSLETTERS/WORKSHOPS AND STAND ALONE TOOLS FOR DISSEMINATION OF PROJECT RESULTS

Igor V. Tetko9, Willie J.G.M. Peijnenburg,1 Mojca Kos Durjava,2 Paola Gramatica,3 Magnus Rahmberg,4 Ullrika Sahlin,5 Nina Jeliazkova,6 Mark A.J. Huijbregts,7 Mike Comber,8
1Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany, 3Public Health Institute Maribor, Slovenia, 4QSAR Research Unit in Environmental Chemistry and Ecotoxicology, University of Insubria, Via J.H. Dunant 3 - 21100 Varese, Italy, 5IVL Swedish Environmental Research Institute, Box 210 60, SE- 100 31 Stockholm, Sweden, 6Linnaeus University, School of Natural Sciences, 391 82 Kalmar, Sweden, 7Ideaconsult Ltd., Sofia, Bulgaria, 7Radboud University Nijmegen, The Netherlands, 8Mike Comber Consulting, Bruxelles, Belgium, 9National Institute for Public Health and the Environment, Bilthoven, The Netherlands

The main activities within WP5 had its central role for collection of all results generated within other workpackages and disseminating them outside of the project by means of website, reports, presentations, publications, stand-alone tools, etc.). These goals were reached by means of the establishment of the public website, database of molecules and their properties, development of on-line and standalone tools for development, upload and publishing QSAR/QSPR models, the organisation of two CADASTER workshops, the publication of the CADASTER newsletters.

The major outcomes of WP5 are the development of a public database of molecules and their properties and dissemination of models, tools and results developed within the project to the external users.

Below, we briefly overview them:
Dissemination of information by the project website: A web site for CADASTER project http://www.cadaster.eu was established. This site provided a dissemination of information about the activities of the project, publications, presentations at the conferences and meetings as well as contains all deliverables (public deliverables are open to all users) of the project. The web site was continuously supported during 4 years of the project and served as the central dissemination point. The CADASTER publications, announcement of events, agenda and materials of workshops, project publications and deliverables were timely and continuously upload there by the HGMU team. The project web site was attracting more than 2000 visitors per month. It was also used to organize "Environmental Toxicity Prediction Challenge" http://www.cadaster.eu/node/65, which was co-organized with International Conference on Neural Networks (ICANN) http://www.kios.org.cy/ICANN09. The challenge attracted 518 submissions from 108 participants from more than 25 countries world-wide. The website also provided dissemination of newsletters and materials and tutorials of two CADASTER Workshops (in Maribor and Munich), which provides a comprehensive coverage of tools available at the web site.

Dissemination of information by publications and newsletters: The publication of materials of the project in peer-reviewed scientific journals was an essential part of the dissemination of the project results. Up to know 49 articles with project results have been published or submitted for publications in scientific literature (a couple of articles are under preparation). The project results were presented as 64 posters and 65 oral presentations at 51 conferences, meetings and workshops. The project results were also summarized in newsletters, which were sent to more than 6,000 registered users. The project has been highlighted in national publications, including newsletters of RIVM, HMGU and of other partners.

Dissemination of information by on-line database: An on-line QSPR-THESAURUS database http://www.qspr-thesaurus.eu, which stores properties of chemical compounds and QSAR/QSPR models was developed. It has been used by the CADASTER partners to upload, verify and curate experimental data and to upload models developed during the project. The database currently contains 5,5k data points collected for greater than 800 molecules from four analyzed classes. More than 95% of data points in the database were collected by HMGU, PHI and UI groups, which contributed 2.9k, 1.6k and 0.9k datapoints respectively. All this information is publicly available to the external users. The external users can access, search, upload and download data available in the database for the four analyzed classes of molecules. The users can also perform calculations using models contributed by the individual partners to new chemical structures and use predictions of these models for the risk assessment. Currently the database contains 30 models, majority of which (greater than 85%) were contributed by UI and HMGU groups. The user can access the developed tools either through the web-interface or with help of web services and standalone tools, which provides remote calculations on the QSPR-THESAURUS web site (developed by HMGU) or through http://toxpredict.org web site of OpenTox project (contributed by IDEA). The users can also upload and publish their own models. The (Q)SAR Model Reporting Format (QMRF) format was integrated as a part of the developed tools. This feature allows the CADASTER partners and external users to report information about published models, which is required for the use of models in the REACH assessment according to “OECD principles for QSAR validation and application in regulation”.

The optimization of molecular structures http://mopac.cadaster.eu for CADASTER project is done by liberally user-contributed CPU time, who are thus aware about the project. Thus, it provided another channel to disseminate the information about the project activities. Since start of the project more than 1600 users have contributed 11.9 years of CPU time to optimize molecules.

The developed tools incorporate methods for applicability domain (AD) assessment, experimental deign and risk assessment, as summarized below.

