Meeting the Google Expectation for Chemical Safety Information:
Supporting Chemical Risk Assessment in Academic Research and Teaching

Leah McEwen, Cornell University, lrm1@cornell.edu
Ralph Stuart, Keene State College, ralph.stuart@keene.edu

Imagine a scenario in an academic research lab where you are working on a chemical system that has not yet been well characterized, perhaps looking at new catalysts, perhaps trying new reagent combinations, perhaps scaling up a new reaction for more sample. How might you gauge the reactivity potential against the equipment and procedures in place to ensure control is maintained? How do you identify critical points in the system where subtle or even inadvertent changes in condition or operation could send things awry, possibly becoming dangerous? How do you assess potential hazards of untried chemistry? Is there any helpful information readily available at your fingertips? Would a search as accessible as Google be able to turn up appropriate data for such scenarios?

Understanding this situation for research chemists has taken increased prominence over the last five years after a series of highly publicized incidents in academic laboratories in the United States and elsewhere. Government[1], scientific[2], professional[3] and academic[4] organizations have raised concern about the planning and oversight of chemical laboratory operations in the academic setting. Academic institutions and individual chemists have been subject to criminal prosecution in response to specific incidents where the primary consideration was adequacy of planning and supervision of the work being conducted in the laboratories. The immediate impact of these events has been a rapid increase in the level of documentation expected to support chemistry research in the academic setting. For example, following the death of a researcher in the University of California system[5], principle investigators of chemistry research laboratories on all 10 campuses are now required to identify which of 45 different hazards are associated with the work in their laboratories, so that the appropriate Personal Protective Equipment (PPE) can be issued to lab workers[6]. However, this approach does not address the broader range of issues that chemists face when planning experiments and managing labs that may contribute to hazardous situations.

Studying lesser-characterized chemistry is a prime focus in chemistry research and all chemical reactions involve energy changes. If too much energy is released too quickly for the immediate environment to absorb it, material damage and dangerous circumstances can result. Known reactivity and interactions of materials as well as documented hazards and handling protocols can help inform planning for prevention or mitigation of runaway and even unanticipated reactions. Incident investigations also indicate that many additional factors may be involved in the escalation of adverse chemical lab events, including process conditions, equipment limitations, laboratory setup, and training of operators, among others, and there is an increasing focus on assessing risks against a range of considerations in local context. It can be a significant challenge to identify and locate the relevant data and information needed to inform these analyses.

Potentially useful information is widely dispersed across a spectrum of scholarly, supplier, agency and other institutional data collection, reporting and publication venues. The most readily available data collections focus on a small number of well-characterized compounds and involve significant time and expense to maintain. Other initiatives aimed more broadly are hampered by ambiguous data provenance. The diversity of data formats intended for different audiences and applications often means that the quality of the information is difficult to evaluate, organize and apply in assessing risks associated with using reactive chemicals. The challenge is particularly pressing in the highly decentralized academic sector, characterized by novel chemistry, novice researchers and very little consistency of practice and resources within and among institutions. How would a research lab manager working with a new protocol apply the exposure control information on an M/SDS that was intended for large-scale hazmat clean-up situations? How would a chemistry student in a teaching lab discern the hazards associated with different physical forms of a compound, perhaps in a variety of dilutions?

What new ideas might an intersection of chemistry service professionals bring to bear on this conundrum: educators, hygiene officers, health and safety directors, librarians, information scientists?

Braving a record snowstorm and sustained by a small exploratory grant, 25 members of the Chemical Health and Safety (CHAS), Chemical information (CINF) and Chemical Education (CHED) Divisions of the American Chemical Society (ACS) and the ACS Committee on Chemical Safety, Environmental Health and Safety Task Force teamed up at Cornell University in Ithaca, NY in the winter of 2014 to pool their expertise and characterize the opportunities and challenges facing chemical safety information in the academic context. The primary goal of this initial meeting was to identify an overarching paradigm within the chemistry domain that could provide a framework for describing the lab safety process. The education community provided a model for risk assessment and management in the academic laboratory characterized by a series of actions necessary to work towards a safety culture: **R**ecognize the hazards, **A**sses the risks, **M**anage the hazards, and **P**repare for emergencies (RAMP, Figure 1). The group added, "Protecting the environment" and the initiative was dubbed "iRAMP" to reflect the iterative nature of both the RAMP process and the project[7].

Figure 1: The RAMP paradigm for managing chemical safety in the laboratory. Reprinted by the authors from http://www.irampp.org/blog.

To support this process, the workshop identified emerging tools from the safety profession and regulatory agencies that can be deployed to support a range of safety planning and management activities. These include the Global Harmonization System of Hazard Classification (GHS)[8], Safety Data Sheets (SDS, f.k.a., MSDS)[9], systems of exposure controls[10], research scale laboratory standards[11], prudent practice guidance for academic research labs[12], and a variety of risk assessment methods[13]. Working on the information front, librarians can support this process by focusing on the scholarly processes for critical inquiry, information evaluation techniques and documentation best practices. Similarly, information scientists are developing new technologies for structuring and interlinking data across sources.

