The Canadian Geographer
Le Géographe canadien
Canadian Association of Geographers
Association Canadienne des Géographes

Our GIS is too small

Mark Gahegan
Centre for eResearch / Department of Computer Science, The University of Auckland

Key Messages

- As *R* is to statistics, the GIScience community needs an innovation platform to deploy and validate new ideas, and empower geographical research more broadly.
- GISystems and Science can become limiting if they end up defining how we see the world and what we believe are worthy and tractable research problems.
- At a deep level, our ontology has favoured the primacy of objects over processes: we have chosen to model the map and not the processes that give rise to it.

GIScience and GISystems have been successful in tackling many geographical problems over the last 30 years. But technologies and associated theory can become limiting if they end up defining how we see the world and what we believe are worthy and tractable research problems. This paper explores some of the limitations currently impacting GISystems and GIScience from the perspective of technology and community, contrasting GIScience with other informatics communities and their practices. It explores several themes: (i) GIScience and the informatics revolution; (ii) the lack of a community-owned innovation platform for GIScience research; (iii) the computational limitations imposed by desktop computing and the inability to scale up analysis; (iv) the continued failure to support the temporal dimension, and especially dynamic processes and models with feedbacks; (v) the challenge of embracing a wider and more heterogeneous view of geographical representation and analysis; and (vi) the urgent need to foster an active software development community to redress some of these shortcomings. A brief discussion then summarizes the issues and suggests that GIScience needs to work harder as a community to become more relevant to the broader geographic field and meet a bigger set of representation, analysis, and modelling needs.

Keywords: GIScience, technical critique, limitations of GIS

Notre SIG est trop restrictif

La science de l'information géographique et les SIG ont réussi à aborder de nombreux problèmes géographiques au cours des 30 dernières années. Mais les technologies et les théories connexes peuvent être limitatives si elles contraignent notre perception du monde et ce que nous croyons être des questions de recherche méritant d'être abordés. Cet article examine certaines des limites qui ont actuellement des répercussions sur la science de l'information géographique du point de vue de la technologie en la confrontant à d'autres communautés informatiques et à leurs pratiques. Il examine plusieurs thèmes : (i) la science de l'information géographique et la révolution informatique ; (ii) l'absence d'une plate-forme d'innovation communautaire pour la recherche en géomatique ; (iii) les limites informatiques imposées par la bureautique et l'incapacité de varier l'échelle des analyses ; (iv) l'insuffisance continue de représentation de la dimension temporelle et plus particulièrement des processus et des modèles dynamiques avec des rétroactions ; (v) le défi d'adopter une vision plus large et plus hétérogène de la représentation et de l'analyse géographiques ; et (vi) l'urgence de promouvoir une communauté du développement de logiciels active afin de combler certaines de ces lacunes. Une brève discussion résume ensuite les problèmes et suggère que la

Correspondence to / Adresse de correspondance: Mark Gahegan, Centre for eResearch, Building 302:585, 23 Symonds St, Auckland, 1142, New Zealand. Email/Courriel: m.gahegan@auckland.ac.nz

The Canadian Geographer / Le Géographe canadien 2018, 62(1): 15–26
DOI: 10.1111/cag.12434

communauté scientifique de l'information géographique doit investir plus d'efforts afin d'accroitre sa pertinence pour la géographie et de répondre à une plus grande diversité de besoins sur le plan de la représentation, de l'analyse et de la modélisation.

Mots clés : Science de l'information géographique, critique technique, limites des SIG

Introduction

Creating computational systems that both enable and provide a home for GIScience research is a bit like driving a car at the same time as various parts of the car are being proposed, added, or modified. Vannevar Bush (1967), the great science visionary and leader who oversaw the creation of the National Science Foundation (NSF) in 1945, described the process by which a research field moves forward using the following analogy in his essay, "The Builders".

