

Deutsche Ausgabe: DOI: 10.1002/ange.201710493
 Internationale Ausgabe: DOI: 10.1002/anie.201710493

Robuste Chemie: die Bedeutung des Teilens von Daten und Methoden

Mattias Björnmalm und Frank Caruso*

Robustheit, Reproduzierbarkeit, Glaubwürdigkeit, Transparenz und „Open Science“ werden immer mehr als elementar für wirkungsvolle Forschung anerkannt, und sie sind eng mit unserem Denken über Wissenschaft sowie unserer Art, zu forschen und unsere Ergebnisse zu beschreiben, verknüpft. Verbesserungen hier könnten helfen, Forschung wirkungsvoller zu machen. Die Chemie kann auf eine stolze Geschichte des Eintretens für Transparenz und Rigorosität zurückblicken. Beispiele sind früh geschaffene offene Depots wie die Cambridge Structural Database ebenso wie schon lange geltende Traditionen und Richtlinien (formell und informell) dafür, wie die Entdeckung, Synthese und Charakterisierung neuer Verbindungen beschrieben wird. Doch obwohl die Chemie im Allgemeinen als „zentrale Wissenschaft“ bezeichnet wird, stammen jüngere Bemühungen um eine größere Robustheit der Forschung aus Nachbargebieten wie der Biologie, Medizin und Psychologie.^[1] Zu diesen Bemühungen zählen neue und bessere Wege, Daten und Methoden zu teilen.

Der Einbau dieser neuen Konzepte in die Chemie führt zu einer „robusten Chemie“. Neuere Beispiele sind die Entwicklung von optimalen Verfahren des Veröffentlichens ebenso wie der zunehmende Einsatz von Videos, um wichtige Experimente und Verfahren zu zeigen. In einem vor kurzem in der *Angewandten Chemie* erschienenen Editorial setzen sich Bergman und Danheiser mit den Themen unbewusste Befangenheit von Forschern und Reproduzierbarkeit von Ergebnissen auseinander, und wir haben uns kürzlich Gedanken über eine wirkungsvollere Forschung in den Bio-Nanowissenschaften gemacht.^[2] Im Folgenden beschreiben wir Fortschritte und Neuerungen, die zu einer höheren Robustheit in der Chemie durch das Teilen von Daten und Methoden führen sollen.

Auch wenn vielen von uns bewusst ist, dass sich die Möglichkeiten, Informationen zu handhaben und zu speichern, seit Jahren exponentiell verbessern, werden die Chancen, die



M. Björnmalm



F. Caruso

solche Verbesserungen bieten, oft nicht richtig gewürdigt. Mehr als ein Jahrhundert war das Veröffentlichende in wissenschaftlichen Zeitschriften der Weg, Ergebnisse zu verbreiten. Die Kunst, Ergebnisse – oft aus Forschung, die sich über Monate oder gar Jahre erstreckt hatte – so knapp und klar zu beschreiben, dass sie auf wenigen Zeitschriftenseiten Platz finden, war ein Markenzeichen der Wissenschaft. Auch wenn heute viele Wissenschaftler PDFs statt der gedruckten Versionen lesen, das Prinzip ist das gleiche: Das Veröffentlichende in Zeitschriften war entscheidend dafür, dass die Wissenschaft das geworden ist, was sie heute ist, denn es ermöglichte Forschern, auf dem Laufenden zu bleiben und auf den Erkenntnissen von Kollegen auf der ganzen Welt aufzubauen. Die Bedeutung des Zugangs zu einer weltweiten Forschergemeinschaft und der daraus zu gewinnenden Inspiration und Orientierung kann nicht hoch genug eingeschätzt werden. Wie Newton sagte, „If I have seen further, it is by standing on the shoulders of giants.“

Eine Veröffentlichung in einer Zeitschrift wird als ein Hauptresultat von Forschungsprojekten angesehen, was als unbeabsichtigte Folge hat, dass alles, was nicht ins Manuskript kommt, oft gar nicht veröffentlicht wird; dazu gehören 1) die Rohdaten, 2) Experimente, die sich als nebensächlich herausstellten oder „nicht funktionierten“ und 3) implizites Wissen (z.B. schwierige experimentelle Schritte, die sich schlecht in einen kurzen Text bündeln lassen). Da das Projekt oft als „abgeschlossen“ angesehen wird, gibt es wenig Motivation, diese Informationen zu teilen. Grund dafür ist kein böser Wille, sondern z. B., dass Forscher sehr beschäftigt sind oder dass einige möglicherweise die Vorteile solcher Daten nicht zu schätzen wissen. Solche Informationen sind dann für die Forschergemeinschaft verloren und ebenso künftige Gelegenheiten, „auf den Schultern von Riesen zu stehen“.

