From Reproducible to Explainable GIScience

Mark Gahegan ✉
School of Computer Science / Centre for eResearch, University of Auckland, New Zealand

Abstract

Communicating deep understanding between humans is key to the effective application and sharing of science, and this is critical in GIScience because much of what we do has practical implications in the modelling and governance of societal and environmental systems. Reproducible and explainable science is needed for public trust, for informed governance, for productivity and for global sustainability [20]. This article summarises some of the more recent research on reproducibility from outside of GIScience, gives practical guidance to current best practice from a GIScience perspective, provides a clearer road-map towards reproducibility and adds in the additional step of explainable GIScience as our final goal.

2012 ACM Subject Classification Information systems → Geographic information systems

Keywords and phrases GIScience, Reproducible, Explainable, discoverable

Digital Object Identifier 10.4230/LIPIcs.GIScience.2023.32

Category Short Paper

1 Introduction

The ‘Reproducibility Crisis’ [2] sent shock waves through both Psychology and Medical Science has changed expectations around how experimental scientists report their research. Apart from the obvious risk of eroding public trust in science if researchers cannot be trusted to behave honestly, reproducibility is critical for two very distinct reasons:

1. For the individual researcher and team, the goal is to discover, access, reuse and build on the work of others, knowing that it can be trusted (efficiency).
2. For the research community as a whole, the goal is to compare new methods, data-sets and theories so we can learn which ones work best, and in what circumstances they can be applied, and to move forward with the best of them (evolution).

Experiments in reproducibility show us that even well-intentioned researchers often fail to provide a complete-enough account of their experiments to allows others to reproduce their results accurately [11]. The bigger issue, then, is not bad actors, but bad practices. The issue has received some good attention of late from the GIScience community [21] including a critical assessment of the reproducibility of GIScience papers published in conference proceedings [14, 15].

Beyond reproducibility is another even more important goal: that of *explainability*. Communicating deep understanding between humans is key to the effective application and sharing of science, and this is critical in GIScience because much of what we do has practical implications in the modelling and governance of societal and environmental systems. Being able to reproduce someone’s research is not enough to ensure it can be successfully repurposed. Repurposing requires that we understand not just the work that was done, but also the situations in which it can be used reliably, and the situations where its underlying assumptions no longer hold. Explainable science is science that can be explored, queried, tested, understood and repurposed, as well as reproduced.

2 The journey to Reproducible and Explainable GIScience

We can view the journey towards explainable science as a series of stepping stones, each one taking us a bit closer. A useful starting point is the concise pathway to reproducibility from the Physiome journal [16]. Their ideas are expanded in Table 1 below, and a GIScience slant added. Explainability was not included in their text, it has been added in here. Note that there are other definitions in use for some of the terms below, this set has strong traction in the wider sciences.

Replicable Re-running the source code produces a result with reported research. In this case, literally a digital replica of the original experiment produces the same answers. *For example*, source code distributed with a research article (runs and) provides exactly the same results as those documented in the article.

Reproducible When research can, by means of an underlying representation based on domain theory (mathematics, logic or a mix of both), be successfully reproduced in some new system. Source code can be ambiguous and opaque. Logic and mathematics is more precise and often provides more clues as to the semantics. *For example*, a new Geographically Weighted Regression method is successfully re-implemented from a set of equations in a published article. Though not efficient, there are benefits from separately re-implementing methods: it demonstrates that the original description of the method is accurate.

Reusable This requires that the model is well documented, the source code is available and that it is licensed for reuse, so that limitations and appropriate use are clear. Licenses do not remove rights, they add them. In most legal jurisdictions, the absence of any statement about reuse of data or code means that no rights whatsoever are extended. See `https://creativecommons.org/licenses/` for details of which licenses to use. *For example*, code and documentation are managed in a software repository such as github (`https://github.com/`) and the program contains a license statement that enables reuse. The OSGeo Docker image `https://wiki.osgeo.org/wiki/DockerImages` library contains over 70 GIScience applications, ready to install, with documentation and licensing information.

