FAIR Models

Hut, Rolf

DOI
10.1111/gwat.13180

Publication date
2022

Document Version
Final published version

Published in
Groundwater

Groundwater

Guest Editorial/

FAIR Models

by Rolf Hut

The geoscientific modeling community has encountered its fair share of changes over the last decades, of which I want to single out two in particular. Under the flag of "Open Science," scientists are encouraged to share as much of the scientific process as possible. Open Science gives decision makers more insight into how the scientific conclusions they rely on were established. Open Science gives underprivileged access to scientific knowledge hitherto inaccessible. And certainly not least: Open Science gives fellow scientists access to the data and software underlying scientific results, which makes it much easier to review and build on those results. A second development is the continuing integration of different fields within geoscience. Smart groundwater models are needed as part of climate models that predict our human impact on the entire climate system: from groundwater availability to global climate change.

Looking back at my own first geoscientific paper, the study of groundwater flow around dams in Kenya, we did what was standard back then: we built our own model from scratch and in the article provided all of the equations and results, but none of the code or data. Since our model was written for a version of Matlab that is now 15 years old, it is highly unlikely that it will still run, even if someone can get their hands on it. We have come a long way since then. The scientific community has introduced the FAIR principle for Findable, Accessible, Inter-operable and Re-usable science. FAIR was introduced for research data (Wilkinson et al. 2016) and has been extended to include research software (Lamprecht et al. 2020). The idea behind FAIR software and data is that if science is FAIR it should be possible for others to both reproduce as well as build on it. Luckily research data and software are increasingly published alongside articles. Journals and science funding agencies have started to demand that scientific results are published openly, including data and software (Nature Geoscience 2019).

Department of Water Resources Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands; r.w.hut@tudelft.nl

doi: 10.1111/gwat.13180

Sharing my code and data does not automatically make my results reproducible and thus my science re-usable. Anyone who has ever tried to get someone else's model to run on their own computer recognizes this. Different operating systems, different programming language versions, or more obscure: different versions of common libraries, all make it difficult to be really "FAIR." Ideally, I want to give other scientists (access to) my computer, so they can work in exactly the same environment. But I am not that idealistic and I really need my computer for work. The next best thing is to use software containers: small virtual machines that contain everything needed for a specific task. A container running a groundwater model contains the code of the model as well as all the libraries and compilers needed to run that specific model. To make our geoscientific modeling work re-usable I highly recommend we use containers as runtime environments for our models.

Containers may solve the "R" of FAIR, but there is still an issue with the "I" of inter-operable. Our 15-year-old Matlab model was never built with the intention to couple it with another model: no parameters are exposed, no way to interact with it other than reading the output. Luckily this has also changed over the years. Model interfaces such as OpenMI (Harpham et al. 2019) and BMI (Hutton et al. 2020) are becoming community standards to incorporate into models. MODFLOW 6 has BMI, which makes coupling easier than ever before.

Since Hutton questioned whether computational hydrology is even science when it cannot be replicated (Hutton et al. 2016), part of the modeling community has taken up the challenge to provide us all with tools that make it easier to do our work in a FAIR manner (see Hall et al. (2021) for an overview). Personally, I've been working on the eWaterCycle project (Hut et al. 2021) where we have combined the container and interface technology mentioned above to provide computational hydrologists with a framework where they can more easily work with each other's data and models. eWaterCycle provides a clear separation between the model, and the scientific experiment done with the model. Interaction with models is through a language agnostic version of BMI, making it easy to do experiments such as swapping one model for another one, or coupling one model to another one, even if these models are written in different programming languages.

The tools provided in eWaterCycle and similar efforts throughout the community have made it easier to do our work in a "FAIR" way. I would love it when in the next review of a modeling study I can run the experiment myself, play with the results and thus better assess the science. I would love it when my own studies are both re-done and built upon with greater ease by others in our community. I call upon journal editors and reviewers to ask authors to make their work FAIR. But above all, I call upon scientists themselves: make your work FAIR so others can reproduce and build upon your work and ultimately you can have more impact with your science.

Acknowledgments

The author would like to thank Ty Ferre for suggesting that the author should write this editorial and Mark Bakker and Leonard Konikow for constructive comments on the text of this editorial.

References

Hall, C.A., S.M. Saia, A.L. Popp, N. Dogulu, S.J. Schymanski, N. Drost, T. van Emmerik, and R. Hut. 2021. A hydrologist's guide to open science. *Hydrology and Earth System Sciences* 26, no. 3, 647–664. https://doi.org/10.5194/hess-26-647-2022

Harpham, Q.K., A. Hughes, and R.V. Moore. 2019. Introductory overview: The OpenMI 2.0 standard for integrating numerical models. *Environmental Modelling and Software* 122: 104549. https://doi.org/10.1016/j.envsoft.2019.104549

Hut, R., N. Drost, N. van de Giesen, B. van Werkhoven, B. Abdollahi, J. Aerts, T. Albers, F. Alidoost, B. Andela, J. Camphuijsen, Y. Dzigan, R. van Haren, E. Hutton, P. Kalverla, M. van Meersbergen, G. van den Oord, I. Pelupessy, S. Smeets, S. Verhoeven, M. de Vos, and B. Weel. 2021. The eWaterCycle platform for open and FAIR hydrological collaboration. *Geoscientific Model Development Discussion*. https://doi.org/10.5194/gmd-2021-344

Hutton, E.W.H., M.D. Piper, and G.E. Tucker. 2020. The basic model Interface 2.0: A standard interface for coupling numerical models in the geosciences. *Journal of Open Source Software* 5, no. 51: 2317. https://doi.org/10.21105/joss.02317

Hutton, C., T. Wagener, J. Freer, D. Han, B. Duffy, and B. Arheimer. 2016. Most computational hydrology is not reproducible, so is it really science? *Water Resources Research* 52, no. 10: 7548–7555. https://doi.org/10.1002/2016WR019285

Lamprecht, A.-L., L. Garcia, M. Kuzak, C. Martinez, R. Arcila, E. Martin Del Pico, V. Dominguez Del Angel, S. van de Sandt, J. Ison, P.A. Martinez, P. McQuilton, A. Valencia, J. Harrow, F. Psomopoulos, J.L. Gelpi, N. Chue Hong, C. Goble, and S. Capella-Gutierrez. 2020. Towards FAIR principles for research & software. *Data Science* 3, no. 1: 37–59. https://doi.org/10.3233/DS-190026

Nature Geoscience. 2019. FAIR play in geoscience data. *Nature Geoscience* 12: 961. https://doi.org/10.1038/s41561-019-0506-4

Wilkinson, M.D., M. Dumontier, J. Ij, G. Aalbersberg, M. Appleton, A. Axton, N. Baak, J.-W. Blomberg, L.B. da Boiten, P.E. Silva Santos, J. Bourne, A.J. Bouwman, T. Brookes, M. Clark, I. Crosas, O. Dillo, S. Dumon, C.T. Edmunds, R.F. Evelo, and B. Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018. https://doi.org/10.1038/sdata.2016.18