Final submitted version of
Lane, S.N., 2020. Editorial: Data from nowhere. *Earth Surface Processes and Landforms*, 45, 5-10
Available at
https://onlinelibrary.wiley.com/doi/full/10.1002/esp.4775

Editorial 2020 Part II: Data from nowhere ?

Stuart N. Lane
Institute of Earth Surface Dynamics, University of Lausanne
November 2019

Abstract

This short editorial reflects upon and explains ESPL's policy with respect to data sharing. Whilst there are clearly a number of important reasons why data should be shared, notably to support transparency in science and to support the long-term monitoring of geomorphological change, data sharing comes with some concerns. Data sharing cannot become an alternative to field data collection as it produces a scientific understanding of a very particular type. When shared data are combined in extensive analyses we run the serious risks of inter-comparability, bias and incompleteness. Automated harnessing and analysis of any kind of geomorphological data runs the risk of combining data that are not comparable and producing conclusions skewed to the datasets available. Shared data allow for none of the serendipity that comes from being in the field and being confronted by what we see and experience; they might as well be data from nowhere. We must be particularly careful not to reduce the perceived value of field data collection in an academic system that places ever more emphasis on speed and efficiency in scientific knowledge production. It is for these reasons that ESPL's policy on data sharing is to support it and to encourage it, notably through requiring authors to provide a data statement, but we will not insist that data should be published as a condition of manuscript acceptance.

Keywords

Data, Data Sharing, Publication, Fieldwork, Extensive Analysis, Chaotic Conception, Scientific Knowledge Production, Geomorphology

Introduction

During its November 2019 meeting, the ESPL Editorial Board reflected on its policy with regard to data sharing. The speed with which data has become a new element of academic publishing is quite astonishing and many journals now require the formal publication of data at article submission or before final acceptance. In this short Editorial I reflect on the position taken by ESPL. Our current policy is that an author must make data available to Editors and reviewers upon request; that authors are encouraged to make their data available to others upon request; but not that those data are formally published, whether linked to the paper or independently. With effect from the 1st January we will require all manuscripts to have a data statement that explains either where data may be acquired or gives a reasonable justification for not making them available. The arguments that follow explain why we do not plan to require data to be published formally in order to publish in *Earth Surface Processes and Landforms*, even if we strongly encourage it. We will not take into account the data statement in evaluating the scientific merit of manuscript, the latter being the basis of all decisions we make.

The case for sharing data

At the beginning, it should be noted that there are good reasons why publishing data might be considered a valuable component of an academic journal's activities. First, most scientific journals, for some time, have required that authors make their data available to reviewers upon request. The idea is that assessing the data could or even should be an important element of the peer review process. In practice, this rarely happens. In my 12 years as Managing Editor at ESPL I have received one single request from a reviewer to see data during review. By making data sharing a formal part of a publication, reviewers may be encouraged to evaluate data more frequently and this is important. A well-written paper founded on poor or incorrect data in some senses is more problematic than a poorly-written paper founded on good data. Peer review of data strengthens the principle of peer review, of growing importance as an indicator of quality in an era when almost anyone can publish almost anything somewhere in the cloud. If publishing data strengthens the status of peer review, it must be a good thing.

Second, collecting and analysing data requires significant expense, whether in terms of time or money or both. Realising the value of those data by making them available to others could carry significant benefits. In the field of geomorpholohy, the societal value of data is perhaps somewhat less tangible than in the field of a discipline like medicine or pharmacology. Yet, there may be examples of where the data of others are valuable in an applied context (e.g. in flood risk management) and certainly many examples in the academic system of where data have a wider use than their original purpose (e.g. the validation of numerical models). Being able to confront our ideas more effectively with a wider set of data is likely to be beneficial to scientific understanding and the evolution of geomorpholohy as a discipline. The capricious nature of the environments we study mean that we can only partially control what it is we can measure. Fieldwork is risky in both health and safety terms, but also scientific terms, as we can never be sure that fieldwork will be successful. Making high quality datasets available allows the wider community to profit from successful data collection campaigns (Allen and Berghuijs, 2018). When the data are fit for purpose and of a high quality, further analysis may lead to valuable new knowledge and publication (Blume et al., 2017; Allen and Berghuijs, 2018); that is "more for less".

