Advancing Science with VGI: Reproducibility and Replicability of Recent Studies using VGI

Frank O. Ostermann*

Faculty of Geo-Information Science and Earth Observation (ITC)

University of Twente

PO Box 217, 7500 AE Enschede, The Netherlands

f.o.ostermann@utwente.nl

Carlos Granell

Institute of New Imaging Technologies

Universitat Jaume I of Castellón

E-12006 Castellón, Spain

carlos.granell@uji.es

* Corresponding author

Abstract.

In scientific research, reproducibility and replicability are requirements to ensure the advancement of our body of knowledge. This holds true also for VGI-related research and studies. However, the characteristics of VGI suggest particular difficulties in ensuring reproducibility and replicability. In this paper, we aim to examine the current situation in VGI-related research, and identify strategies to ensure realization of its full potential. To do so, we first investigate the different aspects of reproducibility and replicability and their impact on VGI-related research. These impacts are different depending on the objectives of the study. Therefore, we examine the study focus of VGI-related research to assess the current body of research and structure our assessment. This work is based on a rigorous review of the elements of reproducibility and a systematic mapping and analysis of 58 papers on the use of VGI in the crisis management field. Results of our investigation show that reproducibility issues related to data are a serious concern, while reproducibility issues related to analysis methods and processes face fewer challenges. However, since most studies still focus on analyzing the source data, reproducibility and replicability are still an unsolved problem in VGI-related research. Therefore, we show initiatives tackling the problem, and finally formulate strategies to improve the situation.

1 Motivation and problem statement

In the past two decades, many scientific disciplines have been affected by the improved availability of data. The development has been described using terms like "data deluge", "information flood", or the widely used "Big Data". Furthermore, it has sparked a new scientific paradigm, that of data-driven science (Hey et al. 2009). While the majority of data in terms of volume is created by electronic sensors (e.g. earth observation through remote sensing, experimental physics using particle accelerators, biotechnology analyzing genomes, to name just a few), a significant part is created by direct human input. Excluding the automatic collection of user data like credit card transactions or web surfing behavior, the user-generated content in social networks, photograph or video sharing platforms, and crowdsourced data collection provides semantically rich, up-to-date information on many phenomena that are of interest to the public in general, but also to specific stakeholders (Craglia and Shanley 2015). Location-based service providers try to use it to refine the information they can offer to their clients, planners and governments attempt to improve participatory processes, and scientists search for ways to work effectively together with citizen scientists. In Geographic Information Science (GIScience), user-generated content that has a geographic component is known under several terms, such as user-generated geographic content (Craglia et al. 2012), contributed geographic information (Harvey 2013), ambient geographic information (Stefanidis et al. 2012), or volunteered geographic information (Goodchild 2007). In this paper, we will use the established term of VGI, but understand it to include also information that is not explicitly geographic or volunteered, e.g. social media containing place names or coordinates.

Data and information generated by "human sensors" (Goodchild 2007) require different approaches to calibration and error correction than electronic sensors. The various aspects of VGI uncertainty (vagueness of categories, imprecision of coordinates, unknown lineage of data) are often difficult to ascertain and assess, which has been the prime concern of potential users of VGI, well before issues of privacy or liability. This is reflected by research on VGI that focused on data handling and processing

(Roick and Heuser 2013; Neis and Zielstra 2014), e.g. issues of collection, storage, and quality assessment. Many approaches exist, including social network analysis (Cheong and Cheong 2011), supervised machine learning (Ostermann and Spinsanti 2012), geographic contextualization (Spinsanti and Ostermann 2013), or manual peer-reviewing (Liu 2010). While not all results have yet found their way into application, the trend is promising and - while many challenges to data quality remain - it seems that significant progress is possible.

We think therefore that it is necessary to broaden our scope again, to reflect on the progress made so far, and focus on an underexplored issue, that of the impact of using VGI on the basic principles of scientific investigation, and thereby on the validity of VGI-related research. We think that the search for general principles and transferable knowledge is fundamental to science. A cornerstone of scientific validity and scientific method is replicability or reproducibility of study results: Any single occurrences that cannot be reproduced bear little significance to science (Popper 1992). In other words, any effect that cannot successfully be replicated with another data sample is possibly idiographic and hence not suitable for describing general principles.

While this is not an entirely new issue, it is of particular importance for data- and algorithm-driven science. It has already been acknowledged in several scientific disciplines. International initiatives like the Research Data Alliance[1] (RDA) are supporting activities to address replicability and reproducibility challenges, e.g. those developed by RDA working groups such as Data Citation[2], Data in Context[3] and Reproducibility[4]. There is also concern about the reproducibility of studies in particular scientific domains such as psychology, where a crowdsourced effort (the Open Science Framework's Reproducibility

---

[1] https://www.rd-alliance.org
[2] https://www.rd-alliance.org/groups/data-citation-wg.html
[3] https://www.rd-alliance.org/groups/data-context-ig.html
[4] https://www.rd-alliance.org/groups/reproducibility-ig.html

Project[5]) involved 270 authors who attempted to reproduce 100 studies, with alarming results. The problem is thus being recognized, with calls for journals to unite for reproducibility (McNutt 2014), or computer and climate science, where a Science Code Manifesto lists good practice[6].