The methods for the assessment of the applicability domain (AD) of models were proposed, developed and benchmarked for regression and classification methods as, reported in WP3 and WP4 and were made publicly available through the QSPR Thesaurus http://www.qspr-thesaurus.eu site. All the models published on the QSPR-Thesaurus web site provide information about the AD and accuracy of estimation (i.e. uncertainty in QSAR prediction assessed by one of the approaches implemented in the web tool). The estimation of the AD was integrated by embedding OpenTox web services, thus integrating tools developed in our partner project.

A number of modern experimental design methods, including original developments, were implemented and evaluated, as reported in WP3. The major results were summarized in PhD thesis of Mr. Brandmaier (HMGU). The approaches providing the best results were made available to the external users as part of tools provided at the QSPR-THESAURUS http://www.qspr-thesaurus.eu web site.

The approach for QSAR-integrated hazard and risk assessment developed in WP4 have been implemented into the web tool http://www.qspr-thesaurus.eu/risk. It includes calculation of the fate assessment of chemical compounds using pre-defined emission scenario implemented within SimpleBox Excel sheet-based program contributed by RIVM scientists. The use of individual QSAR models in species sensitivity distributions (SSDs) is done using a methodology developed in WP4. Intermediate calculations are stored and can be easily downloaded for external assessment by the end-users. Issues related to the treatment of uncertainty, especially associated to QSAR predictions, in fate and effect assessment were reported by the PhD thesis by M Sarfraz Iqbal (LNU).

The next part of the summary provides an overview of the major activities performed from January 2009 to December 2012, within the CADASTER Project by WP5 project partners. Activities are summarized in the following paragraphs according to the timeline of the CADASTER Project, and are organized in 7 tasks.

- Task 5.1 Development of a prototype of the http://www site (experimental database).
- Task 5.2 Development and testing of a prototype of the http://www site (models).
- Task 5.3 Implementation of tools to estimate the Applicability Domain of Models and Experimental Design.
- Task 5.4 Public Release
- Task 5.5 Development of a stand-alone version of the tools.
- Task 5.6 Workshop on the use of QSAR models in REACH
- Task 5.7 Final Workshop and other disseminations of information.

The detailed analyses of the activities performed within each task were reported in the respective deliverables.

Task 5.1 Development of a prototype of the http://www site (experimental database).

This task was to develop a prototype of the QSPR Thesaurus database, which would be suitable for download and storage of data required for the project. The QSPR Thesaurus was based on the OCHEM platform developed by HMGU group, which was adapted to the requirements of the CADASTER project. HMGU team also provided training and continuous support of other project members with respect to the use of the web site during the project. The database allowed the CADASTER users to submit, store and annotate molecular records collected from the literature. It stored data in original units, tracked users and modifications to the data that were performed by them, allowed to introduce new units, new properties. The database automatically checked for duplicates, allowed editing of single or several records simultaneously, performed batch upload of data as Excel and/or SDF files, and allowed exporting data as Excel files. A concept of TAGs was used to identify areas of interests and to restrict work of users only to CADASTER classes. Moreover, HMGU group also developed tools for calculation of MOPAC descriptors following structure

optimization using AM1 method. HMGU and IDEA have also contributed a number of web services to extend CADASTER web-database by integrating AMBIT functionality for substructure search.

Task 5.2 Development and testing of a prototype of the http://www site (models).
The goal of this task was the development of an infrastructure and a prototype of the database of models that could be used by CADASTER participants to upload their previously developed models, publish them and make them available on-line to CADASTER and, later on, to external participants. The task was completely achieved. The QSPR Thesaurus also incorporated tools for linear model upload, that were validated for the upload of models by the CADASTER participants. It was found that reproducing of models requiring 3D structures is a challenging task. In order to have a possibility to provide sustainable platform for such models on the CADASTER web site, a pipe-line for optimization of molecules using MOPAC 7 program was developed and integrated as part of the web site. A graphical interface to submit and access optimized conformation of molecules was developed http://www.cadaster.eu/mopac. The optimization of molecules was implemented based on the BOINC software http://boinc.berkeley.edu (which allows usage of user-contributed calculations during screen saving mode of computers) and the respective web site http://mopac.cadaster.eu to handle this development was developed and was supported during the project by the HMGU group.

