Title

RRIDs: A Simple Step toward Improving Reproducibility through Rigor and Transparency of Experimental Methods

Journal

Neuron, 90(3)

ISSN

0896-6273

Authors

Bandrowski, Anita E
Martone, Maryann E

Publication Date

2016-05-01

DOI

10.1016/j.neuron.2016.04.030

Peer reviewed

HHS Public Access

Author manuscript

Neuron. Author manuscript; available in PMC 2018 March 15.

Published in final edited form as:

Neuron. 2016 May 04; 90(3): 434–436. doi:10.1016/j.neuron.2016.04.030.

RRIDs: A Simple Step toward Improving Reproducibility through Rigor and Transparency of Experimental Methods

Anita E. Bandrowski[1,*] and **Maryann E. Martone**[1]

[1]Center for Research in Biological Systems, UCSD, La Jolla, CA 92093, USA

Abstract

With the call for more rigorous scientific reporting, authentication, and transparency from the scientific community and funding agencies, one critical step is to make finding and identifying key resources in the published literature tractable. We discuss here the use of Research Resource Identifiers (RRIDs) as one tool to help resolve this tricky problem in reproducibility.

Reproducibility in science and the scientific literature has been a hot topic these last few years, especially with Begley's explosive report showing that Amgen, with all of its resources, could not reproduce many of the top cancer studies (Begley and Ellis, 2012). The heads of NIH have called on scientists to improve reporting and publishing preclinical work (Landis et al., 2012; Collins and Tabak, 2014; see also Nature, 2016). Recently, the NIH Office of the Director implemented new guidelines that affect most individual awards and asks grantees to account for their experimental rigor and transparency, particularly with respect to authentication of key biological resources including (but not limited to) cell lines, specialty chemicals, antibodies, and other biologics (see http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-004.html, http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-011.html, and http://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-012.html). These key biological resources are integral to research, are a major source of experimental variability, and may differ from lab to lab or over time. According to these criteria, one could make the case that genetically modified organisms should also be included. With this call for change in scientific reporting and authentication, it is an opportune time to reflect on practices of science and science publication and rethink our own experimental and publishing procedures to make them serve rigor and transparency more effectively. In this NeuroView, we highlight a practical approach our working group, the Resource Identification Initiative, developed and recommends to help improve reproducibility: the Research Resource Identifiers (RRIDs).

What Is an Authenticated Key Biological Resource?

The NIH requires that authentication of key biological reagents be explained, but offers little in terms of concrete guidelines. Thankfully, for many of these resources community groups already exist that take care of authentication and validation. For organisms, this is a function that is handled by the model organism databases or stock centers that form the nuclei of

*Correspondence: abandrowski@ucsd.edu.

those communities. For antibodies, multiple groups have come together to create or adapt older standards to the new technologies now widely available.

However, one simple step should also become common practice when publishing results: research resources used to produce the findings of a study should be identified unambiguously and in a manner that allows search engines to easily retrieve evidence of their use. It is rather shocking that in the year 2016, in the midst of the information revolution, scientists cannot ask two very simple questions of the scientific literature: What resources were used to produce the results of this study, and how many other papers were published using this same resource?

The editors of *Neuron* have taken this on as a cause and joined the Research Resource Identification Initiative (Bandrowski et al., 2016) as a practical and easy method of improving research reproducibility by not just removing outdated practices, but replacing them with the current best practices for identifying and tracking research resources used in studies.

Current Problems with Citing Research Resources

Currently, the standards for citing research resources, such as they exist, are outdated and do little to support search. For example, the practice of citing materials by the company name and address with little to no other information is so common that everyone recognizes statements like this: “antibody against Dectin-1 (BioLegend, San Diego, CA, USA),” used in a 2016 publication. This style is how we are taught to cite reagents by our mentors in graduate school, and while it is no longer present in most instructions to authors, it continues to persist in papers. Why is this not sufficient? First of all, physical location of a vendor is probably the most irrelevant information about the material used. This information may have been useful when one needed to dial directory assistance to contact a vendor, but this is clearly an outdated vestige of the past in this age of globalization and e-commerce.

But the main problem with this commonly used citation style is the lack of sufficient identifying information about what antibody was used. In this particular case, which unfortunately is fairly typical (Vasilevsky et al., 2013), the antibody description can point to multiple products at the manufacturer. There are four products that currently fit the description at the manufacturer’s web site (http://www.biolegend.com/index.php?page=pro_sub_cat&action=search&criteria=Dectin-1+), and if you visit next year, it may be a different set. Although we can always contact our colleagues for more information, as the years pass and the cohort of scientists in the lab changes, reliance on this type of informally shared memory becomes more and more difficult and tenuous. Moreover, by not presenting identifying information for reagents in the paper, we lose the capability to use modern information technology to find other studies that use the same reagent.

