Development of an Improved Pesticide Properties Database for Risk Assessment Applications

K. A. Lewis, A. Green and J. Tzilivakis

University of Hertfordshire, Hatfield, UK, k.a.lewis@herts.ac.uk

Abstract

There is a need for a single authoritative comprehensive database of pesticide properties that is available as both an online resource and in a portable electronic format suitable for embedding in or interrogating from other software applications such as risk assessment systems. This paper introduces the Pesticide Properties Database that holds physicochemical, environmental fate, ecotoxiciological and human health data for, currently, over 700 pesticides and 350 associated metabolites. The results from a comparison exercise between this database and other commonly used resources are reported and information on current usage statistics is also provided.

Key words: Pesticides, Risk assessment, Database

1 Introduction

The demand for detailed data on the physicochemical and toxicological parameters of pesticides has grown considerably over the last couple of decades. This is, in part, due to more stringent regulatory controls and the use by regulators, policy makers and others of data-hungry modelling and risk assessment approaches.

Although it is difficult to be precise there may be as many as 1000 pesticides available globally. Many have been withdrawn from the European market by the ongoing review process (Directive 91/414) but it is not unusual for an individual European State to have around 300 pesticides registered for use at any given time. For example, approximately 330 are currently registered in the UK, 380 in France and 250 in Germany. Depending on the type and purpose of the application being used a wide range of parameters are required. It would not be unreasonable to suggest that 6-10 parameters are required to calculate predicted environmental concentrations. In addition, ecotoxicological threshold values are needed for mammals, birds, earthworms, honey bees and aquatic species (e.g. fish, invertebrates, lower and higher aquatic plants). Therefore one might need up to 18 parameters for each pesticide studied, 6000 for all registered pesticides within one European State and potentially 18000 to study the full range.

It would appear, at first glance, that this data is readily available via a diversity of sources including government departments, manufacturers, universities and environmental bodies. One may also assume that the Internet has made access to this data simple, fast and unproblematic. In fact both assumptions are incorrect. Most of the data sets currently available are extremely limited regarding the range of pesticides they cover. As may be expected, national data sets are usually constrained to those pesticides registered for use in that particular country. Table 1 summarises the main resources (as of late 2006). From the table it can be seen that most data sets have around 200-400 records. However, there is a great deal of overlap in the active substances covered and most suffer from large gaps in information and presentation inconsistencies. Consequently, other sources of information, of which there are many, are essential.

Some data sets are comprehensive regarding the types of data they contain (e.g. environmental fate, human health). However, for the majority of pesticides, different data sets must be interrogated for different data types. For example, IPCS INCHEM might be used for basic chemical information such as molecular mass and structure, the ECOTOX database might be searched for ecotoxicological information and medical databases interrogated for information on human health. There are also commercial databases that contain comprehensive data sets but these require annual subscriptions and are rarely available for direct linking to software applications without payment of significant licence fees.

Table 1: Summary of the main pesticide data resources (2006 data)

