



Report from 2015 OHBM Hackathon (HI)

Advancing Open Science through NiData

Project URL: <http://github.com/nidata/nidata>

Ben Cipollini^{1*} and A. Rokem²

1 Introduction

The goal of this project is to improve accessibility of open datasets by curating them. “NiData” aims to provide a common interface for documentation, downloads, and examples to all open neuroimaging datasets, making data usable for experts and non-experts alike.

2 Approach

Open datasets promise to allow more thorough analysis of hard-to-collect data and re-analysis using state-of-the-art analysis methods. However, open datasets are not truly open unless they are easy to find, simple to access, and have sufficient documentation for use. Currently, publicly available data in neuroscience are scattered across a number of websites and databases, without a common data format, no common method for data access, and varying levels of documentation. Datasets are being uploaded to public databases through a number of initiatives, including OpenfMRI [1] and NITRC [2]. In addition, there are funded efforts for collecting data explicitly for the purpose of public sharing – most visibly in the Human Connectome Project (HCP) [3] – but also in the Pediatric Imaging, Neurocognition and Genetics (PING) study [4]. There are a number of funded efforts to collect old data and re-release as public databases, notably the INDI [5] efforts (which include the popular ABIDE [6] and functional connectomes 1000 datasets [7]). The BRAIN initiative [8] aims to collect data that will be a challenge to store, let alone analyze. There are even online journals focused on publishing datasets (e.g. Nature Scientific Data), or with options to release data (e.g. F1000 “Data Notes”).

NiData is a Python package that provides a single interface accessing data from a variety of open data

sources. The software framework makes it easy to add new data sources, simple to define and to provide access to multiple datasets from a single data source. Software dependencies are managed on a per-dataset basis, allowing downloads and examples to use any public packages without requiring installation of packages required by unused datasets. The interface also allows selective download of data (by subject or type) and caches files locally, allowing easy management of big datasets.

3 Results

We focused on exposing new methods for downloading data from the HCP, supporting access via Amazon S3 and HTTP/XNAT [9]. We were able to provide a downloader that accepts login credentials and downloads files locally. We created an example that interacts with DIPY [10] to produce diffusion imaging results on a single subject from the HCP. We also worked at collecting common data sources, as well as individual datasets stored at each data source, into NiData’s “data sources” wiki page. We incorporated downloads, documentation, and examples from the Nilearn package and began discussion of making a more extensible object model.

Since the hackathon, we have created such an object model and migrated all code to use it, and a Sphinx-based website is under development. The current object model makes it easier to write general-purpose fetchers (e.g. HTTP, XNAT, Amazon S3) that can be extended to access specific databases (e.g. COINS [11], LORIS [12], ADNI [13]).

4 Conclusions

Projects like NiData improve curated data access and increases the effectivity of big data projects with open source data.

*Correspondence: bcipolli@ucsd.edu

¹University of California, San Diego, La Jolla, 9500 Gilman Dr, 92093, California, USA

Full list of author information is available at the end of the article

Availability of Supporting Data

More information about this project can be found at:

<http://github.com/nidata/nidata>. Further data and files supporting this project are hosted in the *GigaScience* repository doi:10.5524/100217.

Competing interests

None

Author's contributions

BC and AR wrote the software and the report.

Acknowledgements

The authors would like to thank the organizers and attendees of the 2015 OHBM Hackathon.

Author details

¹University of California, San Diego, La Jolla, 9500 Gilman Dr, 92093, California, USA. ²University of Washington, Seattle, 3910 15th Ave NE, 98195, Washington, USA.

