INFORMATION SOCIETY TECHNOLOGIES
(IST)
PROGRAMME

OpenMolGRID

ECOTOX-AQUIRE DATA SPECIFICATION

Contract Reference:	**IST-2001-37238**
Document identifier:	**OpenMolGRID-1-D1.1c-0103-2-1-AquireDataSpec**
Date:	**07/01/2004**
Work package:	**WP1: Grid Data Warehousing of Molecular Structure – Property (Activity) Information**
Partner:	**UU, Negri**
Lead Partner:	**UU**
Document status:	**DRAFT**
Classification:	**PUBLIC**
Deliverable identifier:	**D1.1c**

Abstract: This document is to specify what transformations must take place to data before it is placed into the OpenMolGRID data warehouse MOLDW.

Delivery Slip

	Name	Partner	Date
From	Damian McCourt Jean Jing Werner Dubitzky	UU	15/09/03
Verified by	WPM	All	17/11/02
Approved by	G.H.F.Diercksen (TC) R.Ferenczi (QE)	OMC CGX	11/11/03 10/11/03

Document Log

Issue	Date	Comment	Author
0-0	29/08/03	First Version	Damian McCourt Jean Jing Werner Dubitzky
1-0	09/08/03	First Stable Version	Damian McCourt Jean Jing Werner Dubitzky
2-0	15/09/03	Submitted for Authorisation	Damian McCourt Jean Jing Werner Dubitzky
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
1-0	Grammatical Changes	Comments made by individual Partners
2-0	Version status	Submitted for approval
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-D1.1c-0103-2-1-AquireDataSpec

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 5
1.1. PURPOSE AND SCOPE 5
1.2. OVERVIEW 5
1.3. DOCUMENT STRUCTURE 5
2. IDENTIFICATION OF CHEMICALS 7
3. TOXICITY 8
4. CARCINOGENICITY 12
5. PHYSICO-CHEMICAL PROPERTIES(PCP) 13
6. ADME RELATED PROPERTIES AND DESCRIPTORS 14
7. REFERENCES 15

1. Introduction

1.1. Purpose and Scope

The purpose of this document is to specify what transformations must take place to data before it is placed into the OpenMolGRID data warehouse MOLDW. This document is intended to accompany deliverable D1.1a [1]. The information to be integrated into MOLDW is based on the information presented in deliverable D1.3 [2]

1.2. Overview

Currently specifications for data transformations have not yet taken place. This document intends to initiate this process by providing a mapping of the data required as stated in deliverable D1.3 [2] to what is available in the Acquire data source available from Exotox. A description of Acquire is provided in document D1.4a [3]. Where possible data transformations have been stated, but other areas require further specification by other partners in the consortium.

1.3. Document Structure

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

- Section 2 – identification of chemicals
- Section 3 – Toxicity
- Section 4 – Carcinogenicity
- Section 5 – Physico-Chemical Properties(PCP)
- Section 6 – ADME Related Properties and Descriptors

2. Identification of Chemicals

Name	Type	Description	Present In Source	Format	Conversion Details
CAS (unique identifier)	Integer	The Chemical Abstract Service (CAS) number. This is an integer without hyphens or spaces	Present as 'Test CAS' in acquire.txt	As number without hyphens	No change
Molecular Weight	Float	The Atomic Mass Unit (amu) associated with the CAS.	Not Present		
Chemical Formula	String	The chemical formula associated with the CAS	Not present		
Chemical Names	List of Strings	The names by which the chemical is known	Referenced from chemicals.txt in common_validation directory	Only as one name	Store as found

3. Toxicity

Name	Type	Description	Present In Source	Format	Conversion Details
CAS (unique identifier)	Integer	The Chemical Abstract Service (CAS) number. This is an integer without hyphens or spaces	Present as 'Test CAS' in acquire.txt	As number without hyphens	No change
Target Species Names	List of Strings	A list of names associated with the target species	Present in acquire.txt as 'Species Number'	reference to species_common_names.txt and species_latin_names.txt files	Merge target species common names and Latin name from species_common_names.txt and species_latin_names.txt files
End Point Type	String	The type of toxicity measure used in this protocol e.g. LC_50, LD_50.	Present 'Endpoint' in acquire.txt	e.g. BAF/ EC10 EC50 LC50 LD50 LOEC NOEC T1/2 NR	No change Need spec of endpoints to be considered