Could a collective approach be developed that would support the use of these varied tools and information sources, connecting relevant data and official guidance into varied local management systems and planning workflows?

From the Ithaca meeting we were able to take away a set of general goals for the project. The diversity of academic science and laboratory settings argues for a safety culture of prudent practices where not all scenarios can be anticipated, planned for and addressed in established policies. Such dynamic systems suggest that a formulaic risk management “wizard” is not appropriate; assessing risk is site specific and involves educated judgment. How can we support iterative information discovery and reuse for a range of use-cases in a sustainable, transferable and scalable manner? Laboratory work evolves in a variety of local contexts under diverse regulatory requirements using loosely determined types of information management systems. This situation calls for a flexibly structured ecosystem of data, domain expertise and workflow tools mapped to essential commonalities of chemistry research and safety planning processes. A functioning information ecosystem relies on strong data management practices, including robust description of relevant chemistry research processes; collation and indexing of currently published data; organization and visualization tools for evaluation and use of data in risk assessment; codification of output formats; cross-walks for incorporation into local systems; and archiving and markup of method and analysis documentation for future reference.

What are the opportunities for amalgamating and opening up chemical data and information relevant to hazard recognition and safety planning and connect them to create a flexible and functional information ecosystem?

Late in the winter of 2014, another diverse group of chemical professionals from the education, laboratory safety, academic librarianship and information science sectors convened at the offices of the Royal Society of Chemistry in Cambridge UK. They reviewed information sources and specific environmental health and safety field cases in the context of current information management technologies. The field cases considered a variety of applications, including Lessons Learned type factor analyses based on reported laboratory incidents, control banding of laboratory ventilation rates, and risk assessment of common higher education teaching laboratory processes. The discussion re-emphasized the need when assessing risk to consider both hazard potential and exposure potential and that hazards associated with chemical research involve process conditions as well as chemical reactivity potential. These considerations can help inform local decisions for integrating hazard controls and prudent laboratory management practices.

A key challenge for connecting relationships among multiple types of factors is the predominant organization of chemical information by chemical entity. The safety literature is no exception, focusing on the hazard related properties of individual chemicals or substances without reference to specific experimental context or surrounding laboratory conditions. Scale, concentration, temperature, pressure, flow rate, and many other chemical, process, operator and environmental factors have the potential to trigger a runaway hazardous situation, including the presence of other compounds in the experiment or immediate vicinity. A more complete risk assessment process, as described by the RAMP model, involves a holistic laboratory level approach to managing risks beyond hazard identification. Complementing the "object based" focus on specific chemical entities with "process based" indexing could better identify information and data buried in the published literature on how these chemicals are being used under various conditions and combinations and the potential for subsequent unintentional interactions to arise.

There are currently a number of exploratory efforts to develop semantic description for indexing and linking disparate information for a variety of scientific, management and educational applications. Semantic description of data elements such as the Resource Description Framework (RDF)[14] can support flexible querying across a variety of information types. These technologies are just moving beyond describing the basic bibliographic information types into developing approaches for more specified domain data types.[15] Chemical health and safety management is traditionally an information intensive discipline and has created many data structures for supporting safe practices and incident analysis in chemical laboratories. Looking across the domain reveals many information management practices to build on, including the SDS format and GHS labeling system; clearinghouses of reviewed chemical hazard data such as the European Committee Agency portal; and information exchange formats and algorithms such as the International Uniform ChemicaL Information Database protocol (IUCLID)[16], and the IUPAC International Chemical Identifier (InChI)[17].

Developing safety management tools and models useful for the academic sector will necessitate tapping into digitally curated data in ways that are relevant to the decision-making processes of research chemists, safety professionals, institutional administration and other stakeholders. For example, a researcher might be looking at two known chemicals in a proposed reaction scheme and want to know of any conditions that might trigger an adverse outcome, if there are any known procedures for minimizing the likelihood of these conditions and how to mitigate potential harm if something untoward did occur. The relevant data and information may come from a diverse set of sources covering physical properties, synthesis protocols and previously reviewed incidents. However, these sources vary widely in format and much of the data is available only in text format, often not parsed with the necessary granularity and not recognized by computers in the implied scientific context. This "text-locked" state of the data limits the options for searching, analysis, tagging and linking into information management systems such as electronic laboratory notebooks (ELNs). Similarly, institutional Environmental Health and Safety staffs need to collect information about different chemical properties to support campus-wide laboratory operations, including coordinating emergency response planning, addressing ventilation needs and providing waste disposal services.