References

- Poldrack, R.A., Barch, D.M., Mitchell, J.P., Wager, T.D., Wagner, A.D., Devlin, J.T., Cumba, C., Koyejo, O., Milham, M.P.: Toward open sharing of task-based fMRI data: the OpenfMRI project. *Front Neuroinform* **7**, 12 (2013)
- Buccigrossi, R., Ellisman, M., Grethe, J., Haselgrove, C., Kennedy, D.N., Martone, M., Preuss, N., Reynolds, K., Sullivan, M., Turner, J., Wagner, K.: The Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC). *AMIA Annu Symp Proc*, 1000 (2008)
- Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Van Essen, D., Barch, D., Corbetta, M., Goate, A., Heath, A., Larson-Prior, L., Marcus, D., Petersen, S., Prior, F., Province, M., Raichle, K., Schlaggar, B., Shimony, J., Snyder, A., Adeyemo, B., Archie, K., Babajani-Feremi, A., Bloom, N., Bryant, J.E., Burgess, G., Cler, E., Coalson, T., Curtiss, S., Danker, S., Denness, R., Dierker, D., Elam, J., Evans, T., Feldt, C., Fenlon, K., Footer, O., Glasser, M., Gordon, E., Gu, P., Guilday, C., Harms, M., Hartley, T., Harwell, J., Hileman, M., Hodge, M., Hood, L., Horton, W., House, M., Laumann, T., Lugo, M., Marion, S., Miezin, F., Nolan, D., Nolan, T., Power, J., Ramaratnam, M., Reid, E., Schindler, J., Schmitz, D., Schweiss, C., Serati, J., Taylor, B., Tobias, M., Wilson, T., Ugurbil, K., Garwood, M., Harel, N., Lenglet, C., Yacoub, E., Adriany, G., Auerbach, E., Moeller, S., Strupp, J., Smith, S., Behrens, T., Jenkinson, M., Johansen-Berg, H., Miller, K., Woolrich, M., Andersson, J., Duff, E., Hernandez, M., Jbabdi, S., Robinson, E., Salimi-Khorshidi, R., Sotiropoulos, S., Romani, G.L., Della Penna, S., Pizzella, V., de Pasquale, F., Di Pompeo, F., Marzetti, L., Perruci, G., Bucholz, R., Roskos, T., Kiser, T., Luo, Q.-J., Stout, J., Oostenveld, R., Beckmann, C., Schoffelen, J.M., Fries, P., Michalareas, G., Sapiro, G., Sporns, O., Nichols, T., Farber, G., Bjork, J., Blumensath, T., Chang, A., Chen, L., Feinberg, D., Kull, L., Wig, G., Xu, J.G., Basser, P., Bullmore, E., Evans, A., Gazzaniga, M., Glahn, D., Hawrylycz, M., Hennig, J., Parker, G., Poldrack, R., Salmelin, R.: The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79 (2013)
- Jernigan, T.L., Brown, T.T., Hagler, D.J., Akshoomoff, N., Bartsch, H., Newman, E., Thompson, W.K., Bloss, C.S., Murray, S.S., Schork, N., Kennedy, D.N., Kuperman, J.M., McCabe, C., Chung, Y., Libiger, O., Maddox, M., Casey, B.J., Chang, L., Ernst, T.M., Frazier, J.A., Gruen, J.R., Sowell, E.R., Kenet, T., Kaufmann, W.E., Mostofsky, S., Amaral, D.G., Dale, A.M.: The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository. *Neuroimage* **124**(Pt B), 1149–1154 (2016)
- Mennes, M., Biswal, B.B., Castellanos, F.X., Milham, M.P.: Making data sharing work: the FCP/INDI experience. *Neuroimage* **82**, 683–691 (2013)
- Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D.A., Gallagher, L., Kennedy, D.P., Keown, C.L., Keyser, C., Lainhart, J.E., Lord, C., Luna, B., Menon, V., Minshew, N.J., Monk, C.S., Mueller, S., Muller, R.A., Nebel, M.B., Nigg, J.T., O'Hearn, K., Pelphrey, K.A., Peltier, S.J., Rudie, J.D., Sunaert, S., Thioux, M., Tyszka, J.M., Uddin, L.Q., Verhoeven, J.S., Wenderoth, N., Wiggins, J.L., Mostofsky, S.H., Milham, M.P.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* **19**(6), 659–667 (2014)
- Biswal, B.B., Mennes, M., Zuo, X.N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., Dogonowski, A.M., Ernst, M., Fair, D., Hampson, M., Hoptman, M.J., Hyde, J.S., Kiviniemi, V.J., Kotter, R., Li, S.J., Lin, C.P., Lowe, M.J., Mackay, C., Madden, D.J., Madsen, K.H., Margulies, D.S., Mayberg, H.S., McMahon, K., Monk, C.S., Mostofsky, S.H., Nagel, B.J., Pekar, J.J., Peltier, S.J., Petersen, S.E., Riedl, V., Rombouts, S.A., Rypma, B., Schlaggar, B.L., Schmidt, S., Seidler, R.D., Siegle, G.J., Sorg, C., Teng, G.J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.F., Zhang, H.Y., Castellanos, F.X., Milham, M.P.: Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* **107**(10), 4734–4739 (2010)
- Insel, T.R., Landis, S.C., Collins, F.S.: Research priorities. The NIH BRAIN Initiative. *Science* **340**(6133), 687–688 (2013)
- Marcus, D.S., Olsen, T.R., Ramaratnam, M., Buckner, R.L.: The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* **5**(1), 11–34 (2007)
- Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M., Nimmo-Smith, I.: Dipy, a library for the analysis of diffusion MRI data. *Front Neuroinform* **8**, 8 (2014)
- Scott, A., Courtney, W., Wood, D., de la Garza, R., Lane, S., King, M., Wang, R., Roberts, J., Turner, J.A., Calhoun, V.D.: COINS: An Innovative Informatics and Neuroimaging Tool Suite Built for Large Heterogeneous Datasets. *Front Neuroinform* **5**, 33 (2011)
- Das, S., Zijdenbos, A.P., Harlap, J., Vins, D., Evans, A.C.: LORIS: a web-based data management system for multi-center studies. *Front Neuroinform* **5**, 37 (2011)
- Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., Dale, A.M., Felmlee, J.P., Gunter, J.L., Hill, D.L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C.S., Krueger, G., Ward, H.A., Metzger, G.J., Scott, K.T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J.P., Fleisher, A.S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., Weiner, M.W.: The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J Magn Reson Imaging* **27**(4), 685–691 (2008)