Third, geomorphological enquiry addresses landscapes that are a product of both the immanent and the configurational (Simpson, 1963); of processes that may be generic and occur across landscapes and times (the immanent); and their manifestation in particular landscapes at certain times (due to configurational influences), what we observe and eventually measure. To avoid a descent into detailology we have to think about how it is we can extract the general from the particular, including generalising from a singular case study (Richards, 1996). It is possible that the combination of data points from different case studies in an extensive analysis may lead to fundamental scientific findings that cannot be reached on the basis of individual studies alone (Sivapalan and Blöschl, 2017). Herman et al.'s (2013) synthesis of thermochronological data to estimate global mountain erosion rates is an excellent example in geomorpholohy. Such synthesis requires data to be shared in a form that can be re-used and this is behind the growth of the FAIR data sharing principles (Force11, no date). First, I should note what the FAIR principles are not. They only implicitly make the assumption that it is "fair" to share data (the "Open Science" agenda's view that publicly-funded science should be publicly-accessible at no cost). The **FAIR** principles are actually more specific than this, concerned with archiving data such that they are (the quotes are taken from Force11, no date): (1) **F**indable ("*a data object should be uniquely and persistently identifiable*"); (2) **A**ccessible ("*it can be always obtained by machines and humans*"); (3) **I**nteroperable ("*data is (sic) machine action-able*") ; and (4) **R**e-usable ("*data should be sufficiently well-described and rich that it (sic) can be automatically (or with minimal human effort) linked or integrated, like-with-like, with other data sources*". FAIR data are those which have been published " … *to support both manual and automated deposition, exploration, sharing, and use to support machines as well as humans* …". FAIR principles, then, take generalisation one stage further through encouraging a transition to the automated acquisition and analysis of data from a large number of studies with a view to finding new kinds of generalisations. The emphasis is upon harnessing the potential for machines to harness the data and to learn what they can tell us, something already proven valuable in geomorphological applications (e.g. Stumpf and Kerle, 2011; Tien Bui et al., 2016). For contributing to such extensive analyses, there is a clear importance in sharing data.

Fourth, there is a pedagogic role associated with making data available, especially in the broader sense if sharing includes not only data but also images and videos. Since the growth of quantitative methods in geomorpholohy, the analysis of data has been brought progressively into the classroom. Working with datasets is an alternative means of learning to that associated with the conventional lecture. For instance, in the same way that researchers can use datasets in a comparative way to develop new understanding (Sivapalan and Blöschl, 2017), so students now use datasets (broadly defined) to develop their own understanding of particular geomorphic questions and to develop their critical engagement skills (Litherland and Stott, 2012). I now teach fluvial hydraulics (e.g. the relationship between velocity and shear stress) through visualisation (e.g. of velocity profiles) and quantification (e.g. using the law of the wall) based upon student-acquired datasets. This kind of flipped learning is in marked contrast to telling students (or asking them to read) about relevant theory. Faced with pressures to do "more with less" (Litherland and Stott, 2012), along with growing issues surrounding the management of health and safety in fieldwork settings and student pressures on funding, sharing data is a means of sustaining methods like flipped learning. Shared data can itself be used to prepare students for fieldwork, making fieldwork more effective (Warburton and Higgitt, 1997) and even replacing it in some kinds of settings (Stumpf et al., 2008). Sharing data, then, can serve an important pedagogic purpose.

Finally, and perhaps most importantly, in an era of rapid environmental change, the data collected today have the potential to become a crucial resource for generations in the future; the benchmarks against which we can evaluate just how much the environment is changing. Compared with cognate disciplines like climatology and hydrology, geomorphic processes are not typically accorded the privilege of routine monitoring. Even if datasets designed for operational activities (e.g. discharge measurements in hydrology) are not always suited to scientific analyses (Kirchner, 2006), they still provide at least some basis for change detection and attribution. There are examples of the elegant use of hydrological data to infer geomorphic processes (such as quantification of the effect of sediment-transport driven changes in channel capacity on flood risk, Slater, 2016). Of course, this is changing, whether through ever more sophisticated use of the depositional record to infer geomorphological processes, or new ways of quantifying erosion rates over a range of timescales using cosmogenic nuclides. Complementary to such emerging techniques could be the effective archiving of data that allow their future use for change detection and attribution.

