

Open Research: Working with Open Software and Open Data in Academic-based GIS Research

Peter Mooney, Blazej Ciepluch, Ricky Jacob, and Padraig Corcoran¹

¹*Department of Computer Science, National University of Ireland Maynooth, Co. Kildare. Ireland*

In this paper we describe our experiences from university-based academic research while maintaining an “open” ethos: using open source software, using and implementing open standards, protocols and interfaces, and using, managing, collecting, and advocating open data. Initially this approach required the input of both significant time and collaborative energy resources from the research team members. However this proved successful: new skills were learned, ideas and research agendas formed, and crucial a mode of working where the old problems data incompatibility, software integration, etc are practically non-existent.

INTRODUCTION AND MOTIVATION

Research firm Gartner believes that 80 percent of all commercial software applications will include open-source components by 2012. Gartner notes that the value and robustness of open-source software offers compelling opportunities for commercial software developers to reduce development costs Gartner Research [4]. The advantages for organisations and individuals in using open source software are well publicised and documented. Anderson and Moreno-Sanchez [2] find that using open source software has a number of advantages for organisations with scarce resources: no software costs, software tools are easily learned by personnel with general IT backgrounds, small software footprints, no need to commit to proprietary software, freedom to extend the software with functionality not present in commercial software, and compatibility with existing IT infrastructure. Steiniger and Bocher [8] closes his paper with the statement that “while open source software may not be suited for everyone if the price tag argument is set aside. But it is the best choice for research in GIS”. In this paper we describe our experiences from university-based academic research while maintaining an “open” ethos: using open source software, using and implementing open standards, protocols and interfaces, and using, managing, collecting, and advocating open data. In our presentation at OSGISUK 2011 we will provide examples of how successfully this approach has worked for our group. Carmichael and Honour [3] argues that the education sector has, for the most part, “failed to provide any form of training on open source software to IT coordinators or other staff”. We agree with Carmichael and Honour [3] when they state that such training would yield tangible long term benefits far out-weighting the initial investment costs in training. Subsequently many of the problems experienced by the academic research community such as the ‘proprietary lock-in’ issue “would gradually wane”. It is easy to lose sight of the fact that for most academic researchers the software they use is a means to an end. Software is a tool required to do their job.

In Figure 1 we provide a flowchart of the open source components and open data flows between the four major research projects in our research group. Each research project operates independently yet interacts and collaborates easily with the other projects without obstacles surrounding the types of software used. There are some obvious examples of open source choices made: PostgreSQL (PostGIS) over Oracle or Microsoft SQL Server, Java Processing over Flash, OpenOffice over Microsoft Office, SciPy and NumPy in Python over Matlab, Java-based Android development over iPhone development, Quantum-GIS over ESRI GIS software, gdal/OGR over FME, etc. HTML5, while not yet an officially accepted standard of HTML,

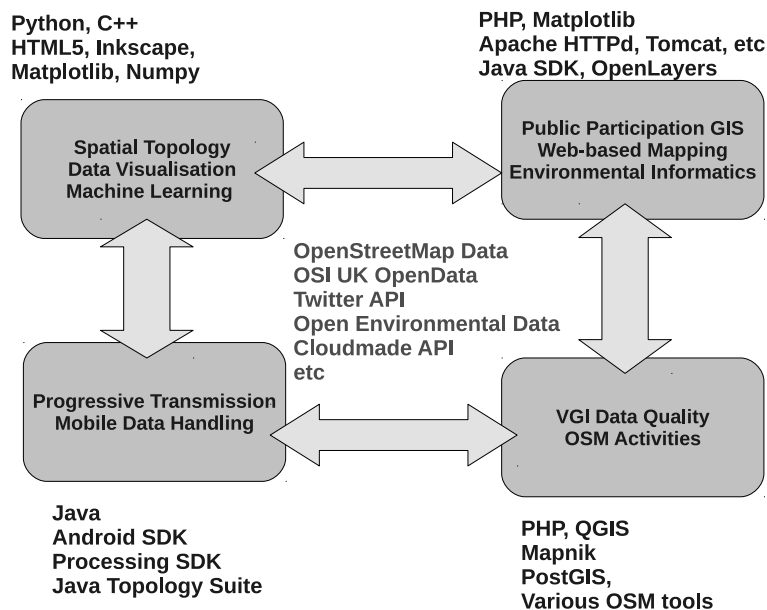


FIG. 1: A flow chart illustrating the four principal projects in our research group. The software used by each group is specified. The key open data sources are shown and we indicate where they are shared commonly by the group

offers a next-generation web development option where vector mapping, client side databases, etc can all be successfully implemented in the browser without the need for proprietary third-party plugins.