Resource name	No. records	Data type[#]	Comments
91/414 EU / EFSA Evaluation Dossiers	140 / 110	A, B, C, D, E	• ec.europa.eu/food/plant/protection/evaluation/index_en.htm • www3.efsa.europa.eu/DAR/displaySubstance.cfm
ACP Evaluation Documents UK	170	A, B, C, D, E	• www.pesticides.gov.uk/publications.asp?id=202 • Some documents are quite old & have limited data
AGRITOX, France	370	A, B, C, D*, E	• www.inra.fr/internet/Produits/agritox/php/fiches.php?PHPSESSID=ae0e236e3de4999ae0f50e9ef45eada7
UK Pesticide Manual 14th Edition	880	A, B, C, D*, E	• Commercial, priced publication • Variable useful content. Large & significant data gaps.
PAN Pesticides Database, USA	Large data set	A, B, C, D*, E	• www.pesticideinfo.org/Index.html • Not user friendly. Mainly portal links, large data gaps
EXTOXNET, USA	180	A, B, C, D, E, F	• extoxnet.orst.edu/ghindex.html • Poor presentation which can undermine data integrity
US EPA Pesticide Factsheets	77	A, B, C, D, E	• www.epa.gov/opprd001/factsheets • Mainly new actives
US EPA ECOTOX Database / Fate Database	Large data set / 189	E / A, B, C	• cfpub.epa.gov/ecotox/ • cfpub.epa.gov/pfate/Chem_details.cfm • Needs careful choice of data best fit for purpose
PMRA Pesticide Factsheets, Canada	70	A, B, C, D*, E	• www.pmra-arla.gc.ca/english/pubs/prdd-e.html • Limited data set
PIC Decision Guidance (DGD) Documents, FAO	350	A*, B*, C*, D, E*	• www.pic.int/en/Table7.htm • DGD produced for chemicals listed in Annex III, Rotterdam Convention on Hazardous Chemicals & subject to a PIC. • Mainly banned or severely controlled pesticides
Pandora, NL	225	A*, B, C, E	• Research report RIVM no. 679101014 (1994) • Quite old, report may not still be available
KingTai Chemicals Datasheets	60	A, B §	• www.kingtaichem.com • Very slow website but does contain some useful data
ChemIDPlus, USA	Unknown	A §	• Chem.sis.nlm.nih.gov/chemidplus • Need CAS number to search, Useful for basic information
Pesticide Data Tables, Danish EPA	400	A, B, C, E†	• www.mst.dk/udgiv/publications • Good data source but not referenced
European Chemicals Bureau	Unknown	D	• http://ecb.jrc.it • Only EU registered pesticides available – many gaps in data
IPCS INCHEM	Unknown	A	• www.inchem.org/

Definition of data types: A – general data, B – physicochemical, C – fate, D – human health, E – ecotoxicological
* Limited range of parameters held, § Very limited range of other data types held, † Aquatic ecotoxicolgical endpoints only

Presentation format is also very variable ranging from tabular to narration. The latter are common and these are useful as they place the data in context. However, researchers needing specific parameters must search each document and collate their own data sets, especially if they wish to make the data available to software applications. This process is very slow and invites typographical errors.

The Internet has dramatically altered the way information is distributed and shared and has become an important research tool. However, due to its loose, forever changing structure, free of content management it can be very difficult to locate information relevant to a specific topic. Unlike bibliographic databases it does not provide an index or contents page. Language barriers can also hinder the use of potentially valuable sites. Whilst many pesticide common names are similar from one language to the next, slight variations can cause search tools to fail. The Internet has no structural boundaries, no standard method of organisation and information available today may not be available tomorrow or necessarily stay at the same URL. There is also little peer review control and judging data credentials, that is fitness for purpose and accuracy, can be problematic. For example, most of the physicochemical and toxicological data needed has a natural variability often dependent on the conditions under which it is measured. In some instances the natural data range can be large and the value selected may have a significant influence on the risk assessment outcome. This is not to mention the unavoidable errors that creep into data sets such as typographical mistakes or problems associated with units of measurement. There is also the question of data maintenance - some resources were collated some time ago and are not

updated. An important issue is that the quality and utility of data should not be separated from its purpose and intent. Any simulation is only as good as the data that drives it.

Consequently, there is a serious need for a single authoritative source to capture, archive, validate, manage, maintain and provide access to data for the majority of pesticides available world-wide for pesticide risk assessments. The Pesticide Properties Database (PPDB) recently launched solves many of the problems discussed here.

2 Database purpose, design and development

The origins of the PPDB can be traced back to 1994 to the development of the award winning Environmental Management for Agriculture (EMA) software (Lewis and Bardon, 1998). This software package included a suite of decision support tools to help UK farmers improve their environmental performance and included a pesticide risk module with an embedded pesticide database (Lewis et al., 2003). The pesticides within the database were restricted to those active substances registered for use in the UK and included both physicochemical parameters and ecotoxicological data. The EMA software and the pesticide database has been constantly maintained and updated and is still available today. However, recent EU funding for a new approach to pesticide risk assessment (FOOTPRINT – Functional Tools for Pesticide Risk Assessment and Management. See www.eu-footprint.org/home.html) has provided the opportunity to re-assess the database in terms of the range of pesticides covered and the data stored. It has also provided the resources to reassess the existing data in terms of accuracy and 'fitness for purpose'.