Thus, like evolutionary leftovers such as human tail bones and goose bumps, an outdated practice for citing resources remains ingrained in the scientific literature.

Organisms Have Nomenclature Authorities; Surely We Do Not Have Problems with Those!

Unfortunately, outdated and unclear standards are not restricted to antibodies alone. Model organisms, a key tool in translational research, suffer from the same problems. It's not that organisms do not have proper identification or authoritative community repositories. What becomes difficult with many identification tags is that proper nomenclature runs against a core human impulse to shorten unnecessarily long names, e.g., Edward or Edmund, into shorter ones, Ed or Eddie. For transgenic organisms, the proper name is often an unruly and unpronounceable string of letters and numbers such as w[1118]; P{y[+t7.7] w[+mC] = GMR32A03-GAL4}attP2 that screams "shorten me!" Satisfying this normal human desire wreaks havoc in the published literature, and unfortunately, using the proper name presents issues as well because search engines have particular difficulties with special characters.

This issue becomes clear when looking at a real example in the published literature. While this example comes from published papers, to keep our focus on the citation of the resource we'll refer to the papers by their publication year so as not to call out a particular paper or author, as this is not an uncommon situation and is an all too familiar scenario that all scientists have encountered when reading the literature. In our example, in a paper published in 2016, the authors cite a mouse line as "J20 C57BL/6 $^{APPK670N/M671L,V717F}$" and refer to a paper published by a different author in 2000 as the source for the line. Delving into this paper published in 2000, the manuscript describes several APP mice, but none of them are named exactly "APP$^{K670N/M671L,V717F}$." So what mouse line is it? Unfortunately, using this nomenclature to search two highly authoritative mouse websites, MGI and JAX, produces zero results. Therefore, the citation itself doesn't bring clarity and requires those interested in this information to conduct extensive digging and likely directly query the authors to answer the simple question of which mouse was used in the study. This situation is unfortunately quite common, as described in Vasilevsky et al. (2013), across biomedicine and adds to the difficulty in reproducing studies for a reason that should be easily fixable (see also Chawla, 2015).

One Practical Solution: Research Resource Identifiers

With its roots in addressing this fundamental problem in citing resources literature, the Resource Identification Initiative introduced a simple convention for citing key research resources in 2014, building on extensive interaction with different stakeholders, including publishers, editors, curators, commercial resource suppliers, government representatives, and, most importantly, the scientific community. Our practical solution asks authors to provide more complete metadata as well as an RRID: a citation convention that provides a simple prefix, RRID, pre-pended to an alpha numeric string. These strings come from community databases that have been aggregating information for many years. Every time scientists register a new entity, e.g., a new antibody, it gets its own "social security number" in the form of an accession number. Why create a number? Machines run on numbers (Bourne et al., 2015), as we experience daily when dialing a string of numbers into a telephone to be connected to the right person anywhere in the world. We liken it to the DNS

server system registering unique names across the entire world to particular numbered servers, e.g., the 130.14.29.110 server numeric address resolves to NCBI (home of PubMed). Research resources also have a set of numbers that can point to each individual resource.

After our successful pilot (Bandrowski et al., 2016), many additional journals have adopted use of RRIDs, and *Neuron* has joined this effort by changing their instructions to authors and requesting inclusion of RRIDs in their publications. *Neuron* authors are now asked to follow resource citation guidelines (see *Neuron* RRID guidelines, http://www.cell.com/neuron/rrid) such that a resource citation would be reported as follows: BioLegend, cat# 101230, RRID: AB_2129374 (vendor, vendor ID, machine-readable ID).

Obtaining an RRID is fairly simple. The Resource Identification Portal community hosts a website (https://scicrunch.org/resources) that displays the proper citations for common research resources. From this website, authors are able to find RRIDs by (1) searching for the catalog number of their research resources, (2) clicking on the “cite this” button, and (3) copying and pasting the resulting text into their methods section. For resources that are generated de novo and available as a stable stock, authors are asked to deposit the organism with a stock center or register the information about it with the proper authority appropriate for the organism currently available in the Resource Identification Portal (https://scicrunch.org/resources/about/guidelines#organism). These model organism authorities will give authors the proper short name for the new resource that can then be used to construct its RRIDs.

Why Not Just Use a Catalog Number or the Company Website?

The RRID syntax is superior to catalog numbers in many ways. First, it is unique and stable. Individual products that a company sells may change over time, and many companies do not keep information about older products, may change the names of products, or go out of business, while the citation in the article does not ever change. As a result, in just a few years many catalog numbers referenced in an article may point to products that are no longer available or may have shifted to being offered by another company under a new URL and catalog number. In fact, at publication 8% of URLs are already broken, and in 5 years over a quarter of the URLs used in papers are expected to break (Ozyurt et al., 2016).

