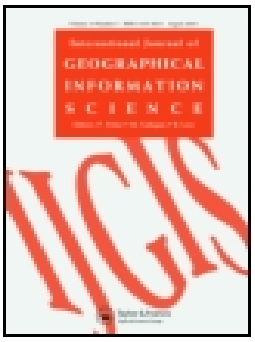

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

# Reproducibility and replicability: opportunities and challenges for geospatial research

Peter Kedron , Wenwen Li , Stewart Fotheringham & Michael Goodchild

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2020.1802032

REVIEW ARTICLE

# Reproducibility and replicability: opportunities and challenges for geospatial research

Peter Kedron[a,b], Wenwen Li[a,b], Stewart Fotheringham[a,b] and Michael Goodchild[b,c]

[a]School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, USA; [b]Spatial Analysis Research Center, Arizona State University, Tempe, AZ, USA; [c]Department of Geography, University of California – Santa Barbara, Santa Barbara, CA, USA

**ABSTRACT**
A cornerstone of the scientific method, the ability to reproduce and replicate the results of research has gained widespread attention across the sciences in recent years. A corresponding burst of energy into how to make research more reproducible and replicable has led to numerous innovations. This article outlines some of the opportunities for geospatial researchers to contribute to and learn from the broader reproducibility literature. We review practices developed in related disciplines to improve the reproducibility and replicability of research and outline current efforts to adapt those practices to geospatial analyses. The article then highlights the open questions, opportunities, and potential new directions in geospatial research related to R&R. We stress that the path ahead will likely require a mixture of computational, geospatial, and behavioral research that collectively addresses the many sides of reproducibility and replicability issues.

**ARTICLE HISTORY**
Received 19 March 2020
Accepted 22 July 2020

**KEYWORDS**
Reproducibility; replicability; geographic Information Science; spatial analysis

## 1. Introduction: a paradox and an opportunity

The ability to reproduce and replicate the work of other researchers has always been an essential part of scientific inquiry (Merton 1973, NASEM 2019). *Reproducibility* – obtaining results consistent with a prior study using the same materials, procedures, and conditions of analysis – and *Replicability* – obtaining consistent findings across studies that aim to answer the same question but with each study collecting and using its own data – are central to the skeptical evaluation of claims, the identification and correction of errors, and the appraisal of scientific explanations (Bollen *et al*. 2015, National Academies of Sciences, Engineering, and Medicine (NASEM) 2019). Repeated replication of a result contributes to the credibility of the underlying claims. When researchers can replicate results across time, space, or populations, they build toward the generalizability of an explanation. Reproducible and replicable research can also accelerate scientific progress by making it easier for researchers to build on the work of others. Across a range of scientific fields, work to improve the reproducibility and replicability (R&R) of research is

**CONTACT** Peter Kedron Peter.Kedron@asu.edu
A detailed discussion of the definition and use of terms can be found in Barba (2018) and Plesser (2018). We adopt the most common definitions and those used by the National Academy of Science, Engineering, and Medicine (2019) in their consensus study report on reproducibility and replicability.

already underway (National Academies of Sciences, Engineering, and Medicine (NASEM) 2019). This article outlines some of the opportunities for geospatial researchers to contribute to and learn from those ongoing efforts.

As with other forms of scientific inquiry, geospatial research currently faces a challenge when it comes to the reproducibility and replicability (R&R) of research. To facilitate self-correction, researchers must document the provenance of data and results and make that information available to others. However, as geospatial research becomes more collaborative, computationally intensive, and data-intensive, it can be challenging to maintain transparency. Geospatial researchers as a community and GIScience as a field are ideally positioned to address this challenge since both groups examine the issues that arise during the production and analysis of geospatial information and the use of geospatial technologies (Goodchild 1992, Duckham *et al*. 2003). It is somewhat surprising then that work on the R&R of geospatial research is only beginning (Brunsdon 2016, Kedron *et al*. 2019, Kedron Forthcoming, Singleton *et al*. 2016). Analyses of the geospatial literature suggest that it is currently challenging to reproduce published work (Konkol *et al*. 2019, Nust *et al*. 2018, Ostermann and Granell 2017), which has led geospatial scholars to call for changes in research and teaching practices (Arribas-Bel & Reades 2018, Brunsdon and Singleton 2015, Holler 2019, Muenchow *et al*. 2019).

This paper is organized into four remaining sections. The following section describes how changes in the practice of science have made it more challenging to reproduce and replicate geospatial research. The third section reviews practices developed in related disciplines to improve the R&R of research and outlines current efforts to adapt those practices to geospatial analysis. The paper then highlights the open questions, opportunities, and potential new directions in geospatial research before concluding with a discussion of the path ahead.

## 2. Collaboration, computation, big data, and scientific paradigms

The ability to reproduce or replicate research minimally requires the existence and availability of the *provenance* of that research – an adequate record of how researchers produced a result. For research involving computation, that record forms a *research compendium* made up of a set of *research artifacts* that include details about research design, data collection and transformation, analytical workflow, and computational environment, along with the original data and code (Nüst *et al*. 2017). Indeed, there is growing recognition that it is the combination of these artifacts, along with the published manuscript, that collectively comprises the scientific contribution of a study (Buckheit and Donoho 1995, Brinckman *et al*. 2019). However, tracking and recording research provenance can be a time-consuming and challenging task, and sharing all elements of a research compendium is not the norm in most disciplines (National Academies of Sciences, Engineering, and Medicine (NASEM) 2019).

At least two changes in the practice of science, broadly, and geospatial research, specifically, have brought the need to track and share research provenance to the forefront. First, research has become a collaborative enterprise that often involves teams of researchers with specialized knowledge working on problems that cross traditional disciplinary boundaries. Labeled *convergence research* by the National Science Foundation (Roco and Bainbridge 2013, Bainbridge and Roco 2016), these collaborations encourage scientists to integrate theories, methods, and data to create new conceptual and analytical frameworks.

Existing at the intersection of geography, information science, computer science, and numerous other disciplines, geospatial researchers have always engaged in convergence research. However, working in teams presents a need to coordinate workflows in a way that can make it challenging to document research provenance, particularly when collaborations cross disciplines. The more people who work on a project, the more opportunities exist for someone to fail to record a change made to data or an analytical decision. When the results of a study depend on combining outputs from instruments or models in which only some members of the team have expertise, it can be similarly challenging to develop a record-keeping strategy or to recognize when team members are failing to fulfil their roles. Similar issues can arise when researchers include groups not traditionally involved in the research process (e.g. citizen scientists). Those groups may lack experience in recording provenance, or their involvement in research may restrict the components of the research compendium that can be shared. At the same time, these challenges may also create the pressure and environment needed to cultivate innovation and develop of practices that foster R&R. Those developing convergence knowledge may recognize that making research reproducible is one way to resolve the communication challenges that inhibit their progress and work toward practice improvements. That process might be facilitated by the co-development of research objectives, plans, and practices by researchers, practitioners, and stakeholders that lies at the heart of convergence research. For example, in geospatial research involving public participation, members of the public might participate with the team in all aspects of the research, from the initial problem definition and planning to the final analysis and inference. These interactions could create an environment for the development of research practices that are domain-spanning and reproducible.