> The process by which the boundaries of knowledge are advanced, and the structure of organized science is built, is a complex process indeed. It corresponds fairly well with the exploitation of a difficult quarry for its building materials and the fitting of these into an edifice; but there are very significant differences. First, the material itself is exceedingly varied, hidden and overlaid with relatively worthless rubble, and the process of uncovering new facts and relationships has some of the attributes of prospecting and exploration rather than of mining or quarrying. Second, the whole effort is highly unorganized. There are no direct orders from architect to quarrymaster. Individuals and small bands proceed about their businesses unimpeded and uncontrolled, digging where they will, working over their material, and tucking it into place in the edifice. (Bush 1967, 11[1])

In a field that is still evolving and expanding, it is difficult to recognize a sound overall structure from what has currently emerged and to construct it into a coherent analytical system, such as a GISystem. So we should expect that, from time to time, the structure and related systems that have emerged so far will be shown to be inadequate and will require some reworking.

This paper explores GIScience and GISystems from a *research process* perspective, questioning how the two are linked and why it is that so much useful GIScience research does not seem to make the transition into a working GISystem. Dawn Wright and colleagues (1997) present a useful overview on the distinctions between GIScience and GISystems. We use this distinction here: GIScience is the research field and related community; GISystems are the computer programs and tools that are used to represent, analyze, and visualize geospatial information. As a GIScientist who has morphed into a Director of eResearch (also known as eScience in Europe and CyberInfrastructure in the United States) I have had the opportunity over the last 15 years to work with informatics researchers from many different fields. I can attest to the fact that all of these domains also contain a large amount of theory and new science. But by and large their corresponding informatics communities see their role differently: to not only propose new analytical theory and methods, but also to operationalize, evaluate, and organize them within a coherent and functioning ecosystem. However, and to be clear, these are personal observations and I do not claim this to be a scientific study. If emoticons were acceptable in research journals, I would place a smiley here.

The paper also reflects on why GISystems do not empower more geography research and why the GIScience community is often not seen as a key partner in geographical research. Much has been said already from a *critical* point of view on the shortcomings of GIScience and GISystems (e.g., Pickles 1995; Schuurman 2000; O'Sullivan 2006; Crampton 2010; Goodchild 2015; Thatcher et al. 2016). However, this paper is not a methodological or epistemological critique of GIScience, but rather a series of reflections on problems with the way we do GIScience and the GISystems we have settled for as a result. It concentrates on computational and representational challenges, not because they are the only challenges, but because this is the perspective that I am best equipped to take. Note also that the arguments made below are generalizations; there are, thankfully, worthy exceptions to all of them!

[1] A more accessible reprint of the essay can be found online in *The Physics Teacher* 8 (266): http://doi.org/10.1119/1.2351487. Although published in 1967, this essay was likely written around 1945.

The name *GIScience* has been the preferred designation for our research community since the early 1990s, emboldened by Goodchild's (1992) visionary rallying cry. But the generic term given to the study of the analytical methods, tools, and information architecture associated with a given research domain is *x-informatics,* where the "*x*" denotes a variable into which a discipline name can be substituted. Hence *bioinformatics* is the study of these topics as they relate to biology and *health informatics* plays the same role for a broad section of health and medical sciences (Tolliver 2008). There are in fact many other such communities—for example, *astro-informatics*, *eco-informatics, cheminformatics* (Augen 2002; Borne 2010; Michener and Jones 2012)—a consequence of the huge shift towards digital research that has changed the nature of many disciplines over the last 25 or so years (e.g., Baker et al. 2008). The x-informatics course syllabus taught at Rensselaer Polytechnic Institute (Fox 2016) gives a practical sense of the generic aspects of x-informatics that are relevant across multiple research fields.

Just like GIScience, these other x-informatics communities have their own mature journals, conferences, and research agendas and they all trace their roots back to the theoretical foundation of their discipline (e.g., Hogeweg 2011). In this company, GIScience is somewhat of an anomaly for two reasons. First, as noted above, it claims to be a science (in preference to computational geography or the more conventional geo-informatics—both names that have been used to define our field at various points in time) and perhaps this sets a bias towards the discovery and development of science theory? Second, by and large the GIScience community does not see its primary role as providing the platform and methods that support geographical enquiry for the rest of the geographical research community, preferring instead to pursue a more independent research quest.

