

TELEMAC model archive: Integrating open-source tools for the management and visualisation of model data

S. L. Mouradian*, A. Avdis*, M. D. Piggott*, C. T. Jacobs†, C. Villaret‡, D. R. de Mijolla§ and J. Lietava*

*Department of Earth Science and Engineering
Imperial College London, South Kensington, UK SW7 2BP
Email: simon.mouradian06@imperial.ac.uk

† Faculty of Engineering and the Environment
University of Southampton

‡ EDF-LNHE, France

§ Department of Physics
Imperial College London

Abstract—

Reliable and robust archival of model set-ups and outputs, including input data is a major challenge. Failure to do this in a robust and sustainable manner could lead to increased set-up costs for new studies, duplication and possible inconsistencies in model outputs. In addition, comprehensive archival of input and output data is an important step in ensuring reproducibility of results. In order to address these challenges, the utilities presented in this paper facilitate the automatic archival of model data, and publication in a citable repositories.

The software is designed to offer data management, visualisation and database-entry functions for modelling with the TELEMAC suite. A purpose-built SALOME plugin is also presented, aimed at automatic file archival and publication. Visualisation is also provided through dedicated lightweight utilities using the crossplatform Google Earth desktop application.

The practice of transparent publication of model set-up and results is increasingly dictated by national legislation and will also lead to the collection of large databases, enabling the analysis of simulation set-ups as well as output. The ability to filter and analyse simulation set-ups can lead to the development of best-practice guidelines, as well as the opportunity to identify areas for the improvement of future simulations. Visualising all simulation set-ups together provides the big picture of research activities, and can offer the ability to identify trends in (perceived) optimal option choices, as well as identify regions where future opportunities might arise. Furthermore, as simulation set-ups from various parties can be collected together and visualised in a single framework, the ability to collaborate and spread knowledge internally within, and between, organisations is made possible through the presented tools.

I. INTRODUCTION

Over the past 30 years, numerical coastal modelling capabilities have increased substantially, reaching a point where they are an integral part of the planning phase for many coastal engineering projects. Ranging from the prediction of coastal flooding to sediment transport and environmental/ecological impact assessment studies, with many applications in between, the use of numerical coastal ocean modelling has become of paramount importance [3]. In such cases, it is generally the

responsibility of the modeller (often in an ad hoc manner) to ensure archival of the model and its results. The (collectively) terabytes of data produced annually are often discarded, apart from a few summarising figures in reports and/or published articles, that can often not be used for recomputation and/or reanalysis at a later date. For many applications, archival of model results and data provenance records must be maintained for many years after the modelling phase of the project, as legislated by government bodies [2].

Within academic institutions, researchers were (and still are) traditionally credited based upon publications and citation rates, with minimal or no credit given for ensuring results are archived, reproducible and easily retrieved. Recently, rules published by the Research Council UK (RCUK) place particular emphasis on public accountability and open access [15], and on making all publically funded research data, samples and models accessible. It is therefore increasingly more important for institutions to adopt a research data management protocol.

The archival solution proposed here, by leveraging the PyRDM library [9], aims to make it easy for modellers to archive their work, as well as the option for providing each simulation setup and its results with a Digital Object Identifier (DOI). The DOI allows each simulation to be cited in reports and/or research papers, which can result in researchers being credited based on the amount of citations their simulations receive.

Large engineering companies often have groups spread geographically, or have several (academic or industrial) sub-contractors, all of whom produce numerical model setups and output data. This can result in inconsistent model setups and thus an inability to perform meaningful model comparisons, or can result in efforts being duplicated across groups. A centralised database that allows for a quick overview of current and previous numerical modelling activities can be a very useful tool used to streamline and optimise the process of setting up new simulations and comparing / benchmarking against similar studies, either by other research groups/teams, possibly based in different geographic locations.

When setting out to develop an archival system, we considered the following requirements: a) The entirety of files should be archived, to allow for the output data to be recomputed and reproduced; b) It must allow the results and setup to be readily retrieved; c) It must be very easy to use, with minimal user steps required, and fit within the workflow of the modeller. This is essential if the archival system is to be adopted and continually used. This is why we suggest our coastal model archiving tool be integrated in the SALOME modelling platform, in the form of a plugin (see section II-D for further details).