To our knowledge, there has been no systematic investigation so far on the state of reproducibility and replicability of VGI-related research, and its potential to actually move science forward. We argue that it is about time to reflect on the characteristics of VGI-related research and examine the impact of using VGI as a data source on reproducibility and replicability. It is important to note here that we do not intend to advocate a return to a strictly positivistic approach which deems unsuitable for capturing and describing the richness of human geospatial and environmental interactions. We acknowledge that many phenomena are likely to be too complex or entropic to be captured in formulae and described in deterministic laws. However, there is reason to assume that VGI poses particular challenges to reproducibility and replicability, which require adapted research strategies to mitigate.

This paper's objective is to provide a first step by identifying the potential and challenges for reproduction or replication within the VGI-related body of research. We argue that the issue of reproducibility and replicability of VGI-related research is of particularly pressing concern for the advancement of GIScience, but we will relate our finding to the broader context of scientific research and discuss some data publishing strategies including open data.

To examine the current situation, we first develop a simple classification scheme to assess the different aspects of reproducibility and replicability. We then apply this classification in a systematic analysis of over 58 papers using VGI as a data source. This qualitative study allows us to describe the past and current trends regarding reproducibility, replicability, and from there to identify open issues, and to

---

[5] https://osf.io/ezcuj/
[6] http://sciencecodemanifesto.org/

develop suggestions for strategies to improve on any deficits. Additionally, this investigation allows us to assess the analysis objectives of the examined studies and discuss in which regard VGI has contributed to advancing the body of GIScience knowledge.

The research questions that have guided this study are:

1. How can we measure reproducibility and replicability in VGI studies?
2. What is the situation on reproducibility and replicability in VGI studies, and are there any past or future trends?
3. What is the focus of the VGI studies, and how do they contribute to GIScience?

The structure of the paper is as follows: In the following section, we briefly discuss relevant work and develop our measures for reproducibility, replicability and type of study focus. In the section thereafter, we apply our measures and classification on a carefully selected dataset of VGI publications. In the penultimate section, we discuss our results, while the last section concludes and looks forward.

2 Replicability, reproducibility or nothing?

In this section, we approach the first research question, by examining relevant literature on the issues of scientific reproduction and replication, and developing the criteria with which we investigate the state of the art in a sample of current literature on VGI (see next section).

At a first glance, being able to replicate a study seems to be synonymous with being able to reproduce it. Most of the scientific discourse focuses on reproduction and reproducible research (Fomel and Claerbout 2009) rather than replication. However, they have distinct connotations (Peng 2011), and through that, implications for advancing science (or not).

We posit that replicability and reproducibility are actually two distinct pillars of scientific validity, and that it makes sense to distinguish between the two to assess a study's scientific merit and potential. We draw on fine semantic differences between reproduction and replication, as described in Merriam-Webster's and Wiktionary's definitions: A reproduction is always an exact copy or duplicate, with exactly the same features and scale, while a replication resembles the original but allows for variations in scale for example. In a scientific context, reproducibility is therefore concerned with the validity of the results of that particular study, i.e. the possibility of readers to check whether results have been manipulated, by reproducing exactly the same study using the same data and methods. Replicability is more concerned with the overall advancement of our body of knowledge, as it enables other researchers to conduct an independent study with different data and similar but not identical methods, yet arriving at results that confirm the original study's hypothesis. This would be strong evidence that the original study has uncovered a general principle through inductive research, which now another study has proven in deductive research design. Ideally, a study is replicable as well as reproducible, but there are obvious examples where it would be only one of the two, as we will explain next.

We operationalize these two pillars by developing criteria for the two dimensions of *Data* and *Methods*, against which we measure the investigated studies. If a study fulfills all criteria for both *Data* and *Methods*, it supports reproduction and/or replication fully. If it fulfills only some of them, the support is limited, and no criteria fulfilled in one of the dimensions result in an overall lack of support for replication and/or reproduction. The latter is the "Do nothing" approach: The investigators undertake no discernable effort to provide any information that allows other researchers to replicate or reproduce the findings. The results of the study have to be taken at face value, and any interpretation or discussion beyond the one offered by the authors is practically impossible. Clearly, this is not an approach that advances science, but rather it is an approach that (if successful) advances individual scientific careers. With an increased level of competitiveness in the academic world, there are voices (Ioannidis 2014) that express concern over the possible increase of such approaches.

On the other end of the spectrum is the ability to replicate or reproduce *fully* the results of the study. In the case of reproducibility, this means that the study provides the full set of data and tools, so that independent researchers can validate the results by following the same steps, reproducing the experiment and arriving at the same results. While in principle the full set of parameters for data collection could be sufficient for the *Data* dimension if the data source is publicly available, this is practically not sufficient, as the data source might change or data might become unavailable. For example, the terms of service of the popular micro-blogging platform Twitter do not allow the free sharing of collected Tweet samples. Since the API is publicly available, providing precise query parameters would in principle allow other researchers to collect the same data set again. However, since the Twitter Search API only provides the last 5-7 days of Tweets, this is a serious obstacle to the reproducibility of studies involving Twitter. Even full reproducibility does not imply that the study is replicable. For example, although the full data set might be provided, the data collection method might not be disclosed, or the methodology might sit in the black box of a closed-source software tool.