Should we require data sharing in publicly accessible permanent archives?

Clearly, there are very good reasons why we should encourage the sharing of data. However, whether or not we should *insist* on the sharing of geomorphic data needs more careful thought. Some of the reasons for questioning obligatory data sharing require us to think about the kinds of knowledge that it is going to produce; other reasons are more practical.

My first point is a strongly philosophical one. Whether in student classes on research design or in the advice that I give to Early Career Researchers in author workshops (Lane, 2014), I place strong emphasis on thinking through the underlying motivation for undertaking the research that is to be done and the results that are to be written up. I commonly use Harré's (1981) reasons for doing science, of which he identifies 13. The kinds of extensive data analysis implicit in data sharing and in the FAIR principles can sustain some of these reasons (notably B4, demonstration of underlying unity within apparent variety; also A3, finding the form of a law inductively; and A1 through exploring the characteristics of a naturally occurring process). Thus, my argument is less against extensive data analysis, but more that there some of Harré's motivations for doing science that do not lend themselves well to extensive analysis at all (A5, exploiting an accidental occurrence; B3 decomposition of an apparently simple phenomena, C2 demonstrating the power and versatility of apparatus) or only with some serendipity (B1, finding the hidden mechanism of a known effect, serendipitous as it depends on whether available datasets capture the hidden mechanisms). There is a kind of geomorpholohy that involves the intensive study of particular landscapes or landforms (see Richards, 1996) that makes a fundamental contribution to knowledge through our work in the field. A number of geomorphologists have challenged the growing emphasis on (global) Earth System Science in the academy, to the detriment of understanding more contingent systems (i.e. the combined product of the immanent and the configurational) at the local scale (Rhoads *et al.*, 1999; Phillips, 2001; Clifford and Richards, 2005; Richards and Clifford, 2008; Görg, 2007; Ashmore and Dodson, 2016). Of course, sharing data does not necessarily exclude such work, but given some of the external drivers of academic practice in the 21st century we have to be very careful about the signals that are sent out as to what matters in academic enquiry in pushing the obligatory sharing of data as necessary; "the only data matter are those that can be shared".

I have argued elsewhere that the academy (in the geosciences) is now in a state of crisis as a result of an academic model that is not sustainable (Lane, 2017) and which has led to one great big academic publication bubble. The bubble is not sustainable because we have more to read than is humanly possible; we have increasingly to compete with one another to get our work read; and the suite of 21st century metrics with which we have become complicit (I see a worryingly large number of young researchers including citation counts and H indices in their C.V.s) is forcing us to cite rather than to sight (Bunge, 1979; Lane, 2017). The emphasis of the FAIR principles on harnessing automation in data assimilation and analysis, on efficiency, risks making certain kinds of research activity more attractive than others simply because it seems to be more productive. The philosophical question becomes what kind of science do we wish geomorpholohy to become and what kind of questions and approaches do we and should we value? Data sharing invokes certain kinds of approaches to geomorpholohy but with, as far as I can see, almost no reflection as to the broader consequences of advocating it. Hydrologists have worried about the reduced status according to fieldwork for some time (Burt and McDonnell, 2015; Vidon, 2015; Allen and Berghuis, 2018) and I see no reason to challenge the same concern in relation to geomorpholohy. I return to some of the challenges facing the status accorded to fieldwork below.

If the above causes us to reflect on the status given to different approaches in geomorpholohy, we also need to reflect upon the kind of knowledge that comes from shared data, especially in the name of extensive searches for generic understanding. As Burt and McDonnell (2015) remind us, progress in science comes through new techniques, new discoveries and new ideas and that, certainly in relatively mature disciplines like geomorpholohy, innovation in techniques is particularly important. This is why the data that we collect today (including their resolution and frequency) are likely to be less valuable than those that we collect in the future, when we have new technologies. In parallel with what is done in disciplines like climatology and hydrology, as data collection methods change, so we will need to homogenise geomorphological data to make them comparable through time. However, homogenisation only deals with the influence of measurement methods on inter-comparison, it does not deal with the removal of configuration itself. Given the configurational influences on geomorphological measurements, at least part of the information behind a data point is unique to the choice of location (in space and time) that is the focus of the research and this should cause us to be cautious of extensive analyses based upon data sharing.