Archival of all of an institution's or community's coastal modelling activities in a cohesive, structured database also opens up the possibility for further analysis. In particular, it allows for a macro-scale analysis of these simulations, to obtain insight in the principles of setting up a model. For example, one can ask questions relating to how many sediment models use the two-dimensional shallow water formulation as their hydrodynamic solver, or which tidal simulations employ wind-forcing at the surface, or regularly used numerical or turbulence modelling choices.

Further insight can be obtained from an institution's research or consultancy activity by visualising coastal model domains on a map (for example, as in [4]), laid over satellite imagery. This allows for an analysis of the spatial distribution of coastal modelling activities. An organisation is able to, amongst other things, identify geographic hotspots of numerical modelling activity that could be consolidated, or identify geographic areas that may form new investment opportunities.

Motivated by the above, we describe an archival solution for coastal models that performs three distinct, yet integrated functions:

- 1) Packages a simulation setup and archives it on a file server hosted by a data repository service or institution. If the archived package is made public on an online repository, a DOI is generated for citation and easy retrieval.
- 2) Parses the simulation setup and populates a database with important parameters (and metadata, such as owner and time) that can be searched and analysed at a later date.
- 3) Converts the domain extent and result to a file format suitable for viewing in Geographical Information Systems (GIS) clients, such as Google Earth.

The resulting archival tool is a stand-alone, cross-platform desktop application with a graphical user interface, which we name AVoCadO (Archiving and Visualisation of Coastal Ocean models).

II. METHODOLOGY

The archival framework has been built upon the PyRDM library. PyRDM is a research data management library which features the ability to curate and publish data (and software) objects using various repository services, for example Figshare and Zenodo [9], [10]. The library implements the relevant API for each service, allowing data publication to be accomplished using a single codebase. These services, in turn, return a citable DOI that can be used for citing the relevant output.

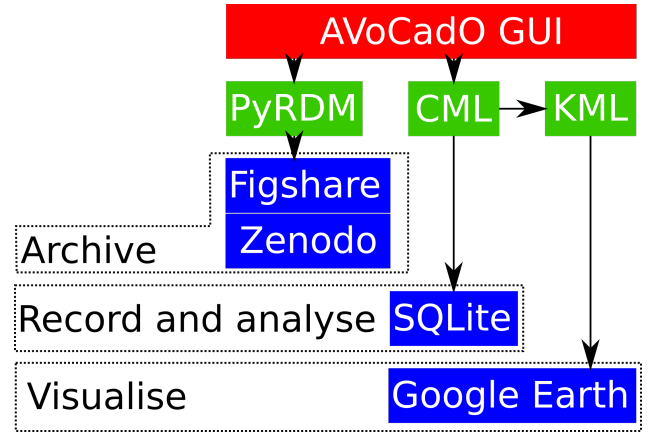


Fig. 1: AVoCadO toolchain. The user is exposed only to the graphical user interface. AVoCadO interfaces the PyRDM library [9] to archive software and data, and the CML library to generate SQL database entries and convert the mesh and results to the KML file format, suitable for visualisation in Google Earth.

Furthermore, AVoCadO is built on the CML library, which was developed specifically for this project, to handle the parsing of configuration files for adding entries to an SQL database, as well as generating GIS objects, to be visualised in software such as Google Earth. An overview of the toolchain is shown in Figure 1.

A. Archival

A large amount of automation is offered by PyRDM. We incorporate PyRDM's curation capabilities into the archiving workflow, which in turn allows AVoCadO to publish data to online, persistent repositories in a straight-forward, 'click of a button' manner. Such benefits of automated publishing tools have been demonstrated by e.g. [9], [10], and [8]. As a result of such advances in digital curation technology, this 'opening up' of data objects promotes more open research workflows and reproducibility best practices, and can also help ensure that research funding policies or government legislations are properly complied with [1], [11], [12], [14], [16].

We note that Figshare and Zenodo do not take ownership of any data published on their servers. The user uploading data is given the option of several licenses under which to publish their data. Figshare and Zenodo are data repositories that ensure the persistence and longevity of published output, and provide users with the ability to search for and obtain the published data. They also collect analytics for the user, such as views, downloads and citation rates.