Task 5.3 Implementation of tools to estimate the Applicability Domain of Models and Experimental Design.
The main goal of this task was to implement tools to estimate the Applicability Domain of Models and Experimental Design, including both the implementation of new approaches for the AD estimation, developed in WP3, and the integration of available AMBIT tools to estimate the applicability domain (AD) of models. All these goals were achieved. Moreover, the (Q)SAR Model Review Format (QMRF) was extended to enable the display of validation results for published models. Users can add several predefined parameters (R2, RMSE, Q2Loo, Q2ext, etc.) for both internal and external validation of their models. The possibility to add any kind of graphical representation (graphs, tables, images) to support or explain the validation results as well as applicability domain of models was also implemented. Thus, the models could be published on the CADASTER web site together with their QRMF as required for their use by the regulators. Experimental design methods were developed and made available on the web site of the project. The methods integrated already established algorithms as well as new step-wise approaches, which take into consideration the target property. The idea of such approaches was first proposed as conclusions of the deliverable of this task. During next two years of the projects, it was further developed and resulted in several publications (Brandmaier et al, 2012a/b, 2013) as well as in PhD thesis of Mr. Brandmaier (2013).

Task 5.4 Public QSPR-THESAURUS site
The goal of this task was to make the database and tools developed within the CADASTER project publicly available to web users. The deliverable of this task contained an overview of the main features of the database and the additional tools that were developed and made available to the external users. The upload of data to the database and development of models were further continued during the project. The functionality of QSPR Thesaurus was demonstrated to the participants of two workshops, which were organized by CADASTER project.

Task 5.5 Development of a stand-alone version of the tools.
The goal of this task was to provide tools to remote access models developed and published within the CADASTER project on the QSPR-THESAURUS web to the external users. While it was expected that the major access to the models and the results would be accessed through the web interface, the remote standalone tools could provide an easy way to integrate the CADASTER models into other software packages by the external users. Within this task we did a number of developments that were requested by CADASTER users, e.g. incorporation of Dragon 5.4 and 5.5 packages (in addition to Dragon 6.0); authorization to download descriptors to the users that have the respective license (however, all users can use the developed models for free on the web site); storage of predicted values for molecules with 3D descriptors, development of enhanced web interface and web SOAP services to allow remote calls to the models developed by CADASTER project participants. A standalone version of tools to remotely access CADASTER web models based on Perl and Java code were developed. The models published on the web site of the project were integrated and made available through the OpenTox thus validating our development. We have also presented CADASTER models to OECD QSAR Toolbox meeting (Dr Tetko and Dr. Rorije in October 2011 and Dr. Jeliazkova and Dr. Rorije in Spring 2012) but no decision about implementation of the models within the OECD QSAR tool was taken by them. Indeed, since CADASTER models are publicly and freely available on-line, the users can easily access them, predict new molecules and use predicted values for the purpose of their analysis with OECD QSAR Toolbox.

Task 5.6 Workshop on the use of QSARs models in REACH
The main goal of this task was to organize a workshop
http://www.cadaster.eu/node/116 which would provide a dissemination about the project development and results and to assist the risk assessors and national chemicals authorities, particular in Eastern European countries, with the use of the QSAR tools for the environmental risks assessments in REACH. The workshop involved 35 participants, including invited speakers from JRC - Institute for Health and Consumer Protection (Italy), ECHA – Evaluation Unit (Finland), Douglas Connect (Switzerland) and University of North Carolina USA). It was organized by the Public Health Institute Maribor (PHI) in Maribor, Slovenia and took place from September 1st to September 2nd 2011. The program of the CADASTER workshop was a combination of plenary sessions to provide a common perspective to all of the attendees and of training lessons for on-line tools that were developed during the project and can be used to estimate REACH endpoints for chemical compounds. The workshop was closed with a panel discussion, which was important to understand different positions of chemical regulator, academy and industry.
Task 5.7 Final workshop and guidelines for model development
The second CADASTER workshop (see http://www.cadaster.eu/workshop online) was organized by the HMGU and took place in its premises from October 7th to 9th 2012. The main goal of this workshop was to provide a tutorial to all interested partners, including industry and SMEs, on how to develop new models for the assessment of REACH-end points (in particular for new scaffolds of compounds for which there are no reliable QSAR models) and how to use the software developed by the project participants. The workshop was attended by 52 participants, including invited speakers from JRC - Institute for Health and Consumer Protection (Italy), EPA – Environmental Protection Agency (USA), Umwelt Bundesamt (Germany) and coordinators of the Seventh Framework Programme (FP7) funded projects OSIRIS (Prof. G. Schüürmann) and COSMOS (Prof. M. Cronin).

There were two main thematic areas of the workshop:
1) Data collection and QSAR Model development for REACH (October 8th) and
2) Case studies and use of QSARs in the risk assessment (October 9th).

The lectures of project participants were followed by lecture of invited speakers. A poster session with 20 posters was organised during the second day of the workshop. The materials of the workshop that summarize expertise of all participants, main results and lessons are under publication in a special open-access issue of ATLA journal. The decision to publish proceedings in an open-access issue was taken by the General Assembly of the CADASTER project (February 2012) to optimize the dissemination of the project results.

