Supporting Information Review and Data Analysis at *Organic Letters*

by Angela M. Hunter

In an era where research data is readily accessible during the peer review process, several obstacles persist in verifying the data required for publication. While Supporting Information (SI) for all American Chemical Society (ACS) journals is freely available, there is no standardized review process, and errors frequently linger, despite peer review. The SI for submitted manuscripts routinely contains missing or incorrect data, as well as inconsistencies between the submitted manuscript and the SI. While rare, issues of data manipulation, including digitally altered spectra, and plagiarism have been identified. [1] For the past five years at the ACS journal Organic Letters, the SI for all manuscripts has been reviewed by a data analyst to ensure compliance with author guidelines and to allow authors to address missing data and/or data inconsistencies prior to publication.

Organic Letters works with authors to provide education and resources to improve manuscript and SI preparation. In collaborating with authors, our office has developed SI checklists that can be used prior to submission to ensure complete and accurate SI. [2] Included in the checklists are the data requirements, necessary synthetic details, and guidance to properly format data. The SI review process is continually evolving as new standards and criteria are incorporated. For example, more rigorous safety standards are being enforced, as ACS publications will require authors to report "experimental details to address and emphasize any unexpected, new, and/or significant hazards or risks associated with the reported work."[3] Furthermore, to increase reproducibility, the Organic Letters guidelines were updated in 2017 to require that one-step organic transformations include a detailed synthetic method at the 1 mmol scale for at least one representative example. These standards ensure the SI is complete, improve reproducibility, maintain a consistent presentation of data, and minimize the likelihood of incorrect and/or manipulated data.

While instances of manipulated data are rare, they are still found in 2-3% of all submitted manuscripts at Organic Letters, although the severity varies. If the data is found to be questionable or unusual during SI review, the Corresponding Author is contacted for an explanation and asked to provide original data for review. Often, authors are unable to provide the original data for their own submission. Explanations vary as to why the original data is not available and have included the following: improper storage of data, deleted data with no back-up, misplaced data, or inaccessible data (*i.e.*, former students have the data and no contact information is available). In dealing with requesting and reviewing data for original submissions, there are several continuing issues: How long should a publisher wait to receive the data from an author? What happens if an author cannot provide the original data? Is newly generated data acceptable to replace questionable data? For an Addition/Correction to a publication, if new data is submitted, how will the data be verified?

One recommendation to enhance the quality of the manuscript and SI is to have additional guidance for peer reviewers. This may be beneficial to the review process and could lead to a higher quality publication. A thorough review requires careful examination and a systematic approach to determine if the experiments, data, and conclusions provided by the authors are worthy of publication. Since the SI is typically longer than the manuscript, it can be overwhelming for reviewers to determine how best to give useful feedback to an author. Ideally, a reviewer should follow these steps:

- Before starting an evaluation of the manuscript and SI, review the journal guidelines, focusing on the sections pertaining to the SI. Each journal has different criteria and it is helpful to understand what data is required before SI review begins.
- Determine that all necessary information is included in the SI and detail any data omissions, inconsistences between the reported methods and the data presented, unusual data, or errors in data reporting in the review.
- Moreover, include any potential ethical concerns detected in the review, such as possible data manipulation or duplicate submission by the

authors. This information is invaluable to the Editor in determining a manuscript's outcome.

A systematic SI review prior to publication improves the quality of the published SI by having a consistent presentation of data that benefits both authors and readers. Organic Letters' continually examines best practices for reviewing the SI, resolving issues encountered when requesting data from authors, and advising authors on submitting data for publication.

References:

1. Amos B. Smith, III. "Data Integrity". *Org. Lett.* **15**(12):2893–2894, 2013. https://doi.org/10.1021/ol401445g
2. Angela M. Hunter and Amos B. Smith, III. "Review of Supporting Information at Organic Letters". *Org. Lett.* **17**(12):2867–2869, 2015. https://doi.org/10.1021/acs.orglett.5b01700
3. Carolyn R. Bertozzi. "Ingredients for a Positive Safety Culture." *ACS Cent. Sci.* **2**(11):764–766, 2016. https://doi.org/10.1021/acscentsci.6b00341

Angela M. Hunter <angela.hunter@orglett.acs.org, > is Data analyst, Organic Letters in the Department of Chemistry at the University of Pennsylvania, Philadelphia, USA. ORCID.org/0000-0002-3474-525X

Update on ThermoML

by Kenneth Kroenlein

ThermoML is the XML-based IUPAC Standard for Thermodynamic Property Data, first developed as an IUPAC project in 2006, and extended in 2001. [1] At present, ThermoML is being used by a number of process design packages as a data input format, as well as more broadly on the Internet as a data dissemination format. The freely-accessible ThermoML archive maintained by the Thermodynamics Research Center (TRC) is the primary source of these data and has new address https://www.nist.gov/mml/acmd/trc/thermoml-archive. We fully expect this archive to continue to grow as new materials as published.

The most recent development impacting ThermoML is the expansion of TRC's data collection activities into metal-based systems. Several elements of metals-based systems reporting requirements, around concepts such as phase description and differing expectations for sample characterization, made substantial update of the ThermoML schema necessary. Accordingly, TRC is developing an updated schema, version 5.0, and associated forward and backward conversion software in the context of IUPAC.

Previous recommendations related to ThermoML:

- Frenkel, M. *et al.* XML-based IUPAC standard for experimental, predicted, and critically evaluated thermodynamic property data storage and capture (ThermoML) (IUPAC Recommendations 2006). *Pure Appl. Chem.* **78**(3):541-612, 2006, https://doi.org/10.1351/pac200678030541
- Frenkel, M. *et al.* Extension of ThermoML: The IUPAC standard for thermodynamic data communications (IUPAC Recommendations 2011). *Pure. Appl. Chem.* 2011, **83**(10):1937-1969, https://doi.org/10.1351/PAC-REC-11-05-01

[1] http://www.iupac.org/namespaces/ThermoML

Kenneth Kroenlein <kenneth.kroenlein@nist.gov> is Group Leader, Thermodynamics Research Center, National Institute of Standards and Technology, in Broomfield, Colorado, USA.