Second, the expanding amount and availability of data coupled with rising computing power have changed traditional research processes. Presented as the third and fourth scientific paradigms by Hey *et al*. (2009), researchers now use computing resources to numerically model complex systems (3rd paradigm), or combine data- and computationally-intensive techniques (e.g. deep learning) to identify patterns in large, complex datasets (4th paradigm). Using computers has moved many of the details related to data acquisition, transformation, and analysis out of publications and into code. Similarly, studies now regularly involve a heterogeneous mixture of computational environments and quantities of data and code that are too large to share through traditional publishing mechanisms. Reproducing and replicating research relies on the tracking and availability of this information, which makes sharing code essential. Even when code is shared, a researcher may find it challenging to recreate the computational environment used during analysis without details about system and software parameters. Technical solutions such as containers, which package code together with libraries and dependencies so they can be run in different computer environments, can help mitigate these issues. However, in many instances, code and information about the computational environment(s) are not shared, as researchers responding to academic incentive systems may view this information as a resource or product and seek to maximize the value they derive from their work through publication (Baker 2016).

More broadly, computation is a tool that has reduced the cost and time to complete many forms of analysis and made it easy to run many alternative analyses. How we use computing power will determine its impact on R&R. As a negative, computation can be used to facilitate specification searching (Leamer 1983, Humphreys *et al*. 2013) and selective reporting (Rosenthal 1979). A specification search occurs when a researcher, seeking to explain a phenomenon through the construction of a model, tests alternative variable combinations and functional forms. It remains common practice for that same researcher to report some, or only one, of these many analyses. Selective reporting means that only a portion of the evidence generated by that researcher is included in the published record. Combined with review-based incentives to publish results that meet certain thresholds (e.g. statistical significance), omitting the alternative results that did not reach those thresholds increases the chances that a reported result is a false-positive (Ioannidis 2005). Geospatial research is susceptible to specification search because spatial analyses require a well-documented set of apriori decisions that set the parameters of analysis (e.g. the scale of analysis, representation of spatial relationships). While those parameters can affect the results of a study, researchers typically do not have complete knowledge of their actual values. This lack of knowledge widens the range of reasonable specifications a researcher might explore.

As a concrete example, consider an analysis of the relationship between urbanization and air quality in 289 Chinese cities conducted by Fang *et al*. (2015). Using a spatial regression approach, the authors model air pollution levels as a function of several urbanization variables. To conduct this analysis, the authors not only needed to select the set and combination of explanatory variables; they also needed to choose a weight matrix that defines the spatial relationships that exist among neighboring cities and which variables are subject to those spatial relationships. Ultimately, Fang *et al*. report single specifications for an ordinary least squares, a spatial lag, and a geographically weighted regression model, each executed with a single spatial weights scheme. While the authors may have only examined this small set of specifications during their analysis, it is also possible that they explored several other variable combinations and spatial weights matrices. Such an approach would be reasonable if the authors did not have strong prior beliefs about the appropriate structure of the spatial weights matrix. However, if multiple matrices were tested, each iteration represents an additional set of hypothesis tests conducted under slightly different functional forms. If only one of these many spatial regression specifications is reported, these many other tests are not included in the literature and cannot be replicated or reproduced. Two points stand out. First, from the published result, we have no way of knowing if other analyses were conducted and not reported. Second, without that knowledge, we cannot adequately assess the published result and, if additional analyses were conducted, we cannot reproduce the research.

Alternatively, the development and increased use of computation in research can improve R&R in at least two ways. First, computational power reduces the effort needed to conduct and record some analysis, which makes it easier to conduct and share exploratory analyses. Extending the above example, if Fang *et al*. did test numerous specifications, adding those analyses to the published knowledge base could be accomplished by sharing the relevant code and data. Second, computational power facilitates repeated analysis, which can also provide insight into why reproductions and replications fail. For example, by repeating the air pollution analysis for different sets of cities, researchers could detect changes in context that impact observed relationships. Similarly, re-running the analysis with minor variations can help researchers identify ambiguities within their workflow that hinder reproduction.

## 3. Existing efforts to improve the reproducibility and replicability

Efforts to improve the R&R of scientific research have taken different forms in different disciplinary and institutional contexts. Nonetheless, all efforts focus on the two central causes of non-reproducibility and non-replicability: the inadequate tracking of research provenance and the need to share that record along with all artifacts of the research compendium. We organize our discussion of existing efforts along these two dimensions and highlight how geospatial researchers are beginning to adapt or mirror these developments.

### *3.1 Practices developed to improve the documentation of research provenance*

Experimental science has a well-established procedure to record and share research information dating back to Francis Bacon and the development of the written article (Stodden *et al*. 2014). However, in areas such as geospatial research, where computation is an essential part of the analysis, it is often difficult to document the provenance of research in a written article alone. As an alternative, this information can be captured and shared as code. Consequently, efforts to address R&R in computationally-intensive fields related to geospatial research have focused on improving the capture of research provenance in code and the other digital artifacts that collectively make up the research compendium of an analysis.

The development of interactive computational notebooks that allow researchers to combine code and descriptive text is one change that has made it easier to record and share code in a format interpretable by those without expert knowledge. Two of the most popular applications supporting computational notebooks, Jupyter Notebooks and RMarkdown, have seen broad adoption within the geospatial research community. CyberGIS-Jupyter (Yin *et al*. 2017) is adapting the Jupyter framework to spatial analysis applications in a scalable, cloud-based computing environment. This effort is one part of a more extensive project to address computational sources of non-reproducibility through the development of scalable cyberinfrastructure and the conceptualization of geospatial software standards (Wang 2010, 2016). Esri's 2019 release of ArcGIS Notebooks (MacDonald and Kalisky 2019) and the development of various GRASS GIS Notebooks (see GRASS Wiki 2020) mirror these efforts and are attempts to improve the reproducibility in commercial and industrial practice.