I have recently been reflecting on this issue in relation to the extent to which glaciers are capable of evacuating the sediment that they erode. This is a classic "cover and tool" problem (Turowski et al., 2008): a glacier (as an erosional tool) can only continue to erode its bed as long as it is possible to evacuate the sediment it has already eroded (so there is no cover). Subglacial drainage systems are generally (Nienow et al., 1998) either distributed when surface melt rates are low (and so the melt water needed to move sediment at the glacier bed is not available) or canalised into drainage networks when melt is high (and so whilst there might be melt water at the bed, its coverage is geographically restricted). Thus, whilst subglacial rivers have very high transport capacity (Hallet et al., 1996) they are not necessarily able to access subglacially-eroded sediment at the times at which this capacity is highest. Evacuation of sediment is then a function of other basal sediment transport processes including subglacial deformation of till, regelation, hydraulic supercooling and so on (Alley et al., 1997). These processes depend on the environmental setting of the glacier, the shape of its bed, ice thickness etc. If we then measure the sediment exported from the glacier, at whatever time-scale, the records obtained will reflect these configurational elements which (even if measurable) may not have been measured. Thus, what is the meaning of inter-comparison between glaciers of rates of sediment evacuation measured in proglacial streams given that they represent signals produced by very different combinations of configurational influences? In an experimental sense, we could deal with this problem through *a priori* research design such that we sample glaciers across different axes of configurational influence and then model these influences statistically through an extensive analysis. Logistically, given the extreme difficulty of measuring sediment evacuation rates, it would be a severe challenge to get a large enough sample size to meet normal experimental requirements. If we think of some kind of factorial design where the number of glaciers needed is 2^k and k is the number of factors, we might identify the following factors; cold versus warm-based; steep ice surface versus shallow; thick versus thin; positive versus negative bed slopes (due to hydraulic supercooling effects); laterally-confined versus unconfined; land-terminating versus lake-/marine-terminating; and more erodible versus less erodible substrates. This means we would need 128 glaciers. We would need to double this to 256 glaciers, to guarantee some kind of repetition. We would need to make sure that each glacier was studied for four field seasons (to capture the combinations of warm versus cold and wet versus dry). Perhaps we could reduce the sample size through some kind of Latin hypercube sampling but there may be other factors that need to be included and the difficulty of quantifying sediment evacuation (wash load, suspended load and bedload) in just a single melt season makes it highly unlikely to be economically or scientifically feasible. If we invert the issue, and imagine a compilation of evacuation rates from different glaciers across the world (e.g. Hallet et al., 1996) the problem becomes more severe as the erosion rates available (e.g. Hallet et al. (1996) compiled data from 61 glaciers) may be; based upon different measurement methods or only partial data (true data on bedload evacuation, for example, remain extremely rare); will not have been sampled according to an *a priori* research design such that some factors may be under-represented, others over-represented; may have uncertain or unknown configurational aspects (e.g. subglacial glacier bed morphology); and may have been sampled in years with different, even incomparable, climate forcing.

This example shows what I think is a general geomorphological trait that configurational aspects may dominate geomorphological signals such that the statistical basis of generalisation from data compilations may be extremely weak. Sayer (1984, 138) introduced the notion of a 'chaotic conception' to describe a process of 'bad abstraction'. In the latter, elements of a system are measured, associated and generalised either because it is convenient to do so (e.g. ease of measurement; availability of data) or because there is some sort of emergent pattern that merits simple generalisation. Examples abound in the environmental sciences. Pattinson and Lane (2012) label attempts to relate changing flood frequency to land-use management as a chaotic conception. The long-established debate over desertification and its relationship to both climate change and over-grazing provides a second example; these processes may correlate but they are not necessarily and directly causally-related. Piers Blaikie's (1985) demonstration that even if soil erosion and "land mismanagement" are correlated it was actually the effects of unequal terms of trade upon farmer behaviour that forced them to adopt soil eroding land use practices is a more geomorphological example. Any kind of data analysis that bases itself on the automated search for associations between complied data risks making chaotic conceptions, especially as it is the data availability that is driving the association being sought. Yet, as others have argued (e.g. Burt and McDonnell (2015) in hydrology) there are other kinds of understanding that come not from the data that fit, but the data points that don't fit. The measurements we make in a place and time that don't fit our theory causing us to pause, to reflect and to think; "is this error?"; or "a unique case?"; or "is there something else that I don't quite understand or that I have not yet thought about?". This is where the field, and fieldwork, play a crucial role in slowing us down, forcing us to question not only our answers but also the very questions that we are asking themselves (Stengers, 2005 ; Lane, 2017). The questions we can ask of shared data are self-limiting. You can't modify your data collection as the analysis progresses as the data are already collected. It is hard to let the landform or the process "speak back" and say "measure me differently" and to capture the wider set of empirical experiences that bring meaning to the data we collect. What you can learn is restricted to the data that you have. It cannot be challenged by the data that you don't have but which you could have had if you had been involved in field data collection. Thus, shared data may play a role in flagging what we don't know. They may also play a role in allowing us to upscale the findings from a single case-study, by putting the case-study into a wider context through comparison with data from different sites. However, these roles are limited as compared with what can come from carefully-selected field sites and the associated fieldwork, whether undertaken extensively (across different sites) or intensively in a single one.