A second archival option is provided within AVoCadO, whereby the data is archived on a private server. An encrypted connection is established using the secure shell protocol, and data is transferred from the modeller's workstation to an institution's file server. Using this archival method, a DOI is not generated, and thus the database holds a path to the hostname and location of the archived data on the file server. Whether a modeller archives simulation data and/or software to a public repository or a private server, the data is sent as an

archived package. A unique filename is obtained by naming each archived package with the *Simulation Name* provided by the user, appended by the MD5 hash.

B. Database

For the purposes of rapid prototyping the AVoCadO application, an SQLite database [7] is used, rather than a hosted, process based database server. We note the implementation is similar and only minor further development would be necessary to implement a MySQL or PostgreSQL database. These SQL servers would offer access control with user authentication, not currently offered by the embedded SQLite database.

A database entry is generated for each model archived. The database holds information that relates to retrieval of the archived input/output data (either a DOI or the hostname and location of the private file server). The database also holds information that is parsed from the model configuration file. For example, a database entry for the TELEMAT model would include information on whether a simulation used the finite element or finite volume discretisation method, or whether tidal flats were enabled.

A separate SQL table is used for each coastal modelling solver implemented in AVoCadO (currently only TELEMAT [6] (both 2D and 3D) and Fluidity [13]). Each table contains entries representing a single, archived model. For TELEMAT, the configuration (.cas) file is parsed using the parser utility that is provided in the TELEMAT source code. AVoCadO therefore has all the information that can be extracted from the configuration file at its disposal, and can record it in the database. A drawback of recording all the information includes inconsistencies between different versions of TELEMAT. A further significant note regarding the database entry, is that differences present in supplementary Fortran subroutines, or modifications made to the software source, are not recorded in the database (although the supplementary files and software source can be archived, if the modeller chooses to do so).

C. Visualisation

Visualisation is also provided by the CML library, developed as a part of AVoCadO. The CML library takes as input the directory in which the simulation resides and performs the following functions:

- *Establish numerical solver used:*
Given the directory of model setup and results, CML attempts to determine whether the model uses TELEMAT or Fluidity as the numerical solver. This is established by checking the types of files within the directory. For example, the presence of a .cas file would suggest a TELEMAT model, whilst the presence of an .fml file the use of Fluidity.
- *Parse directory for files:*
The directory is parsed to determine the configuration file, mesh file, results files and any supplementary files such as Fortran files, wind or tidal forcing. The modeller archiving the simulation can choose to override the automatically detected files, to only publish

specific files, or replace certain files with files from a different directory. This AVoCadO feature of auto-selection of files ensures that minimal user input is required before a model can be archived. An example of the auto-populated interface is shown in Figure 2, where fields are populated with the TELEMAT example of the Monai Valley test case.

- *Extract mesh triangulation and results:*
The input mesh is parsed to determine the boundary of the domain, as well as the triangulation. The points along the domain boundary are used to define a new line path in the KML format used by Google Earth, and the triangulation is used to define a network of lines, in the KML linestring format. The results file of the final timestep is parsed and used to produce an image in the .png file format. The domain boundary is used to *mask* the image, ensuring that the visualised result is only produced within the domain boundaries. Figure 3 shows example boundaries visualised in Google Earth.

For performance reasons, the initial visualisation only includes the outlines of archived models, as well as a place-mark labelled with the name. This provides users with an overview of all archived simulations. A user is further able to enable/disable the visualisation of the mesh and/or the snapshot images generated from the results files. We note, however, that the memory requirement of Google Earth increases substantially when the mesh for multiple archived simulations is viewed at the same time. Figure 4 shows an example of the functionality of enabling/disabling field images, as well as how AVoCadO takes advantage of Google Earth's 'regioning' features. Regioning is the property whereby certain image overlays are only loaded and shown on screen when at low altitudes (zoomed in), thus avoiding densely populating the screen and limiting the memory required by Google Earth.