Discoverable Research artifacts can be made discoverable via a metadata description of the content that is accessible to a search engine. As research artifacts have moved online, metadata has been increasingly used to describe the 'container' for these artifacts in progressively richer ways. Discovery can be improved by adding in terms that describe the domain and application semantics of artifacts. *For example*, a repository of global landcover maps uses schema(.org) metadata, augmented with the UN's GlobalLand30 international land-cover categories (`https://www.un-spider.org/links-and-resources/data-sources/land-cover-map-globeland-30-ngcc`) to allow content-based search. State-of-the-art for packaging research artefacts for discoverability and reuse is RO-Crate: `https://www.researchobject.org/ro-crate/`

Validated A method can be considered validated when its predictions under specified conditions match experimental observations. In other words, validation requires that we test a model against real-world observations, not just for consistency within own internal logic or mathematics. Models are typically validated within a range of 'safe' operating conditions (such as a scale interval, or between two temperature values). Data-sets can also be validated, or fit-for-purpose. *For example*, a new climate circulation model is validated by several research teams against observed data [18]. We rarely validate in GIScience: we propose new methods, demonstrate that the method works on a test data-set, but push any comparison to future work. Where a comparison is present, it is often very limited. As a community, we have no real sense which methods are better, nor in what circumstances we should, or should not, use them [8].

Explainable Explanation requires that we can interrogate a model to find out more detail, to clarify our understanding, or to test our assumptions. Such questions could target the data, the code, the theory, the workflow, as well as the more mundane aspects such as the software license or the data-sets used and their reliability and suitability.

For example, a model for political redistricting can reveal to the users relevant details of likely bias and quality issues in underlying data and explain the theory behind the analytical methods employed. No examples exist yet in GIScience.

These six aspects of reproducible science are somewhat entangled. For example, a model can be discoverable without being reusable simply because it does not have an appropriate license information to allow the data or code to be reused.

3 Explainable GIScience – a road-map

The first 5 stepping stones above each add in some useful aspects or hints of explanations, for example by a more provable formal description, or by adding in meta-data. But providing a more complete understanding exactly what has been done in a piece of research, and how, and why, remains challenging. Theory may tell us whether a model is valid, but not how or when to use it; semantics help us to share our ideas and concepts, but does not anchor them into our workflows. Explanations require a complex blend of formal theory, semantics and pragmatics [3, 10, 13] for which there is no conveniently simple packaging.

The challenge in building GIScience explanations is the difficulty in 'grounding' geography, that is, to find some scaffolding that is solid enough to build our formal representations upon. The data and concepts used in GIScience are often loaded with complex meaning and abstraction; they can be far removed from physical measurements (though not always). This abstraction also helps explain why it is difficult to come up with laws and theory for GIScience – the data we use are already filtered through so many conceptual lenses that patterns arising from actual measurements are easily lost [9]. So how do we proceed?

3.1 Theory: Connecting Symbolics to Semantics

Symbolic reasoning uses logic, mathematics and other formal theory to represent meaning, with an emphasis on internal consistency and provability, rather than a grounding into semantics. Where the research conducted can be expressed mathematically (e.g. spatial statistical methods), or symbolically using formal logic (some qualitative spatial reasoning methods), then symbolic reasoning provides an excellent grounding into something that is not itself subject to further abstraction – it is foundational. Formal representations seem to have a high currency in the GIScience community, we value the formal grounding of our ideas into symbolic reasoning. (Less so the grounding of our data into suitable ontologies)

But symbolic reasoning by itself is a house of cards. The abstract symbols and functions used are not anchored into any domain semantics, nor into any implementation in a computer program. The reader can often understand them in this way, sometimes with effort, but the process is subjective. Similarly, we can be taught how to translate a symbolic notation from a mathematical form into a computer program. However, translation between a provable abstract representation, the semantics of the domain and the implementation in (say) a program is prone to error. Inconsistencies and translation errors can lead to failures of reproducibility; misunderstanding and confusion can lead to a failure of explanation. For explanations based only on symbolics there may still be a significant semantic gap and there is no guarantee that the code perfectly implements the equations.

The good news is that symbolics can be tied more closely to both domain semantics and to code, as the following example demonstrate.

LinguaPhylo (LPhy) is a framework to precisely define phylogenetic models (as used for example to understand virus evolution). As the authors state: *"We present a new lightweight and concise model specification language, called 'LPhy', that is both human and machine readable. 'LPhy' is accompanied by a graphical user interface for building models and simulating data using this new language, as well as for creating natural language narratives describing such models. These narratives can form the basis of manuscript method sections.."* [6]. The code and model examples are here: `https://linguaphylo.github.io/`. LPhy is a programming language designed specifically for a given domain – its operators are those directly used in the domain – rather than abstract types and methods of a traditional programming language. Behind the scenes, and using some clever markup, LPhy creates English language descriptions of the models a user creates. LPhy essentially provides an immutable mapping between the methods that phylogenetics researchers use, the implementation of these methods in code and human-readable descriptions of the resulting workflow.