The objectives of the new database is to provide:

- a single, comprehensive resource of reliable, consistently presented pesticide data having common syntax, units and semantics
- a portable format for direct linking to software applications such as risk assessment systems
- on-line access using a simple tabular format supported by layperson interpretations and user tools.

If the new database was to provide a solution to many of the problems experienced by users of pesticide data then it was essential that the data stored was suitable for a wide range of applications and different user types. To this end the data stored covers:

- **General information.** For example common and chemical names, language translations, chemical group, formula, structures, pesticide type, CAS/EC numbers and data related to country registration.
- **Physicochemical data.** Including solubility, vapour pressure, density, dissociation constants and information on degradation products.
- **Environmental fate data**. For example the octanol-water partition constant (Log P), Henry's law constant (Hd), degradation rates in soil, sediments and water (DT_{50}), the Freundlich coefficient and the organic-carbon sorption constant (K_{oc}).
- **Human health information.** This includes World Health Organisation toxicity classifications, Acceptable Daily Intakes (ADI), toxicity to mammals, other exposure limits and toxicity endpoints, plus the EC risk and safety classifications.
- **Ecotoxicology**. For example acute and chronic toxicity data for a range of fauna and flora plus information on bioaccumulation.

Two distinct user types have been identified: (i) those requiring large data sets for software applications such as risk assessments and (ii) users seeking specific data items or data on a specific chemical. MS Access 2000 Database format has been selected as the primary storage method. It is probable that this format will satisfy the requirements of users needing to embed data sets into software applications. The database is then streamed through various 'data filters' for formatting into HTML pages for online access and translation from English language into a range of EU languages. This process helps simplify the updating and maintenance process.

The online version of the database has various user tools available including a search mechanism that can identify a pesticide record by active substance, common name (in any of the EU languages offered), alternative names or the chemical registration number (CAS RN). There is also a full index page and direct electronic links from parent chemicals to their metabolite data pages. To enable laypersons to use the on-line resource standard interpretations of the data are offered as hazard classifications. In most cases the thresholds used are those used for regulatory purposes or are 'rules of thumb' in wide and general use

such as the guidelines used by the UK pesticide industry for developing Pesticide Environmental Information Sheets. Two risk indicators have also been calculated from the available data: (i) the GUS Index for groundwater leaching potential (Gustafson, 1989) and (ii) a measure of particle bound transport which indicates the pesticides risk of being transported with runoff (Goss and Wauchope, 1990).

3 Data compilation

Probably the best sources of information currently available for pesticide properties are the monographs produced as part of the EU review process and data within these documents has been first choice for populating the database. Where EU documents are not available, alternative sources are used including:

- Databases and documents from the EU and national government departments including the UK's PSD, Germany's Federal Environment Bureau, the Danish Environmental Protection Authority and the US EPA
- On-line databases e.g. ARIS, EXTOXNET, ARS/OSU, PAN, AGRITOX (see Table 1)
- Manufacturers safety datasheets and environmental fact sheets, on- and off-line
- Hardcopy, peer reviewed scientific publications and data derived from research projects.

As the primary purpose of the database is for risk assessment data has been selected for this particular use. The values quoted for physico-chemical properties is usually a mean of the various studies identified. Where data is particularly sensitive, to climate or soil for example, information on the data range has been added. Where data is just naturally very variable, we have attempted to select that most appropriate for EU conditions. For ecotoxicological data the 'worst case' value has been selected unless it appears wildly out of character with the majority of studies published. The data relates to specific species and endpoints, where possible, to ensure a harmonised and balanced data set.

The PPDB has been extensively cross-referenced against other data sets as an evaluation exercise. This has helped to identify erroneous and inconsistent data highlighting the need for further investigation and further consultation of the original data source (rather than the data set).