Therefore, reproducibility is only a first step to determine whether a study has the potential to advance science. Replicability means that all employed methods are fully disclosed and complete meta-data is provided, i.e. the precise way the data was collected, and full pseudo-code plus algorithms, or the full source-code of all software employed. It does not require reproducibility, i.e. it does not mean that the full data set needs to be available, but the data source as such should be freely available and not restricted (e.g. data on crime or telecommunications are usually not openly available). If that is the case and the precise method of accessing the API was provided, other studies can try to replicate the results. Obviously, there is some grey area between complete failure and success to fulfill the criteria, resulting in partial fulfillment of the criteria and limited support of replication and or reproduction.

The following Table 1 sums up the criteria for full support, as described above:

[insert table 1]

The following Table 2 sums up the levels for the two dimensions, as described above:

[insert table 2]

Another angle to assess the advancement of science through VGI is to examine the studies' objectives. Are they trying to improve the handling of VGI, or are they already using it to investigate concrete phenomena? With VGI being a new data source, we can expect the majority of early research to be concerned about understanding this data source and handling it. However, for the advancement of science, VGI should only be a means to improve our understanding of human-environmental processes. As long as research exclusively revolves around handling VGI, the only potential for advancing science is in developing methods to handle VGI that can be applied to other data sources as well. We are interested in the objectives of the studies, since preliminary exploration of the literature has shown that most studies have obviously a goal, but many fail to define explicitly an initial scientific question that shapes the analysis strategy and guides the subsequent data analysis. The definition of an initial research question in a data analysis study is fundamental because it determines the overall analysis approach or strategy, which in turn should determine the pool of analytical methods to employ. From our literature exploration, we distinguish the following five broad types of study objectives:

- Descriptive, which describes data sets or current situations.
- Exploratory, which tries to find hidden relationships or patterns and develop ideas for follow-up studies
- Inferential, which is based on a relative small sample of data to generalize results to a bigger population.

- Predictive, which utilizes the value of some variables (predictors) to predict values for another object/variable.
- Causal, which tries to find out the conditions under which associations and correlations amid variables can be interpreted as causality.

In the following section, we will apply these classes to our set of VGI-related studies.

3 Current trends in VGI-related research

In order to respond to research question 2 -- What is the situation on reproducibility and replicability in VGI studies, and are there any past or future trends? --, we conducted a purposeful sampling of VGI-related studies. The criteria for inclusion were:

1. Publication in scientific journals, magazines, conferences, symposia or workshops with full text being accessible.
2. Written in English.
3. The title, abstract or keywords explicitly mention the utilization of VGI for analysis, i.e. no purely conceptual papers about VGI, state-of-the-arts, or research agendas.

Initially, we obtained more than 400 papers as a result of several bibliographic search queries in major specialized and general databases engines such as ISI Web of Science, Scopus, ACM, IEEE, and DBLP, as well as thematic repositories like the Humanitarian Computing Library. We also sought for relevant conferences and workshops for which VGI was a central topic.

In order to check for the fulfillment of the selection criteria and to better understand the set, we performed an initial exploratory analysis. First, titles and abstracts of all papers retrieved from the literature were screened to remove duplicates and to assess whether they fulfill all eligibility criteria, reducing the set of potential papers to about 100. Further reduction was achieved by eliminating papers not directly using

VGI as a data source. This results in a set of 58 papers. We are aware that we cannot claim completeness in our survey, but we are confident that the sample size is sufficient and representative for making conclusions. Furthermore, future studies can easily adopt our approach, widen the literature base, and support (or refute!) our findings.

In a next step, we examined the studies more closely along the two dimensions of *Data* and *Methods*, categorizing their support for reproducibility and replicability along the criteria described in the previous section. Each study thus receives a two-letter code indicating the level of support (None, Limited, Full) for reproducibility and replicability respectively, e.g. a paper categorized as "L-F" provides limited support for reproducibility and full support for replicability. It is important to emphasize here that we conducted the investigation for data sets and tools to the best of our ability. However, we cannot exclude that for some of the studies, more information is available than we were able to discover, e.g. a study might be replicable because the necessary information is available in some supplementary document not publicly available. Likewise, some datasets or code might be available upon request from the authors. Nevertheless, we argue that if such information is difficult to obtain, this acts already as a barrier and has similar effects to complete unavailability. Any such information should be made available openly in order to prevent unnecessary barriers for the advancement of science.

