International Edition: DOI: 10.1002/anie.201710493
German Edition: DOI: 10.1002/ange.201710493

Robust Chemistry: The Importance of Data and Methods Sharing

*Mattias Björnmalm and Frank Caruso**


M. Björnmalm


F. Caruso

Robustness, reproducibility, reliability, transparency, and "open science" are areas that are increasingly recognized as foundational for impactful research. These topics are intimately connected to how we think about science, how we conduct research, and how we report results, and improvements in all of these areas may help advance and accelerate impactful research.

Chemistry has a proud history of championing transparency and rigorousness. Examples range from the early establishment of open repositories such as the Cambridge Structural Database to long-established traditions and guidelines (formal and informal) for reporting the discovery, synthesis, and characterization of new compounds. However, although chemistry is commonly referred to as the "central science", more recent efforts towards increasing robustness in research have been led by neighboring fields such as biology, medicine, and psychology.[1] These efforts include new and improved ways to share data and methods.

Integration of these emerging concepts into the chemical sciences results in "robust chemistry". Recent examples range from the development of "best reporting practices" to the increasing use of videos to demonstrate important experiments and procedures. In a recent Editorial in *Angewandte Chemie*, Bergman and Danheiser discussed challenges involving unconscious investigator bias and the replication of results, and we have recently discussed ways to achieve increasingly impactful research within bio-nano science.[2] Herein, we highlight recent advances and innovations towards increasing robustness in chemistry through the sharing of data and methods.

Whilst many of us are aware that capabilities to handle and store information have been improving exponentially for years, it is not easy to appreciate the full opportunities that such improvements provide. For more than a century, scientific journal publishing has been the main route for disseminating results. The art of succinctly and clearly describing results—often from research spanning many months or years—so that they can be printed onto a few journal pages has been a hallmark of science.

While many scientists nowadays read PDFs instead of the hard-copy versions, the principle is the same: journal publishing has been foundational for science to become what it is today, as it has allowed researchers to stay up to date with developments around the globe and to build on each other's discoveries and insights. The importance of having access to and drawing inspiration and guidance from a global research community cannot be overstated. As Newton said, "If I have seen further, it is by standing on the shoulders of giants."

In the past, the practice of condensing months and years of research into a few printed pages was necessary to allow for widespread circulation of discoveries. However, this is no longer the case because of the advent of new information-storing and -sharing technologies.

The publication of journal articles is considered a main outcome of research projects. An unintended consequence of this is that anything that does not make it into a manuscript will often not be reported, which often includes 1) the raw data, 2) experiments that turned out to be tangential or "did not work", and 3) tacit knowledge (e.g., tricky experimental steps that are difficult to condense into a short text). And as the project is often considered "finished", there is little incentive to share this information. The fact that these details are unlikely to be shared is not due to malice, but for a number of other reasons, including the fact that researchers are busy people or that some may not appreciate the benefits that such data may bring. Such information is then lost to the broader community as well as further potential opportunities to "stand on the shoulders of giants".

Nowadays, it is easy to sign up for an online storage service (e.g., Dropbox, Google Drive, and OneDrive), and the cost is typically less than a cent per gigabyte per month (or even free!). Additionally, many institutions offer free data repositories, and a wide range of public data repositories are also available. The point is that presently, it is easier and cheaper

[*] Dr. M. Björnmalm, Prof. F. Caruso
ARC Centre of Excellence in Convergent Bio-Nano Science and Technology, and the Department of Chemical Engineering
The University of Melbourne
Parkville, Victoria 3010 (Australia)
E-mail: fcaruso@unimelb.edu.au
Dr. M. Björnmalm
Department of Materials, Imperial College London (UK)
The ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/anie.201710493.

than ever to store and share data. This opens up opportunities for sharing results in much more comprehensive ways than those currently being utilized. Underutilization of data sharing is a loss for two reasons.

One is illustrated by the emerging field of "big data", which has demonstrated that large data sets (for example, aggregated from multiple studies) can be used to reveal patterns, trends, and associations that cannot be detected in smaller (e.g., single-study) data sets. To tackle more complex questions, for example in interdisciplinary fields, these types of integrative approaches may not only be an "added bonus", but also central for advancing understanding and discovery.