While computational notebooks facilitate the recording and communication of research provenance, they do not themselves contain the materials (e.g. data) needed to re-execute an analysis. To compile all the digital artifacts and provenance information required to reproduce a computational result together with a snapshot of the computational research environment, researchers have developed programs to generate executable research compendia, colloquially known as 'containers' (Boettiger 2015, Nüst *et al*. 2017). When containers are correctly compiled using applications such as Binder (http://mybinder.org) or WholeTale (http://wholetale.org), researchers can re-execute an initial study under the exact conditions in which the original study was computed with minimal additional effort. Konkol *et al*. (2020) provide a concise review of ten such applications and how to integrate these tools into the academic publication process. Within the geospatial community, the Opening Reproducible Research project (http://o2r.info) is leading the development of open standards and software to create executable research compendia for reproducible research. In a similar stream of work, the Open Geospatial Consortium (OGC, https://www.ogc.org/) continues to develop and release community standards for web-based geospatial data sharing and data processing designed to facilitate interoperability across geospatial processing systems. These efforts are fundamental to R&R because they promote interoperability among distributed and heterogeneous systems through standardization and facilitate the reuse of research artifacts. The ability to reuse research artifacts is the foundation for the cumulative progression of geospatial research and our understanding of geospatial phenomena. By linking code to an API that follows certain standards, the OGC makes it possible for a second user to invoke a function and parse the results. The benefit of doing this is that even if the second user does not have the original code, the user can call a module remotely and reproduce the results. Moreover, this invocation does not require any configuration of the software environment by that second user, because these details are already set in the cloud by the API provider.

While executable notebooks and containers can improve the recording of computational components of research, they may not capture non-computational steps in the broader scientific workflow of a project. As Bowers and Ludäscher (2005) note, tracking scientific workflows requires not only a record of dataflow but also a record of task coordination and conceptual decision-making. Some of this information can be tracked using version control systems like Git (https://git-scm.com/), or through the Open Science Framework (http://osf.io). However, how well those systems map to different forms of computationally-intensive spatial analysis has not been systematically examined and is one avenue for future geospatial research. As one way forward, researchers could test the adequacy of these systems by scrutinizing the provenance records they create for geospatial workflows that involve multiple locations, diverse research groups, and mixed methodologies that use both computational and non-computational analyses. Examining this type of geospatial research may be particularly fruitful because it is likely to present challenges the systems may not have been designed to address. Such workflow studies could be built on an adaptation of the Open Provenance (PROV) Model (https://www.opmw.org/model/OPMW/), which organizes workflow tracking within a formal data model designed to accommodate information produced in heterogeneous research environments. Accommodating agent-, entity-, and activity-centered provenance information, the PROV Model may be well suited to recording geospatial workflows. Outlining and tracking geospatial workflows with the PROV Model offers an opportunity to extend and operationalize research into geospatial ontologies.

As an example, the PROV Model could be used to communicate the workflow procedures captured in the geospatial cyberinfrastructure (GeoCI) platform recently introduced by Shao *et al*. (2020). GeoCI links the web-based and open-source Python Spatial Analysis Library (WebPySAL) with existing geospatial data search engines like PolarHub (Li *et al*. 2016) and allows users to execute spatial analyses using non-local computing resources while also recovering provenance information. Specifically, GeoCI automatically records software versioning, analytical parameters, and metadata following the standards of the

Open Geospatial Consortium and returns that information to users in a bundle with results. Features that are particularly useful for researchers using open software undergoing continuous development. The entity, activity, agent types, and relationship definitions of the PROV Model could be used to organize information recorded by GeoCI as part of a larger workflow that also captures team member roles and actions outside the computer environment. For example, a Moran's I statistic derived from a census dataset (entity) could be linked to the processing functions of WebPySAL (activity), to the research assistant (agent) that executed the procedures, and to the investigator (agent) that oversaw the production. Highlighting some of the key relationships of this workflow, the PROV model could link the Moran's I statistic to the original dataset through the wasDerivedFrom relation, the statistic to the research assistant through wasAttributedTo relation, and the research assistant to the investigator through the actedOnBehalfOf relation. Because the PROV Model can capture information that exists outside of the computational environment, this approach can be extended to more complex spatial analyses. For example, who collected and performed what forms of processing on samples during a field visit could be modeled using the relationships of the model along with any analysis. Interview information could be similarly recorded, as could information about the process of coding and theme extraction common in many qualitative studies.

An alternative approach to the documentation of research provenance is to present in as much detail as possible the steps and decision criteria of a scientific workflow in the form of a pre-analysis plan. No single agreed-upon template exists that outlines the information that should be included in a pre-analysis plan (Glennerster and Takavarasha 2013), but Christensen *et al*. (2019) present a list of ten items around which consensus appears to be forming. This list includes details about study design, sampling procedures, adjustments for multiple hypothesis testing, statistical methods, and a registered timestamp of when the plan was created. Pre-analysis plans perform their function best when they are placed in registries or filed with funding agencies. The pre-registration of experimental design, data protocols, and analysis plans is now the norm in medical research, and this practice is becoming more regular in other fields (Christensen *et al*. 2019). Pre-registration facilitates replicability by limiting specification searching and selective reporting, but presents the negative trade-off of constraining the chances of unexpected and useful results emerging during exploratory analysis. Olken (2015) reviews other positives and negatives of pre-registration for experimental designs, and Dal-Re *et al*. (2014) offer similar treatment for the pre-registration of observational research. Both authors also address practical questions related to the design of pre-analysis plans.

To our knowledge, the geospatial community has yet to deeply explore the possibilities of pre-analysis plans and pre-registration or the practicalities of their implementation. The existence of spatial autocorrelation and spatial non-stationarity in nearly all forms of geographic data and processes will likely place additional demands on any geographic pre-analysis plan (see Anselin *et al*. 2014 for some related results). For example, a geographic pre-analysis plan would likely need to include an explicit statement of the scale(s) used in the analysis to restrict the possibility of MAUP-induced false-positives. Similarly, research using spatial statistical methods to explore or adjust for the impacts of spatial autocorrelation would need to outline and justify the range of spatial weights matrices examined. Geographic analyses also face uncertainty related to the conceptualization, measurement, and representation of phenomena in space. When possible, an estimate of the spatial uncertainty expected in an analysis and how the researchers plan to account for this issue should be included in any pre-analysis plan.

Geospatial researchers may be able to draw inspiration from the computer science community, which has already begun to discuss pre-analysis standards and study registries. The examination by Cockburn *et al*. (2018) of the pre-registration of research that investigates the human–computer interface may be a fruitful starting point for geospatial researchers as the field share features; with studies of spatial cognition. Several psychology journals have taken pre-analysis plans a step further and have adopted a result-blind, peer-review process (Chambers 2013). During this process, the pre-analysis plan is peer-reviewed before the authors undertake any data collection or analysis. If reviewers decide the project has sufficient scientific merit, the research receives in-principle acceptance and is published irrespective of the results as long as the authors follow the original plan. Result-blind peer-review facilitates R&R by simultaneously ensuring the transparency of research decisions and the full reporting of the findings and evidence. The closest practice we are aware of in geospatial research is the peer review of funded research proposals. However, in most cases, the plan of work set out in proposals is not made publicly available, and funding, of course, does not guarantee adherence to the plan or publication of results.

