

RDA/IUPAC Interest Group Draft Charter Template

Name of Proposed Interest Group: Chemistry Research Data Interest Group

Introduction (A brief articulation of what issues the IG will address, how this IG is aligned with the RDA mission, and how this IG would be a value-added contribution to the RDA community):

We are proposing a Chemistry Research Data Interest Group under the auspices of the Research Data Alliance (RDA), to foster diverse professional exchange on issues particular to data originating from the field of chemistry. Chemistry, as one of the central sciences, has fundamental impact on the fields of health, pharmaceuticals, materials, energy and many other applied sciences. There is a wealth of chemical data in various heterogeneous formats, distributed across a myriad of systems with endless potential for reuse in chemistry research and many related domains. However, many social, technical and administrative factors have limited the opportunities for open sharing and interoperable exchange.

The high reuse value of chemical information has sparked decades of innovative technologies addressing various challenges in handling chemical specific data, but very few approaches have persisted, are extensible beyond specific data types and/or are operable at scale. There is demonstrable need for coordinated development of updated and scaled infrastructures, hard and soft, for enabling chemical data exchange and connecting data providers with data users across sources and applications. The RDA mission is to build the social and technical bridges that enable open sharing of data. Organizing a forum for professional exchange directed at addressing opportunities and challenges for chemistry data management within the RDA framework will support international participation across a broad range of stakeholders and foster connections with data types and user scenarios in many disciplines. Bringing in IUPAC (International Union of Pure & Applied Chemistry) as co-sponsor of the group would clearly bridge the activities of this group between those of RDA and the responsible standards body for chemistry.

User scenario(s) or use case(s) the IG wishes to address (what triggered the desire for this IG in the first place):

In response to many scientific, technical, and socioeconomic drivers, research chemists, chemical educators and chemical information specialists are recognizing the necessity to move forward with infrastructures, best practices, and cultural shifts to support consistent data management and sharing practices. Research funding agencies are increasingly requiring openly accessible research data and are looking to the scientific research communities to develop domain-appropriate criteria. Professional societies recognize the benefits in encouraging chemistry professionals to be experts in handling electronic data and documentation, and supporting these skills in professional education. Increasing opportunities for low-barrier technical solutions are opening up the market for electronic based information and data flow through electronic notebooks, automated data collection

and analysis, data repositories and citation networks.

The importance of chemical data has long been recognized by science communities and centuries-old efforts in indexing and repackaging chemical data from primary literature into expansive collections that support innovation across many disciplines. However, there are many challenges to meet increasing demands for open research data deposit and maximizing machine operable data exchange. Working with chemistry research data often involves extensive consideration of contextual factors and layers of interpretive technologies. Divergent high-touch workflows have evolved to manage data in the existing collections. Long traditions of small laboratory culture and strong proprietary and commercial value impact the overall adoption and incorporation of open data exchange and high performance computing directly in research chemistry outside of a few sub-disciplines (e.g. drug discovery). As already experienced in many networking venues amidst chemical information professionals, an international Interest Group that spans a range of professional perspectives and expertise can provide much needed opportunity for fostering convergent and informed discussions.

Objectives (*A specific set of focus areas for discussion, including use cases that pointed to the need for the IG in the first place. Articulate how this group is different from other current activities inside or outside of RDA.*):

At some level, chemistry information is ubiquitous to every wet science laboratory and many theoretical research problems as well. The high value and wide applicability of chemical data generally has ensured a landscape of numerous and scattered, thoughtful and variously adopted “best practices”. Many venerable research and scientific publishing institutions and disciplinary data projects are involved in reviewing and managing data of high utility and have influential roles in long-standing community standards of practice around data use in the discipline. To maximize on the knowledge potential of the discipline, we are interested in approaching the functionality of data from several angles, including domain scope, infrastructure, and community practice. Specifically we propose to:

1. Characterize different chemical data types of interest, identify critical points in the data life-cycle from instrument to publication, compile data management criteria in practice, map gaps in interoperability and opportunity potential for standards and other infrastructures, and prioritize outreach approaches and tools for researchers, primary publishers, data compilers, and others who manage chemistry research data.
2. Leverage effort from all parties to establish metadata standards, ontologies and other soft infrastructures for chemical data that are adaptable for different application purposes
3. Examine current research workflows in various research domains that interact with chemical data to support minimal disruption, encourage development of best practices and lower barriers to adoption. Particular attention will be given to engaging instrument manufacturers in the discussions, as they represent a good target to reduce the barriers to storing both data and metadata early in the research workflow.
4. Cultivate sharing culture among researchers working in chemistry related fields by

demonstrating potential innovations based on reusable chemical data.

Participation (*Address which communities will be involved, what skills or knowledge should they have, and how will you engage these communities. Also address how this group proposes to coordinate its activity with relevant related groups.*):

There is increasing interest within RDA to engage with domain-based initiatives and data-driven organizations. The International Union of Pure and Applied Chemistry (IUPAC) is a long-standing professional international organization with vested interest in supporting broad dissemination and usability of chemical data through development of standards and recommended practices. IUPAC engages members from adhering organizations in over 50 countries and is associated with over 30 international scientific organizations. Positioning this initiative as a joint RDA/IUPAC interest group will enable us to leverage the mechanisms and infrastructure of both international working member organizations to facilitate global input, dissemination and practical implementation of initiatives.