Potential Impact:

Potential impact and main dissemination activities
The REACH legislation requires chemical industries to demonstrate the safe manufacture and use of their chemicals throughout the supply chain. The major innovation of CADASTER concerns REACH-targeted assessment of environmental effects and exposure of chemicals belonging to four chemical classes of emerging compounds, whilst considering the European diversity from the viewpoint of the stakeholders who are primarily responsible for carrying out safety, hazard and risk assessment, and who face the task of actually integrating all tools that are made available for application within REACH for 'their' classes of compounds. This is to be done whilst minimizing animal testing and costs of assessment. CADASTER merged high-costs endpoints of environmental toxicology, intelligent combinations of in silico techniques which in turn were combined with other alternatives to animal testing (like read across), risk-targeted decision support systems and economic valuation of substitution of chemicals from within chemical classes, and it integrated probabilistic precautionary approaches to chemical risk assessment with science-based assessment and management. CADASTER thus provided the tools and scientific and pragmatic insights to enable Europe to become a stronger partner within the international area of regulating industrial chemicals. The innovation achieved in science and in regulatory risk assessment will improve the position of Europe in influencing decision-making processes about testing strategies that take place in international bodies such as the OECD. The active involvement of industry and SMEs in CADASTER foster their international competitiveness through early exploitation of innovative know how to design and develop low hazard and low risk chemicals. Thereupon, special effort was put in dissemination of project results and project tools towards Eastern European countries, with the aim of improving the chemical assessment skills of the regulatory bodies in these countries. CADASTER assisted these stakeholders in the implementation of REACH and will serve to guide these stakeholders through all aspects of ITS in its widest sense, including the economic valuation of substitution of chemicals from within chemical classes.

To proper implement REACH, concerted action and intensive efforts are needed to operationalize all possible alternatives into a workable, consensually acceptable, and scientifically sound strategy for hazard and risk assessment of large numbers of chemicals. The production of guidance and (web-based) tools is essential in this respect. So far, the use of non-testing methods in the European regulatory context is quite limited and fragmented. Reasons include the lack of distinct application criteria and guidance, and the fact that uncertainty has not been addressed rigorously. Practical guidance is needed on how to apply the elements of newly derived testing strategies in a consistent manner to assist industry, SMEs, and regulators in carrying out the risk assessments. CADASTER provides practical guidance to integrated hazard and risk assessment procedures by exemplifying a hazard and risk assessment for chemicals belonging to four specific compound classes by integrating the various tools that are made available within the project for each of the four compound classes. The tools and the underlying data and models are made available via the project website http://www.cadaster.eu as an on-line and standalone tool for development, publishing and use of QSAR models for REACH. The CADASTER tools predict physico-chemical properties and toxicities for four analysed classes. The predictions provided are compatible with the OECD QSAR Application Toolbox and the EPI SuiteTM

Toolbox developed by the EPA's Office of Pollution Prevention Toxics and the Syracuse Research Corporation (SRC).

In exploiting and disseminating the results of the CADASTER project, the emphasis was on making the results publically available. A central role in this respect was played by the publically available project website, and it was made sure that this website is continuously updated to contain all up to date results of the CADASTER project in the widest sense. The website will also in future be available for developing alternative tools to animal testing, including chemicals and chemicals classes that are not within the domain of the four compounds classes that were used to exemplify the use of alternative methods to animal testing in the CADASTER project. On top of the central role of the website in dissemination of project results, the impact of the project was increased by various additional dissemination activities like organisation of workshops, oral and poster presentations at scientific meetings, publications in peer reviewed (partly open access) journals, reports, publication of newsletters, direct exchange of information with stakeholders, development of stand-alone internet based tools, and active interaction with related projects like the Seventh Framework Programme (FP7)-sponsored project OpenTox. The main activities in this respect were performed within Workpackage 5 dealing explicitly with gathering all information generated within the project and subsequent dissemination. The ultimate goal of optimal dissemination of project results was reached by means of the establishment of the public website, development of the database of molecules and their properties, development of on-line and standalone tools for development, upload and publishing of QSAR/QSPR models, the organisation of two CADASTER workshops, and the publication of the CADASTER newsletters.

A brief overview of the major dissemination activities and their outcomes is given below:
Dissemination of information via the project website: A web site for the CADASTER project http://www.cadaster.eu was established. This site provided a dissemination of information about the activities of the project, publications, presentations at the conferences and meetings as well as contains all deliverables (i.e. restricted to all public deliverables that are open to all users, non-public deliverables are made available to the project partners via a user-restricted part of the website) of the project. The web site was continuously supported during 4 years of the project and served as the central dissemination point. The CADASTER publications, announcement of events, agenda and materials of workshops, project publications and deliverables were timely and continuously upload there by the HGMU team. The project web site was attracting more than 2000 visitors per month. It was also used to organize the "Environmental Toxicity Prediction Challenge" http://www.cadaster.eu/node/65, which was co-organized with the International Conference on Neural Networks (ICANN) http://www.kios.org.cy/ICANN09. The challenge attracted 518 submissions from 108 participants from more than 25 countries world-wide. The website also provided dissemination of newsletters and materials and tutorials of two CADASTER Workshops (in Maribor and Munich), which provides a comprehensive coverage of tools available at the web site.
Dissemination of information by publications and newsletters: The publication of materials of the project in peer-reviewed scientific journals was an essential part of the dissemination of the project results. Up till now 49 articles with project results have been published or submitted for publication in scientific literature (a couple of