Even if there are examples of research questions where data sharing is valuable there are more practical issues. One in particular has concerned hydrologists and begins with what Allen and Berghuis (2018) describe as a "free ride" effect; rather than incurring the time and expense needed to produce field data a researcher chooses to focus on the use of others' data. Making your data available clearly contributes to collective benefit. However, we work in a system that prioritizes individual gain. The award of research funding, the appointment to stable positions and academic promotion are generally based around the assessment of individual contributions and, in particular, publications. Even with the move away from naïve quantification of performance using journal impact factors and H indices (see the San Francisco Declaration on Research Assessment, DoRA), and a growing focus on quality in research output, we still judge what an individual has achieved. Publication remains the academic currency and as long as it does so the collective benefits of making data available risk being undermined by the disincentives to collect data that come from a system that emphasises individual academic output. This is a practical issue; it might be dealt with through giving credit for making data available. Indeed, community surveys (e.g. Blume et al., 2017, for hydrology) suggest that this is a necessary component of an effective data sharing system. It is perhaps not surprising that we are currently witnessing the commodification of data sharing, for example, by giving published data a digital object identifier such that it can more readily accrue credit via citation. When a publication has data linked to it, it has been claimed that this can make a paper more highly cited in the ever-growing pressure on us to show that we have some kind of "impact". But this is where we also need to pause for thought. Citation says nothing about the quality of the work we do, nor the quality of the datasets that we produce. The drivers of citation may reflect quality, but they also reflect a growing number of factors that at best reflect natural differences in the size of academic disciplines, in a more neutral sense reflect activities used to get our work read (social media, for example) and at worst reflect the kinds of manipulation that researchers can embark upon to get their work cited (e.g. when requesting citation of their own work during the review process, and there have sadly been some extreme examples of this within the field of geomorpholohy). Data can be shared in other ways, however. It can be shared through contact between the user of the data and the original producer of the data. This kind of contact allows issues surrounding data homogenisation and configurational influences to be brought into the discussion. Data producers can be involved in the intellectual exercise of putting different datasets together in a way that seeks both communalities and differences, and ultimately produces new research questions, proposals and field data collection. In requiring data to be made available without any contact with the original producers of the data (as long as they remain contactable), even if those shared data can accrue citation, we lose a critical element of what makes sharing data so very valuable and academic life so enjoyable; discussion and debate within an academic community.