Potential Interest Group members hail from a range of professions and sectors that intersect chemistry research data, including experimental and theoretical researchers, educators, data and information scientists, librarians, publishers, database providers, and many others in academic, industrial, private and public sectors worldwide. Many are active in professional groups with expertise in chemistry data, including the American Chemical Society (ACS) Division of Chemical Information (CINF), the Royal Society of Chemistry Society (RSC) Chemical Information and Computer Applications Group (CICAG), the Chemical Structure Association (CSA Trust), the German Chemical Society, the Chemical Society of Japan, the Chinese Chemical Society, among others. Opportunities exist to participate regularly in the technical programming and social networks of these organizations to further engage chemistry researchers and information professionals.

Outcomes (*Discuss what the IG intends to accomplish. Include examples of WG topics or supporting IG-level outputs that might lead to WGs later on.*):

Understanding current data management practices (in the broadest sense) and perceived gaps across the chemistry discipline is key for targeted action. Suggested documentation projects of potential interest and value for the community include:

1. Collect top five priority outcomes from members with rationale; from these identify commonalities and diversities for the group and the chemical information and data management professions writ large; collect at discussion events and consider a survey question for new members
2. Identify and characterize existing systems and solutions relevant to chemistry and the interests that arise in the survey, including existing disciplinary data repositories, ontologies, and other community data projects
3. Identify and compare funding agency requirements internationally that potentially involve chemistry data
4. Determine what chemistry analysis instruments are already doing in taking data, what

file formats are in use? What metadata commonalities, diversity? Proprietary data, format issues?

5. Survey top chemistry publishers of the various types of data that are included with manuscripts as supplemental information
6. Others as they arise from discussion

Mechanism (Describe how often your group will meet and how will you maintain momentum between Plenaries.):

Discussions of interest initiated at the ACS meetings in March 2015 and August 2015 sparked a proposal for a BoF session at the September RDA Plenary in Paris to seek input on formulating a mechanism for a group. Further international outreach is planned through meetings and technical symposia at the multi-national chemical societies Pacificchem meeting in December 2015 and the ACS meeting in March 2016. Additional programming will be proposed with other societies and meetings. Monthly virtual meetings and regular inclusive communication channels will be established in the fall. Additional meetings focused on specific outcomes will be scheduled as needed.

Timeline (Describe draft milestones and goals for the first 12 months):

Outreach – first 6 months

1. Discussion with the IUPAC Committee on Publications and Cheminformatics Data Standards – August 2015
2. BoF session at the RDA Plenary in Paris – September 2015
3. Reach out to other pertinent RDA groups, such as the RDA/CODATA Materials Data, Infrastructure & Interoperability IG, the Data Citation IG, and others - start at the Paris Plenary, September 2015
4. Establish communication structure – Fall 2015
5. Outreach and increase group member list through planned symposia and networking highlighting a broad range of data initiatives at other various domain meetings – ongoing, started March 2015
6. Continued brainstorming for issues and outcomes of potential interest for further discussion – ongoing, started March 2015

Roadmap – second 6 months

1. Focus on 3-5 documentation activities first year, primarily focusing on professional and scientific community information gathering to develop a roadmap of challenges and opportunities for chemistry data management
2. Identify deliverables and establish working groups by the end of the year for 1-3 problems in the community for the common good of all / most stakeholders

Potential Group Members (Include proposed chairs/initial leadership and all members who have expressed interest):

FIRST NAME	LAST NAME	EMAIL	TITLE
Elsa	Alvaro	elsa.alvaro@northwestern.edu	Northwestern University, US
Evan	Bolton	bolton@ncbi.nlm.nih.gov	National Institutes of Health (NIH), US
Ian	Bruno	bruno@ccdc.cam.ac.uk	Cambridge Crystallographic Data Centre (CCDC), UK
Stuart	Chalk	stuartjchalk@gmail.com	University of North Florida, US, convening organizer
Alex	Clark	aclark.xyz@gmail.com	Molecular Materials Informatics, Canada
Simon	Coles	s.j.coles@soton.ac.uk	University of Southampton, UK
Judith	Currano	currano@pobox.upenn.edu	University of Pennsylvania, US
Erin	Davis	erinsdavis@gmail.com	Cambridge Crystallographic Data Centre (CCDC), UK
Jeremy	Frey	J.G.Frey@soton.ac.uk	University of Southampton, UK
Evan	Hepler-Smith	evan.heplersmith@gmail.com	Princeton University, US
Brynn	Hibbert	b.hibbert@unsw.edu.au	University of New South Wales, Australia, IUPAC Division V
Richard	Kidd	KiddR@rsc.org	Royal Society of Chemistry (RSC), UK
Mark	Kinnan	mkkinna@sandia.gov	Sandia National Laboratories, US
Ye	Li	liye@umich.edu	University of Michigan, US
Bonnie	Lawlor	chescot@aol.com	IUPAC, US
Peter	Linstrom	peter.linstrom@nist.gov	National Institutes of Standards and Technology (NIST), US
David	Martinsen	D_Martinsen@acs.org	American Chemical Society (ACS), US
Leah	McEwen	lrm1@cornell.edu	Cornell University, US, convening organizer
Mirek	Sopek	info@chemicalsemantics.com	Chemical Semantics, Poland
Suzanna	Ward	ward@ccdc.cam.ac.uk	Cambridge Crystallographic Data Centre (CCDC), UK
Brian	Westra	bwestra@uoregon.edu	University of Oregon, US
Antony	Williams	tony27587@gmail.com	Environmental Protection Agency (EPA), US
Jeffrey	Wiseman	wisemaj2@ohio.edu	Ohio University, US

NOTE: The convening group is actively pursuing a number of outreach opportunities through connections with IUPAC and other Chemistry Societies to expand membership globally and engage experts across professional and industrial sectors, including research & development, manufacturing & distribution, education, and regulation.