We feel that our use of open source (GIS) software and the use of open data sources has helped to drive our focus towards the *open science* paradigm. Open science broadly describes science carried out and communicated in a manner which allows others to contribute, collaborate, and add to the research effort. All kinds of data, results, and protocols are made freely available at different stages of the research process. Supporters of this approach argue that such collaboration will lead to more efficient research and innovation [6].

USING OPEN DATA IN OUR RESEARCH

The definitions of “open data” does not necessarily imply that this is spatial data. Rather it offers insight into the issue of openness and the conditions for re-use of the data. Open Data enhances the visibility and scope for engagement of data, offers opportunities for wider engagements, across the research community and other, broader, communities, including new possibilities for citizen science and volunteered geographic information and for public engagement with the processes and results of research NESTA [7]. While it could be argued that university-based researchers and academics have had a significant impact on the development of open source software, etc - the open data movement benefits industry, government, academia, and the general public. All can be consumers and producers (“prosumers”). An interesting observation we have made over the past number of years with the emergence of Volunteered Geographic Information (VGI) is how National Mapping Agencies have begun to provide open access/APIs etc to their spatial data. Rather than the VGI approach being a direct competitor both sides of the “spatial data divide” are deriving benefits from an open data approach. This approach is enhancing collaboration and community-

building. This provides new opportunities for collaboration across institutional, national, and disciplinary boundaries, and encourages the open sharing of knowledge and expertise. In our presentation at the conference we will outline how each of the projects illustrated in Figure 1 have become stronger candidates for successful potential collaboration and research-network building as a result of the removal of barriers related to software and data access. The focus of collaborative efforts can now be fully concentrated on problem solving, generating ideas, and finding new research areas. The eventual implementation of the EU INSPIRE directive might add additional momentum to increased open access to government, national, and European spatial data. In closing we can draw many positives from our experience of GIS within a fully open source software approach which uses, in as far as is possible, (spatial) data from open data repositories/available under open license. At this point in time we can confidently speculate that even as yet unknown future research needs and directions will not provide reason for our research group to stray outside this “Open Research” paradigm. There are some issues which must be addressed urgently relating to open data in academia. Anand et al. [1] summarises the problem very well by stating that new sources of (spatial) data are becoming available and easy-to-use toolkits are opening up spatial analysis beyond the traditional user, offering considerable opportunities throughout academia. However, as data availability increases, especially in terms of crowdsourced data, those using this data must be aware of both the risks and benefits of such approaches. “Crowdsourcing” may be an acceptable route for data collection as long as standards are put into place to ensure high quality survey, sampling, and collection methods. Ghosal and Cross [5] advises that more effort be placed into developing INSPIRE-capable open source GIS software. The authors state that the advantages of this could “led to significantly more effective adoption of directives such as INSPIRE”. Overall these issues, outlined by Anand et al. [1] and Ghosal and Cross [5], will inturn provide important research questions for the GIS community now and into the future.

-
- [1] Anand, S., Batty, M., Crooks, A., Hudson-Smith, A., Jackson, M., Milton, R. and Morley, J. [2010]. Data mash-ups and the future of mapping, *Technical Report Horizon Scanning Report 10*₀₁.
 - [2] Anderson, G. and Moreno-Sanchez, R. [2003]. Building web-based spatial information solutions around open specifications and open source software, *Transactions in GIS* **7**(4): 447–466.
 - [3] Carmichael, P. and Honour, L. [2002]. Open source as appropriate technology for global education, *International Journal of Educational Development* **22**(1): 47 – 53.
 - [4] Gartner Research [2008]. Key predictions for it organisations and users in 2008 and beyond, Online press release - <http://www.gartner.com/it/page.jsp?id=593207>. Last checked January 2011.
 - [5] Ghosal, S., L. N. and Cross, I. [2010]. The role of open source software for handling INSPIRE-compliant data, *Geoinformatics* **13**: 16–22.
 - [6] Kousha, K. and Thelwall, M. [2007]. The web impact of open access social science research, *Library & Information Science Research* **29**(4): 495–507.
 - [7] NESTA [2010]. Open to all? case studies of openness in research, Published online as a PDF from National Endowment for Science Technology and the Arts (UK) at <http://www.apo.org.au/node/22622>.
 - [8] Steiniger, S. and Bocher, E. [2009]. An overview on current free and open source desktop gis developments, *International Journal of Geographical Information Science* **23**(10): 1345 – 1370.