### 3.2 Practices to improve the transparency and availability of research artifacts

Many of the applications and practices created to track and record research provenance also enhance the availability of research artifacts. Github can automatically render any Markdown file produced using an executable notebook and share the detailed history of development through its version-control system. The Open Science Framework provides an open-source project management software that researchers can use to record the provenance of their project. More broadly, digital repositories allow researchers to share the digital artifacts and workflow information. There is no single set of guidelines and standards for the content of digital repositories (Sandve *et al*. 2013, Stodden *et al*. 2014). However, one widely adopted set of criteria is the FAIR standard, which requires that research artifacts be findable, accessible, interoperable, and reusable (Wilkinson *et al*. 2016). Individual repositories generally house discipline- or organization-specific content and can be searched through the Directory of Open Access Repositories (OpenDOAR) and the Registry of Open Access Repositories (ROAR).

Subject searches of OpenDOAR and ROAR for repositories containing the artifacts of geospatial research identified a total of 198 and 105 repositories, respectively, suggesting some level of adoption within the discipline. However, a cursory review of these repositories indicates that they currently appear to primarily house geographic datasets, digital copies of physical artifacts (e.g. scanned copies of maps), and manuscripts. Provenance models, code, and pre-analysis plans appear to be lacking. A systematic evaluation of the contents of digital repositories housing geographic data could clarify what they contain and support an investigation of the extent to which these repositories can be used to facilitate successful reproductions and replications of geospatial research. Such a review does not currently exist.

While repositories facilitate the sharing of research artifacts, their impact on R&R can be limited when privacy and ethical considerations restrict data sharing. It may not be feasible to share data in some areas of geospatial research. Geospatial analyses involving human subjects often requires that collected data not only be anonymized but that data also not be shared to preserve confidentiality. For example, cognitive research focused on understanding spatial reasoning and improving the usability of geographic information systems collects individual performance on mapping tasks (Montello 2005), eye movements (Kiefer *et al*. 2017), and even functional Magnetic Resonance Imaging of brain activity during spatial reasoning tasks (Moen *et al*. 2020) that cannot be easily shared for privacy reasons. Participatory mapping exercises conducted during a study may similarly create datasets that could adversely impact participants if they were shared. The need to protect research subjects can be particularly difficult in a geospatial analysis because even when the names and other identifying characteristics of participants are removed, spatial attributes can often be used to identify research subjects (Armstrong and Ruggles 2005, Giannotti and Pedreschi 2008).

Research artifacts needed to reproduce a result may also be unavailable because of a direct prohibition on sharing by a data provider or research partner. Geospatial researchers studying the spatial organization of industry or patterns of disease often rely on data sourced from companies or healthcare organizations that not only prohibit data sharing but never allow data to leave their secure sites. Stodden (2014) suggests that dual-licensing agreements that distinguish between commercial and research uses of code and data may be one way to overcome industry-imposed restrictions on sharing for competitive reasons. However, these types of agreements may not apply to data whose release is restricted by HIPAA.

When geospatial data sharing is restricted for any of the reasons outlined above, reproduction may not be possible. However, researchers in other disciplines have proposed several different approaches to improve the reproducibility of research that relies on confidential data. It may be possible to adapt those practices for geospatial analysis. When geospatial data is held by a large institution; such as a government agency or large healthcare provider network, one approach is to grant data access to a designated third party to conduct certified reproductions on behalf of the geospatial research community. Perignon *et al*. (2019) outline such a third-party certification scheme being implemented in France by the Certification Agency for Scientific Code and Data (CASCAD, www.cascad.tech). Under this scheme, research conducted using confidential data maintained by the French Statistical Institute and several French ministries can be reproduced and reviewed by CASCAD and awarded a reproducibility certificate backed by these ministries. An author can then include this certificate with their publication as proof of reproducibility. A similar scheme could be implemented with industry partners. Alternatively, geospatial researchers working with industry partners to develop code or analytical procedures could adapt existing systems used to check the robustness and scalability of computer code as part of the commercial software development process.

Another approach is to increase the accessibility of confidential data by creating schemes that grant individual researchers access to selected data stored in repositories. In collaboration with the University of Michigan's Inter-university Consortium for Political and Social Research (ICPSR), Richardson and Kwan (funded by NSF award BCS-1,832,465) are developing standards, practices, and a Geospatial Virtual Data Enclave that will allow individual researchers to access and analyze remotely hosted, confidential spatial data (Richardson 2019). As important, these researchers are creating a credential system that will allow researchers to access different types of restricted data and track their access. These efforts represent a first step toward overcoming an impediment to R&R: access to original, confidential data. However, both the ICPRS and CASD initiatives only address situations in which data is stored in a repository.

Researchers have also developed strategies for situations in which confidential data cannot be shared for ethical reasons. For example, Shepherd *et al*. (2017) propose that researchers release the code used in an analysis, but rather than releasing the original data, create and release a simulated dataset with characteristics that match the original data and an analysis of that dataset using the original code. The authors argue that another researcher could then reproduce the analysis of the matched simulated data using the original code and that this would at least increase the transparency of the analysis and allow for critical evaluation of analytical procedures. In geospatial research, such simulated data could be created to match the spatial structures (e.g. autocorrelation) and attribute relationships in the original data. If a researcher is working within a large team or working with a data-providing organization that has related ongoing collaborations with other researcher groups, an arrangement could be made in which another team member or research group conducts a reproduction to verify the results of the original researcher. This arrangement is an imperfect solution as it raises questions related to the independence of the reproduction and the incentives to undertake this work. Nonetheless, it does offer one possible route toward reproduction.

Even without original data, the assessment of research facing industry or ethical restrictions can be facilitated by focusing on the other dimension of reproducibility: the transparency and sharing of other digital artifacts (e.g. code), details of analysis, and provenance as far as possible under the restrictions of the data-providing partner. Transparency makes it easier for another researcher to compare published results with related studies. When results such as estimated effect sizes are reported along with analytical parameters, meta-analytical techniques can be used to place a result within the broader literature of results and simultaneously add new evidence to the estimate of underlying effects. Similarly, transparency facilitates replication with other organizations or industry partners willing to provide similar data. Examining the economics literature, Coffman and Niederle (2015) argue that even a small number of replication studies that use different data can correct the inaccurate beliefs.

Research data and code that could be shared may not be shared because researchers have little incentive to do so. Researchers may view data and code as a resource that should be sheltered to maximize publication numbers before being released (National Academies of Sciences, Engineering, and Medicine (NASEM) 2019). In this instance, improving R&R requires addressing such perceptions and changing those incentive systems in collaboration with a wide range of stakeholders, including universities, funding institutions, publishing outlets, and practitioner groups. As an example, journal editors can change the standard for publication in their respective journals, which will secondarily affect how scholars practice geospatial research (McNutt 2014, Stodden *et al*. 2018). To move default practices toward reproducibility, editors could minimally require that authors share data, code, and information about computational environments of published work in addition to traditional methodological descriptions. More generally, reproducibility could be a review criterion, and reviewers could be asked to assess whether a study is, in principle, reproducible. Journal editors could also recruit reproducibility editors tasked specifically with assessing reproducibility of submitted work, just as cartographic editors ensure the quality of published figures and maps. Publications could then be assigned ratings or badges, which certify the level of R&R achieved (Kidwell *et al*. 2016). For example, the Association for Computing Machinery (2018) recommends a three-badge system in which research is certified based on the level of evaluation, artifact availability, and whether results have been replicated or reproduced.

