Experimental Neurology 295 (2017) 135–143

Contents lists available at ScienceDirect

Experimental Neurology

journal homepage: www.elsevier.com/locate/yexnr

Review Article

Developing a data sharing community for spinal cord injury research

Alison Callahan [a], Kim D. Anderson [b,c], Michael S. Beattie [d], John L. Bixby [b,e,f], Adam R. Ferguson [d,g], Karim Fouad [h], Lyn B. Jakeman [i], Jessica L. Nielson [d,g], Phillip G. Popovich [j,k], Jan M. Schwab [k,l,m], Vance P. Lemmon [b,e,*], the FAIR Share Workshop Participants [1]

[a] Stanford Center for Biomedical Informatics Research, Stanford University, Stanford 94305, CA, USA
[b] Miami Project to Cure Paralysis, University of Miami School of Medicine, Miami 33136, FL, USA
[c] Department of Neurological Surgery, University of Miami Miller School of Medicine, Miami 33136, FL, USA
[d] UCSF Brain and Spinal Injury Center, University of California, San Francisco 94110, CA, USA
[e] Center for Computational Science, University of Miami, Coral Gables 33146, FL, USA
[f] Department of Cellular and Molecular Pharmacology, University of Miami School of Medicine, Miami 33136, FL, USA
[g] San Francisco VA Medical Center, San Francisco 94121, CA, USA
[h] Department of Physical Therapy, Neuroscience and Mental Health Institute, University of Alberta, Edmonton T6G2G4, Alberta, Canada
[i] National Institute of Neurological Disorders and Stroke, The National Institutes of Health, Rockville 20852, MD, USA
[j] Center for Brain and Spinal Cord Repair, The Neurological Institute, Ohio State University Wexner Medical Center, Columbus 43210, OH, USA
[k] Department of Neuroscience, The Neurological Institute, Ohio State University Wexner Medical Center, Columbus 43210, OH, USA
[l] Department of Neurology, The Neurological Institute, Ohio State University Wexner Medical Center, Columbus 43210, OH, USA
[m] Department of Physical Medicine and Rehabilitation, The Neurological Institute, Ohio State University Wexner Medical Center, Columbus 43210, OH, USA

ARTICLE INFO

Article history:
Received 14 April 2017
Received in revised form 24 May 2017
Accepted 29 May 2017
Available online 30 May 2017

Keywords:
FAIR data principles
Reproducibility
Neuroscience
Informatics
Workshop proceedings
Open Data Commons

ABSTRACT

The rapid growth in data sharing presents new opportunities across the spectrum of biomedical research. Global efforts are underway to develop practical guidance for implementation of data sharing and open data resources. These include the recent recommendation of 'FAIR Data Principles', which assert that if data is to have broad scientific value, then digital representations of that data should be Findable, Accessible, Interoperable and Reusable (FAIR). The spinal cord injury (SCI) research field has a long history of collaborative initiatives that include sharing of preclinical research models and outcome measures. In addition, new tools and resources are being developed by the SCI research community to enhance opportunities for data sharing and access. With this in mind, the National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health (NIH) hosted a workshop on October 5–6, 2016 in Bethesda, MD, in collaboration with the Open Data Commons for Spinal Cord Injury (ODC-SCI) titled "Preclinical SCI Data: Creating a FAIR Share Community". Workshop invitees were nominated by the workshop steering committee (co-chairs: ARF and VPL; members: AC, KDA, MSB, KF, LBJ, PGP, JMS), to bring together junior and senior level experts including preclinical and basic SCI researchers from academia and industry, data science and bioinformatics experts, investigators with expertise in other neurological disease fields, clinical researchers, members of the SCI community, and program staff representing federal and private funding agencies. The workshop and ODC-SCI efforts were sponsored by the International Spinal Research Trust (ISRT), the Rick Hansen Institute, Wings for Life, the Craig H. Neilsen Foundation and NINDS. The number of attendees was limited to ensure active participation and feedback in small groups. The goals were to examine the current landscape for data sharing in SCI research and provide a path to its future. Below are highlights from the workshop, including perspectives on the value of data sharing in SCI research, workshop participant perspectives and concerns, descriptions of existing resources and actionable directions for further engaging the SCI research community in a model that may be applicable to many other areas of neuroscience. This manuscript is intended to share these initial findings with the broader research community, and to provide talking points for continued feedback from the SCI field, as it continues to move forward in the age of data sharing.

Abbreviations: AES, American Epilepsy Society; ASIA, American Spinal Injury Association; CDE, common data element; BD2K, Big Data To Knowledge; FORE-SCI, Facilities of Research Excellence in Spinal Cord Injury; FAIR, Findable, Accessible, Interoperable, Reusable; FITBIR, Federal Interagency Traumatic Brain Injury Research; fMRI, functional magnetic resonance imaging; FORCE 11, Future of Research Communications and e-Scholarship; ISI, Institute for Scientific Information; ILAE, International League Against Epilepsy; ISCoS, International Spinal Cord Society; ISRT, International Spinal Research Trust; MIASCI, Minimum Information About a Spinal Cord Injury; MASCIS, Multicenter Animal Spinal Cord Injury Study; NINDS, National Institute of Neurological Disorders and Stroke; ODC-SCI, Open Data Commons for Spinal Cord Injury; PB, petabyte; PCR, polymerase chain reaction; RCT, randomized controlled trial; RDF, resource description framework; RNA, ribonucleic acid; SCI, spinal cord injury; SfN, Society for Neuroscience; SOP, standard operating procedure.

* Corresponding author.
E-mail addresses: acallaha@stanford.edu (A. Callahan), VLemmon@med.miami.edu (V.P. Lemmon).