H heute ist es ganz einfach, sich bei einem Online-Speicherdienst (z.B. Dropbox, Google Drive und OneDrive) anzumelden, und die Kosten liegen meist bei weniger als einem Cent pro Gigabyte und Monat (oder der Dienst ist sogar kostenlos!). Zudem gibt es kostenlose Datenrepositorien bei vielen Einrichtungen und in öffentlicher Form. Das bedeutet, es ist heute einfacher und billiger als je zuvor, Daten zu speichern und zu teilen. Damit wird das Teilen von Ergebnissen viel breiter möglich als auf den derzeit genutzten Wegen. Das nicht optimale Teilen von Daten ist aus zwei Gründen von Nachteil.

[*] 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 (Australien)
 E-Mail: fcaruso@unimelb.edu.au

Dr. M. Björnmalm
 Department of Materials, Imperial College London (Großbritannien)

ID Die Identifikationsnummern (ORCID) der Autoren sind unter <https://doi.org/10.1002/anie.201710493> zu finden.

Der eine wird durch das sich entwickelnde Gebiet „Big Data“ illustriert, das gezeigt hat, dass große Datenmengen (z. B. aus mehreren zusammengefassten Studien) dazu dienen können, Muster, Trends und Verbindungen aufzudecken, die bei kleineren Datensätzen (z. B. aus einer Einzelstudie) nicht entdeckt werden können. Um komplexere Fragen, z. B. bei interdisziplinären Themen, zu behandeln, sind diese integrativen Ansätze nicht nur eine „nette Zugabe“, sondern können entscheidend für ein besseres Verständnis und weitere Entdeckungen sein.

Der zweite Nachteil ist die Existenz „dunkler Daten“, das sind Daten, die in gut ausgeführten Studien ermittelt, aber nicht mit der Forschungsgemeinschaft geteilt wurden. Dazu gehören Daten aus Experimenten, die es nicht ins Manuskript schafften, ebenso wie Daten aus nicht abgeschlossenen Studien. Idealerweise sollten alle Daten aus gut geplanten und ausgeführten Experimenten verbreitet werden. Damit würden nicht nur die oben erwähnten integrativen Ansätze erleichtert, sondern die Forschungsgemeinschaft könnte auch die Straßen sehen, auf denen man unterwegs war, und nicht nur das erreichte Ziel. Stellen Sie sich vor, Sie würden alle Forschungsprojekte und Experimente vergessen, von denen Sie wissen, dass sie nicht veröffentlicht wurden. Wie viel schwieriger wäre es für Sie zu entscheiden, was als Nächstes zu tun ist? Und wie wahrscheinlich würden Sie etwas ausprobieren, das Sie schon einmal versucht haben (aber nicht in einer Zeitschrift veröffentlicht hatten)? Ein weiterer Vorteil eines Zugangs zu dunklen Daten ist vorstellbar: Sie könnten Wissenschaftler inspirieren; so könnte bei Experimenten, die nicht „klappten“, ein anderer Ansatz oder eine Optimierung neue Einsichten und Forschungsrichtungen liefern. Das Erkunden ist eine wesentliche Grundlage jeder Forschung, und zu wissen, was bereits versucht wurde und mit welchen Ergebnissen, hilft nicht nur, unnötige Wiederholungen zu vermeiden, sondern weist auch künftigen wissenschaftlichen Arbeiten den Weg.

Es gibt bereits exzellente Initiativen und Rahmenpläne (z. B. von Forschern, Fördereinrichtungen und Verlagen), um den Einsatz optimaler Verfahren zum Teilen von Daten zu erleichtern (siehe z. B. <https://opennessinitiative.org/making-your-data-public/>). Diese Verfahren helfen, die Transparenz und Verlässlichkeit von beschriebenen Daten zu verbessern. Ihr vermehrter Einsatz würde sicherlich dazu beitragen, die Forschung zu beschleunigen und die Robustheit in der Chemie zu erhöhen.