It is quite possible that an unintended consequence of FAIR principles is the progressive privatization of data (Allen and Berghuis, 2018). The issue here is that we all rely to greater or lesser degrees on data that have been collected by public, quasi-public or private organisations; digital map data, weather data, hydrological data, borehole logs etc. Under the FAIR principles it is accepted that if such data have been purchased and licensed to an author, then those data need not be made openly available, but other readers should be able to obtain (i.e. purchase) the same rights to those data. A second potential user would have to buy the license to use them. There then exists the grey area between freely available data and commercialised data where the data come from a 3rd party, and this 3rd party is making the data available to authors, normally at no cost. The Coalition for Publishing Data in the Earth and Space Sciences (no date) illustrate the problem that this grey area poses. They note "*We consider third-party data to be data that cannot be legally distributed by the authors … If an author does not have the rights to distribute the data, they must include all necessary contact information in the Data Availability Statement to gain access to the relevant data. If permission is required to use a third-party data set, authors must include the third-party source and verification of permission in the Data Availability Statement, as well as provide proper acknowledgment in the article. Authors are asked to submit information from the data owner that data will be available post-publication, in the same manner as that by which the authors obtained the data.".* This last statement is where the problem lies as a commitment from the data owner at the point of publication can only be guaranteed in the future if the data are placed within a legal framework and with an appropriate and formal specification of copyright. Such third parties may have little or no interest in the data sharing agenda, and so may prefer not to provide their data rather than having to invest time in implementing it. At the other extreme, it may encourage them to see their data as having unrealized value and that it would be better to begin to provide it licensed at a price. There is a risk of damaging the delicate relations that can develop between private (potential) data providers and the academy. I can give you a concrete example. A hydropower company recently made available a 50-year series of 15-minute discharge data from glaciated basins. These data are unique in the world because they are measured at altitude close to the glaciers concerned. Their analysis revealed systematic changes in hydrograph shape following from climate warming (Lane and Nienow, 2019). The hydropower company did not want us to publish the source data, but was happy to make them available to others on request. Of course, this left the company with the right to refuse, which is not ideal. The position of the journal was very clear: we had to publish them in order to publish the paper, even though the data were not our own. If the hydropower company had sold us the data under license we would not have had to do so. We did find a compromise (publication of standardised data, enough to do the analyses in the paper, but not enough to produce some of the more commercially sensitive statistics on changing water yield), and the company was always willing for us to provide the original data for the purpose of academic review. If we had not achieved a compromise, we would have been in a position where the requirement of making data available would have held back the publication of scientific knowledge. At the same time, the hydropower company might see an incentive to commercialise their data, to limit the conditions under which the data can be used through the license as they saw fit, and to decide who they wish to sell the data to. Even if licensed, the company has no obligation to sell them to anyone. I contend that data sharing policies in relation to 3rd party data sources remain largely incoherent even though we, as geomorphologists, are often heavily dependent upon them.

If I return to the FAIR data principles, I want to restate the point made above that FAIR is an unfortunate acronym because of the way in which it links the FAIR model of data sharing with notions of what is and what is not "fair". This again comes back to questions regarding data ownership, but not 3rd party data, rather those we have collected ourselves. I fear that as an academic community we have become too willing to accept the notion that it is only fair to share our data because we are publicly funded, as are many of our research projects. Of course, this is partly true. Traditionally, the argument may have been stronger because the academic system did indeed invest significant public funds in us with relatively little demanded in return. Demands on us were significantly less severe; student numbers were lower; competition between universities for students, status and funding was less; the academic hierarchy in terms of promotion was flatter (with fewer Professors by proportion in many countries); incentives to perform were fewer; and requirements to demonstrate the impact of our work beyond the academy were rarer. As what is known as the neo-liberalisation of the academy has taken place, and the benefits of academic freedom have been replaced by ever-greater academic responsibilities and constraints, justifying data sharing in terms of public funding becomes harder and not easier. Faced with ever more crowded diaries, we increasingly subsidise data collection to a very significant degree through pur personal time and often through enormous personal sacrifice. Many of us work many more hours than we are salaried to do. Geomorphic processes don't stop at 5pm in the evening or at weekends. We don't get paid overtime to keep our field data coming in. Statements like "*Publicly funded data are a public good and should be publicly preserved and sustained over time*" (American Geophysical Union, 2015) are not as clear cut as they might seem in the neoliberal academy of the 21st Century. Enforcing FAIR data principles does not always seem particularly fair.