GIScience has indeed proven to be a very effective rallying call, and evidence suggests it successfully opened up new channels of research funding that helped to coalesce the field and support much of the early progress. And there is nothing wrong at all in striving for GIScience theory (Goodchild 2004); new theory is certainly needed. But we need to be able to clearly establish if new theory is useful, and if it is then it needs to become part of some GISystem for other researchers to use. There has, of course, been much success to celebrate too; many geographical problems have become accessible and tractable with methods developed in GIScience, particularly where the problems involve map operations and some kinds of spatial analysis (more on this later). But we can do better.

This paper outlines five issues with the current state of GIScience and offers a few suggestions to perhaps make us a more effective community, via comparison with other research fields that are also heavily based on informatics. It concludes with a brief discussion.

Issue 1: The innovation platform appears to be missing

> There are those who are quite content, given a few tools, to dig away unearthing odd blocks, piling them up in the view of fellow workers, and apparently not caring whether they fit anywhere or not. Unfortunately there are also those who watch carefully until some industrious group digs out a particularly ornamental block, whereupon they fit it in place with much gusto and bow to the crowd. Some groups do not dig at all, but spend all their time arguing as to the exact arrangement of a cornice or an abutment. Some spend all their days trying to pull down a block or two that a rival has put in place. Some, indeed, neither dig nor argue, but go among with the crowd, scratch here and there, and enjoy the scenery. Some sit by and give advice, and some just sit. (Bush 1967, 11)

By and large, we the GIScience research community do not own or develop our own codebase(s) and systems. This is a tragedy, and contrasts sharply with bioinformatics and other x-informatics disciplines where community-built, open-source, cutting-edge systems are much more common and, in many cases, are widely used by their respective disciplines as their *research engines.* There are of course some excellent counter-examples that we can point to in GIScience, such as the *PySAL* open toolkit, *PostGIS*, *GeoTools,* and the *GRASS* project.[2] Such platforms could be used as the basis to integrate and validate new methods, perhaps even

[2] Descriptions are available here: http://pysal.readthedocs.io/en/latest/#) (PySAL), http://www.postgis.net/ (PostGIS), http://www.geotools.org/ (GeoTools), and https://grass.osgeo.org/ (GRASS).

new data structures and conceptual models in the case of *GRASS* and *PostGIS*. But this is sadly not our usual practice.

So why isn't GIScience leading the development of the methods and systems needed by geography more broadly? Much of reported GIScience research to date has not found its way into GISystems; most of it remains trapped in the hallowed pages of our GIScience journals. There is of course merit in ideas for their own sake, and in places to share and discuss them. But ultimately, I believe that journals should not be their final resting place. They need to take material form, be used, be evaluated, then to be put to work or discarded, depending on their perceived value. Validating new theory and methods is actually a vital stage in the process of science itself (Oreskes 1994; Gauch 2003) and it is perhaps an uncomfortable truth that much of our GIScience research is, by this measure, not very scientific!

Contrast this with bioinformatics, where this community has provided many widely used analysis tools, and even entire workflow systems that represent and facilitate a huge range of analytical activities; for examples *Galaxy*, *Taverna,* and *Triana*, see Leipzig (2016) for a more comprehensive review.[3] One could make similar claims for the statistics field via the *R* open-source community and the huge collection of tools based on *R*[4] that are now used routinely in many fields. There are also, of course, many highly successful commercial analytics environments within use in these two fields, but despite these, the x-informatics research communities have: (i) *created* and (ii) *maintain* and (iii) *own and govern* the software platforms that successfully support open innovation and experimentation across a broad swath of the research discipline.