Accuracy of the data always challenges the accuracy of the model or risk assessment system. No matter how good the model is the results will only ever be as good as the input data. The impact of data quality can be very significant. Due to the wide range of different data sources used for the PPDB a 'code' has been tagged to each parameter that identifies the data source and a confidence score. Rules are invoked which assign scores depending on a number of variables, which reflect the perceived reliability of the data. These values are in the range 0 (no confidence) to 5 (reliable data) and are a function of the data source, publication date, referencing, the match to the desired parameter and fitness-for-purpose. For example, avian toxicity data taken from a recent EU dossier that exactly matches the endpoint required (e.g. acute oral LD_{50} mallard) would be assigned A5. 'A' indicates the source as an EU dossier and '5' indicates high confidence in the data quality. However, a soil DT_{50} extracted from an ad-hoc publication which gives no details of the original data source or the conditions under which the data was collected would be assigned Q1. 'Q' being the reference code for miscellaneous documents and '1' referring to the poor level of data confidence. It should be remembered, however, that the process of confidence scoring is somewhat subjective in nature and only meant as a guide. A low score does not, necessarily, indicate inaccurate or poor data. Full details of the quality and source codes can be found on the database web site.

4 Comparison with other resources

The PPDB currently holds approximately 700 pesticide data records and a further 350 records for associated metabolites. Table 2 provides information on gaps for the main parameters and provides a rough comparison with sub-sets of 100 records taken from two other data sets. Records for the subsets were randomly selected and only the presence of numerical data has been counted. Qualitative statements (e.g. high, low) have been ignored as they can not be used as data input with any accuracy.

Table 2: Data set gaps analysis

Parameter § Data not always relevant	**PPDB**	**AGRITOX**	**EXTOXNET**
Number records examined/total	700/700	100/370	100/180
Presentation style	Tabular	Tabular	Narration
General description & generic name	100%	100%	100%
CAS RN Number	100%	99%	99%
Chemical formula	100%	94%	0%
Molecular mass	100%	94%	100%
Structural picture	97%	51%	0%
Relative density / bulk density	78%	0%	2%
Pka§	33%	32%	0%
Aqueous solubility	97%	90%	94%
Solubility in organic solvents	55%	81%	6%
Partition Coeft Log P	95%	77%	58%
Vapour Pressure	94%	85%	84%
Henry's constant	87%	59%	0%
Soil DT50	86%	64%	85%
Sorption Coeft Koc	80%	44%	84%
Aqueous photolysis DT50	49%	37%	30%
Neutral hydrolysis DT50	66%	86%	34%
Water-sediment system DT50	35%	25%	0%
Bioconcentration data	45%	18%	58%
Ecotoxicity – mammals	99%	97%	97%
Ecotoxicity – birds	87%	82%	88%
Ecotoxicity – acute fish	93%	84%	92%
Ecotoxicity – acute aqueous invertebrates	89%	76%	10%
Ecotoxicity – honeybees	76%	74%	40%
Ecotoxicity – earthworms	56%	46%	9%
Ecotoxicity – algae	75%	61%	0%
Ecotoxicity – higher aquatic plants	31%	14%	0%
Toxicity – oral mammals	99%	97%	97%
Toxicity – WHO classification	100%	0%	0%
Toxicity – EC Risk & safety classifications§	58%	0%	0%
Average % across record set	77.9%	60.6%	43.6
Index value [(Average x no records)/100]	545.3	224.2	78.6

5 Database access and current usage

The MS ACCESS database on CD is available to all subject to an administration fee and, if it is to be used with software that will be distributed for third party or commercial use, a modest licence fee. The fees include database updates for 12 months.

A prototype on-line version of the database was launched on a free-to-all basis during 2006 via the FOOTPRINT website (www.eu-footprint.org). In early 2007 the format and content were finalised and the database was made available more widely via ADLib (www.adlib.ac.uk) and other websites (www.herts.ac.uk/aeru/footprint). Shortly after the launch, processes were put in place to monitor activity on the site and provide user statistics. The results of which have highlighted the fact that the database is being used extensively. Fig 1 shows user statistics for the first 14 days of the official database launch.