How do we better identify and organize chemical hazard and risk management concepts to facilitate discovery and use of what has been reported? Can we classify and define chemical reactivity in the context of hazard identification, evaluation and exposure control?

The iRAMP project is currently collaborating with the PubChem program at the US National Library of Medicine to explore the opportunities and challenges on a practical level. Objectives include connecting the broad range of information sources relevant to chemical safety planning, applying vocabularies of relevant concepts to index and annotate reports, and analyzing indicators to extend classification of potential hazard across lesser characterized chemical interactions. The approach is to develop a data stream based on the community standard format of the Laboratory Chemical Safety Summary (LCSS)[18] described by the National Research Council in *Prudent Practices*[19]. The data stream consolidates safety and health-related data reported in open authoritative literature and can be downloaded or hyperlinked into institutional systems to augment local chemical information, such as chemical inventories, lab-specific personal protective equipment assessments or ELNs. The project is not aiming to develop local chemical management systems or interfaces but to facilitate management of the data stream through user manuals and examples of use in risk assessment protocols.

A critical component that has been identified for scalability and sustainability is the use of robust chemical identifiers to allow linking and validation of information collected. The familiar CAS Registry Number is ubiquitous across many chemical safety information sources but is not designed for linking and does not include important chemical designations related to concentration and purity. IUPAC projects are underway to extend the InChI algorithm to incorporate the GHS designations and support better linking and access to documentation of chemicals from manufacture through waste disposal and emergency response. In response to community input, we are also investigating the development of chemical incompatibility classification that will group potentially similar hazardous interactions, extending on the functionality of the chemical reactivity worksheets available in CAMEO[20].

Next steps include extending the LCSS summaries beyond hazard information about specific chemicals to include additional relevant data such as process variables and to support further annotation of reactivity hazards. It will be possible to leverage domain terminology to more systematically analyze the reported data and construct specific process scenarios such as chemical A + chemical B under X condition could lead to Y type of adverse event. An ongoing effort is the development of a Chemical Safety Ontology (CSO) that establishes the relationships among chemical safety concepts, such as hazard classification and exposure control, and maps these to the semantic information structures used by computers for indexing, classification and further organization through knowledge graphs. These approaches will rely on terminology and descriptions developed for use in the chemistry enterprise.

The chemistry domain boasts a myriad of glossaries and regulatory terminology developed for a wide range of chemical research and applied subdisciplines, including those represented in the IUPAC color books[21]. Ascertaining the applicability and interconnectivity of the available vocabularies to the questions around chemical hazard management and safety planning will require a process of review and curation by chemistry research support professionals, including environmental safety officers and chemistry librarians in collaboration with domain information scientists. A successful model that interlinks and semantically enables a range of authoritative vocabularies is the AGROVOC thesaurus, a project of the Agriculture Information Management Standards (AIMS) initiative under the auspices of the Food and Agriculture Organization of the United Nations[22]. Could such a work model be developed in the chemistry domain to integrate chemistry vocabularies more systematically into chemical information management systems, building on the efforts of the current IUPAC Color Book Data Management project[23]?

Everyone we have connected with in the chemical safety, chemical information and chemical education communities has appreciated the complex nature of the challenge and the potential for a multi-faceted professional community approach. It is a global concern for our discipline and we welcome ideas and contributions from a diversity of stakeholders and locations interested in moving this effort forward.

The laboratory safety data stream provided through the PubChem database is an ideal proof-of-concept to support professional community engagement in managing information relevant to chemical safety planning; large scale, government-based, and open to use. The initial word from the academic chemical safety community is that this approach will add much needed value to the prominent electronic information search scenario for researchers via Google. Stay tuned for the next chapter of this compelling exploration.

---

[1] United States Chemical Safety Board. *Texas Tech University Laboratory Explosion*; Washington, DC, 2011. http://www.csb.gov/assets/1/19/CSB_Study_TTU_.pdf (accessed June 2015).

[2] National Research Council. *Safe Science: Promoting a Culture of Safety in Academic Chemical Research;* Washington, DC: The National Academies Press, 2014. http://www.nap.edu/catalog/18706/safe-science-promoting-a-culture-of-safety-in-academic-chemical (accessed June 2015).

[3] American Chemical Society, Committee on Chemical Safety. *Creating Safety Cultures in Academic Institutions*; Washington, DC, 2012. https://www.acs.org/content/dam/acsorg/about/governance/committees/chemicalsafety/academic-safety-culture-report-final-v2.pdf (accessed June 2015).

[4] Association of Public and Land-grant Universities. *Association of Public and Land-grant Universities Announces Creation of Task Force on Laboratory Safety*. April 23, 2015. http://www.aplu.org/news-and-media/News/association-of-public-and-land-grant-universities-announces-creation-of-task-force-on-laboratory-safety (accessed June 2015).