A sign of their success is that most of these communities can point to significant, open toolsets that not only empower many domain research questions, but also provide a cohesive and effective platform by which to validate and deploy new research—new theory, new methods, and even new data models all have a natural home, and can be independently verified, contrasted with established methods, and rolled out to the broader research community. In such a functioning ecosystem, evolution can work to push the best of these innovations into newer versions of the informatics platforms. Such a platform also allows researchers around the globe to all contribute via development, testing, comparison, and use of the tools. An excellent example of this idea in action is the data and analytics infrastructure developed by the high-energy physics informatics community to support the research carried out on the Large Hadron Collider (e.g., the virtualized analysis environment available for the Compact Muon Solenoid experiment[5]). This infrastructure is in constant use worldwide, and provides the gateway for scientists from all over the world to access and analyze the latest experimental data.

One could argue that perhaps these other fields are less epistemologically diverse or perhaps ontologically simpler, and thus are easier to support because the analytical needs are clearer to define and self-consistent. There may be some truth in this, geography certainly is a complex and multi-faceted field that is difficult to define and full of contradictions. But the more I learn about other fields, the more I see that complexity and uncertainty are the norms in science. Though for the sake of simplifying the arguments that follow, let us imagine for now that our efforts to build a better GISystem are targeted at a set of application domains with enough ontological and epistemological consistency for this to be a feasible goal: for example, the study of dynamic landscape impacts associated with disaster events such as earthquakes, tsunamis, wildfire, flooding, etc. However, we will revisit this question later, and re-broaden this argument in Issue 3.

The absence of such a dedicated platform for research makes progress in GIScience very difficult. Even when the source code is available for new methods, for example, they seldom find their way into an established research platform or toolkit because they are not engineered with this in mind. While this situation is allowed to persist, substantial progress in GIScience will be much harder to achieve, lowering our impact and importance as a

[3] See https://omictools.com/workflow-management-systems-category for a succinct summary of the more popular offerings.

[4] Such as https://www.r-project.org/, https://www.rstudio.com/products/rstudio2/, and http://r.analyticflow.com/en/.

[5] The CMS experiment and the open, computational infrastructure that supports it are described at http://opendata.cern.ch/research/CMS.

necessary partner and enabler of geographical research.

Lack of a shared experimental codebase also makes it very difficult to evaluate new ideas (say new data structures or new qualitative algebras) in any consistent way, so we do not really know which methods are best, or in what situations they perform better than others. As a consequence, we do not move forward in the sense that our better ideas do not end up being incorporated into our shared codebase. Or to draw on the opening quotation, we do not know which blocks are sound enough to build upon. This helps to explain why—despite huge research pushes in the areas of time, or spatial information theory, for example—we still do not have a GISystem that implements the stronger ideas in these fields. There are exceptions, where ideas from academia have successfully migrated into commercial GIS. One example that comes readily to mind is the *ColorBrewer*[6] tool for selecting a perceptually sound colour palette for maps that has found its way into *ArcGIS*, *R*, and several other mapping tools. This particular idea would have been easy to add to an existing GIS as it would have little impact on adopted data models and workflow, but much GIScience research is more challenging to operationalize. Of course, there are also ideas emerging from GIScience that do not require implementation, or that cannot yet be implemented—these should still be encouraged, yet recognized as such. Indeed, the idea for which Vannevar Bush is perhaps most well known is the *Memex*—a hypothesized mechanical device for discovering and connecting knowledge. It resembles the modern Internet in many respects.

Please note that this is not a criticism of commercial GIS; it is in no way *ESRI*'s fault, nor indeed that of any other commercial GISystems vendor, that GIScience research does not easily migrate into its codebase. It is down to us. We got lazy. We should not expect a third party to take responsibility for validating and delivering our ideas into a GISystem that is focused on research.

A home-grown innovation platform has some clear advantages: (i) software interfaces can be created where needed to simplify the addition of certain kinds of functionality, such as spatial analysis methods, cartographic presentation methods and external databases; (ii) it allows the GIScience research community to comparatively evaluate different methods if they all reside in the same systems, so we can determine if a new method really does outperform an established one, or offers some other advantage; (iii) the community can propose and make deeper changes to the core of the system—analytical workflows, even the underlying data structures and conceptual model itself; and (iv) the community can drive the evolution of this platform based on emerging research needs, as an endeavour in common.

