

# Making Environmental Research Data Publicly Available – Experiences from Ireland

Peter Mooney  
Environmental Research Center,  
Environmental Protection Agency,  
Richview, Clonskeagh, Dublin 14. Ireland  
Telephone: +353 (1) 268 0100  
Email: [p.mooney@epa.ie](mailto:p.mooney@epa.ie)

Adam C. Winstanley,  
Department of Computer Science,  
National University of Ireland Maynooth (NUIM),  
Maynooth, Co. Kildare. Ireland  
Telephone: +353 (1) 7083853  
Email: [adam.winstanley@nuim.ie](mailto:adam.winstanley@nuim.ie)

## Abstract

Directive 2003/4/EC “Public Access to Environmental Information” (PAES, 2003) instructs public authorities to make available and disseminate environmental information to the general public to the widest extent possible. It is the responsibility of Member States to determine the practical arrangements under which such information is made available. These arrangements shall guarantee that the information is effectively and easily accessible and made available through publicly accessible channels (ie the Internet). Ireland introduced the legislation for this Directive in 2005 with further guidance published recently (see DoEHLG, 2008). This paper discusses our experiences at the Irish Environmental Protection Agency in our efforts to improve Internet-based public access to environmental data and information generated by environmental research programmes in Ireland. In particular we focus on how public access to these data flows can be maintained efficiently while maximising the potential re-use value of the research resources themselves.

## Context of This Work

The EPA research programme for the period 2007 – 2013 is entitled Science, Technology, Research and Innovation for the Environment (STRIVE). The funding budget for STRIVE is substantial. The current amount of funding available is approximately €108 million. The objectives of STRIVE include: increasing the number of research teams led by internationally competitive principal investigators, upgrade existing research infrastructure to develop new facilities to support research, and to double the number of PhD graduates by 2013. Two goals of the STRIVE programme overlap nicely with Directive 2003/4/EC. These are “to contribute to a better environment by delivering applicable and relevant data, information, and knowledge based on high quality scientific research and technology” and “to disseminate findings of the individual research projects to the widest possible audience in a coherent and timely manner”. The EPA are working pro actively to meet both of these goals. With the recent SEIS (Shared Environmental Information System (SEIS, 2008)) communication from the Commission the STRIVE programme has commenced at an appropriate time and an opportunity to implement the SEIS principles from the earliest stage in the data management of the programme. STRIVE will fund hundreds of researchers, generate vast quantities of new environmental data and information, and involve scientific researchers from almost every environmental research institute in Ireland. The potential influence of STRIVE in delivering a substantial data component to a SEIS should not be underestimated.

## Technical Aspects of This Work

The major technological product developed to meet the requirement for the dissemination of findings from individual research projects is SAFER-Data (the Secure Archive For Environmental Research Data – whose development strategy was outlined in Mooney and Winstanley (2007) at

EnviroInfo 2007). SAFER-Data is a web-based interface to the archive of environmental research data managed by the EPA. Currently the archive is almost exclusively comprised of data resources generated by the previous funding programme (ERTDI) and some early STRIVE projects. A key component of SAFER-Data is the availability of metadata browsing services (including RSS feeds), download services for publicly available research data, and secure private upload of data and metadata from the individual research projects. Metadata creation, data preparation, and final data upload from source aligns with the 1<sup>st</sup> SEIS principle that “information should be managed as close to the source as possible”. Before metadata documentation, upload of data resources, and online availability via SAFER-Data takes place we work closely with the research group to ensure “due consideration is given to the choice of level of aggregation and implementation of appropriate confidentiality constraints” (SEIS principle 6). Going forward the research programme is most likely to generate many of the types of information outlined in Article 7 (2 a-g) of Directive 2003/4/EC. These include (1) data or summaries of data derived from monitoring of activities affecting, or likely to affect, the environment, (2) environmental impact studies and risk assessments concerning environmental elements, and (3) reports on the state of the environment. Besides these data resources exhibiting unique temporal characteristics they are innately linked to detailed local geospatial information. An information space, such as the one outlined in SEIS, is more than just measurements and data exchange – importantly it includes local, regional, national, and global scales (Pick, 2007).

## Key Issues for Consideration

It is a condition of the STRIVE funding award to researchers that “all significant datasets generated during the duration of the project are made available for public access using SAFER-Data” (STRIVE, 2008). Significant datasets are normally evaluated under the following criteria which examines (1) the relative importance or impact of the study, (2) the suitability or integratability with existing data holdings, (3) if the data exhibits potential for further analytical research or re-use, and if (4) the data is key raw data or methodologies necessary for making state of the environment assessments. Several key issues are recurrent and provide significant barriers to full open public access to data and information generated by the research programmes. These issues are summarised as follows:

1. Without an ongoing pro-active approach from the EPA and a user-centered development model for systems such as SAFER-Data large volumes of data may be left undocumented and essentially “withering away on a vine” (Bishop, 2007)
2. Technical change amongst some data providers is difficult to initiate. This can include implementation of new data management practices or variations in data collection regimes. However cultural change in regards to data sharing is much harder still.
3. Most environmental data producers and scientists can see the potential benefits of a SEIS but are often unable to visualise or evaluate their specific role within the SEIS fabric. It is often difficult for the data producers to objectively evaluate what subsets of their overall data holdings are “desirable” and “sharable” for the wider environmental science community.
4. There is an urgent need to capture data generated by the research projects as quickly as possible after the conclusion of the project. Lessons learned from other domains such as social science (Bishop, 2007) shows that secondary users of the data are much more engaged if the primary user or generator is still contactable and/or available
5. We have a situation of conflicting objectives: the EPA and STRIVE on the one hand who look to disseminate the findings of individual research projects to the widest possible audience and the individual researchers on the other who (a) gain little or no academic recognition for making their data available, (b) risk unsolicited third party usage of their datasets as the possibility exists such third parties could further publish the work, and (c) fear losing ownership of their datasets resources. The argument that there already exists a public dissemination point for information, data and results in academic journals “is true but misses the point” (Sauber, 2004). Usually journal articles are available as PDF documents

if one has obtained the correct access authorisations. Analysing the science in this manner requires manual effort. The scientific results and in some cases the raw data measurements should be available in a form that is easier to process and analyse for third parties.