An important step in converting the mesh and results files to formats suitable for visualising in Google Earth, is applying the appropriate projection. Objects visualised in Google Earth are required to be in the World Geodetic System defined in 1984 (WGS 84). The transformation is realised by the use of the Python bindings of the GDAL library [5]. GDAL uses the spatial reference system identifiers as defined by the European Petroleum Survey Group (EPSG), and thus the EPSG code of the input coordinate reference system must be provided by the modeller. This input is labelled 'Coord system' and is given in the bottom right hand corner of the GUI, as seen in Figure 2. Generating a visualisation for Google Earth is only reasonable for numerical models on realistic geographical domain. Models that are performed on idealised domains in an arbitrary Cartesian coordinate system can not be projected for visualisation in Google Earth. An entry of zero (0) in this field will result in AVoCadO not producing the subsequent GIS objects, and will only use PyRDM for archiving.

D. SALOME plugin

The standalone application is currently able to parse input files from TELEMAT and Fluidity simulations, with a view to support more models in the future. The generality of AVoCadO as an archiving framework for coastal ocean models necessitates the use of a separate, stand-alone application. This, in

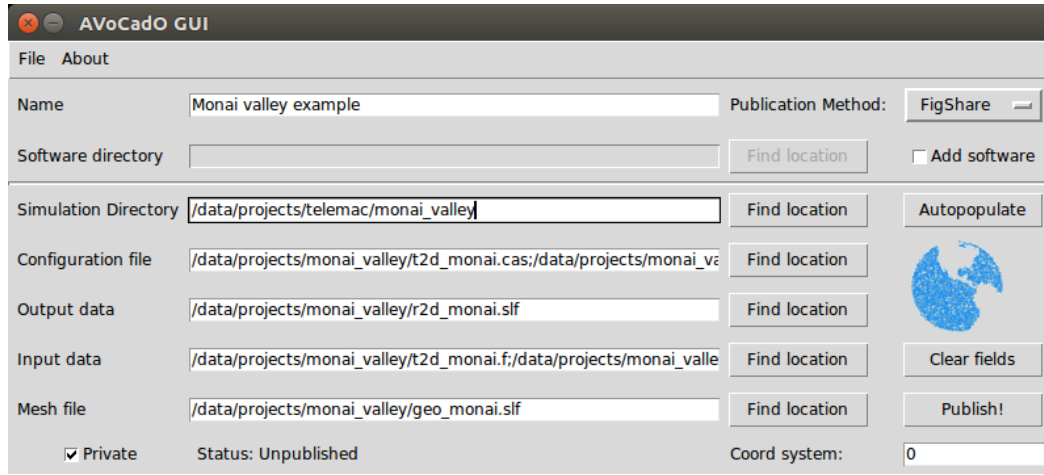


Fig. 2: The AVoCadO GUI after auto-populating with the TELEMAT example directory of the Monai Valley.



Fig. 3: Domain outlines and labeled placemarks for models of the Torness and Severn Estuary, two models archived using AVoCadO, visualised in Google Earth. Users have the ability to select the placemark, after which more summary information about the model will appear on the screen. By default, only boundary lines are shown to ensure responsiveness, even with multiple archived simulations loaded in memory.

itself, could be a barrier to its adoption. Here, a TELEMAT specific archiving solution is also explored through integration within the SALOME framework. This solution is more integrated in the workflow of setting up a TELEMAT simulation and thus is more likely to be of use to the wider TELEMAT community in the longer term. To achieve integration with SALOME, the AVoCadO front-end is replaced by a front-end written in PyQt4 (the same graphical interface library used by SALOME), while using the same underlying Python modules

as AVoCadO. This plugin allows for archival of models set up within the SALOME framework. The toolchain of the archival framework integrated within SALOME is illustrated in Figure 5.

In contrast to AVoCadO, the plugin does not require the modeller to provide the software with the directory in which the model setup resides. This is because the SALOME plugin is able to parse the SALOME configuration that specifies the files required for TELEMAT to run.

