INFORMATION SOCIETY TECHNOLOGIES
(IST)
PROGRAMME

OpenMolGRID

DESCRIPTION OF ECOTOX

Contract Reference:	**IST-2001-37238**
Document identifier:	**OpenMolGRID-1-D1.4a-0109-2-1-DescEcotox**
Date:	**07/01/2004**
Work package:	**WP1: Grid Data Warehousing of Molecular Structure – Property (Activity) Information**
Partner:	**UU, Negri**
Lead Partner:	**UU**
Document status:	**APPROVED**
Classification:	**PUBLIC**
Deliverable identifier:	**D1.4a**

Abstract: This document is to describe the Ecotox data source that will be integrated into the OpenMolGRID data warehouse (MOLDW) and to indicate what information present is useful.

Delivery Slip

	Name	Partner	Date
From	Damian McCourt Jean Jing Werner Dubitzky	UU	15/09/03
Verified by	WPM	All	01/10/03
Approved by	G.H.F.Diercksen (TC) R.Ferenczi (QE)	OMC CGX	16/10/03 15/10/03

Document Log

Issue	Date	Comment	Author
0-0	02/06/03	First Version	Damian McCourt
0-1	05/09/03	New Template	Damian McCourt
1-0	10/09/03	Submitted to WPM	Damian McCourt
2-0	15/09/03	Submitted for Authorisation	Damian McCourt Jean Jing Werner Dubitzky
2-0	16/10/03	Document authorised	Damian McCourt
2-1	07/01/04	Updated due to the change of the document template (version1.3)	Jean Jing

Document Change Record

Issue	Item	Reason for Change
0-1	Document Format change	New document template available
2-0	Document Status	Document Authorised
2-1	The format of the head file is changed	The standard template of the document is changed

Files

Files in this section relate to actual storage locations on the BSCW server located at https://hermes.chem.ut.ee/bscw/bscw.cgi. The URL below describes the location on BSCW from the root OpenMolGRID directory

Software Products	User files / URL
Word 2000/XP	OpenMolGRID/Workpackage 1/Deliverables/ OpenMolGRID-1-D1.4a-0109-2-1-DescEcotox

Project information

Project acronym:	OpenMolGRID
Project full title:	Open Computing GRID for Molecular Science and Engineering
Proposal/Contract no.:	IST-2001-37238
European Commission:	
Project Officer:	Annalisa BOGLIOLO
Address:	European Commission - DG Information Society F2 - Grids for Complex Problem Solving B-1049 Brussels Belgium
Office	BU31 4/79
Phone:	+32 2 295 8131
Fax:	+32 2 299 1749
E-mail	annalisa.bogliolo@cec.eu.int
Project Coordinator:	Mathilde ROMBERG
Address:	Forschungszentrum Jülich GmbH ZAM D-52425 Jülich Germany
Phone:	+49 2461 61 3703
Fax:	+49 2461 61 6656
E-mail	m.romberg@fz-juelich.de

Contents

1. Introduction

1.1. Purpose and Scope

The main purpose of this document is to describe the Ecotox data source that will be integrated into the OpenMolGRID data warehouse (MOLDW) and to indicate what information present is useful. It is to be used as a reference document for other documents that are written.

1.2. Document Overview

The main aim of the MOLDW is to integrate data relevant to molecular engineering from disparate repositories. These repositories are held in different systems, at different locations and in different formats. Integration of each source into the MOLDW will be specific to that source, meaning that the adaptation of formats and access procedures will be required. Each source must therefore be described in detail. This document aims to describe the Ecotox data source.

1.3. Document Structure

In addition to this section the document contains the following sections:

- Section 2 – a description of Ecotox.
- Appendix A-B – Appendices used to support the document.

2. ECOTOX

Ecotoxicology or ECOTOX is an online resource that provides toxicity data for aquatic and terrestrial life. The US EPA, Office of Research and Development (ORD), and the Mid-Continent Ecology Division of the National Health and Environmental Effects Research Laboratory (NHEERL) maintain the database. The resource is available at http://www.epa.gov/ecotox.