The following Table 3 gives a complete overview of the distribution:

[insert table 3]

As we can see, none of the studies provide full reproducibility (F-x), and only around 10% provided limited reproducibility (L-x). The situation is better for replicability, with 43% of the studies being fully (x-F)) and about 29% being partially replicability (x-L). Still, around 28% of the studies do not provide any support to reproduce or replicate the reported findings (N-N). An analysis for trends over time is

difficult to accomplish, because the number of papers is relatively small for most years, and the often long publishing procedures could lead to a significant "spill-off" from one year to the next. There is no clearly discernible trend (as in "the share of fully replicable studies is increasing over the years"), and the high number of papers in 2013 that support neither (N-N) corresponds to the high total number of papers for that year. Figure 1 depicts an alternative visualization to present the data in Table 3. It allows a more detailed look at the combinations for replicability, reproducibility, and year per paper.

[insert figure 1]

The first observation from the diagram is that one third of the studies do not support reproducibility and replicability at all (blue band in Figure 1). These studies are labeled as "None" in Table 2 and take the strategy "doing nothing".

A second group is composed of studies that favor full replicability because data and methods are properly documented (see column "Replicability" in Table 1). A typical example of this type of paper would be a study in which original data set is not available for reasons of privacy and licensing (e.g. Twitter Streaming API). However, enough information is given on data collection parameters to replicate the study in the future, using the same or different case study. The original source code used for the computations is not available. Again, enough information is provided to describe the methods and tools employed in the form of pseudo code and general formulas. Thus, these studies meet the conditions for replicability, which are necessary but not sufficient to carry out ultimately replication (See section 4).

The third most numerous group of papers are those that still describe and document the methods and tools used in the study but aspects of data used and procedures for data collection are missing (limited replicability, green band in Figure 1). Even though they make some effort to make part of their

experiments publicly available to others, the studies cannot meet the necessary conditions for replicability and therefore results and findings cannot be reproduced and validated.

The results support the initial observation that most studies completely fail to meet the reproducibility conditions. None of the papers analyzed provide access to raw data used in the experiments, and analysis code and tools. This may effectively limit the way VGI can be used in scientific contexts and as an enabler to advance science. On the positive side, the few studies that support at least limited reproducibility are all of recent (2013 or 2014) origin.

The problem seems to be that ensuring reproducibility is becoming more and more challenging because some experiments are getting larger and more complex, with more scientists taking part in them, possibly increased multidisciplinarity, and more complex and high-dimensional data. These studies would require prohibitive quantities of money, resources, and time to reproduce these experiments. From our results, it turns to be that replicability conditions seem more affordable than reproducibility conditions. That is why we draw attention to the subset of replicable experiments to observe what types of analysis methods and tools are being used in these studies in order to respond to the research question 3 -- What is the focus of the VGI studies, and how do they contribute to GISience?.

Most studies briefly state vague or broad objectives such as to improve decision making for disaster management and to enhance situational awareness. Often they fail to operationalize these vague research questions into a suitable strategy for data analysis (e.g., descriptive, exploratory), as pointed out earlier in Section 2. This can be interpreted however as relating to the overall exploratory nature of early VGI research. Consequently, the lack of information about the research question that motivated the studies, along with the lack of clearly identified target users of the analysis, made it difficult in many cases to identify or interpret the analysis strategy.

Almost all of the full replicable studies have been categorized as descriptive or exploratory analysis. Descriptive studies often use descriptive statistics (e.g. mean, median, standard deviation, quartiles, etc.) and plots (histograms, box plots, etc.) to help interpret statistic coefficients and errors, and measures of uncertainty of the data. Within the group of exploratory analysis, these studies often relied on well-known analysis methods such as Natural Language Processing (NLP), Machine Learning (ML) techniques, and methods to compute word frequency, word disambiguation, and named entity recognition (NER) to support geo-parsing and geo-location. Also supervised and unsupervised classification and spatial clustering techniques were commonly applied in order to group relevant entities (e.g. emotions, place names) and detect patterns. Other methods amply used were the analysis of spatiotemporal patterns using density surface maps (i.e. heat maps) and traditional algorithms for computing spatial distances (Manhattan, etc.). Other authors have applied social network analysis techniques to study the network structural properties as well as propagation and diffusion models because of transforming interrelated social data into network graphs. These results are also confirmed by a recent study that examined research papers that employed methods for spatiotemporal analyses over Twitter data (Steiger et al. 2015).

What these analysis methods have in common is that they are widely documented because have been extensively used in the literature. If replicability means in the context of the *Methods* dimension that all employed methods, full pseudo-code, and even algorithms, are fully disclosed, then the use of widely recognized, proven and documented analysis methods is an easy, safe path to ensure replicability.

However, the number of studies that attempt to move beyond data handling and processing is still in the minority, resulting in fewer contributions to increased understanding of human-environmental processes. We will discuss possible reasons and implications in the following section.

4 Discussion: VGI research and the advancement of Science

Our investigation into the status of reproducibility and replicability has shown encouraging trends, but also highlighted some unresolved challenges across these studies. In this section, we are going to discuss our findings and strategies to handle these challenges.