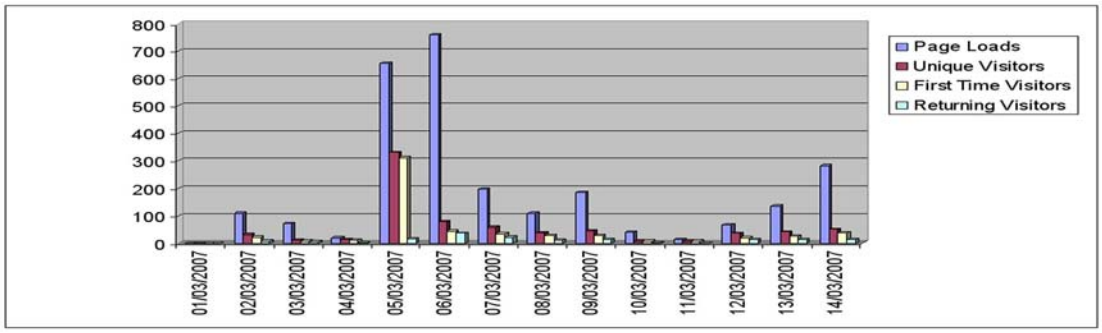

Fig.1 Database use statistics

Fig. 2a Last 100 hits to 08.00 BST

Fig. 2b Last 100 hits to 17.00 BST

The peak shown shortly after the launch coincides with the start of the working week and, as the statistical counters were all set to zero, shows the initial level of interest. Since this time, usage is quite steady at around 200 hits each working day. Initially, the majority of users were finding the database using Internet browsers and not via direct website links. However, current statistics indicate the site is now being bookmarked and users are returning on a frequent basis. Another interesting issue is the wide range of countries being represented by users. Snap shots taken in the first month of the launch are shown in Fig 2a and 2b. Data shows that, as would be expected, whilst most users are clustered in Europe there is also activity on the website from all over the world.

6 Conclusions

The amount of pesticide data required for risk assessment exercises is quite considerable and despite the Internet and greater public availability identifying the right data that is fit-for-purpose can be time consuming and frustrating. The Internet has often been heralded as revolutionising publishing, information sharing and research. This is undoubtedly true but this revolution has introduced its own problems. The new PPDB does not pretend to hold full data sets for all pesticides; indeed, there are still many data-poor pesticides in use. However, basic analysis and comparison of these data sets show it to be an improvement on existing resources bringing together the best data from the most reliable sources. Usage statistics show that the database is in global use and website activity is increasing month by month.

The MS Access database will enable external software applications to extract data directly avoiding the need for users to source and input their own data. This effectively reduces data input time and ensures that the best available data is used, helping to optimising the performance of the risk assessment process. This is the approach taken by the EU FOOTPRINT project that is currently developing a suite of harmonised pesticide risk assessment tools at national, regional and farm level. The on-line database and associated tools provide a more complete service to users. Language translations, comprehensive indices and extensive search facilities plus the layman data interpretations also make the data more user friendly.

7 References

Goss, D. & Wauchope, R.D., 1990. The SCR/ARS/CES Pesticides Properties Database: II Using It with Soils Data in a Screening Procedure. In: Weigmann, D.L. (Ed), Pesticides in the Next Decade: the Challenge Ahead, Virginia Water Resources Research Centre, Blacksburg, VA, USA, 471-493.

Gustafson, D.I., 1989. Groundwater Ubiquity Score: A Simple Method for Assessing Pesticide Leachability. Environmental Toxicology and Chemistry, 8, 339-357.

Lewis, K.A. & Bardon, K.S., 1998. A computer based informal environmental management system for agriculture. Journal of Environmental Modelling and Software, 13, 123-127.

Lewis, K.A., Brown, C.D., Hart, A. & Tzilivakis J., 2003. P-EMA: Overview and application of a software system to assess the environmental risks of agricultural pesticides. Agronomie. 23, 85-96.

8 Acknowledgement

The funding of the FOOTPRINT project by the European Commission through its Sixth Framework Programme is gratefully acknowledged.