Data within Ecotox is available in three main ways:

1) Simple query
2) Advanced query
3) Download data archive (self-extracting compressed zip format)

ECOTOX provides a querying mechanism that is realised in two ways; a quick or basic query and an advanced query. In both instances the user is asked to enter information using forms. In the advanced query section more information can be entered.

Typical information that can be entered in the simple query relate to chemical, species and effect. The chemical can either be the chemical name (exact or partial match) or the CAS number associated with the chemical. The species can be the scientific or common name (exact or partial match) or the species number. The effect is selected by checking a number of different checkboxes. The user can also select information on kingdom and habitat type in order to refine the search. There is also the ability to eliminate results returned based on year of publication. The user also has the ability to select the output format of the information. This can be displayed in a browser window, or can be downloaded in pipe-delimited format that can be easily integrated into spreadsheet applications.

In the advanced query, the user is given the ability to eliminate more data based on their requirements. The input form is divided into six separate categories. In addition to the information that can be entered in the simple query (with more options), the user is also able to choose which dataset the information comes from and from recent modifications only. The user also has more control of what kind of test conditions were present when the toxicity measurements were made. The user also has more output options. They have the ability to eliminate columns that they are not interested in from the output.

The website offers the user some help when querying the database. The help section describes everything the database contains. It offers the user the ability to browse/search based on chemical, species or effect. These facilities only provide a means of locating the exact chemical, species or effect name and do not actually return any data. They are in effect a lookup facility that can be used when actually using the query facility.

The database can also be downloaded in compressed file format (self extracting "zip" archive). It is broken down into two separate downloads – aquatic and terrestrial.

2.1. Aquatic Archive

Details associated with the aquatic archive are as follows:

Update frequency	Last Update	URL
Quarterly	20-12-2002	ftp://ftp.epa.gov/pub/ecotox/aquire_ascii_XX-XX-XXXX.exe

In this table the XX-XX-XXXX relates to the last date on which an archive was made. A listing of the contents of this archive is shown in Appendix A. Within this archive there are three main folders. These are:

1) acquire/
2) acquire/validation

3) acquire/common_validation

The operational data is present in the root `acquire` directory. Within this directory the file `acquire_structure.pdf` contains the information necessary to reconstruct the aquatic archive in a relational database. The `validation` directory contains information to validate data that is specific to the aquatic archive, whereas the `common_validation` directory contains information for validation that is common to the terrestrial archive also.

Each data file in the aquatic archive has the extension `.txt`. Accompanying most `.txt` files is a `.htm` file. This file contains metadata about what the associated `.txt` file contains.

2.2. Terrestrial Archive

Details associated with the terrestrial archive are as follows:

Update frequency	Last Update	URL
Quarterly	20-12-2002	ftp://ftp.epa.gov/pub/ecotox/terretox_ascii_XX-XX-XXXX.exe

In this table the XX-XX-XXXX relates to the last date on which an archive was made. A listing of the contents of this archive is shown in Appendix B. Within this archive there are three main folders. These are:

1) terretox/
2) terretox /validation
3) terretox /common_validation

The operational data is present in the root `terretox` directory. Within this directory the file `terretox_structure.pdf` contains the information necessary to reconstruct the terrestrial archive in a relational database. The `validation` directory contains information to validate data that is specific to the terrestrial archive, whereas the `common_validation` directory contains information that is common to the aquatic download also.

Each data file in the terrestrial archive has the extension `.txt`. Accompanying most `.txt` files is a `.htm` file. This file contains metadata about what the associated `.txt` file contains.

2.3. General Notes

A typical user query is of the simple kind. It involves the input of either a CAS number or chemical name, a species name or number and details of the effects the user is interested in. These details are therefore most important from the user point of view.

For integration into MOLDW the query access mechanisms are unrealistic. The third data access method however is conceivable. The data from Ecotox is available in two main archives that are freely downloadable via FTP. Their format is generally the same.