III. DISCUSSION

AVoCadO is an archiving framework that brings the power of PyRDM [9] to the coastal modelling community. It lowers the complexity and time needed for archiving coastal model setups and their results, thus removing impediments for modellers to do so and encourages a more reproducible and open scientific workflow. Archival of model setups and results on a persistent, searchable, citable repository hosting service such as Figshare, that enables easy retrieval of data, is ‘added value’ to multiple parties. It is added value to the modellers themselves, as well as the modellers’ colleagues, who can look back upon previous simulations. This has the potential of lowering the time required to setup a new model, allows for improvements upon previous setups while also maintaining a historical record of configurations used for previous models. It can also provide the particular model with a citable DOI, which can be used to cite the modellers’ setup and results in reports, articles and presentations. This can potentially increase the exposure of the modeller within their respective community, and enable sharing of models, results and ideas. Such an archive also encourages collaboration across an institution’s departments or groups, often spanning multiple geographic locations, by ensuring a centralised database of archived models is available for everyone to access and review. It ensures that time and money is not wasted by setting up duplicate models and allowing modellers to become familiarised with modelling methodologies employed by colleagues.

A further benefit, realised by archiving the complete setup, is that it enables a more thorough comparison between different models. Very often, unfair comparisons are made between

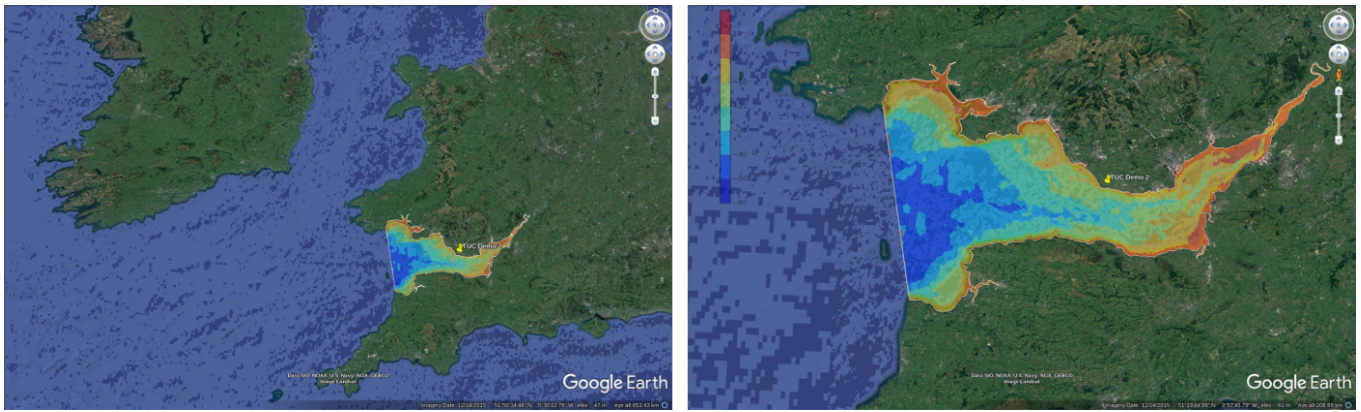


Fig. 4: Google Earth visualisation of a model of the Severn Estuary, as archived by AVoCadO. In these snapshots, the ‘BOTTOM’ field, depicting bathymetry, has been enabled. A .png image of bathymetry was generated and geographically placed by AVoCadO. The masking ensures that the bathymetry is only defined within the outline of the domain. The depiction on the right is zoomed closer to the model, at a level where a colourmap appears. KML files generated with AVoCadO take advantage of Google Earth’s regioning, to ensure responsiveness even when multiple models are being visualised.

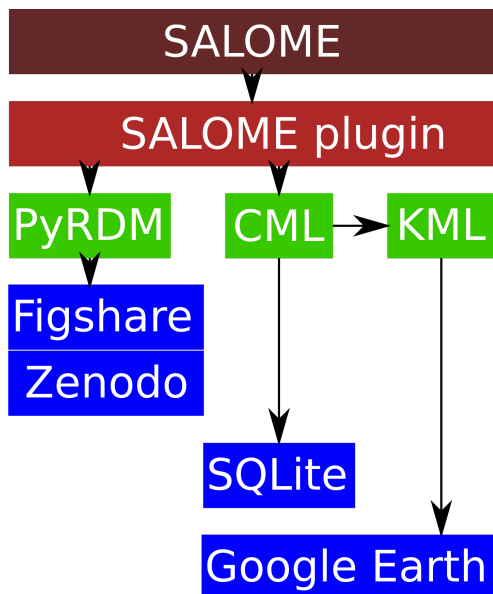


Fig. 5: SALOME plugin toolchain. The graphical user interface is now embedded within SALOME, and makes use of the same underlying Python libraries that AVoCadO does, providing the same capabilities to SALOME users.