First, the majority of studies provide enough information to replicate them, thus in principle not putting any insurmountable barriers in the way of overall advancement of science. However, there are several caveats that need to be taken into consideration. Concerning the *Data* dimension, even if full information on the data collection methods has been provided, attempting to replicate a study may be a daunting task. Many of the potential VGI sources are rapidly evolving platforms that frequently offer new or changed functionality and changing APIs that allow access to the data. Furthermore, the platforms themselves might be short-lived, and some have already ceased to exist completely or lost most of their user-base (e.g. Gowalla, MySpace). We can expect a certain consolidation in the social media universe, but even if the platforms persist and even flourish, their user base might change dramatically. This is an intrinsic challenge of all user-generated content: The population of users that provides the data for a study can evolve within a short amount of time, even if the geographic extent stays the same (Liu et al. 2014). An evolving content provider population and possible edits to existing content are intrinsic to VGI, with potentially severe implications for research, and might put to test the whole objective of finding and describing general principles and laws. Concerning the *Method* dimension, we perceive a lack of support for new opportunities to share any tools and code employed in the study. While a description of the process through pseudo-code and formulas is sufficient for replication, none of the studies provided or shared actual code, which would facilitate replication and is easy to host nowadays with platforms such as Github.

Given the inherent complexities of VGI, ensuring reproducibility seems almost like a simpler task than to ensure replicability. Thus, the general lack of support for reproducibility is surprising, given the technology opportunities to share data and code (González-Beltrán et al. 2015). There are several possible practical reasons for this. Regarding the *Data*, there is the obvious ethical dilemma of privacy vs transparency. Although most of the studies we examined rely exclusively on data that has been collected from publicly available content, the open sharing of processed, analyzed and non-anonymized data raises concerns over privacy. An anonymization is often also problematic from a scientific point of view, since insights and added valued of the research stem from the rich semantics of the content, which will be lost or limited if thoroughly anonymized. Furthermore, many source platforms may have terms of service that forbid the sharing of data. In principle, if the complete meta-data on data collection methods is provided, any study aiming to reproduce the findings could re-collect the data. In practice, this faces the mentioned problems of volatile content (some user might have deleted, added, or changed content) and provider restrictions (e.g. Twitter Search API returns only results from within the past couple of days). For example, Liu et al. (2014) observed that up to 20% of users' publicly issued tweets are no longer available which may significantly impacts the ability for researchers to reproduce prior results. Another reason could be the sheer size of the data sets, for which there is yet no sufficient infrastructure in place for sharing. Regarding the *Methods*, the problem with using proprietary software that is not accessible without a (costly) license can be an obvious problem for reproducibility, but the alternatives are available, however not used. Free and open-source software leaves little functionality to be desired, and also avoids a problem of backward-compatibility: Older versions are readily available.

Having outlined some fundamental and practical issues concerning reproducibility and replicability of VGI-related research, it is now time to discuss in more detail some strategies that might improve the situation and help realize the full potential of VGI-related studies to contribute to GIScience.

The current scale of VGI complexity, heterogeneity and volatility seems to be testing the limits of established channels of scientific communication. One obvious strategy to counter this is a faster dissemination process of study results through different channels like blogs or open platforms. The peer-review would then be post-publication and open through other researchers' comments, similar to the approach by the life sciences platform F1000[7]. While this would seriously threaten existing publishing models, it would still require some form of third-party management of the resources to ensure that no comments are deleted or results and data changed without track records. The measurement of a research output could be the social network of fellow researchers reviewing and sharing the paper, or even some form of voting like popular sites such as Amazon or YouTube have implemented. Participation would require a non-anonymous identification like Orcid.

Such a fast-lane publishing approach would not diminish the need for a sustainable solution to provide and maintain scientific data and code repositories. The concept of "data curation" emerged when data is made publicly available and should be maintained in the long term (Lynch 2014).

The Open Data movement[8] is not only focused on scientific data, but it is very active in the "freeing up" of governmental datasets. In this sense, we recently discussed the need to find synergies between public sector information and research data as these two communities typically use distinct mechanisms for making data open and public (Schade et al., 2015). With regards to research data, it is important to highlight the current approaches to publish data pointed out by Kratz & Strasser (2015). The first approach is the data paper, which describes a dataset along with the data collection methods used. Examples are the Scientific Data[9] publication, which in uniquely devoted to data papers, and the Open Geospatial Data, Software and Standards[10] journal, which accepts contributions of "data papers". A

---

[7] http://f1000research.com/
[8] e.g. https://okfn.org/
[9] http://www.nature.com/sdata/about
[10] http://www.opengeospatialdata.com/about

second approach is to make use of domain-specific repositories to host research data. Some publishers link the scientific paper to associated data hosted in such open data repositories. For example, GigaScience[11] journal already links each published article to a data repository (GigaDB Dataset), which hosts any type of research resources related to the article itself, whether data, scripts, or additional documentation. The entire set of resources is identified with a DOI, which is usually included as a citable reference in the own scientific article. A third approach is exemplified by general-purpose open data repositories that are aimed to create and maintain online repository for scientific data sets such as figshare[12] - which lets researchers preserve, share, discover and eventually give credit to any type of research resource ranging from papers, to presentations and additional documentations, to data sets and figures -, and Academic Torrents[13], a community-maintained distributed repository for data sets and scientific knowledge.