Indeed, one antibody (RRID: AB_2178887) tracked by its originating laboratory was found to have passed through various vendors, including Boehringer-Mannheim, Roche, and Chemicon, in a 15 year period (Slotta et al., 2014), and the process of company name changes shows no sign of slowing. For example, with the next round of merges in the biotech and pharmaceutical industry, an antibody originally offered by Chemicon was passed to Millipore, then to EMB, Merck, and soon Sigma. Researchers keep and use antibodies for many years, and referencing an antibody by a company name like Chemicon is accurate to the particular stock of the antibody but clearly not sufficient for another scientist to obtain the antibody without understanding the current mergers and acquisitions landscape. Fortunately, where it is possible, the Antibody Registry (http://antibodyregistry.org/) keeps track of antibodies’ past locations and current whereabouts,

including information about whether it is discontinued. It also aggregates the data (where it is known) if the same antibody is sold by multiple vendors. However, this is no easy task when business practices allow companies to resell each other's products with different catalog numbers (Baker, 2015a).

The volatility of the commercial market is one reason why groups that are coming together to set standards about what constitutes a validated antibody are recommending that in addition to specifying information that will need to be included in the material data sheet and what sorts of experiments need to be done to verify that an antibody is working within a researcher's lab (IWGAV and GBSI; Baker 2015b; Alm et al., 2016), each antibody should be registered with the Antibody Registry and obtain a stable identifier.

Simple Common-Sense Steps Help Authors Avoid Reproducibility Pitfalls

The RRID syntax, by virtue of its presence in the materials and methods section, denotes substantive use of the research resource (e.g., https://scholar.google.com/scholar?q=RRID). As this gains broader acceptance, the question of which mice were used in the paper or who else used the antibody will be easier to ask with a simple search using an RRID. In fact, proper use of the RRID already allows Elsevier to display the antibody information on a side panel (see Elsevier Antibody App, https://www.elsevier.com/books-and-journals/content-innovation/antibody-data) and link to other papers that used the same product, making the process of tracking down information about the exact antibody used in a study substantially simpler. In this way, use of RRIDs can directly benefit researchers who are seeking more information about what reagents to use and how they perform across systems by providing ready access to the information they need.

The issue of reproducibility in science is a tricky one that is sure to continue to be debated and discussed in the coming years. We hope, however, that at least being clear about what research resources were used to produce the findings of the study is not controversial. The use of numerical identifiers and the RRID syntax provides a simple and search-friendly way to comply with these new directives.

Acknowledgments

CONFLICTS OF INTEREST

Drs. Bandrowski and Martone have founded SciCrunch Inc, which underpins the technical infrastructure necessary to run the Resource Identification Initiative, Neuroscience Information Framework, and other projects.

References

Alm T, Lundberg E, Uhlén M. EuPA Open Proteom. 2016; 10:56–58. http://dx.doi.org/10.1016/j.euprot.2016.01.002.

Baker M. Nature. 2015a; 521:274–276. http://dx.doi.org/10.1038/521274a. [PubMed: 25993940]

Baker M. Nature. 2015b; 527:545–551. http://dx.doi.org/10.1038/527545a. [PubMed: 26607547]

Bandrowski A, Brush M, Grethe JS, Haendel MA, Kennedy DN, Hill S, Hof PR, Martone ME, Pols M, Tan SC, et al. J Comp Neurol. 2016; 524:8–22. http://dx.doi.org/10.1002/cne.23913. [PubMed: 26599696]

Begley CG, Ellis LM. Nature. 2012; 483:531–533. http://dx.doi.org/10.1038/483531a. [PubMed: 22460880]

Bourne PE, Lorsch JR, Green ED. Nature. 2015; 527:S16–S17. http://dx.doi.org/10.1038/527S16a. [PubMed: 26536219]

Chawla DS. Nature. 2015; 29 http://dx.doi.org/10.1038/nature.2015.17652.

Collins FS, Tabak LA. Nature. 2014; 505:612–613. [PubMed: 24482835]

Editorial. Nature. 2016; 529:256. [PubMed: 26791685]

Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, Crystal RG, Darnell RB, Ferrante RJ, Fillit H, et al. Nature. 2012; 490:187–191. [PubMed: 23060188]

Ozyurt IB, Grethe JS, Martone ME, Bandrowski AE. PLoS ONE. 2016; 11:e0146300. http://dx.doi.org/10.1371/journal.pone.0146300. [PubMed: 26730820]

Slotta C, Müller J, Tran L, Hauser S, Widera D, Kaltschmidt B, Kaltschmidt C. J Histochem Cytochem. 2014; 62:157–161. Published online November 19, 2013. http://dx.doi.org/10.1369/0022155413515484. [PubMed: 24255050]

Vasilevsky NA, Brush MH, Paddock H, Ponting L, Tripathy SJ, Larocca GM, Haendel MA. PeerJ. 2013; 1:e148. http://dx.doi.org/10.7717/peerj.148. [PubMed: 24032093]