models that employ different parameters, for example different forcing, or undocumented drag coefficients, or mesh resolution. Having access to the complete setup of previous studies one is comparing to allows for like-for-like comparisons.

A retrievable archive of previous modelling results, that allow recomputation and reanalysis, along with associated provenance (numerical model ‘owner’ and origin) is often required of an institution, either to comply with legislation, or the conditions set by a funding body. An archiving framework such as proposed in this paper can help institutions comply with such requirements.

The generation of a single database that contains entries and information on multiple simulation models makes obtaining insight into general modelling methodologies possible. Once a large number of simulations are recorded, possibly from various groups and/or using different underlying solvers, a statistical analysis of the database could lead to identification of common practices within the modelling community. These could result in guidelines of best practice, either automatically generated by mining the database and statistically identifying model setups, or with the meticulous direction of an expert.

The visualisation method described herein, provides an overview of all of an institution’s coastal modelling activities. With regards to each particular model, the visualisation is coarse, simplified and often incomplete. This is because AVoCadO, in addition to an archiving tool, seeks to encourage collaboration between departments or groups within an institution, and the sharing of data, models and ideas. It does not seek to replace traditional data visualisation tools that are more powerful or better suited for producing animations from a model and have tools for analysis of the results.

The framework described offers the possibility of archiving models and software either publically on Figshare, or privately on a self-hosted file server. A further possibility is the use of the paid-for service ‘Figshare for Institutions’. This provides institutions with an installation of Figshare, either on distributed cloud services such as Amazon Web Services, or on local institutional storage. This solution would offer institutions the benefits of Figshare, such as accurate metrics on shares, downloads, views and citations, but with the added benefit of being able to control where the data is stored and who has access to it. This is often necessary in cases where the data is commercially sensitive.

AVoCadO, as well as the described SALOME plugin, are free software released under the GNU GPL license. They are hosted on bitbucket.org, available to download, modify and redistribute, for both academic and commercial purposes. Installation instructions, along with a list of software requirements, as well as usage instructions, are provided in the manual.

available with the source code.

IV. CONCLUSION

Data curation is often perceived to be an error-prone and time-consuming task and is typically treated as an afterthought leading to inconsistencies in the way data is stored, or even worse incomplete records. Herein we have presented AVoCadO, an archival framework for coastal ocean models, which encourages consistent data archival, aimed to form part of the model set-up and results visualisation processes. Consistency is achieved through automation, for example an automatic identification of the input and output data within SALOME, but also the integration of RDM utilities [9]. In addition, AVoCadO is built around a centralised database that holds information specific to each simulation model. As the database grows, it can be used for macro-scale analysis of various coastal modelling practices and be used as a valuable guide of best-practices. For models in a geographic domain, AVoCadO also offers the capability to illustrate the simulation overlaid on a map. More specifically for TELEMAT users, a similar archival strategy is presented that is embedded within SALOME as a plugin. At the expense of a single solution for multiple numerical solvers, the SALOME plugin offers TELEMAT users a way to archive their simulations with just a few clicks of a button.

ACKNOWLEDGMENT

Avdis and Mouradian would like to acknowledge the support of EPSRC Impact Acceleration Account EP/K503733/1, de Mijolla and Lietava the funding from Summer Research Projects from Imperial College London. The authors would also like to acknowledge the help of Z.Zachariadis for IT support during the project.

