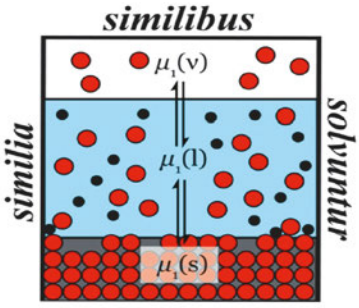
Assessment of Reliability and Uncertainty of Solubility Data

by Simão P. Pinho and Olga Ferreira

Initiated in May 2022, the Project "Assessment of Reliability and Uncertainty of Solubility Data" is supported by the IUPAC Analytical Chemistry Division. The main objectives and expected outcomes are to transfer fundamental and applied knowledge on good practices for phase equilibria measurements (particularly solubility), data analysis, and reporting high-quality data to the whole academic, technical, and scientific communities in the field. The project is largely focused on giving Ph.D. students and young researchers a series of tools to improve the quality of their outputs and their scientific literacy. So far six seminars have been held and recordings are retained online for broader dissemination.

The inaugural session was presented in July 2022 by Brynn Hibbert (University of New South Wales, Australia), entitled "Uncertainties of Solubility Measurements—Dark and Otherwise," where fundamental questions such as how to evaluate, estimate, and express uncertainties were presented, and discussed. The speaker developed the "dark uncertainty" concept, explaining the DerSimonian-Laird approach and its relevance to consensus solubility values from different laboratories considering unknown between-laboratory systematic effects.

In September 2022, we chaired the 20th edition of the International Symposium on Solubility Phenomena and Related Equilibrium Processes (Bragança, Portugal, online). Well-known and recognized by the researchers working in the field, this conference is associated with the Subcommittee on Solubility and Equilibrium Data and gathered about 110 researchers from 25 countries. One of the significant events at the conference was a workshop organized within this project. First, Ala Bazyleva and Vladimir Diky (National Institute for Standards and



Technology, USA) summarized general methods of building uncertainty budgets and conducting thermodynamic data validation, as well as suggested reference systems for various kinds of solubility measurements. An interesting open discussion with the participants evolved with respect to various types of uncertainty, testing experimental methods and equipment, and evaluating uncertainty for own measurements and evaluations. Then, Johan Jacquemin (Mohammed VI Polytechnic University, Morocco) showed an exhaustive data analysis on gas solubility in materials technical evaluation. Novel materials such as metal-organic frameworks, ionic liquids, or porous liquids were considered, and model systems to test the procedures were proposed, along with a set of requested data to support any scientific paper on the topic.

In 2023, Earle Waghorne (University College Dublin, Ireland) conveyed a rather pedagogical and practical approach using the solubility of dioic acids in different solvents, connecting it with the spirit of the first presentation in the project. Several sets of solubility data were checked for consistency, proposing consensus values, and focusing on the standard uncertainty and systematic errors. Subsequently, involving professionals from the consultancy area, Charles S. Oakes (Insight Geochemistry LLC, USA) brought to attention the truth in science in his presentation entitled "Chemical Thermodynamic Measurements That Are Not What They Are Claimed To Be." The tangible and intangible costs due to falsified data were discussed. Several examples showing the weaknesses linked to the action required by the scientific and technical communities were highlighted.

In February 2024, Wolfgang Voigt (Technische Universität Bergakademie, Germany) addressed the study of the solubility of salts in water. Three binary aqueous systems containing ($CaSO_4$, NaCl, or $NaNO_3$), and the ternary system H_2O + $MgSO_4$ + $MgCl_2$ were used to showcase how to deal with uncertainty and evaluation of solubility experimental data in this type of systems. A critical analysis on how statistics contributes more to the robustness than to the accuracy of the data, highlighting the enormous relevance

21st International Symposium on Solubility Phenomena and Related Equilibrium Processes

9-13 September 2024

University of Novi Sad, Serbia

ttps://issp2024.pmf.uns.ac.rs/

in the knowledge of the solution chemistry as well as in the formation of hydrates, and their stability, was presented. A series of questions to select and screen data from different sources was given and developed, as well as the difficulty of thermodynamic models to produce accurate solubility values of salts in water, under conditions not experimentally studied. In a well contemporary approach, perhaps with more significant impact on younger researchers, it was underlined how AI will not participate in obtaining better data.