Appendix A – Acquire Archive Contents

```
acquire/aquire.htm
acquire/aquire.txt
acquire/aquire_structure.pdf
acquire/chemicalinfo.htm
acquire/chemicalinfo.txt
acquire/fieldstudy.htm
acquire/fieldstudy.txt
acquire/remarks.htm
acquire/remarks.txt
acquire/remarksdescriptions.htm
acquire/remarksdescriptions.txt
acquire/watercharacteristics.htm
acquire/watercharacteristics.txt
acquire/waterchardescription.htm
acquire/waterchardescription.txt

acquire/common_validation/chemicals.htm
acquire/common_validation/chemicals.txt
acquire/common_validation/chemical_analysis_methods.txt
acquire/common_validation/chemical_formulations.txt
acquire/common_validation/chemical_grades.txt
acquire/common_validation/concentration_units.txt
acquire/common_validation/control_types.txt
acquire/common_validation/effects.txt
acquire/common_validation/endpoints.txt
acquire/common_validation/ion_codes.txt
acquire/common_validation/measurements.txt
acquire/common_validation/radio_labels.txt
acquire/common_validation/references.htm
acquire/common_validation/references.txt
acquire/common_validation/response_sites.txt
acquire/common_validation/species_common_name.htm
acquire/common_validation/species_common_name.txt
acquire/common_validation/species_data_file.htm
acquire/common_validation/species_data_file.txt
acquire/common_validation/species_latin_names.htm
acquire/common_validation/species_latin_names.txt

acquire/validation/applicationtypes.txt
acquire/validation/concentrationtypecodes.txt
acquire/validation/depthunits.txt
acquire/validation/durationmodifiercodes.txt
acquire/validation/durationunitscodes.txt
acquire/validation/effecttrendcodes.txt
acquire/validation/exposuretypecodes.txt
acquire/validation/geocodes.txt
acquire/validation/habitatcodes.txt
acquire/validation/organiccarbontypecodes.txt
acquire/validation/seasoncodes.txt
acquire/validation/statisticalsignificancecodes.txt
acquire/validation/studytypes.txt
acquire/validation/substratecodes.txt
acquire/validation/testlocationcodes.txt
acquire/validation/watercharacteristicunits.txt
acquire/validation/watertypecodes.txt
```

Appendix B – Terretox Archive Contents

```
Adding terretox/chemical_carriers.htm
Adding terretox/chemical_carriers.txt
Adding terretox/exposure.htm
Adding terretox/exposure.txt
Adding terretox/result.htm
Adding terretox/result.txt
Adding terretox/terretox_structure.pdf
Adding terretox/test.htm
Adding terretox/test.txt

Adding terretox/common_validation/chemicals.htm
Adding terretox/common_validation/chemicals.txt
Adding terretox/common_validation/chemical_analysis_methods.txt
Adding terretox/common_validation/chemical_formulations.txt
Adding terretox/common_validation/chemical_grades.txt
Adding terretox/common_validation/concentration_units.txt
Adding terretox/common_validation/control_types.txt
Adding terretox/common_validation/effects.txt
Adding terretox/common_validation/endpoints.txt
Adding terretox/common_validation/ion_codes.txt
Adding terretox/common_validation/measurements.txt
Adding terretox/common_validation/radio_labels.txt
Adding terretox/common_validation/references.htm
Adding terretox/common_validation/references.txt
Adding terretox/common_validation/response_sites.txt
Adding terretox/common_validation/species_common_name.htm
Adding terretox/common_validation/species_common_name.txt
Adding terretox/common_validation/species_data_file.htm
Adding terretox/common_validation/species_data_file.txt
Adding terretox/common_validation/species_latin_names.htm
Adding terretox/common_validation/species_latin_names.txt

Adding terretox/validation/application_frequency.txt
Adding terretox/validation/cec_units.txt
Adding terretox/validation/duration_units.txt
Adding terretox/validation/exposure_types.txt
Adding terretox/validation/gender.txt
Adding terretox/validation/lifestages.txt
Adding terretox/validation/major_exposure_types.txt
Adding terretox/validation/media_types.txt
Adding terretox/validation/organic_matter_types.txt
Adding terretox/validation/organic_matter_units.txt
Adding terretox/validation/organism_sources.txt
Adding terretox/validation/sample_units.txt
Adding terretox/validation/test_locations.txt
```