Geospatial journals are only beginning to adopt such policies and standards. As a leading example, this journal has a data-sharing policy that requires authors to make data and research artifacts freely available and aligned with FAIR standards. This journal also suggests that authors store data and executable code in a digital repository to facilitate R&R. More broadly, Wilson *et al*. (2020) propose a five-star guide for sharing data and code in geospatial research that could be used as a foundation for the development of similar policies across journals. Modeled after the Berners-Lee (2009) system for publishing open data on the web, the guide is designed to encourage the progressive adoption of R&R practices by researchers.

Geospatial research could benefit from further development and adaptation of these standards to the unique aspects of spatial data analysis. As a start, an initiative within the Association of Geographic Information Laboratories in Europe is developing guidelines to assess the reproducibility of publications, and is creating learning materials to disseminate best practices for achieving the reproducibility of computational geospatial analyses (Nüst *et al*. 2020). A similar initiative here in the United States would be beneficial. However, the impact of any change in submission policies will depend on the response of the geospatial research community. If researchers refuse to share data and code, or simply choose not to publish in outlets with such requirements, institutional changes will have limited impact. A better understanding of community perceptions of sharing requirements and capacity to meet such standards is also needed. If support is not there from the research community, any development aimed at improving R&R will be unlikely to succeed. Equally, if stringent R&R requirements deter researchers from publishing important findings, the loss from implementing an R&R strategy might outweigh the gains.

Finally, a key challenge to fostering R&R will be recognizing when data are being withheld by choice, under the terms of an agreement with a data providing organization, or due to ethical and privacy concerns, and which institutions have influence over the transparency and availability of specific research artifacts. If a geospatial researcher cannot share data for ethical reasons or because an industry partner is restricting the release of data, that research should not be kept from the published record. However, in the latter case, there may be ways to make agreements for limited data sharing with industry partners before analysis. Protocols and templates maintained by university technology transfer offices may be useful as models for such contracts. University foundations that regularly facilitate funded research with private industry partners may be another source of model agreements.

## 4. Emerging questions, open opportunities: the path toward reproducible and replicable geospatial research

Work to improve R&R across the sciences can serve as the building blocks for the improvement of R&R in geospatial research. Building on advances from related fields will be a process of adaptation. To be successful, geospatial researchers will need to consider how the characteristics of spatial data and spatial processes will shape our ability to use the technologies and practices developed in other fields. This will be no small task, as how to conceptualize R&R in geospatial research, how to account for the common characteristics of spatial analysis, and how to incorporate R&R into geospatial training remain open questions.

### *4.1 Conceptualizing R&R in geospatial research*

If making research more reproducible and replicable is to improve the production of geospatial knowledge, it is critical to define and identify the role we expect R&R to play in different types of geospatial research. While it may be reasonable to expect geospatial analysis to be reproducible, it is less clear whether or not geospatial research should be replicable. When does a change in the location, time, or context of a second study represent enough of a difference that we should no longer expect a result to replicate? A related practical question is how do we determine the consistency of results of studies addressing the same problem using different data and methods, possibly collected from different locations. How changes in spatial context affect the operation of spatial processes has been a question at the heart of geography since at least the Hartshorne-Schaefer debate. As the scientific mechanism used to test and re-test results, R&R can be directly linked to this debate and used to empirically examine this long-standing question (Sui and Kedron Forthcoming).

In the geospatial sciences, it is the failure to replicate across space and time that is most in need of a formal framework (Goodchild *et al*. Forthcoming). However, creating a single uniform standard applicable across the discipline will be difficult for at least three reasons. First, some amount of natural variation exists in any geospatial system, and any formal framework defining replication across space and time should account for that variation in its definition of consistency. Second, geospatial researchers study a range of systems that vary in complexity and controllability. The more complex and less controllable a system, the more difficult it is to estimate its natural variation and create a definition of consistency. Third, geospatial researchers use different approaches and methods to examine the systems they study, and it is not clear how to construct standards of comparison across approaches. Even in cases where the definition of consistency focuses on quantitative measures such as effect size and variance, estimates produced using different techniques may not rest on the same system of inference (e.g. machine learning versus conventional regression), complicating their comparison.

A related set of challenges that will need to be addressed during the development of any R&R framework for geospatial research is the need to share the conceptualizations and situational influences that become embedded and fixed in data and code during the research process. While it may be possible to reproduce or replicate a research finding with data and code alone, this will add limited value to the body of knowledge without a clear understanding of the conceptual underpinnings of the research and the specific situations that may affect their objectivity (Pickles 1995, 1999, Shuurman 2008). Sharing the conceptualization of research has been traditionally achieved by the written article, which makes clarifying and preserving links between data, code, and the written explanation of research design and decision making essential to understanding the implications of a reproduction or a replication. Recent technical developments, such as computational notebooks, bindings, and executable research papers, offer new ways to link the written article with code and data. However, these developments primarily improve our capture of what Gahegan and Pike (2006) label the syntax of knowledge production – the mechanics of representation and encoding of knowledge. A complete description of the research process would also capture and share the conceptual structures of research (semantics) and the surrounding situations (pragmatics). These aspects of the research process warrant further attention.

Geospatial ontologies that formalize conceptualizations of geospatial entities and the relationships among them are one means of capturing semantic information. To facilitate R&R, it would be fruitful to develop domain-specific ontologies of geospatial phenomena further but to also link those ontologies with the ontologies and structures of the provenance models and open data platforms that act as the means of sharing geospatial data and code. However, ontological research alone does not address the situational constraints on the creation of geospatial knowledge. One path toward gathering and sharing the situational information that may impact R&R is to expand the study of and development of database ethnographies (Schuurman 2008) that link context to data. More broadly, it may be useful to consider critical portions of data or code created during a study as boundary objects (Star and Griesemer 1989) that temporarily stabilize a concept and allow for communication between groups. This approach would acknowledge, and at least partially record, how concepts are operationalized and help us track conceptual and operational changes through time. Recognizing and developing systems capable of accounting for the situational nature of research and the instability of concepts and definitions as fields evolve remains a crucial challenge to not only R&R but to the reuse of the code and data (Gupta and Gahegan 2020).