Six seminars have been completed, and the videos uploaded to the IUPAC webpage dedicated to the project (https://iupac.org/project/2022-002-2-500). Besides the knowledge produced and shared, all these activities underlined the project's impact. On average, fifty participants from Australia to the USA attended the online presentations, always keen to join the discussions and share their expertise vividly. Still, during 2024, new activities will take place in the framework of the project. In September, a workshop will be organized within the scope of the 21st International Symposium on Solubility Phenomena (Novi Sad, Serbia), and a special issue in the *Journal of Solution Chemistry* is to be co-edited by Magdalena Bendová (University of Chemistry and Technology, Czechia) and Earle Waghorne.

For more information and comment, contact Task Group Chair Simão P. Pinho <spinho@ipb.pt> | https://iupac.org/project/2022-002-2-500

Human Drug Metabolism Database (hDMdb)

The IUPAC's Drug Discovery and Development Subcommittee (D3SC) has long been considering the assembly of a human drug metabolism database (hDMdb) that, in particular, might be a useful tool for interdisciplinary scientists engaged in small molecule drug discovery or toxicology. Ideally, it would be mounted on the Internet and available at no (or minimal) cost similar to the Protein Data Bank. Prominent applications would be to have an established compilation of human data from which each user could confidentially:

(i) Predict the biological disposition of new (e.g. proprietary) compounds relative to metabolic possibilities and, more importantly, relative to statistically-derived and ranked metabolic probabilities;

(ii) Assess the value of preclinical in vitro and in vivo models for predicting human drug metabolism, drug-drug interactions, and off-target toxicity when preceded by one or more competing bioactivation versus detoxification events;

(iii) Develop general and/or proprietary methods to account for 3D chemical structures and derive structure-metabolism/transporter relationships for specific metabophores or structurally-related families of compounds; and,

(iv) Explore machine learning (AI and 'facial recognition' at the molecular level) relative to drug metabolism in both humans and in preclinical in vitro and in vivo models of human drug metabolism.

The accompanying paper describes what we were initially thinking in more technical detail (*J. Current Drug Metabolism*, **2003**, *4*, 411-422). Current interest in exploiting AI within the context of drug discovery suggests that the concept outlined in this article may be even more relevant today than when it was first being contemplated several years ago (*e.g.* see *Drug Metabolism – Databases and High-Throughput Testing During Drug Design and Development* published for IUPAC by Blackwell Sciences Ltd. **1999**; ISBN 0-632-05342-9).

For various reasons we made limited progress past the early testing points noted in this initial paper. Alternatively, we feel that databases and the field of predicting/assessing drug metabolism have, in general, moved forward considerably in several directions over the last twenty years. Thus, we now envision writing a white paper update (pros and cons) about today's status of this topic and where it may be heading in the future. For that we are seeking your assistance in whatever way you may want to help, and specifically toward answering three questions. Depending upon which you prefer, your input can be cited in: (a) A non-identifying, confidential manner for you and your organization; (b) An acknowledgement of you and/or your organization; or (c) As a specifically cited reference including as much detail as you may want. Either way, we are not requesting any proprietary information or details for unpublished methods. The three questions are:

1. Is there still a need for a hDMdb like the one described above or for some other version; and can you please elaborate what resources are available for you to currently use in this regard and indicate any missing features that would add value if they could be incorporated?
2. What other types of resources are available or are you trying to develop internally, for you to address items (i) to (iv) above, especially with regard to the cutting-edge machine learning (AI) arena as it might be applied to the field of small molecule drug metabolism?
3. Given today's interest in small molecule