[1] See Appendix A.

http://dx.doi.org/10.1016/j.expneurol.2017.05.012
0014-4886/

Contents

1. The culture of data sharing

Neuroscientists, including SCI researchers, have a long history of sharing data, traditionally through publications. The Institute for Scientific Information (ISI) Science Citation Index has over 46,000 publications indexed under 'spinal cord injury' from 2000 to 2016, many of which include detailed methods, results, and supplementary data that are used by other investigators in planning experiments and interpreting their own findings. Data shared in publications, however, is usually carefully selected, and represent only a fraction of the data generated by preclinical SCI researchers. Data that do not fit the 'story' of a discovery are often left unpublished, and most primary preclinical research data are accessible and interpretable only by individuals in a shared laboratory or collaborative group. These 'dark data', never made available in repositories or publications, are estimated to make up 85% of all data collected (Ferguson et al., 2014). The inability to access dark data impedes efforts to promote transparency, replication and independent validation of promising findings (Ferguson et al., 2014). Moreover, for the 15% of data that are reported in the scientific literature, inconsistent study design and statistical analysis contribute to complications and bias in interpretations (Burke et al., 2013; Watzlawick et al., 2014).

Informal data sharing occurs at meetings and symposia, where preliminary findings are presented and discussed with colleagues. At the 2016 Society for Neuroscience (SfN) meeting, for example, 2256 presentations had the words 'spinal cord injury' associated with them. Only a subset of these posters and presentations will end up as publications. The informal interchange of ideas, technical approaches, and importantly, knowledge about what experiments are being done in other labs, is therefore highly valuable to the community. However, even at conferences, presenters are often careful to provide only select information to their peers. Many of us remember being admonished as students for enthusiastic sharing of not-yet-ready lab data at conferences and meetings. The free exchange of data and ideas *versus* 'saving' data for curated, peer-reviewed publications in high impact journals are competing interests in the current research landscape, in part responsible for a cultural bias against open data sharing.

In the current era of accountability and transparency, each community must consider how best to share data and seize opportunities afforded by making experimental data more widely available. The culture of sharing pre-publication findings in physics and genomics and the rapid and fruitful evolution of approaches for managing and analyzing big data in scientific research have driven discoveries in these fields. Sharing data necessitates that others can examine entire datasets from which interpretations were made. This can be seen as a challenge to the integrity of the traditional process of neuroscience research, yet it is the most transparent and useful approach to finding the 'truth'. Recently, much attention has been paid to open data sharing as a means to increase rigor and reproducibility in neuroscience research (Ferguson et al., 2014). Effective data sharing practices can be leveraged to improve reproducibility by providing platforms for depositing published and unpublished data, enabling better meta-analyses of research studies, reducing redundancy and waste, and providing large scale resources for analytic approaches to generate new discoveries.

As a consequence, the entire biomedical research enterprise is experiencing a cultural shift in approaches to data collection and data sharing. This shift has been particularly evident in the preclinical research spectrum. In 2011, a meeting of international leaders in data science known as "The Future of Research Communications and e-Scholarship", or FORCE 11, took on the task of creating standard recommendations for data sharing. One product of this effort was the development of "FAIR Data Principles", which describe digital objects that hold value as those that are Findable (with sufficient explicit metadata), Accessible (open and available to other researchers), Interoperable (using standard definitions and common data elements (CDEs)), and Reusable (meeting community standards, and sufficiently documented). The Office of Data Science at NIH has endorsed the FAIR Data Principles, and plans to incorporate these standards in future data sharing recommendations and programs (Wilkinson et al., 2016).

The SCI research community is well-positioned to embark on fruitful data sharing practices and lead by example. Clinical SCI researchers have joined with the International Spinal Cord Society (ISCoS), the American Spinal Injury Association (ASIA) and NINDS to develop standard definitions, case report forms, and CDEs for collection and reporting of clinical research data (Biering-Sørensen et al., 2015; Charlifue et al., 2016). In addition, basic and preclinical SCI researchers have embarked on initiatives and developed resources for data sharing over the past three decades. In the 1990s, NINDS funded a Multicenter Animal Spinal Cord Injury Study (MASCIS) as a consortium to facilitate validation of promising preclinical leads. This led to development of standard models and data collection procedures across several laboratories (Basso et al., 1996, 1995; Young, 2002).

From 2003 to 2013, NINDS executed contract agreements as Facilities of Research Excellence in Spinal Cord Injury (FORE-SCI), which led to additional outcome measures in mice and rats (Aguilar and Steward, 2010; Anderson et al., 2009), established a research training course for investigators new to the field, and completed 18 controlled replication studies in order to identify leads for translation (Steward et al., 2012). The FORE-SCI investment enriched the field with a highly-trained workforce, highlighted the challenges in replication attempts, and contributed to a larger effort across the NIH to enhance transparency, rigor and data quality for all preclinical research (Landis et al., 2012).

Since 2013, four projects have added data resources and tools for the SCI preclinical research community: (1) the VISION-SCI data repository with source data contributed by multiple research laboratories (Nielson et al., 2015a, 2014), (2) a consensus guideline of minimal reporting expectations for preclinical SCI research (MIASCI) (Lemmon et al., 2014), (3) a knowledge base and ontology for integration of SCI research data that is compatible with domain wide terminology standards (RegenBase) (Callahan et al., 2016), and (4) a rapidly-developing open data commons for SCI research. Each of these efforts has been a product of wide collaboration with dozens of contributing SCI scientists and multiple authors and is described in more detail below.

Given the state of readiness of the SCI research community and the availability of these unique resources, NINDS hosted the FAIR Share Workshop to engage stakeholders in discussion of the new challenges and opportunities for data sharing (Fig. 1). The goals of the workshop were to (1) bring together researchers and data science experts with