[5] Kemsley, J. N. Learning From UCLA: Details of the experiment that led to a researcher's death prompt evaluations of academic safety. Chem. Engr. News., 2009, 87 (31) pp 29-31, 33-34. http://cen.acs.org/articles/87/i31/Learning-UCLA.html (accessed Jul 2015).

[6] University of California San Diego, Environmental Health and Safety. Laboratory Hazard Assessment Tool (LHAT). Version 11, 2014. http://www-ehs.ucsd.edu/ih/ppe/LHAT_UCSD.pdf (accessed Jul 2015).

[7] American Chemical Society, Divisions of Chemical Health & Safety and Chemical Information. iRAMP Innovative Project Grant. http://www.irampp.org/blog/ (accessed Jul 2015).

[8] United Nations, Committee of Experts on the Transport of Dangerous Goods and on the Globally Harmonized System of Classification and Labelling of Chemicals (UNCETDG/GHS). Globally Harmonized System of Classification and Labelling of Chemicals. New York and Geneva, 2003. http://www.unece.org/trans/danger/publi/ghs/ghs_welcome_e.html (accessed Jul 2015).

[9] Occupational Safety and Health Standards. OSHA Brief no. 3514, Hazard Communication Standard: Safety Data Sheets (29 CFR 1910.1200(g)). https://www.osha.gov/Publications/OSHA3514.html (accessed Jul 2015).

[10] Department of Labor, Occupational Safety and Health Administration. Preventing occupational illnesses through safer chemical management. https://www.osha.gov/chemicalmanagement/ (accessed Jul 2015).

[11] Toxic and Hazardous Substances: National Research Council Recommendations Concerning Chemical Hygiene in Laboratories (Non-Mandatory); 29 CFR 1910.1450 App A; Occupational Safety and Health Administration, Code of Federal Regulations; U.S. Government Printing Office: Washington, DC, 2012. https://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=STANDARDS&p_id=10107 (accessed Jul 2015).

[12] National Research Council. Prudent Practices in the Laboratory: Handling and Management of Chemical Hazards, Updated Version; Washington, DC: The National Academies Press, 2011. http://www.nap.edu/catalog/12654/prudent-practices-in-the-laboratory-handling-and-management-of-chemical (accessed Jul 2015).

[13] 19. American Chemical Society, Committee on Chemical Safety, Hazard Identification and Evaluation Task Force. Identifying and Evaluating Hazards in Research Laboratories; Washington, DC, 2015. http://www.acs.org/content/dam/acsorg/about/governance/committees/chemicalsafety/publications/identifying-and-evaluating-hazards-in-research-laboratories.pdf (accessed Jul 2015).

[14] World Wide Web Consortium, RDF Core Working Group. Resources Discovery Framework. http://www.w3.org/2001/sw/#rdf (accessed Jul 2015).

[15] Borkum, M. I.; Frey, J. G. What's in a Name? Quite a Lot, as it Happens! Chemistry International, 2015, 37(2) pp 7–9. http://dx.doi.org/10.1515/ci-2015-0231 (accessed Jul 2015).

[16] European Chemicals Agency. International Uniform ChemicaL Information Database, version 5. http://iuclid.eu/ (accessed Jul 2015).

[17] International Union of Pure and Applied Chemistry, Chemical Nomenclature and Structure Representation Division, Subcommittee on the IUPAC International Chemical Identifier. IUPAC International Chemical Identifier (InChI). http://www.iupac.org/home/publications/e-resources/inchi.html (accessed Jul 2015).

[18] National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information. PubChem Laboratory Chemical Safety Summaries. http://testpubchem.ncbi.nlm.nih.gov/LCSS/ (accessed Jul 2015).

[19] National Research Council. Prudent Practices in the Laboratory: Handling and Management of Chemical Hazards, Updated Version; Washington, DC: The National Academies Press, 2011. http://www.nap.edu/catalog/12654/prudent-practices-in-the-laboratory-handling-and-management-of-chemical (accessed Jul 2015).

[20] National Oceanic and Atmospheric Administration. CAMEO Chemicals Database of Hazardous Materials. http://cameochemicals.noaa.gov/ (accessed Jul 2015).

[21]International Union of Pure and Applied Chemistry. Nomenclature and Terminology publications. http://www.iupac.org/home/publications/e-resources/nomenclature-and-terminology.html (accessed Jul 2015).

[22] Food and Agriculture Organization of the United Nations, Agriculture Information Management Standards. AGROVOC Multilingual agricultural thesaurus.

http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus (accessed Jul 2015).

[23] International Union of Pure and Applied Chemistry, Committee on Publications and Cheminformatics Data Standards. IUPAC Color Book Data Management Project. http://www.iupac.org/nc/home/projects/project-db/project-details.html?tx_wfqbe_pi1%5Bproject_nr%5D=2013-052-1-024 (accessed Jul 2015).