There is a useful lesson to learn here. Building a bespoke programming language for a large swath of science or geography problems is intractably hard. But if we take a problem that is small enough to have a consistent epistemology, it is possible to create a domain-oriented programming language that is more consistent, reproducible, self-documenting, and explainable, and that makes programming easier. In GIScience, this idea could be used for geostatistical modelling, or to re-engineer tools such as PySal (`https://pysal.org/`) so that they propel GIScience towards reproducible and explainable goals.

Cao et al [4] demonstrate exactly how geographic processes can be represented using geographic and other foundational ontologies. It is these ontologies, then, that need to form the analytical functions in a GIScience LinguaGeo. Of course, geospatial data can also be connected back into ontologies of observations [12] and from there to ontologies of foundational scientific (SI) measurements [17]. The very same anchoring can be used in the representation of variables representing data in the symbolic logic of our methods.

4 What we can do now to encourage reproducibility in GIScience

Replicable Require the publishing of source code and data by all publications that use them. Encourage journal reviewers to run the code for themselves to establish the truth of the claimed results. Move beyond publishing code to publishing workflows, which also capture additional control flow information.

Reproducible Encourage clear representation of key algorithms in the text of the article. Do not rely solely on source code.

Reusable Insist that all code and data published be made available to other researchers via a permissive license. Ensure that the repositories we use explicitly hold such licensing information (e.g. the OSGeo Docker Repository `https://wiki.osgeo.org/wiki/DockerImages`).

Discoverable Ensure that all data and code are at very least available via a website that is publicly accessible. Use persistent identifiers (DOIs) to ensure longevity. Discoverability is improved by the use of subject-level metadata, so look for repositories that provide this functionality. Even a small amount of metadata is better than none for example see the New York University Spatial Data Repository: `https://geo.nyu.edu/`.

Validated Encourage the validation and comparison of proposed methods, either in the originating article or amongst the wider scholarly community. Use special issues to provide the opportunity for publication of articles that compare methods and that validate published data-sets.

Explainable An open challenge for GIScience is to develop our own LinguaGeo programming language(s) to reduce the gap between code and theory and to automatically generate text descriptions of workflows. In the meantime, we should insist on clear descriptions of methods in text as well as in mathematics or logic. We can also ask for statements that describe any known bias in the data and methods used. For example, if an article examines sentiment analysis in geo-located tweets, what are the socio-demographic biases inherent in these data? Which voices (e.g. ages/genders/ethnicities) are over-represented, and which are not? Where data is being used to train methods, insist that a statement explaining how bias in the data may skew the results obtained. See [19] for more details.

All of these stepping stones, by increasing levels of sophistication, record what was done in precise ways that can survive the process of sharing and so enable researchers to reproduce the findings in a separate computational environment. Some of the responsibility rests with authors, but also some with reviewers and journal editors as well as the scholarly community at large to hold ourselves to a higher standard.

5 The Future: Live and Explainable GIScience?

Perhaps the holy grail of repeatability is a journal article that is itself an executable experiment – that describes an analysis in words, mathematics (or formal logic), semantics and code, but also allows the analysis to be repeated and queried by the reader. A compelling recent example is the *Physiome* journal [16] that encourages authors to submit the analytical models that accompany their more traditional written publications. Physiome evaluates submissions "to determine their reproducibility, reusability, and discoverability. At a minimum, accepted submissions are guaranteed to be in an executable state that reproduces the modelling predictions in an accompanying primary paper, and are archived for permanent access by the community." The journal uses shared method libraries, process and data ontologies (see [5] for more details), common workflow descriptions and packaged data to deliver on its ambitious claims. It is the culmination of many years of collaborative research within a segment of the bioengineering community. A more general solution to this problem that also maintains dynamic links to changing data (thus updating a publication in real time as new data becomes available) is provided by [7].

6 Conclusion

A lot has been said about whether GIScience is in fact a science [22]. The pros and cons of this argument often revolve around whether GIScience has a unique body of theory that might justify the title. And whilst developing theory is important, it is still only one approach to science [1]. Another is experimentation, and GIScience has much to learn from other experimental sciences in terms of how to report science in ways that are reproducible, understandable by others and that can be easily built upon. Or put another way, GIScience would benefit from acting more like a science in the way we conduct and report our experiments! This article describes the pathway to reproducibility and provides a practical summary of improvements we can collectively make now.