A framework for replication can be built by addressing these challenges. One key to progress will be recognizing that the primary purpose of replication is not identifying whether a single study is replicable, but whether the entire set of studies addressing the same question collectively point toward the same answer. In this framing of replication, we can think of any particular study as a single data point within a collection of related studies. It may also be useful for GIScience to develop an assessment framework for individual studies that recognize replicability not as a binary success-failure measure, but as existing along a spectrum. Such a structure would not only adjust for characteristics of the system under study but also assess the degree to which different dimensions (e.g. magnitude, direction) of a replication study are consistent with the results of a set of related studies. If we allow space to be a key variable across the set of studies, this approach creates an opportunity to examine the accumulation and strength of evidence about a question across locations and to use triangulation through the application of different approaches and methods to establish degrees of belief. However, any framework for replication in geospatial research must be supported by the development of practices that capture and communicate how the geospatial phenomena under examination are conceptualized and how the operationalization of those conceptualizations is linked to situational influences. Workshops held at Arizona State University (http://osf.io/gvp3q) and the University of Arkansas (https://cast.uark.edu/events/giscience.php) have begun the effort to synthesize comparable approaches in related disciplines.

### 4.2 Considering the characteristics of geospatial analysis

Research can only be reproduced when information about how an analysis was conducted and the data in that analysis are available. In geospatial research, the details necessary for reproduction include information about how spatial processes were conceptualized, measured through spatial data, and then operationalized during analysis. Because these decisions are imperfect, different forms of uncertainty will enter into any geospatial analysis. To the extent possible, researchers should also address what action they took to account for the spatial uncertainties. The effects of uncertainty and the operationalization of spatial processes are well-documented within the literature. However, these geospatial issues do not appear in the badge systems, publication guidelines, or pre-analysis plan templates used in other fields.

Building publication guidelines, badge systems, and planning templates around the main characteristics of spatial processes (e.g. spatial non-stationarity and spatial dependence) would be a practical step toward developing a formal framework for replication in geospatial research. Greater clarity on how geospatial analyses are constructed will allow others to better set expectations about whether a result should replicate across space. For example, if a researcher selected a weight matrix to capture a dispersion process that is constrained by some physical law around fixed source locations, a second researcher might reasonably expect that process, and the weight matrix that represents it, to be applicable in other places subject to the interference of other confounding effects. However, if a researcher selected a weight matrix to represent a diffusion process through a socio-spatial network, a second researcher may not expect that same matrix to apply in another location. Building guidelines that outline how this type of information should be reported, or applications that automate their recording, would be a practical way to improve the R&R of geospatial research.

### 4.3 Improving R&R through education and training

Ensuring lasting change that will sustain a culture of R&R in geospatial research requires not only cultivating an expectation that research is reproducible but training the next generation of scientists and practitioners to work in reproducible ways. There is an opportunity to develop new geospatial curricula at both the undergraduate and graduate levels to educate students about the importance of sharing and communicating research information by focusing instruction on the reproduction of prior research. Just as the lab sciences teach concepts by reproducing seminal experiments, educators can teach critical lessons and techniques by reproducing results from the geospatial literature. Geospatial problems offer fertile ground for educators to collaborate with students to examine what forms of reproducibility can be expected when analyzing different problems using different methods and tools. Combined with open platforms such as R and QGIS, training students and practitioners to do reproducible and replicable work creates an opportunity to enliven perhaps unexpected areas of research. For example, Holler (2019) argues that teaching open and reproducible research practices creates new lines of inquiry within critical GIScience because reproducible work expands our ability to critically scrutinize all aspects of the research process and resituate that process in different conceptual and methodological frameworks. At the same time, student researchers pursuing replications can fill the need for replications of geospatial research while also learning fundamental concepts and practices. If well documented, this type of educational experience may also produce a series of case studies that could be used to communicate key challenges to R&R in geospatial research and solutions developed to address them.

## 5. The path ahead

Making geospatial research more reproducible and replicable depends on identifying how computation, collaboration, and data availability make it challenging to record and share provenance, and then developing the science and scientific practices needed to create those results. This is a task for GIScience, but not one the field will have to start from scratch. Geospatial researchers can begin to develop the necessary science by building on the conceptual, technical, and institutional advances developed in related fields of study. As the first step in this direction, this article alerts readers to those advances and highlights the early work of those in the geospatial research community. As a second step, this article outlines opportunities and challenges to building on those advances in geospatial research.

Opportunities to expand upon existing work abound, but researchers seeking to contribute to this emerging research area must consider the special nature of R&R in geospatial studies. As a field, we have yet to set expectations about the R&R of geospatial research in space and time, define how to assess R&R, or identify the practices we expect to improve R&R. Recent actions taken by this journal that require authors to share data and code are encouraging first steps, as are attempts to create publication guidelines and platforms to securely share confidential geospatial data. Building on these projects can open a path toward more efficient geospatial knowledge production and the more effective application of that knowledge to real-world problem-solving. At the same time, creating the behavioral changes necessary for a widespread shift toward R&R in geospatial research will be a challenge. The availability of tools and techniques that facilitate R&R research practices does not imply their adoption, and working in an R&R manner does not necessarily mean that research results will be credible or reliable. It is possible to imagine a negative case where expanding R&R requirements, in fact, disincentivizes the undertaking insightful research, particularly exploratory research. This need not be the case if we attend to all sides of the issue. Creating better research practices and better research will depend on moving forward with a balanced approach that addresses the technical issues rooted in computational science, conceptual issues rooted in geospatial problem solving, and practical issues rooted in the behavioral and social sciences.

## Data and codes availability statement

Data sharing is not applicable to this article as no new data were created nor analyzed in this study.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Notes on contributors

***Peter Kedron*** is Assistant Professor in the School of Geographical Sciences and Urban Planning and a core faculty member of the Spatial Analysis Research Center at Arizona State University. Email: Peter.Kedron@asu.edu. His research focuses on understanding how different processes create persistently uneven spatial patterns, developing new methods of spatial analysis, and using those methods to address economic, social, and environmental problems.

***Wenwen Li*** is Associate Professor in GIScience in the School of Geographical Sciences and Urban Planning at Arizona State University. Email: wenwen@asu.edu. Her research interests include cyberinfrastructure, geospatial big data, and machine learning, and their applications in data-intensive environmental and social sciences.

***Stewart Fotheringham*** is Professor of Computational Spatial Science and Director of the Spatial Analysis Research Center (SPARC) in the School of Geographical Sciences and Urban Planning, Arizona State University. Email: stewart.fotheringham@asu.edu. His research interests include spatial analysis, local statistical modeling, spatial interaction modeling, and GI Science.

***Michael Goodchild*** is Professor Emeritus in the Department of Geography at the University of California, Santa Barbara. E-mail: good@geog.ucsb.edu. His research interests include geographic information science and systems, spatial data science, and uncertainty in geographic information.

## ORCID

Peter Kedron http://orcid.org/0000-0002-1093-3416
Wenwen Li http://orcid.org/0000-0003-2237-9499

## References

Anselin, L., Rey, S.J., and Li, W., 2014. Metadata and provenance for spatial analysis: the case of spatial weights. *International Journal of Geographical Information Science*, 28 (11), 2261–2280. doi:10.1080/13658816.2014.917313.

Armstrong, M.P. and Ruggles, A.J., 2005. Geographic information technologies and personal privacy. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 40 (4), 63–73. doi:10.3138/RU65-81R3-0W75-8V21.