Conclusion

Over the last few years, data sharing in the academic publication process has progressively passed from something that is encouraged through to something that is required. This transition merits reflection. Clearly, the peer review process can be strengthened through including data in the manuscript evaluation process, but this in itself does not require public data sharing. There is a role that data sharing may play in the inter-comparison of data acquired across different geographical and historical settings, in helping us to put our own case studies into wider context and in terms of more interactive classroom teaching. My argument, then, is not that sharing geomorphological data is not a good thing. Rather, it is it that data sharing needs to proceed cautiously. We should not become complacent in thinking that applying FAIR data principles in themselves will generate new kinds of scientific understanding. It will generate understanding of a very particular type and has inherent limits. As the data available implicitly define both what has been measured and how, we run the risk of encountering problems of inter-comparability, bias and incompleteness, notably in terms of configurational information. We have to be sure the data are fit for purpose (Kirchner, 2006; Allen and Berghuijs, 2018). We certainly need to be extremely careful of the automated harnessing and analysis of any kind of geomorphological data without thinking through the extent to which the kinds of measurements being combined and their configurational aspects are fully understood. Shared data have none of the serendipity that comes from being in the field and being confronted by what we see and experience; they might as well be data from nowhere. We have to be careful that the compilation and inter-comparison that data sharing allows through extensive data analyses, when taken with wider external influences on the academy, does not undermine the perceived value of field data collection. The replacement of humans by machines in such data compilation and analysis risks removing some of the scientific creativity and reflexivity that interactions between humans motivated by data exchange can sustain. We must also be careful that we don't lead to the over-commodification of data, through commercialisation whether financially or in terms of academic currencies like citation.

These arguments explain why ESPL has decided that it will not insist on data being publicly available prior to a manuscript being accepted for publication. Rather, we will require authors to honour two points: (1), as with current practice, authors must make their data available to reviewers upon request; and (2), as a new requirement, authors must include a data statement that will be published. The latter identifies where data used in the paper may be obtained. An author who wishes to publish their data in a suitable database service will be welcome to do so, as will one who is required to do so via their Funding Body. An author who wishes to make their data available upon request is equally welcome to choose this option, but this is on the understanding that such requests will be honoured as far as is possible. If there are situations where data cannot be made publicly available, an author is will be asked to explain why in their data statement. The data statement will not be used in the scientific evaluation of manuscripts, but the journal Editors may ask authors to make changes to their data statement in the spirit of the value of data sharing. We will continue to support data sharing by allowing data to be published (and permanently archived) as supplementary online information. We note that supplementary information is linked to a paper where, as is always the case, we expect a complete, correct, justified and reproducible methodology. It is this kind of publication of data that is likely to provide information of real value as readers will be able to know not only about the data they might wish to use, but will be shown, through the paper, precisely how it can be used. But, we will not proscribe where or in what form data should be shared, and will leave this to authors to decide.

Acknowledgements
I would like to thank my Editorial Board for their constructive comments on this document and for their valuable discussion of ESPL's data sharing policy.

References

Allen, S.T. and Berghuijs, W.R., 2018. A need for incentivizing field hydrology, especially in an era of open data. *Hydrological Sciences Journal*, 63, 1262-1265

Alley, R.B., Cuffey, K.M., Evenson, E.B., Strasser, J.C., Lawson, D.E. and Larson, G.J., 1997 How glaciers entrain and transport basal sediment: Physical constraints. *Quaternary Science Reviews*, 16, 1017-38

American Geophysical Union, 2015. Position statement on scientific data. https://www.agu.org/Share-and-Advocate/Share/Policymakers/Position-Statements/ESS-Data. Accessed 15th November 2019.

Ashmore, P. and Dodson, B., 2016. Urbanizing Physical Geography. The Canadian Geographer, 61, 102–106

Blaikie, P.M., 1985. *The Political Economy of Soil Erosion in Developing Countries*. London: Longman.

Blume, T., Van Meerveld, I., and Weiler, M., 2017. The role of experimental work in hydrological sciences—insights from a community survey. *Hydrological Sciences Journal*, 62, 334–337

Bunge, W.W., 1979. Perspectives on theoretical geography. *Annals of the Association of American Geographers*, 69, 169–174

Burt, T.P. and McDonnell, J.J., 2015. Whither field hydrology? The need for discovery science and outrageous hydrological hypotheses. *Water Resources Research*, 51, 5919–5928.

Clifford, N. and Richards, K., 2005. Earth System Science: an oxymoron? *Earth Surface Processes and Landforms*, 30, 379–383

Coalition for Publishing Data in the Earth and Space Sciences, no date. Author guidelines. https://copdess.org/enabling-fair-data-project/author-guidelines/, accessed 15th November 2019.