These technical solutions need to be supported by measures to address incompatible or restrictive data policies. Some examples can highlight the importance of having common data policies at work. The Policy RECommendations for Open Access to Research Data in Europe[14] (RECODE) project addresses challenges within the open access and data dissemination and preservation sector and produce policy recommendations for open access to research data based on existing good practice. For doing so, RECODE puts special emphasis in legal, ethical, institutional and policy issues by analyzing cross-domain case studies (e.g. physics, health, earth sciences) to enable open access for research data. Even though the project is addressing different interoperability levels (legal, organizational, semantic, etc.) related to open access, the main stress is on how professional researchers can be incentivized to deposit their data on open access repositories and how to formulate such open access polices, which will also take into account legal and ethical issues, in an easy, unambiguous, and understandable way. This naturally

---

[11] http://www.gigasciencejournal.com/
[12] http://figshare.com/
[13] http://academictorrents.com/
[14] http://recodeproject.eu/

puts the focus on the policy dimension, as rightly suggested by the reviewer, although technology and infrastructure aspects are also of vital importance to overcome interoperability, accessibility and sustainability issues given the ever-increasing volume of research data (Schade et al. 2015)

Along the same lines of RECODE, the goal of the European Data Infrastructure[15] (EUDAT) is to build a solid foundation to allow heterogeneous research communities, in terms of scale, scope, domain, practices, and so forth, to share research data and to preserve it on the long term. To achieve this, EUDAT is organized into dedicated working groups to deal with the management of massive generation of real time data from research facilities, instruments and sensors, the definition of data workflows for sophisticated services, the required semantics to let researchers annotate research data in consistent and interoperable ways, and, most important, the establishment of policies regarding data access and reuse. Again, EUDAT identifies data policies as extremely challenging because of the need of harmonizing diverse open data and data access policies as well as devising a common licensing scheme in order to set pragmatic ways for opening, accessing and sharing cross-domain research data.

Targeted more at e-Government, the Joinup[16] platform created by the European Commission aims at helping professionals to collaborate and share interoperability solutions for public. While not offering particular services for data sharing, its declared aim is to provide interoperability solutions, and this includes knowledge and code repositories to share methods.

While these initiatives aim to improve the situation for scientific research in general, it is worth considering their relevance for VGI in particular, and asking the question whether it even makes sense to attempt to store VGI-related data forever. Will a researcher in 2050 use today's VGI for experimentation? If so, will it be possible to do something with such VGI (in 2050) given how technology changes? Will

---

[15] http://eudat.eu/
[16] https://joinup.ec.europa.eu/

today's VGI be comparable to "VGI" in 2050? To answer these questions right now is not an easy task, and possibly, we cannot. However, answers may be partly connected to the nature and type of phenomena being studied with VGI. For example, VGI on a persistent feature (e.g. a building like a church) will likely remain valid for a long time, while a feature on a short-lived feature (e.g. amenities like a bar) might become invalid within a short time, however, hopefully being replaced by new, updated VGI. For long-term phenomena and processes, ensuring reproducibility of the data *within one study* is hard to achieve because the phenomena being studied may outlive the life cycle of a particular VGI platform. This suggests that VGI seems to be most useful for short-term (ephemeral) events or features, which often correlate with sudden events that last short periods of time. For example, VGI could help us track and map ephemeral features or events that would not appear in a dataset with lower sampling frequency (e.g. traditional surveys of national mapping agencies). Still, scientific disciplines such as sociology or history might attempt to study VGI from a longitudinal perspective, and outdated VGI would be needed to do this. Therefore, saving snapshots of VGI in repositories for long-term storage does seem to be a useful endeavor. However, the sheer amount of VGI forces any potential data curator to make important choices on what to store. Given the interest of VGI for other disciplines and the origins of much VGI from social media platforms, it appears that any such VGI repository requires a multi-disciplinary effort. Such an effort is currently undertaken by the Group on Earth Observation for example by developing data management principles[17].

The difficulties of handling the new data source of VGI are also reflected in the current objectives and analytical focus of many studies. A significant part of VGI-related research is still investigating issues of handling it, instead of methods to apply it for concrete research problems or to improve our understanding of human-environmental processes. After several years of VGI-related research, we would expect this situation to change and the overall analytical focus to shift from descriptive and explorative to inferential and predictive (modelling) studies. The coming years will show whether this is going to happen or

---

[17] https://www.earthobservations.org/geoss_ta_da_tar.shtml

whether VGI is such a complex and fast evolving data source that research is struggling to keep up with its development.