articles are under preparation). The project results were presented as 64 posters and 65 oral presentations at 51 conferences, meetings and workshops. The project results were also summarized in newsletters, which were sent to more than 6,000 registered users. The project has been highlighted in national publications, including newsletters of RIVM, HMGU and of other partners.

Dissemination of information by on-line database: An on-line QSPR-THESAURUS database http://www.qspr-thesaurus.eu, which stores properties of chemical compounds and QSAR/QSPR models, was developed. It has been used by the CADASTER partners to upload, verify and curate experimental data and to upload models developed during the project. The database currently contains 5,5k data points collected for greater than 800 molecules from four analyzed classes. More than 95% of the data points in the database were collected by HMGU, PHI and UI groups, which contributed 2.9k, 1.6k and 0.9k datapoints respectively. All this information is publicly available to the external users. The external users can access, search, upload and download data available in the database for the four analyzed classes of molecules. The users can also perform calculations, using models contributed by the individual partners to new chemical structures and use predictions of these models for the risk assessment. Currently the database contains 30 models, the majority of which (greater than 85%) were contributed by the UI and HMGU groups. The user can access the developed tools either through the web-interface or with help of web services and standalone tools, which provides remote calculations on the QSPR-THESAURUS web site (developed by HMGU) or through http://toxpredict.org web site of the OpenTox project (contributed by IDEA). The users can also upload and publish their own models. The (Q)SAR Model Reporting Format (QMRF) format was integrated as a part of the developed tools. This feature allows the CADASTER partners and external users to report information about published models, which is required for the use of models in the REACH assessment according to “OECD principles for QSAR validation and application in regulation”.

The optimization of molecular structures at the website http://mopac.cadaster.eu for the CADASTER project is done by liberally user-contributed CPU time, who are thus aware about the project. Thus, it provided another channel to disseminate the information about the project activities. Since the start of the project more than 1600 users have contributed 11.9 years of CPU time to optimize molecules. The developed tools incorporate methods for applicability domain (AD) assessment, experimental deign and risk assessment, as summarized below.

The methods for the assessment of the applicability domain (AD) of models were proposed, developed and benchmarked for regression and classification methods as reported in WP3 and WP4 and were made publicly available through the QSPR Thesaurus http://www.qspr-thesaurus.eu site. All the models published on the QSPR-Thesaurus web site provide information about the AD and accuracy of estimation (i.e. uncertainty in QSAR prediction assessed by one of the approaches implemented in the web tool). The estimation of the AD was integrated by embedding OpenTox web services, thus integrating tools developed in our partner project.

A number of modern experimental design methods, including original developments, were implemented and evaluated, as reported in WP3. The major results were summarized in PhD thesis of Mr. Brandmaier (HMGU). The approaches providing the best results were made available to the external

users as part of tools provided at the QSPR-THESAURUS http://www.qspr-thesaurus.eu web site.

The approaches for QSAR-integrated hazard and risk assessment developed in WP4 have been implemented into the web tool http://www.qspr-thesaurus.eu/risk. It includes calculation of the fate assessment of chemical compounds using pre-defined emission scenarios implemented within the SimpleBox Excel sheet-based program contributed by RIVM scientists. The use of individual QSAR models in species sensitivity distributions (SSDs) is done using a methodology developed in WP4. Intermediate calculations are stored and can be easily downloaded for external assessment by the end-users. Issues related to the treatment of uncertainty, especially associated to QSAR predictions, in fate and effect assessment were reported by the PhD thesis by M Sarfraz Iqbal (LNU).

A detailed description of dissemination activities for each of the tasks identified within Workpackage 5 is given below:

Task 5.1 Development of a prototype of the http://www site (experimental database).
This task was to develop a prototype of the QSPR Thesaurus database, which would be suitable for downloading and storage of data required for the project. The QSPR Thesaurus was based on the OCHEM platform developed by HMGU group, which was adapted to the requirements of the CADASTER project. HMGU team also provided training and continuous support of other project members with respect to the use of the web site during the project. The database allowed the CADASTER users to submit, store and annotate molecular records collected from the literature. It stored data in original units, tracked users and modifications to the data that were performed by them, allowed to introduce new units, new properties. The database automatically checked for duplicates, allowed editing of single or several records simultaneously, performed batch upload of data as Excel and/or SDF files, and allowed exporting data as Excel files. A concept of TAGs was used to identify areas of interests and to restrict work of users only to CADASTER classes. Moreover, HMGU group also developed tools for calculation of MOPAC descriptors following structure optimization using AM1 method. HMGU and IDEA have also contributed a number of web services to extend CADASTER web-database by integrating AMBIT functionality for substructure search.