Force11, no date. The FAIR data principles. https://www.force11.org/group/fairgroup/fairprinciples. Accessed 18th November 2019

Görg, C., 2007. Landscape governance: the "politics of scale" and the "natural" conditions of places. *Geoforum*, 38, 954–966

Hallet, B., Hunter, L. and Bogen, J., 1996 Rates of erosion and sediment evacuation by glaciers: A review of field data and their implications. *Global and Planetary Change*, 12, 213–235

Harré, R., 1981. *Great Scientific Experiments*, Phaidon, Oxford.

Herman, F., Seward, D., Valla, P.G., Carter, A., Kohn, B., Willett, S.D. and Ehlers, T.E., 2013. Worldwide acceleration of mountain erosion under a cooling climate. *Nature*, 504, 423-6.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, 42, W03S04

Lane, S.N., 2014. Good practice in authoring manuscripts on geomorphology. *Earth Surface Processes and Landforms*, 39, 126-32.

Lane, S.N., 2017. Slow science, the geographical expedition and critical physical geography. *The Canadian Geographer*, 61, 84-101

Lane, S.N. and Nienow, P.W., 2019. Decadal-scale forcing of Alpine glacial hydrological systems. *Water Resources Research*, 55, 2478-92

Litherland, K. and Stott, T.A., 2012. Virtual Field Sites: losses and gains in authenticity with semantic technologies. *Technology, Pedagogy and Education*, 21, 213-230

Nienow, P. , Sharp, M. and Willis, I., 1998. Seasonal changes in the morphology of the subglacial drainage system, Haut Glacier d'Arolla, Switzerland. *Earth Surface Processes and Landforms*, 23, 825-843

Pattison, I. and Lane, S.N., 2012. The link between land use management and fluvial flood risk: a chaotic conception? *Progress in Physical Geography*, 36, 72-92.

Phillips, J.D., 2001. Human impacts on the environment: unpredictability and the primacy of place. *Physical Geography*, 32, 321-32

Rhoads, B.L., Wilson, D., Urban, M. and Herricks, E.E., 1999 Interaction between scientists and nonscientists in community-based watershed management: Emergence of the concept of stream naturalization. *Environmental Management*, 24, 297-308

Richards, K. and Clifford, N., 2008. Science, systems and geomorphologies: why LESS may be more. *Earth Surface Processes and Landforms*, 33, 1323–1340

Richards, K.S., 1996. Samples and cases: generalisation and explanation in geomorphology. In Rhoads, B.L. and Thorn, C.E. (Eds), *The Scientific Basis of Geomorphology*, Binghamton Symposium in Geomorphology, Elsevier, 171–9

Sayer, A., 1984. Method in social science: a realist approach, Hutchinson.

Simpson, G.G., 1963. Historical science. In Albritton, C. C. (Ed.), *The Fabric of Geology*, Freeman, Cooper and Co., Stanford, CA, 24–48.

Sivapalan, M. and Blöschl, G., 2017. The growth of hydrological understanding: technologies, ideas and societal needs shape the field. *Water Resources Research*, 53, 8137–8146.

Slater, L.J., 2016 To what extent have changes in channel capacity contributed to flood hazard trends in England and Wales? *Earth Surface Processes and Landforms*, 41, 1115– 1128

Stengers, I., 2005. The cosmopolitical proposal. In Making things public, ed. B. Latour and P. Weibel. Cambridge, MA, MIT Press, 994–1003.

Stumpf, A. and Kerle, N., 2011. Object-oriented mapping of landslides using Random Forests *Remote Sensing of Environment*, 115, 2564-2577

Stumpf, R.J. II, Douglass, J., Dorn, R.I., 2008. Learning desert geomorphology virtually versus in the field. *Journal of Geography in Higher Education*, 32, 387-399

Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B. and Revhaug, I., 2016. Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, 13, 361-78

Turowski, J.M., Hovius, N. , Meng-Long, H. , Lague, D. and Men-Chiang, C., 2008. Distribution of erosion across bedrock channels. *Earth Surface Processes and Landforms*, 33, 353-363

Vidon, P.G., 2015. Field hydrologists needed: a call for young hydrologists to (re)-focus on field studies. *Hydrological Processes*, 29, 5478–5480

Warburton, J. and Higgitt, M., 1997. Improving the preparation for fieldwork with 'IT': two examples from physical geography. *Journal of Geography in Higher Education*, 21, 333-47.