A second loss is the existence of "dark data", which is data generated through well-performed studies that are not shared with the research community. This includes data from experiments that did not make the "final cut" for a manuscript and data from unfinished studies. Ideally, all data generated from well-designed and well-executed experiments should be disseminated. This would not only facilitate the integrative approaches discussed above, but also allow the research community to see the "roads travelled" instead of only the "destinations reached". Imagine if you were to forget all research projects and experiments that you know of that have not been published. How much more difficult would it be to decide what to do next? And how likely would you be to try something you have already tried (that did not make it into a journal publication)? An additional benefit of access to dark data is that it could provide inspiration for scientists; for example, for experiments that did not "work out", a different approach or an optimization procedure could potentially provide new insights and research directions. Exploration is a key foundation of research, and knowing what has been attempted already and what the results are not only helps us to avoid unnecessary repeats, but also guides future scientific endeavors.

Excellent initiatives and frameworks (e.g., by researchers, funding organizations, and publishers) already exist to facilitate the adoption of best practices for data sharing (e.g., to share as much data as possible and to use reliable data hosting services such as repositories; for example, see https://opennessinitiative.org/making-your-data-public/). These types of practices help increase transparency and reliability of reported data. Increased adoption would undoubtedly help accelerate research, as well as increase robustness in chemistry.

The sharing of methods is similar, but distinct, to the sharing of data. Although both have the opportunity to benefit greatly from our increasing ability to store and share large amounts of information, they do so in different ways. For data, much of it comes down to the possibility of sharing raw data and dark data, which can be largely covered by tools such as public repositories (discipline-specific or general ones) and best data-sharing practices, as discussed above.

For the sharing of methods, the challenge is, instead, to share "experience" and other tacit knowledge. Detailed and accurate methods sections in publications are critical, but this goes beyond that. Much of the knowledge in the chemical sciences is difficult to express concisely and precisely in words. Consider this simple example: when one of the authors recently wanted to sew a button on their shirt, they did not try to find written instructions, but rather searched YouTube and watched a short step-by-step guide. As many procedures in chemistry are arguably more complex than the sewing of a button, imagine the possibilities if we could easily and routinely access photos and video guides that would demonstrate the steps involved in a synthesis or how to best operate an instrument for a specific experiment.

Ways to achieve this include the use of action cameras such as GoPro cameras.[3] The idea is that because they can be worn hands-free (like a headlamp), one has the possibility to record, unencumbered, all steps of an experiment with first-person video and audio. For someone watching the video, it could be like having an expert guiding you through all of the steps, which, we would argue, is close to the "gold standard" for gaining expertise nowadays! Before using a new instrument or attempting a synthesis that you have not done before, ideally an experienced person would show you the steps involved. Using these types of cameras (that are becoming cheaper and increasingly ubiquitous), an expert of that instrument, synthesis, or procedure could show you the steps. And you could pause, rewind, and see it again as often as desired. Similar types of videos are already being disseminated through video journals, such as JoVE (see https://www.jove.com), and videos can often be submitted as supporting information in other journals. The idea is not to replace conventional face-to-face training (the "gold standard"), but to complement it. The widespread use of engaging videos for entertainment, education, and information is an integral part of today's society. We believe that, if widely adopted, these types of videos can have a profound, positive impact on the development of research.

Robustness in chemistry can be enhanced by increasing the adoption of best data-sharing practices and the use of photos and videos for sharing methods and practical knowledge. Journals and journal articles are an integral part of the foundation of scientific endeavor. The key points discussed herein (i.e., sharing data and methods) are complementary practices that, if widely adopted, could lead to a step change in how we conduct and report research, especially if combined with new and improved tools to effectively engage with such valuable information (e.g., through specialized search engines). Yes, cultural changes will be required, which can be challenging, but the potential rewards are enormous. This includes accelerating and streamlining current research, as well as enabling completely new research directions.

[1] M. R. Munafò et al., *Nat. Hum. Behav.* **2017**, *1*, 21.
[2] M. Björnmalm, M. Faria, F. Caruso, *J. Am. Chem. Soc.* **2016**, *138*, 13449.
[3] M. Björnmalm, M. Faria, F. Caruso, *Chem. Mater.* **2016**, *28*, 8441.