REFERENCES

- [1] A. A. Alsheikh-Ali, W. Qureshi, M. H. Al-Mallah, and J. P. A. Ioannidis, "Public Availability of Published Research Data in High-Impact Journals," *PLoS ONE*, vol. 6, no. 9, p. e24357, 2011.
- [2] A. Avdis *et al.*, "Efficient unstructured mesh generation for marine renewable energy applications," , Manuscript in preparation.
- [3] A. Avdis, C. T. Jacobs, S. L. Mouradian, J. Hill, and M. D. Piggott, "Meshing ocean domains for coastal engineering applications," in *VII European Congress on Computational Methods in Applied Sciences and Engineering*, 2016. [Online]. Available: <https://www.eccomas2016.org/proceedings/pdf/7712.pdf>
- [4] G.-T. Chiang, T. O. White, M. T. Dove, C. I. Bovolo, and J. Ewen, "Geovisualization fortran library," *Computers and Geosciences*, vol. 37, no. 1, pp. 65 – 74, 2011, virtual Globes in Science. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0098300410002177>
- [5] GDAL Development Team, *GDAL – Geospatial Data Abstraction Library, Version 2.1.0*, Open Source Geospatial Foundation, 2016. [Online]. Available: <http://www.gdal.org>
- [6] J.-M. Hervouet, *Hydrodynamics of Free Surface Flows*. John Wiley and Sons, Ltd, 2007. [Online]. Available: <http://dx.doi.org/10.1002/9780470319628>
- [7] R. Hipp *et al.*, "SQLite Development Team: SQLite," <https://www.sqlite.org/download.html/>, 2015.
- [8] C. T. Jacobs and A. Avdis, "Git-RDM: A research data management plugin for the Git version control system," *Journal of Open Source Software*, vol. 29, 2016.
- [9] C. T. Jacobs, A. Avdis, G. J. Gorman, and M. D. Piggott, "PyRDM: A Python-based library for automating the management and online publication of scientific software and data," *Journal of Open Research Software*, vol. 2, no. 1, p. e28, 2014.
- [10] C. T. Jacobs, A. Avdis, S. L. Mouradian, and M. D. Piggott, "Integrating Research Data Management into Geographical Information Systems," in *Proceedings of the 5th International Workshop on Semantic Digital Archives*, 2015. [Online]. Available: <http://hdl.handle.net/10044/1/28557>
- [11] R. J. LeVeque, I. M. Mitchell, and V. Stodden, "Reproducible Research for Scientific Computing: Tools and Strategies for Changing the Culture," *Computing in Science & Engineering*, vol. 14, no. 4, pp. 13–17, 2012.
- [12] E. C. McKiernan, P. E. Bourne, C. T. Brown, S. Buck, A. Kenall, J. Lin, D. McDougall, B. A. Nosek, K. Ram, C. K. Soderberg, J. R. Spies, K. Thaney, A. Updegrave, K. H. Woo, and T. Yarkoni, "How open science helps researchers succeed," *eLife*, vol. 5, p. e16800, jul 2016. [Online]. Available: <https://dx.doi.org/10.7554/eLife.16800>
- [13] M. Piggott, G. Gorman, C. Pain, P. Allison, A. Candy, B. Martin, and M. Wells, "A new computational framework for multi-scale ocean modelling based on adapting unstructured meshes," *International Journal for Numerical Methods in Fluids*, vol. 56, no. 8, pp. 1003–1015, 2008.
- [14] V. Stodden, D. Bailey, J. Borwein, R. J. LeVeque, W. Rider, and W. Stein, "Setting the Default to Reproducible: Reproducibility in Computational and Experimental Mathematics," Institute for Computational and Experimental Research in Mathematics (ICERM), Tech. Rep., 2013. [Online]. Available: <http://www.davidhbailey.com/dhbpapers/icerm-report.pdf>
- [15] R. C. UK, "RCUK Policy on Open Access and Supporting Guidance," Tech. Rep., 2013. [Online]. Available: <http://www.rcuk.ac.uk/documents/documents/rcukopenaccesspolicy-pdf>
- [16] T. H. Vines, R. L. Andrew, D. G. Bock, M. T. Franklin, K. J. Gilbert, N. C. Kane, J.-S. Moore, B. T. Moyers, S. Renaut, D. J. Rennison, T. Veen, and S. Yeaman, "Mandated data archiving greatly improves access to research data," *The FASEB Journal*, vol. 27, no. 4, pp. 1304–1308, 2013.