An underexplored opportunity might also be volunteered geographic analysis of VGI or traditional, authoritative data sources. VGI is frequently discussed in conjunction with concepts such as neogeography, participation and citizen science. However, as Haklay (2013) critically argues, the current developments are far from heralding a democratization of geospatial information processing. Apart from a gender and socio-economic bias in the contributing communities, even the "technophile elite" that contributes the majority of "true" VGI has limited proficiency in using it for analysis purposes. While VGI can contribute to more advanced forms of citizen science, this potential is currently largely untapped. The interpretation and execution of analytical processes along with data rely strongly on the availability of supporting software tools (Granell, 2014). Such tools are necessary to get tasks such as data collection, processing, filtering and so on automated. Furthermore, accessing analyses and data without proper supporting tools may be seen as an obstacle by expert and non-expert users alike (Beniston et al. 2012), thereby preventing VGI reproducibility. VGI needs suitable tools to lowering entry barriers to VGI reproducibility by letting researchers assemble data and analysis together and to understand them in a new context. None of the studies used any scientific workflow management software (e.g. VisTrails[18], Taverna[19], Kepler[20], etc.). Doing so would facilitate both replication and reproduction. If these tools would be made simpler to lower access barriers, they could also contribute to enhance opportunities for collaborative volunteered analysis.

Closing the discussion, we would like to comment on this paper's replicability and reproducibility. Although not using VGI as data source, our aim was of course to enable both reproduction and replication. While the source data (published papers) is fully disclosed and will remain available for

---

[18] http://www.vistrails.org/index.php/Main_Page
[19] https://taverna.incubator.apache.org
[20] https://kepler-project.org

future researchers to investigate, much of its content is behind a paywall, thus outside of easy reach for the majority of laypersons. Similarly, we cannot provide the precise steps for arriving at the sample used for the study. Thus, the *Data* dimension of this study is certainly replicable, but not reproducible. Concerning our *Methods*, a qualitative investigation such as this one is almost impossible to reproduce. While we have tried to provide clear criteria for our classification, many studies are of limited or partial reproducibility or replicability, requiring interpretation, which is always subjective. Therefore, a conservative assessment would lead to a score of N-F (not reproducible, but fully replicable), although we would like to think that we have provided enough information to make it at least partially reproducible (L-F).

5 Conclusion and Outlook

In this paper, we argue that despite changing scientific paradigms (data-driven science, open science), scientific validity is still based on the two pillars of replicability and reproducibility. The new data sources available to research offer great opportunities for innovative research, but they also pose new and particular challenges for both reproducibility and replicability.

We developed a simple classification of reproducibility and replicability based on the two dimensions of *Data* and *Methods*, and then applied this in a literature review of VGI-related studies. The selection is a purposeful sample from GIScience journals and conference proceedings with a focus on crisis management related studies. The results show that the majority of papers provide enough information to allow replication of the study; however none of the studies provide enough material to reproduce them.

The reasons for the latter are likely based on ethical concerns (privacy of the users), legal concerns (terms of services of the platform providers, liability), and possibly personal careers concerns (fear of unattributed use of data and methods by other researchers). While the high ratio of replicable studies is encouraging, the lack of reproducibility is still a serious concern, because the nature of VGI makes it very

difficult to reproduce studies: VGI is heterogeneous, localized and volatile, greatly impeding on meaningful replications and thereby the discovery of general principles and laws.

Furthermore, we argue also that VGI research is mature enough to move from purely data-centric objectives, i.e. methods to handling of VGI, to more fundamental as well as applied research.

How can we therefore counter the threat of purely idiographic studies? We argue that a successful strategy is likely for Science to become both faster and more sustainable. Becoming more sustainable with respect to storing data and methods is not exclusive to VGI-related research. We have shown some concerted efforts in providing long-term curation effort for both. Becoming faster with respect to the sharing of research results, in order to counter the volatility of VGI, could mean that instead of the traditional model of paper-based output, the research needs to be published without prior peer-review, and instead reviewed and discussed post-publication, with the option to create updates when new results are available. Since this would require a substantial change in the traditional author-publisher-consumer model and new ways to measure individual scientific output, it would require a substantial effort, the first prerequisite to this a wide acknowledgement and discussion of the problems at stake. This paper aims to contribute towards this goal.

References

Beniston M, Stoffel M, Harding R, Kernan M, Ludwig R, Moors E, Samuels P and Tockner K 2012 Obstacles to data access for research related to climate and water: Implications for science and EU policy-making, Environmental Science & Policy 17: 41-48

Cheong F and Cheong C 2011 Social Media Data Mining: A Social Network Analysis of Tweets during the Australian 2010-2011 Floods. In Proceedings of the 15th Pacific Asia Conference on Information Systems: Quality Research in Pacific, Brisbane, Australia

Craglia M, Ostermann F and Spinsanti L 2012 Digital Earth from Vision to Practice: Making Sense of Citizen-Generated Content. International Journal of Digital Earth 5(5): 398–416.

Craglia C and Shanley L 2015 Data democracy – increased supply of geospatial information and expanded participatory processes in the production of data. International Journal of Digital Earth, DOI: 10.1080/17538947.2015.1008214

Fomel S and Claerbout JF 2009. Guest Editors' Introduction: Reproducible Research. Computing in Science & Engineering 11 (1): 5–7

González-Beltrán A, Li P, Zhao J, Avila-Garcia MS, Roos M, Thompson M, … Rocca-Serra P 2015. From Peer-Reviewed to Peer-Reproduced in Scholarly Publishing: The Complementary Roles of Data Models and Workflows in Bioinformatics. PLOS ONE, 10(7), e0127612.

