



"Where did my data layer come from?" The semantics of data release

Adam Leadbetter (1) and Justin Buck (2)

(1) Marine Institute, Oranmore, Ireland (adam.leadbetter@marine.ie), (2) British Oceanographic Data Centre, National Oceanography Centre, Liverpool, United Kingdom (juck@bodc.ac.uk)

In his lecture, "Theory of Creative Fitting" (Margullis, Corner & Holt, 2006), Ian McHarg introduced his vision for cross-disciplinary data and information sharing networks with the end goal of producing detailed overlay maps for the purposes of ecological architectural planning. Within McHarg's networks, experts in various fields, such as hydrology or surface geology, would provide data layers to the final overlay map with full provenance, such that the users of the overlay maps would know the originator of the data, the "value systems" by which the data were created and could place their trust in the outcomes.

In the light of McHarg's statements and in order to allow the encoding of value systems in a cyber-GIS, analyses of: data quality (Giarlo, 2013); data publication networks (Reinsfelder, 2012); trust in collaborative research networks (Leadbetter, 2015); and the metaphors of data publication, data release and data ecosystems (Parsons & Fox, 2013) have been synthesised into a logical model of the data release lifecycle. This model concerns the actors in the data release process; the data-information-knowledge ecosystem through the various stages of the data release process and the impact of data release on perceptions of trust through the data release lifecycle.

The data-information-knowledge ecosystem described how the collection of data can be presented in new ways to form information products, and how these information products can inform conversations amongst information-consumers who integrate the information into new knowledge. The actors concerned in the process comprise:

- researchers
- data publishers
- academic publishers
- & academic administrators

Finally, the lifecycle of data release involves the initial release of a data-layer, possibly with a Persistent Identifier (PID) more generic than a Digital Object Identifier (DOI). A data description paper can be written about the dataset, which then necessitates the assignment of a DOI to the datasets; the DOI can be seen as an indicator of trust through "benevolence". A technical document citing the dataset may then be informed by the dataset release or the dataset description paper. These citations may show the "competence" (in terms of a trust model) of the original datasets, and the dataset description papers or other technical articles show the integrity of the dataset.

The synthesised logical model has been represented in freely available ontologies, such that data layers can be annotated with metadata about their provenance and stage within the data release lifecycle before incorporation into a cyber-GIS, in which distributed data providers provide for a collaborative research environment.

References

- Giarlo, M. (2013). Academic libraries as data quality hubs. *Journal of Librarianship and Scholarly Communication* 1(3): eP1059. doi: 10.7710/2162-3309.1059
- Leadbetter, A. (2015). Examining trust in collaborative research networks. In P. Diviacco, P. Fox, C. Pshenichny, and A. Leadbetter (Eds.) *Collaborative Knowledge in Scientific Research Networks*, Hershey, PA: IGI Global. doi: 10.4018/978-1-4666-6567-5.ch002
- Margulis, L., Corner, J. and Hawthorne, B. (Editors), 2006. Ian McHarg: conversations with students. *Dwelling with nature*. New York, NY: Princeton Architectural Press.
- Parsons, M., and Fox, P. (2013). Is data publication the right metaphor? *Data Science Journal* 12, 32-46. doi:10.2481/dsj.WDS-042

Reinsfelder, T. (2012). Open access publishing practices in a complex environment: conditions, barriers and bases of power. *Journal of Librarianship and Scholarly Communication* 1(1): eP1029. doi: 10.7710/2162-3309.1029