Task 5.2 Development and testing of a prototype of the http://www site (models).
The goal of this task was the development of an infrastructure and a prototype of the database of models that could be used by CADASTER participants to upload their previously developed models, publish them and make them available on-line to CADASTER and, later on, to external participants. The task was completely achieved. The QSPR Thesaurus also incorporated tools for linear model upload, which were validated for the upload of models by the CADASTER participants. It was found that reproducing of models requiring 3D structures is a challenging task. In order to have a possibility to provide sustainable platform for such models on the CADASTER web site, a pipe-line for optimization of molecules using MOPAC 7 program was developed and integrated as part of the web site. A graphical interface to submit and access optimized conformation of molecules was developed http://www.cadaster.eu/mopac. The optimization of molecules was implemented based on the BOINC software http://boinc.berkeley.edu (which allows usage of user-contributed

calculations during screen saving mode of computers) and the respective web site http://mopac.cadaster.eu to handle this development was developed and was supported during the project by the HMGU group.\

Task 5.3 Implementation of tools to estimate the Applicability Domain of Models and Experimental Design.
The main goal of this task was to implement tools to estimate the Applicability Domain of Models and Experimental Design, including both the implementation of new approaches for the AD estimation, developed in WP3, and the integration of available AMBIT tools to estimate the applicability domain (AD) of models. All these goals were achieved. Moreover, the (Q)SAR Model Review Format (QMRF) was extended to enable the display of validation results for published models. Users can add several predefined parameters (R2, RMSE, Q2Loo, Q2ext, etc.) for both internal and external validation of their models. The possibility to add any kind of graphical representation (graphs, tables, images) to support or explain the validation results as well as applicability domain of models was also implemented. Thus, the models could be published on the CADASTER web site together with their QRMF as required for their use by the regulators. Experimental design methods were developed and made available on the web site of the project. The methods integrated already established algorithms as well as new step-wise approaches, which take into consideration the target property. The idea of such approaches was first proposed as conclusions of the deliverable of this task. During the next two years of the projects, it was further developed and resulted in several publications (Brandmaier et al, 2012a/b, 2013) as well as in PhD thesis of Mr. Brandmaier (2013).

Task 5.4 Public QSPR-THESAURUS site
The goal of this task was to make the database and tools developed within the CADASTER project publicly available to web users. The deliverable of this task contained an overview of the main features of the database and the additional tools that were developed and made available to the external users. The upload of data to the database and development of models were further continued during the project. The functionality of QSPR Thesaurus was demonstrated to the participants of two workshops, which were organized by CADASTER project.

Task 5.5 Development of a stand-alone version of the tools.
The goal of this task was to provide tools to remote access models developed and published within the CADASTER project on the QSPR-THESAURUS web to the external users. While it was expected that the major access to the models and the results would be accessed through the web interface, the remote standalone tools could provide an easy way to integrate the CADASTER models into other software packages by the external users. Within this task we did a number of developments that were requested by CADASTER users, e.g. incorporation of Dragon 5.4 and 5.5 packages (in addition to Dragon 6.0); authorization to download descriptors to the users that have the respective license (however, all users can use the developed models for free on the web site); storage of predicted values for molecules with 3D descriptors, development of enhanced web interface and web SOAP services to allow remote calls to the models developed by CADASTER project participants. A standalone version of tools to remotely access CADASTER web models based on Perl and Java code were developed. The models published on the web site of the project were integrated and made available through the OpenTox thus validating our development. We have also presented CADASTER models to OECD QSAR Toolbox meeting (Dr Tetko and Dr. Rorije in October 2011 and Dr. Jeliazkova and Dr. Rorije in

Spring 2012) but no decision about implementation of the models within the OECD QSAR tool was taken by them. Indeed, since CADASTER models are publicly and freely available on-line, the users can easily access them, predict new molecules and use predicted values for the purpose of their analysis with OECD QSAR Toolbox.

Task 5.6 Workshop on the use of QSARs models in REACH
The main goal of this task was to organize a workshop
http://www.cadaster.eu/node/116 which would provide a dissemination platform about the project development and project results, and to assist the risk assessors and national chemicals authorities, particular in Eastern European countries, with the use of the QSAR tools for the environmental risks assessments in REACH. The workshop involved 35 participants, including invited speakers from JRC - Institute for Health and Consumer Protection (Italy), ECHA – Evaluation Unit (Finland), Douglas Connect (Switzerland) and University of North Carolina USA). It was organized by the Public Health Institute Maribor (PHI) in Maribor, Slovenia and took place from September 1st to September 2nd 2011. The program of the CADASTER workshop was a combination of plenary sessions to provide a common perspective to all of the attendees and of training lessons for on-line tools that were developed during the project and can be used to estimate REACH endpoints for chemical compounds. The workshop was closed with a panel discussion, which was important to understand different positions of chemical regulator, academy and industry.