Goodchild MF 2007 Citizens as sensors: The world of volunteered geography. GeoJournal 69(4): 211-221

Granell C 2014 Robust Workflow Systems + Flexible Geoprocessing Services = Geo-enabled Model Web? In: E. Pourabbas (Ed). Geographical Information Systems: Trends and Technologies. Boca Raton: CRC Press, 172-204

Haklay, M. 2013. Neogeography and the Delusion of Democratisation. Environment and Planning A 45 (1): 55–69.

Harvey F 2013 To Volunteer or to Contribute Locational Information? Towards Truth in Labeling for Crowdsourced Geographic Information. In Crowdsourcing Geographic Knowledge: Volunteered

Geographic Information (VGI) in Theory and Practice, edited by Daniel Sui, Sarah Elwood, and Michael Goodchild, 31–42. Berlin: Springer.

Hecht B, Hong L, Suh B and Chi E.H 2011 Tweets from Justin Bieber's Heart: The Dynamics of the "Location" Field in User Profiles. Proceedings of CHI, New York: 237-246

Hey T, Tansley S, Tolle K (eds) 2009 The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond: Microsoft Research

Ioannidis JPA 2014 How to make published research true. Plos Medicine 11(10): e10011747

Kratz JE and Strasser C 2015. Researcher Perspectives on Publication and Peer Review of Data. Plos ONE 10(2): e0117619

Liu SB 2010 The Rise of Curated Crisis Content. In Proceedings of the 7th International ISCRAM Conference, Seattle.

Liu Y, Kliman-Silver C and Mislove A 2014 The tweets they are a-changin': Evolution of Twitter users and behavior. In Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM'14), Ann Arbor, MI

Lynch C 2014 The Next Generation of Challenges in the Curation of Scholarly Data. In: Research Data Management: Practical Strategies for Information Professionals. Purdue University Press. 395-408

McNutt M 2014 Journals unite for reproducibility. Science 346 (6210) p. 679

Neis P and Zielstra D 2014 Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap. Future Internet 6:76-106

Ostermann FO and Spinsanti L 2012 Context Analysis of Volunteered Geographic Information from Social Media Networks to Support Disaster Management: A Case Study On Forest Fires. International Journal of Information Systems for Crisis Response and Management 4(4): 16–37

Peng RD 2011 Reproducible Research in Computational Science. Science 334(6060): 1226-1227

Popper K 1992 The Logic of Scientific Discovery. London: Routledge.

Roick O and Heuser S 2013 Location Based Social Networks – Definition, Current State of the Art and Research Agenda. Transactions in GIS 17(5): 763–784

Schade S, Granell C and Perego A 2015. Coupling Public Sector Information and Public-Funded Research Data: A Vision of an Open Data Ecosystem. In: C.G. Reddick, L. Anthopoulos (Eds). Information and Communication Technologies in Public Administration: Innovations from Developed Countries. Boca Raton: CRC Press, 271-293

Spinsanti L and Ostermann FO 2013 Automated Geographic Context Analysis for Volunteered Information. Applied Geography 43: 36–44.

Stefanidis A, Crooks A and Radzikowski J 2013 Harvesting Ambient Geospatial Information from Social Media Feeds. GeoJournal 78 (2): 319–38

Steiger E, Porto de Albuquerque J and Zipf A 2015 An Advanced Systematic Literature Review on Spatiotemporal Analysis of Twitter Data. Transaction in GIS.

Table 1: Dimension and criteria for reproducibility and replicability

Dimensions	Reproducibility	Replicability
Data	Full data set	Full meta-data on collection, data source accessible
Methods	Executable tools or precise step-by-step information	Pseudo-code and formulas OR source-code

Table 2: Completeness of dimensions and support for replicability and reproducibility

Level of completeness (Data and Methods)	Pillar (Replicability or Reproducibility)
Both Complete	Full
Any partial or none	Limited
Both none	None

Table 3. Distribution of the analyzed papers (N=58); N/A not applicable, N = None, L = Limited, F = Full (see Table 2). Column headings are encoded as follows: the first letter refers to reproducibility, and the second to replicability.

Year / Support for Reproducibility - Replicability	N-N	L-N	N-L	L-L	F-N	N-F	L-F	F-L	F-F	Total
2008	1									1
2009						2				2
2010	2		1			3				6
2011	3		2			2				7
2012	2		3			4				9
2013	7		6	2		4	2			21
2014	1		3			5	2			11
2015						1				1
Total	16	0	15	2	0	21	4	0	0	58
Percentage of total (rounded)	28%	0%	26%	3%	0.00%	36%	7%	0.00%	0.00%	100%

Figure 1: Parallel set diagram to show how replicability and reproducibility categorical conditions are related. For an interactive version, where users can select individual bands, and drag and each dimension to reorder the chart, visit: http://www3.uji.es/~canut/parsets/Parsets.html. The chart is built upon the Parallel Sets JS library by Jason Davies (http://www.jasondavies.com/)