6. “*The data is available if one asks in person*”. This classic and often ad-hoc data sharing and exchange model is incompatible with proven data management practice. It often creates multiple copies of the same datasets, version incompatibility, and essentially hides datasets from more general third party exploration. The data can be transferred onto CD or DVD media and delivered in this fashion. However this is usually more effort than both the data owner and the third party enquirer would like to expend.
7. Metadata regardless of its format, standard, or implementation remains a virtual outcast of everyday data management amongst the scientific community in general. Some authors remark that it is ironic that at a period where the volume of data generated in scientific research is at an all time high the practice of documenting and accessing these resources is at an all time low (Bulterman, 2004). Metadata carries the unfortunate stigma as being “boring” (Comber et al, 2005). Consequently many researchers see little or no academic or workplace recognition for the task of providing or maintaining metadata (Mooney & Winstanley, 2007). Given the strategic importance of metadata for INSPIRE and SEIS the metadata *problem* must be cracked. The development of elaborate software systems “shiny front-ends” (Pollock, 2007) for these types of problems is not working. The secret to communicating the multidimensional value of metadata effectively to data producers remains as elusive as ever.

None of the issues outlined above are unique to the environmental science domain. Nor are the issues unique to environmental research in Ireland. We feel that overcoming these problems or at minimum finding some acceptable middle-ground is critical to the successful integration of data and information from environmental research in all Member States into any Shared Environmental Information System.

## Acknowledgments

Dr. Peter Mooney's work on this project is funded by the STRIVE programme (Project Code: 2008-FS-DM-14-S5) under the National Development Plan 2007-2013.

## References

- Bulterman, D.C.A. **Is it time for a moratorium on metadata?** IEEE Multimedia Volume 11, Issue 4, Page(s): 10 - 17 Oct.-Dec. 2004
- Bishop, L. **A Reflexive Account of Reusing Qualitative Data: Beyond primary/secondary dualism**, Sociological Research Online [Online], Special Section on Reusing Qualitative Data, 12(3). Available at: <http://www.socresonline.org.uk/12/3/2.html> (Last Accessed March 18<sup>th</sup> 2008).
- Comber, A., Ahlqvist, O., Fisher, P., Harvey, F., Gahegan, M. and Wadsworth, R., (2005). **Can Metadata ever be interesting?: the case for expanded metadata**. Pp. 53-57 in *Proceedings of the GIS Research UK 13<sup>th</sup> Annual Conference*, University of Glasgow 6-8 April 2005
- DoEHLG, **Access to Information on the Environment Regulations 2007 Guidelines**: Issued by the Minister for the Environment, Heritage and Local Government under Article 14 of the European Communities (Access to Information on the Environment) Regulations 2007, (S.I. No. 133). <http://www.environ.ie/en/Publications/Environment/Miscellaneous/FileDownload,2479,en.pdf>
- Mooney, P. and Winstanley, A.C **Improving Environmental Research Data Management**. EnviroInfo 2007. Proceedings of 21<sup>st</sup> International Conference on Informatics for Environmental Protection. Warsaw Poland. Pages 473—476 September 2007
- PAES, **Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information** *Official Journal L 041* , 14/02/2003 P.

0026 - 0032

Pick, T. **From Århus to Inspire: Putting Environmental Information on the Map.** Proceedings of 21<sup>st</sup> International Conference on Informatics for Environmental Protection. Warsaw Poland. Pages 21—33 September 2007

Pollock, R. **Give us the raw data, and give it to us now.** An entry to the Open Knowledge Foundation Weblog. Available online at permalink <http://blog.okfn.org/2007/11/07/give-us-the-data-raw-and-give-it-to-us-now/> (Last Accessed March 18<sup>th</sup> 2008)

SAFER-Data, **The Secure Archive For Environmental Research Data:** Environmental Protection Agency Ireland: Service available online at <http://erc.epa.ie/safer> March 2008

Sauber, P. **The Taxpayer Argument for Open Access** – The SPARC Open Access Newsletter Issue 65. Available online at <http://www.earlham.edu/~peters/fos/newsletter/09-04-03.htm> (Last Accessed March 18<sup>th</sup> 2008)

SEIS, Commission to the European Communities: **Towards a Shared Environmental Information System (SEIS)** – COM(2008) 46 Final. Brussels 1.2.2008

STRIVE, **Science, Technology, Research and Innovation for the Environment** Funding Programme 2007 – 2013 Environmental Protection Agency Ireland, PO Box 3000, Johnstown Castle Estate, Co. Wexford. Ireland. STRIVE Documentation available on line at <http://www.epa.ie/researchandeducation/research/research/> (Last Accessed March 18<sup>th</sup> 2008)