Arribas-Bel, D. and Reades, J., 2018. Geography and computers: past, present, and future. *Geography Compass*, 12 (10), e12403. doi:10.1111/gec3.12403.

Association for Computing Machinery (ACM), 2018. *Artifact review and badging* [online] Available from: https://www.acm.org/publications/policies/artifact-review-badging.

Bainbridge, W.S. and Roco, M.C., eds., 2016. *Handbook of science and technology convergence*. Switzerland: Springer International Publishing.

Baker, M., 2016. 1,500 scientists lift the lid on reproducibility. *Nature News*, 533 (7604), 452. doi:10.1038/533452a.

Barba, L. 2018. Terminologies for reproducible research. *arXiv preprint: 1802.03311*.

Berners-Lee, T. 2009. *Linked Data: design issues* [online] Available from: https://www.w3.org/DesignIssues/LinkedData.html

Boettiger, C., 2015. An introduction to Docker for reproducible research, with example from the R environment. *ACM SIGOPS Operating Systems Review*, 49 (1), 71–79. Special Issue on Repeatability and Sharing of Experimental Artifacts. doi:10.1145/2723872.2723882.

Bollen, K., *et al*. 2015. Social, behavioral, and economic sciences perspectives on robust and reliable science. Report of the subcommittee on replicability in science advisory committee to the national science foundation directorate for social. *Behavioral, and Economic Sciences*. Available from: https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf.

Bowers, S. and Ludäscher, B. 2005. Towards automatic generation of semantic types in scientific workflows. In *International Conference on Web Information Systems Engineering* (pp. 207–216). Springer, Berlin, Heidelberg.

Brinckman, A., *et al*., 2019. Computing environments for reproducibility: capturing the "Whole Tale". *Future Generation Computer Systems*, 94, 854–867. doi:10.1016/j.future.2017.12.029

Brunsdon, C., 2016. Quantitative methods I: reproducible research and quantitative geography. *Progress in Human Geography*, 40 (5), 687–696. doi:10.1177/0309132515599625.

Brunsdon, C. and Singleton, A., 2015. Reproducible research: concepts, techniques and issues. *In*: C. Brunsdon and A. Singleton, eds. *Geocomputation: a practical primer*. London: SAGE, 254–264.

Buckheit, J.B. and Donoho, D.L., 1995. Wavelab and reproducible research. *In*: A. Antoniadis, ed. *Wavelets and statistics*, of Lecture Notes in Statistics, Vol. 103. New York: Springer, 55–81. doi:10.1007/978-1-4612-2544-7_5.

Chambers, C.D., 2013. Registered reports: a new publishing perspective at Cortex. *Cortex*, 49 (3), 609–610. doi:10.1016/j.cortex.2012.12.016.

Christensen, G., Freese, J., and Miguel, E., 2019. *Transparent and reproducible social science research: how to do open science*. Oakland: University of California Press.

Cockburn, A., Gutwin, C., and Dix, A., 2018. HARK no more: on the preregistration of CHI experiments. In *Proceedings of the 2018*. CHI Conference on Human Factors in Computing Systems, Montreal, Canada, 1–12.

Coffman, L. and Niederle, M., 2015. Pre-analysis plans have limited upside, especially when replications are feasible. *Journal of Economic Perspectives*, 29 (3), 81–98. doi:10.1257/jep.29.3.81.

Dal-Re, R., *et al*., 2014. Making prospective registration of observational research a reality. *Science Translational Medicine*, 6, 224cml. doi:10.1126/scitranslmed.3007513.

Duckham, M., Goodchild, M.F., and Worboys, M., 2003. *Foundations of geographic information science*. London: Taylor and Francis.

Fang, C., *et al*., 2015. Estimating the impact of urbanization on air quality in China using spatial regression models. *Sustainability*, 7 (11), 15570–15592. doi:10.3390/su71115570.

Gahegan, M. and Pike, W., 2006. A situated knowledge representation of geographical information. *Transactions in GIS*, 10 (5), 727–749. doi:10.1111/j.1467-9671.2006.01025.x.

Giannotti, F. and Pedreschi, D., eds., 2008. *Mobility, data mining and privacy: geographic knowledge discovery*. Berlin: Springer Science & Business Media.

Glennerster, R. and Takavarasha, K., 2013. *Running randomized evaluations: A practical guide*. Princeton: Princeton University Press.

Goodchild, M.F., 1992. Geographical information science. *International Journal of Geographical Information Systems*, 6 (1), 31–45. doi:10.1080/02693799208901893.

Goodchild, M.F., *et al*., Forthcoming. Introduction: forum on reproducibility and replicability in Geography. *Annals of the Association of American Geographers*.

GRASS Wiki. 2020. *GRASS GIS Jupyter notebooks* [online]. GRASS. Available from: https://grasswiki.osgeo.org/wiki/GRASS_GIS_Jupyter_notebooks.

Gupta, P. and Gahegan, M., 2020. Categories in flux, but their computational representations are fixed: that's a problem. *Transactions in GIS*, 24 (2), 291–314. doi:10.1111/tgis.12602.

Hey, T., Tansley, S., and Tolle, K., eds., 2009. *The fourth paradigm: data-intensive scientific discovery*. Redmond: Microsoft research.

Holler, J., 2019. Teaching critical open GIS. *The Canadian Geographer/Le Géographe Canadien*. doi:10.1111/cag.12521.

Humphreys, M., De la Sierra, R.S., and Van der Windt, P., 2013. Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis*, 21 (1), 1–20. doi:10.1093/pan/mps021.

Ioannidis, J.P., 2005. Why most published research findings are false. *PLOS Medicine*, 2 (8), e124. doi:10.1371/journal.pmed.0020124.

Kedron, P., *et al*., 2019. Reproducibility and replicability in geographical analysis. *Geographical Analysis*. doi:10.1111/gean.12221.

Kedron, P., Forthcoming. Replicability and reproducibility. *In*: D. Richardson, *et al*., eds. *International encyclopedia of geography: people the earth, environment, and technology*. Malden: Wiley Blackwell AAG.

Kidwell, M.C., *et al*. 2016. Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, 14 (5), e1002456. doi:10.1371/journal.pbio.1002456.

Kiefer, P., *et al*., 2017. Eye tracking for spatial research: cognition, computation, challenges. *Spatial Cognition & Computation*, 17 (1–2), 1–19. doi:10.1080/13875868.2016.1254634.

Konkol, M., Kray, C., and Pfeiffer, M., 2019. Computational reproducibility in geoscientific papers: insights from a series of studies with geoscientists and a reproduction study. *International Journal of Geographical Information Science*, 33 (2), 408–429. doi:10.1080/13658816.2018.1508687.

Konkol, M., Nüst, D., and Goulier, L. (2020). Publishing computational research–A review of infrastructures for reproducible and transparent scholarly communication. *arXiv preprint arXiv:2001.00484*.