Dose Metric	Float	The dose of chemical studied (for aquatic toxicity the compound is in the tank, and the fish exposed)measured in milligrams per kilogram (mg/kg) or milligrams per litre (mg/l).	Present as 'Concentration Mean' in acquire.txt	String e.g. 125 NR	Float
Dose Metric Units	String	The units associated with the Metric dose	Present as 'Concentration Units' in exposure.txt	ng/g ug/l ul/L g/Ha mg/kg	Unit standardisation
Dose Mol	Float	The dose of chemical studied measured in millimols per kilogram (mmol/kg) or millimols per litre (mmol/l).	As before		
Dose Mol Units	String	The units associated with the Mol dose	As before	mM uM mmol/m3	Unit standardisation
Exposure Time	Float	The amount of time the target species was exposed to the chemical. This is measured in hours.	Available as 'test duration' in acquire.txt	String e.g. 6 NR	Change to Float in accordance with unit standardisation
Protocol Details	Text	General information associated with the protocol. There is no particular format associated with this.			

Mode of Action USA	String	The Mode of Action of the chemical. These are classes based on the EPA (Duluth, USA) standard. These are as follows: • Non Polar Narcosis (Base Line Narcosis) • Polar Narcosis (Narcosis II) • Narcosis III (Ester/ Acrylate compounds) • Oxidative Phosphorilation uncoupling • Respiratory inhibition • Electrophile and proelectrophile reactivity • Acetilcholinesterase inhibition • Central nervous system seizure responses	Present as 'Effect' as acquire.txt, not USA or EU	e.g. MOR REP The full list presents in effect.txt	
Mode of Action EU	String	The mechanism of action of the chemical. These are classes based on the EU (Netherlands) standard. These classes are as follows: • Non Polar Narcosis • Polar Narcosis • Reactive • Receptor Mediated	As for USA		
Author(s)	List of Strings	The author who reported the toxicity measure.	Present as 'Reference Number' in acquiret.txt, reference to 'Author' in reference.txt	e.g. Veldhuizen-Tsoerkan, M.B., D.A. Holwerda, and D.I. Zandee	Multiple authors

Year	Integer	The year in which the author reported the toxicity measure.	Present as 'Reference Number' in acquire.txt, reference to 'PublicationYear' in reference.txt	String	Change to Integer
Database	String	The database from which the toxicity measure was obtained.		String	ACQUIRE(ECOTOX)

4. Carcinogenicity

Name	Type	Description	Present In Source	Format	Conversion Details
CAS (unique identifier)	Integer	The Chemical Abstract Service (CAS) number. This is an integer without hyphens or spaces	Present as 'Test CAS' in acquire.txt	As number without hyphens	No change
Carcinogenicity	String	Classification of carcinogenicity according to classes proposed by the World Health Organisation (International Agency for Research on Cancer- IARC). These classes are as follows: 1 – Carcinogenic to human 2a – Probably Carcinogenic to human 2b – Possibly Carcinogenic to human 3 – Unknown 4 – Non-carcinogenic to human			

5. Physico-Chemical Properties(PCP)

Not present in data Source

6. ADME Related Properties and Descriptors

Not present in data source

7. References

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.

[1] D. McCourt, "Data Warehouse software Specification," /OpenMolGRID/Workpackage 1/ Deliverables/OpenMolGRID-1-D1.1a-0101-2-0-MOLDW, 15/09/03.

[2] E. Benfenati and A. Papp, "Properties and priorities of the data for pharmaceutical and phytopharmaceutical compounds," /OpenMolGRID/Workpackage 1/ Deliverables/OpenMolGRID-12-D1.3a-0108-1-0, 15/09/03.

[3] D. McCourt, J. Jing and W. Dubitzky, "Description of Ecotox," /OpenMolGRID/Workpackage 1/ Deliverables/OpenMolGRID-1-D1.4a-0109-1-0, 15/09/03.