Task 5.7 Final workshop and guidelines for model development
The second CADASTER workshop (see http://www.cadaster.eu/workshop online) was organized by the HMGU and took place in its premises from October 7th to 9th 2012. The main goal of this workshop was to provide a tutorial to all interested partners, including industry and SMEs, on how to develop new models for the assessment of REACH-end points (in particular for new scaffolds of compounds for which there are no reliable QSAR models) and how to use the software developed by the project participants. The workshop was attended by 52 participants, including invited speakers from JRC - Institute for Health And Consumer Protection (Italy), EPA – Environmental Protection Agency (USA), Umwelt Bundesamt (Germany) and coordinators of the Seventh Framework Programme (FP7) funded projects OSIRIS (Prof. G. Schüürmann) and COSMOS (Prof. M. Cronin). There were two main thematic areas of the workshop: 1) Data collection and QSAR Model development for REACH (October 8th) and 2) Case studies and use of QSARs in the risk assessment (October 9th). The lectures of project participants were followed by lecture of invited speakers. A poster session with 20 posters was organised during the second day of the workshop. The materials of the workshop that summarize expertise of all participants, main results and lessons are under publication in a special open-access issue of ATLA journal. The decision to publish proceedings in an open-access issue was taken by the General Assembly of the CADASTER project (February 2012) to optimize the dissemination of the project results.

ADVISORY BOARD
In the design of the CADASTETR project it was judged essential that an Advisory Board consisting of representatives of the major stakeholders in REACH would supervise progress of the project and optimal dissemination of the project results. The Advisory board was chaired by a senior scientist with long-term experience in industry in chemical risk assessment, and contained representatives from SMEs, regulatory agencies, JRC, and research institutes. As the members of the Advisory Board had

ample experience in risk assessment, and exploitation and dissemination of risk assessment related activities, the input of the Advisory Board was continuously used to steer the activities within the various workpackages and to warrant optimal output of the project. To illustrate the opinions of the Advisory Board on the progress of the project and the project results, a brief overview of the input received is given below, arranged per individual Workpackage:

ADVISORY PANEL INPUT TO CADASTER

The Advisory Panel, chaired by Mike Comber, comprised of Dr Barry Hardy, Professor Gerrit Schüürmann, Dr Andrew Worth, Dr Theo Traas, Dr Chris Watts and Dr Ian Doyle. Although the involvement of the Advisory Panel was limited due to various reasons, the worst of which was REACH, the following was input received during the project.

Work Package 2: Collection of Data and Models

The Advisory Panel were asked for their opinion on the approach used as part of WP2 for selecting the chemicals to develop the QSARs and whether they had any comments on processes used for selecting chemicals in such testing programs. The Advisory Panel felt that the selection of substances for study is always a difficult task as there are many thousands of substances that could be chosen, and only limited resources for the research project are available. A balance has to be struck between regulatory interest, academic interest, availability of data, use and occurrence [in the environment], available resources and many other factors. This makes the choice of substances more of a subjective than an objective decision. The Advisory Panel agreed that the substances chosen for the Cadaster project covered a wide range of structures of current interest about which regulators are increasingly requesting data, with respect to the occurrence and effects of these groups of substances in the environment.

The Advisory Panel members were asked to comment on the scientific work and the utility of the work in WP2 for Industry and Government. The biggest driver for regulatory assessment of chemicals in Europe is the REACH regulation and this has been exercising the minds of Industry and Governments for several years now. One of the major tenets of REACH was the desire to limit the use of additional testing of chemicals on vertebrate species and to do this by exploring alternatives. The use of QSARs represents one of the main alternative approaches for generating data on the effects of chemicals on particular target organisms but this use has been limited by the number and coverage of rigorously tested QSARs currently available. Any research program that seeks to address this dearth of information is to be welcomed. However, the extent of the additional rigorously assessed QSARs produced is relatively modest [though commensurate with the resources available] and their availability comes at a time when the REACH process is already almost half completed. Consequently the outputs will only be primarily of use for the third tier of the REACH program.

Scientifically, the work has been of a very high order and reflects the intellectual quality of the scientists who have been involved in the project.