Leamer, E., 1983. Model choice and specification analysis. *In*: Z. Griliches and M.D. Intriligator, eds. *Handbook of Econometrics*. Amsterdam: Elsevier, 285–330.

Li, W., Wang, S., and Bhatia, V., 2016. PolarHub: A large-scale web crawling engine for OGC service discovery in cyberinfrastructure. *Computers, Environment and Urban Systems*, 59, 195–207. doi:10.1016/j.compenvurbsys.2016.07.004

MacDonald, S. and Kalisky, S. 2019. *Introducing ArcGIS notebooks* [online]. ESRI. Available from: https://www.esri.com/arcgis-blog/products/arcgis-enterprise/analytics/introducing-arcgis-notebooks/

McNutt, M., 2014. Journals unite for reproducibility. *Science*, 346 (6210), 679. doi:10.1126/science.aaa1724.

Merton, R.K., 1973. *The sociology of science: theoretical and empirical investigations*. Chicago: Chicago University Press.

Moen, K., *et al*. 2020. Strengthening spatial reasoning: elucidating the attentional and neural mechanisms associated with mental rotation skill development. *Cognitive Research*, 5, 20.

Montello, D., 2005. Navigation. *In*: P. Shah and A. Miyake, eds. *The Cambridge handbook of visiospatial thinking*. Cambridge: Cambridge University Press, 257–294.

Muenchow, J., Schäfer, S., and Krüger, E., 2019. Reviewing qualitative GIS research—Toward a wider usage of open-source GIS and reproducible research practices. *Geography Compass*, 13 (6), e12441. doi:10.1111/gec3.12441.

National Academies of Sciences, Engineering, and Medicine (NASEM), 2019. *Reproducibility and replicability in science*. Washington, DC: The National Academies Press. doi:10.17226/25303

Nüst, D., *et al*. 2017. Opening the publication process with executable research compendia. *D-Lib Magazine*, 23 (1/2). Available from: http://www.dlib.org/dlib/january17/nuest/01nuest.html.

Nüst, D., *et al*., 2018. Reproducible research and GIScience: an evaluation using AGILE conference papers. *PeerJ*, 6, e5072. Available from: https://peerj.com/articles/5072/.

Nüst, D., *et al*., 2020. AGILE Reproducible Paper Guidelines. Available from: 10.17605/OSF.IO/CB7Z8.

Olken, B.A., 2015. Promises and perils of pre-analysis plans. *Journal of Economic Perspectives*, 29 (3), 61–80. doi:10.1257/jep.29.3.61.

Ostermann, F.O. and Granell, C., 2017. Advancing science with VGI: reproducibility and replicability of recent studies using VGI. *Transactions in GIS*, 21 (2), 224–237. doi:10.1111/tgis.12195.

Perignon, C., *et al*., 2019. Certify reproducibility with confidential data. *Science*, 364 (6449), 127–128. doi:10.1126/science.aaw7086.

Pickles, J., 1995. Representations in an electronic age: geography, GIS, and democracy. *In*: J. Pickles, ed. *Ground truth: the social implications of geographic information systems*. Guilford: New York, 1–30.

Pickles, J., 1999. Arguments, debates, and dialogues: the social theory debate and the concern for alternatives. *In*: P. Longley, *et al*., eds. *Geographical information systems: principles, techniques, applications, and management*. Second ed. New York: John Wiley and Sons, 49–60.

Plesser, H., 2018. Reproduciblity vs. replicability: A brief history of a confused terminology. *Frontiers on Neuroinformatics*, 11, 76. doi:10.3389/fninf.2017.00076

Richardson, D., 2019. Dealing with geoprivacy and confidential geospatial data crossing borders. *ARC News*, 41, 2.

Roco, M.C. and Bainbridge, W.S., eds., 2013. *Converging technologies for improving human performance: nanotechnology, biotechnology, information technology and cognitive science*. Dordrecht: Springer Science & Business Media.

Rosenthal, R., 1979. The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86 (3), 638–641. doi:10.1037/0033-2909.86.3.638.

Sandve, G.K., *et al*., 2013. Ten simple rules for reproducible computational research. *PLOS Computational Biology*, 9 (10), e1003285. doi:10.1371/journal.pcbi.1003285.

Schuurman, N., 2008. Database ethnographies using social science methodologies to enhance data analysis and interpretation. *Geography Compass*, 2 (5), 1529–1548. doi:10.1111/j.1749-8198.2008.00150.x.

Shao, H., *et al*., 2020. When spatial analytics meets cyberinfrastructure: an interoperable and replicable platform for online spatial-statistical-visual analytics. *Journal of Geovisualization and Spatial Analysis*, 4 (2). (in press). doi:10.1007/s41651-020-00056-5.

Shepherd, B.E., *et al*., 2017. A pragmatic approach for reproducible research with sensitive data. *American Journal of Epidemiology*, 186 (4), 387-392. doi:10.1093/aje/kwx066.

Singleton, A.D, Spielman S, and Brunsdon, C., 2016. Establishing a framework for open geographic information science. *International Journal of Geographical Information Science*, 30 (8), 1507–1521.

Star, S. and Griesemer, J., 1989. Institutional ecology, 'translations', and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-1939. *Social Studies of Science*, 19 (3), 387–430. doi:10.1177/030631289019003001.

Stodden, V., 2014. Enabling reproducibility in big data research: balancing confidentiality and scientific transparency. *In*: J. Lane, *et al*., eds. *Privacy, Big Data, and the Public Good: frameworks for Engagement*. New York: Cambridge University Press. 112–132.

Stodden, V., Leisch, F., and Peng, R.D., eds.. 2014. *Implementing reproducible research*. Boca Raton, Florida: CRC Press.

Stodden, V., Seiler, J., and Ma, Z., 2018. An Empirical Analysis of Journal Policy Effectiveness for Computational Reproducibility. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (11), 2584–2589. doi:10.1073/pnas.1708290115.

Sui, D. and Kedron, P., Forthcoming. Reproducibility and replicability in the context of the contested identities of Geography. *Annals of the Association of American Geographers*.

Wang, S., 2010. A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers*, 100 (3), 535–557. doi:10.1080/00045601003791243.

Wang, S., 2016. CyberGIS and spatial data science. *GeoJournal*, 81 (6), 965–968. doi:10.1007/s10708-016-9740-0.

Wilkinson, M.D., *et al*. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3 (1), 160018. doi:10.1038/sdata.2016.18.

Wilson, J., *et al*., Forthcoming. A five-star guide for achieving replicability and reproducibility when working with GIS software and algorithms. *Annals of the Association of American Geographers*.

Yin, D., *et al*. 2017. A CyberGIS-Jupyter framework for geospatial analytics at scale. In *Proceedings of the Practice and Experience in Advanced Research* Computing *2017 on Sustainability, Success and Impact* (p. 18). New Orleans, LA, USA: ACM.