Das Teilen von Methoden ähnelt dem Teilen von Daten, unterscheidet sich aber auch davon. Beide können erheblich von unserer zunehmenden Fähigkeit, große Informationsmengen zu speichern und zu teilen, profitieren, aber auf unterschiedliche Art. Bei den Daten geht es letztlich vor allem darum, Roh- und dunkle Daten zu teilen, wofür – wie oben diskutiert – öffentliche Repositorien (fachspezifische oder allgemeine) und optimale Verfahren des Datenteilens weitgehend ausreichen. Bei Methoden dagegen ist die Herausforderung, „Erfahrung“ und anderes implizites Wissen zu teilen. Ausführliche und genaue Beschreibungen der Methoden in einer Veröffentlichung sind wichtig, aber nicht aus-

reichend. Viel chemisches Wissen lässt sich mit Worten kaum knapp und genau wiedergeben. Hier ein einfaches Beispiel: Als einer von uns vor kurzem einen Knopf am Hemd annähen wollte, suchte er nicht nach einer schriftlichen Anleitung, sondern ging auf YouTube und sah sich eine kurze Schritt-für-Schritt-Anleitung an. Da viele Verfahren in der Chemie mit Sicherheit schwieriger sind als das Annähen eines Knopfes, kann man sich ausmalen, wie nützlich es wäre, wenn wir einfach und routinemäßig auf Fotos und Video-Anleitungen zugreifen könnten, die die Schritte einer Synthese erklären oder zeigen, wie man ein Instrument bei einem bestimmten Experiment am besten einsetzt.

Dies kann beispielsweise mithilfe von Action-Kameras wie den GoPro-Kameras erreicht werden.^[3] Da diese (wie eine Stirnlampe) die Hände frei lassen, kann man völlig unbelastet alle Schritte eines Experiments als Ich-Erzähler aufzeichnen. Jemand, der das Video anschaut, würde quasi von einem Experten durch alle Schritte geleitet, was unserer Meinung nach dem heutigen „Goldstandard“ für das Erwerben von Fachkenntnissen sehr nahe kommt! Bevor man ein neues Instrument einsetzt oder eine noch nie durchgeführte Synthese anpackt, würde im Idealfall ein Experte einem die beteiligten Schritte zeigen. Mit Action-Kameras (die immer billiger und verfügbarer werden) können diese bei einem bestimmten Instrument oder Verfahren oder bei einer bestimmten Synthese dem Interessenten alle Schritte zeigen. Zudem könnte der Interessent das Video beliebig oft anhalten, auf ihm zurückgehen, und es so oft ansehen, wie er möchte. Derartige Videos werden bereits von Videozeitschriften wie JoVE (siehe <https://www.jove.com>) verbreitet, und bei anderen Zeitschriften können Videos oft als Hintergrundinformationen eingereicht werden. Das Ziel ist nicht, das eingeführte direkte Training („Goldstandard“) zu ersetzen, sondern, es zu ergänzen. Der verbreitete Einsatz emotional ansprechender Videos zur Unterhaltung, Bildung und Information ist integraler Teil unserer heutigen Gesellschaft. Wir sind uns sicher, dass diese Arten von Videos, breit eingesetzt, einen tiefgreifenden, positiven Einfluss auf die Entwicklung der Forschung haben können.

Die hier zentral behandelten Punkte (das Teilen von Daten und Methoden) sind komplementäre Verfahren, die bei breitem Einsatz zu einer deutlichen Änderung der Art führen könnten, wie wir Forschung betreiben und beschreiben, vor allem in Kombination mit neuen und besseren Hilfsmitteln (z. B. spezialisierten Suchmaschinen), um solche wertvollen Informationen effektiv zu nutzen. Ja, dafür braucht es eine Änderung unserer Kultur, was nicht einfach ist, aber der mögliche Nutzen wäre enorm. Dazu gehören das Beschleunigen und Rationalisieren der gegenwärtigen Forschung sowie die Möglichkeit, ganz neue Forschungsrichtungen aufzutun.

[1] M. R. Munafò et al., *Nat. Hum. Behav.* **2017**, *1*, 21.

[2] M. Björnmalin, M. Faria, F. Caruso, *J. Am. Chem. Soc.* **2016**, *138*, 13449.

[3] M. Björnmalin, M. Faria, F. Caruso, *Chem. Mater.* **2016**, *28*, 8441.