DOIs for DICOM Raw Images: Enabling Science Reproducibility¹

Philip E. Bourne, PhD

Of concern to all stakeholders in the scholarly enterprise is the ability to reproduce scientific research to confirm their findings and/or move to the next level of inquiry. The apparent failure to reproduce published findings has received substantial attention of late and requires action. For example, a 2012 Amgen study (1) was unable to reproduce 47 of 53 studies described in “landmark” publications from top journals and reputable academic laboratories. That notion was further elaborated by publications in the scientific press (2) and by national leadership at the National Institutes of Health (NIH) (3), as well as the U.S. Institute of Medicine (4). One element of that concern has been attributed to the systemic unavailability of online data that form the basis of articles published in major journals (5). Some progress seems to have been made by routine use of supplementary data (maintained by the publisher or a third party) that is linked to the research article, and in some cases, such data availability is a prerequisite for publication (6).

A critical technical keystone of data accessibility is to identify the associated data sets and be sure they are resolvable (findable) online at the original source. The ability to do this already exists for the research article itself in the form of a digital object identifier (DOI) (7). DOIs originate from the International DOI Foundation and provide a means of uniquely registering online content. Thus, the DOI system provides a framework for persistent identification, managing intellectual content, and managing metadata. So far, most DOIs registered are for scholarly articles, but they clearly have a broader potential use. When applied to data, DOIs make their provenance trackable and citable and therefore allow interoperability with

existing reference services. An important side effect of using this identifier scheme for data is to potentially elevate data to a form of scholarship in its own right.

For what I believe is the first time for this journal, and perhaps for medical imaging at large, DOIs have been applied to Digital Imaging and Communications in Medicine (DICOM) raw image data sets. The DOIs linked to an article in this issue of *Radiology* (8) exemplify this new model for medical imaging. In this instance, DOIs permit an online reader to download Health Insurance Portability and Accountability Act–compliant patient DICOM magnetic resonance (MR) images in the 99 patients with glioblastoma in that study. Those MR images were the principal focus of the image-processing analysis that forms the basis of the investigators’ conclusions. Sharing these original images with other investigators through this DOI gives the wider science community an opportunity to rigorously test the article’s conclusions independently. Hence, and by example, this journal’s use of DOIs stands as a new model for clinical cancer imaging–genetics science investigation. These actions are consistent with NIH’s concern with science reliability through independent reproducibility and are solidly in alignment with its new Data Science emphasis (9), as well as with the NIH funding program that encourages “Big Data to Knowledge” (BD2K) (10). It speaks to the NIH as a digital enterprise—an enterprise where research objects (eg, data, software, narrative, published articles) exist in a digital environment and must be uniquely identified if their full value is to be unleashed.

By being a front-runner, the imaging community has everything to gain, because original DICOM raw data exposure to the wider science audience

Published online

10.1148/radiol.15150144 Content code:

Radiology 2015; 275:3–4

¹ From the Office of the Director, the National Institutes of Health, 1 Center Dr, Building 1, Room 228, Bethesda, MD 20892. Received January 19, 2015; revision requested January 19; revision received January 24; final version accepted January 26. Address correspondence to the author.

Conflicts of interest are listed at the end of this article.

See also the article by Colen et al in this issue.

© RSNA, 2015

is likely to speed standardized image acquisition as well as engender greater confidence in the clinical imaging literature. As the field of radiogenomics gathers momentum, the discipline of biomedical imaging is well placed to position itself as an exemplary investigative community solidly grounded in reproducible science. Other investigators planning to perform cross-disciplinary research in the NIH open-access, open-science resources offered by the Cancer Genome Atlas (11) and the Cancer Imaging Archive (12) (and further enriched by same-case clinical data and digitized pathologic images [13]) are similarly encouraged to construct such DOIs simply by following the method found on the Cancer Imaging Archive Web site (14).

Acknowledgment: This work benefited from communication with C. Carl Jaffe, MD.

Disclosures of Conflicts of Interest: P.E.B. disclosed no relevant relationships.

References

1. Begley CG, Ellis LM. Drug Development: raise standards for preclinical cancer research. *Nature* 2012 Mar 28;483(7391):531–533. PubMed doi: 10.1038/483531a.
2. Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2(8):e124.
3. <http://www.nature.com/news/policy-nih-plans-to-enhance-reproducibility-1.14586>. Accessed January 22, 2015.
4. <http://www.iom.edu/Activities/Research/SharingClinicalTrialData.aspx>. Accessed January 22, 2015.
5. Evangelou E, Trikalinos TA, Ioannidis JP. Unavailability of online supplementary scientific information from articles published in major journals. *FASEB J* 2005;19(14):1943–1944.
6. Data Access for the Open Access Literature: PLOS's Data Policy. <http://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/>. Accessed January 22, 2015.
7. Digital object identifier. Wikipedia. http://en.wikipedia.org/wiki/Digital_object_identifier. Accessed January 22, 2015.
8. Colen RR, Wang J, Singh SK, Gutman DA, Zinn PO. Glioblastoma: imaging genomic mapping reveals sex-specific oncogenic associations of cell death. *Radiology* 2015;275(1):215–227.
9. NIH Names Dr. Philip E. Bourne First Associate Director for Data Science. National Institutes of Health. <http://www.nih.gov/news/health/dec2013/od-09.htm>. Accessed January 22, 2015.
10. NIH Big Data to Knowledge (BD2K). <http://bd2k.nih.gov/#sthash.NyXUCU3o.aZn-Nolu5.dpbs>. Accessed January 22, 2015.
11. The Cancer Genome Atlas. National Cancer Institute. <http://cancergenome.nih.gov/>. Accessed January 22, 2015.
12. Cancer Imaging Archive. <http://www.cancerimagingarchive.net/>. Accessed January 22, 2015.
13. Cancer Digital Slide Archive. <http://cancer.digitalslidearchive.net/>. Accessed January 22, 2015.
14. TCIA Digital Object Identifiers. Cancer Imaging Archive. <https://wiki.cancerimagingarchive.net/display/DOI/TCIA+Digital+Object+Identifiers>. Accessed January 22, 2015.