References

1 S. T. Arundel and W. Li. *The Evolution of Geospatial Reasoning, Analytics and Modeling In: The Geographic Information Science and Technology Body of Knowledge, John P. Wilson (Ed.).* UCGIS, 2021.

2 M Baker. 1,500 scientists lift the lid on reproducibility. *Nature*, 533:452–454, 2016. doi:10.1038/533452a.

3 M. Bunge. How does it work? the search for explanatory mechanisms. *Philosophy of the social sciences*, 34(2):182–210, 2004.

4 Y.H. Cao, C.J. Yi, and Y.H. Sheng. Geographic process modeling based on geographic ontology. *Open Geosciences*, 10(1):782–796, 2018.

5 M. Clerx, M. Cooling, J. Cooper, A. Garny, K. Moyle, D. Nickerson, P. Nielsen, and H Sorby. Cellml 2.0. *Journal of Integrative Bioinformatics*, 17(2):2020–0021, 2020. doi:10.1515/jib-2020-0021.

6 A. J. Drummond, K. Chen, F. K. Mendes, and D. Xie. Linguaphylo: a probabilistic model specification language for reproducible phylogenetic analyses. *bioRxiv*, 2022.08.08.503246, 2022. doi:10.1101/2022.08.08.503246.

7 A. Ellerm, B. Adams, M. Gahegan, and L. Trombach. Enabling livepublication. In *IEEE 18th International Conference on e-Science, Salt Lake City, UT, USA*. IEEE Computer Society, 2022. doi:10.1109/eScience55777.2022.00067.

8 M. Gahegan. Our gis is too small. *The Canadian Geographer / Le Géographe canadien*, 62:15–26, 2018. doi:10.1111/cag.12434.

9 M. F. Goodchild. Commentary: general principles and analytical frameworks in geography and giscience. *Annals of GIS*, 28(1):85–87, 2022. doi:10.1080/19475683.2022.2030943.

10 R Harris. *The semantics of science.* A-and-C Black, 2005.

11 T. Hothorn and F. Leisch. Case studies in reproducibility. *Briefings in bioinformatics*, 12(3):288–300, 2011.

12 K. Janowicz, A. Haller, S. Cox, D. Phuoc, and M. Lefrancois. Sosa: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics*, 2018. doi:10.2139/ssrn.3248499.

13 S. Levinson. *Pragmatics.* Cambridge University Press, 1983.

14 D. Nüst, C. Granell, B. Hofer, M. Konkol, F.O. Ostermann, R. Sileryte, and V. Cerutti. Reproducible research and giscience: an evaluation using agile conference papers. *PeerJ*, 13(6):p.e 5072, 2018. doi:10.7717/peerj.5072.

15 Frank O. Ostermann, Daniel Nüst, Carlos Granell, Barbara Hofer, and Markus Konkol. Reproducible research and giscience: An evaluation using giscience conference papers. In *International Conference Geographic Information Science*, 2020.

16 Physiome. The physiome journal. *Physiome*, 1(1):1, 2019.

17 H. Rijgersberg, M. Van Assem, and J.L. Top. Ontology of units of measure and related concepts. *Semantic Web*, 4:3–13, 2013.

18 F.J. Tapiador, A. Navarro, V. Levizzani, E. García-Ortega, G.J. Huffman, C. Kidd, P.A. Kucera, C.D. Kummerow, H. Masunaga, W.A. Petersen, and R. Roca. Global precipitation measurements for validating climate models. *Atmospheric Research*, 197:1–20, 2017.

19 N. Turner Lee, P. Resnick, and G. B. Wednesday. *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms.* Brookings Institute, 2019.

20 R. Vicente-Saez, R. Gustafsson, and C. Martinez-Fuentes. Opening up science for a sustainable world: An expansive normative structure of open science in the digital era. *Science and Public Policy*, 48(6):799–813, 2021. doi:10.1093/scipol/scab049.

21 J.P. Wilson, K. Butler, S. Gao, Y. Hu, W. Li, and D.J. Wright. A five-star guide for achieving replicability and reproducibility when working with gis software and algorithms. *Annals of the American Association of Geographers*, 111(5):1311–1317, 2021.

22 D. J. Wright, M. F. Goodchild, and J. D. Proctor. Demystifying the persistent ambiguity of gis as “tool” versus “science”. *The Annals of the Association of American Geographers*, 87(2):346–362, 1997.