WP 3: Development and validation of QSAR models

One of the problems with developing and validating QSARs for specific groups of chemicals is variability in the experimental data and only

having a small data set. The Advisory Panel commented that small datasets and variable test data quality have been a problem for development of QSARs since this area of research started and there hasn't been significant improvement in availability of quality datasets over that period. The release of test data [produced for regulatory purposes] to the wider scientific community would have been one major way to address this problem. Of course, regulatory testing has to be paid for [by the company registering the chemical] and provides information that may be commercially sensitive and there is an understandable reluctance to make it available in the public domain. Hopefully, the approach adopted by ECHA in relation to test data supplied for the REACH registration will enable more test data to be made widely available. The only alternative is for research programs to be funded to generate test data, but that would mean that more vertebrate testing would be needed and that is not a good option.

The Advisory Panel was impressed by the amount of information available in the Cadaster QSAR database for the four compound types and it was fairly straightforward to access. Enhancing acceptance will always be a difficult process, but incorporation of the Cadaster information into the [widely used and accepted] QSAR Toolbox will help to achieve their acceptance. Users require a format whereby all they have to do is to enter the structure of their substance and the software picks the most appropriate QSARs for the endpoints of interest.

The Advisory Panel thought that the quality of the scientific work was very high and has been confirmed by the number of peer reviewed publications that have resulted from the Cadaster project. The Industry representative from the Advisory panel felt that the work has produced a number of QSARs that will prove useful for some substances that are still to be registered under REACH. There is still a lot of interest from regulators in all four of the compound types selected by Cadaster, so the Advisory Panel would expect that the outputs produced will also be of interest to Governments.

WP 4: Integration of QSARs within hazard and risk assessment

With respect to uncertainty the problem considered by the Advisory Panel was that Regulators need a clear outcome from a risk assessment that either says that a chemical is safe [for the environment, workers and consumers] for use in the amounts and specific purposes assessed, or that it is not safe and those areas where use needs to be limited or banned to provide safe use are indicated. Unfortunately, providing such 'black and white' outcomes from a risk assessment is limited by the quality [uncertainty] of the input data and the uncertainty of the risk assessment model. When reporting uncertainty, breaking it down into uncertainty of input data and uncertainty of model prediction is useful, as this allows the user to determine whether more effort put into improving input data quality will provide a sufficiently level of uncertainty for the overall risk assessment. One tool that has been found useful is a probabilistic approach to assessing the bounds of the risk as this enables a quantitative assessment. The Advisory Panel was pleased to see that this has been adopted in the Cadaster project and that the Cadaster project has also been developing probabilistic approaches and has commented on the need to link predictive toxicology models to fate and behaviour models.

Part of CADASTER requires the development of a Decision Support System (for risk assessment). When asked about the key elements that would be

in such a system, the Advisory Panel observed that risk assessment is a complex process that usually relies on a large amount of data and a large number of [usually mathematical] models of the world. As noticed previously, there are uncertainties associated with most parts of the risk assessment process and hence the output is seldom a clear 'it's safe' or 'it's not safe'. The risk assessor and risk manager need to know [and understand and be comfortable with] all of these uncertainties in order to be able to take appropriate risk management measures where they are needed. Hence the 'decision support network' comprises uncertainty information that comes with an associated level of confidence. The outputs seen, particularly from the presentations at CADASTER workshops, have tried to focus on the usability of the models and approaches developed. However, some of the tools developed do require a certain level of knowledge of modelling techniques for appropriate use. While some users would find this to be a drawback, the Advisory Panel felt it is important that users have at least a basic understanding of how the [predictive] models that they are using, actually work.

When asked whether a comparison between the risks in use of a chemical derived from Probability Risk Assessment based on testing data versus non-testing data be useful and enhance the acceptability of QSAR based approaches, they responded that such a comparison wouldn't hinder the acceptance of QSAR derived data and may actually enhance its acceptance. The danger is that the QSAR derived data produces a different risk assessment outcome [probably OK if it's more 'protective'] and that there is no overlap between the two outcomes when the uncertainties associated with each are used to produce confidence limits for the risk assessment.

The Advisory Panel felt that WP4 was the key to gaining greater acceptance of QSAR based data in risk assessments. Making QSARs easier to use in risk assessment and making more information available on the uncertainties of their use, will make Regulators more comfortable with their use. Cadaster has gone a long way towards this goal in the development of the approaches used and particularly with the QSPR database.

WP 5: Outreach via development of website, newsletters/workshop(s) and standalone tools for dissemination of project results

The Advisory Panel requested that all of the CADASTER reports from meetings be made readily available on the website together with all progress reports. Copies of all scientific publications resulting from the Cadaster project should be available in full on the website. The QSPR thesaurus was felt to be useful.

The availability, ease of use and transparency of the models developed will be the key to their widespread application, so the CADASTER website needs to be user friendly and structured in a way that makes its use logical. The incorporation of the QSAR models into the QSAR Toolbox is a must and doesn't yet appear to have been done. There could usefully be more information on the Cadaster website on how to use the models and what the outputs mean and this would encourage use by less experienced potential users in Industry and government.

List of Websites:
http://www.cadaster.eu