The problem is, it is in nobody's short-term interests to address this glaring omission. Taking a short-term view, all of us are better off sticking to small, tractable research questions that can easily be published, and avoiding these bigger and more complex issues. And in writing this, I *out* myself as part of the problem too. It is also true that funding long-term development platforms is not seen as a priority for many funding agencies, and although this is lamentable and short-sighted, it should not be used as an excuse for our inaction.

The following ideas may help us reset our course a little.

- Insist that data and code be published for all new research that uses them, in open and accessible formats using ISO and OGC interface specifications where appropriate, and preferably hosted in repositories that are owned by the community (and not by the journal publishers). There may of course be circumstances where this is not possible (commercially funded research for example), but these should be the exception, not the rule.
- Encourage or insist that new ideas be directly compared with established ones where feasible. To move forward, we need to establish which *unearthed blocks* make the most solid foundations.
- Work towards developing community research infrastructure to provide an open codebase for GIScience research. This implies getting funding agencies on board, which is challenging. The NSF-funded *EarthScope* and EU-funded *Inspire* projects[7] both demonstrate that they can occasionally be persuaded to fund community infrastructure into the longer term.

[6] See http://colorbrewer2.org/.

[7] See http://www.earthscope.org/ and http://inspire.ec.europa.eu/ for more details.

- Begin to value the reuse of research outcomes other than journal articles, such as methods and datasets, track their uptake, and give credit for them.
- Ensure that GIScience education gives adequate attention to scientific programming, information science, and mathematics, so that postgraduates are equipped to implement and validate their ideas.

In short, we have let the available software set the limits around the research that we—and our colleagues in other fields of geography—can do, and we need to fix this.

Issue 2: Not all problems conveniently fit within the capabilities of a personal computer

Whereas many science disciplines have translated their modelling problems successfully onto High Performance Computing (HPC) and data-intensive platforms to allow them to scale up to bigger datasets, higher precision, or more analytical complexity, current GISystems and most related research seems to still be forever stuck in a single, serial thread! This seems to lead to the view that if a research problem does not fit in a GISystem, because it is too computationally demanding in terms of cores needed or memory footprint or data throughput, then it is somehow not a GIS problem. The most worrying aspect of this is that it has become so accepted that hardly anyone questions it any more. If the problem can be solved on a typical desktop computer, then it belongs to GIS, otherwise it is a problem for some other community. But in taking this line, we have created an artificial barrier between GISystems and much useful analytical functionality. We have perhaps also pushed some of the geographical research community away who need more performant analysis methods (such as erosion modelling and predator-prey dynamics). Many sub-disciplines of geography and environmental science need the analysis and visualization methods that are in current GIS, but they also need more advanced analysis and modelling tools than GIS can provide (a topic that is taken up in the next section).

Scaling up analysis capabilities is by no means a trivial exercise, and requires attention to both the algorithms we use for analysis, and the data infrastructure that supports these algorithms. Scaling up the data infrastructure can be achieved using a data distribution framework such as *Spark* or *Hadoop* and can be facilitated via the Cloud.[8] There are several good examples of success using this technology in the context of GIScience research, for example: Evans et al. (2013), Yang et al. (2017), and Wang (2013).

Scaling up algorithms, so that they can utilize several cores or threads, is perhaps the more challenging of the two problems. Much time and effort has been spent optimizing HPC codes in fields such as climate forecasting, genome sequencing, earthquake modelling, early universe cosmology, and the like. As a result of this work, there are excellent HPC templates available that support many different research problems: for example, sparse matrix linear algebra, *N*-body methods, and structured grids, all of which have strong applicability to many geographical problems. Indeed, these templates underpin most of the computing used throughout the sciences: from modelling wildfire, stream dynamics, gentrification, transportation network analysis, to tsunami and earthquake simulation. The challenge is therefore to map our algorithms to these established templates. An excellent starting point to understand these HPC challenges is provided by Asanovic and colleagues (2006). It is time that GIS scaled up: the opportunity has been understood for some time (Armstrong 1995); the HPC templates (including code) needed to address bigger and more complex problems are readily available.[9]

There is some useful research currently underway in this area, and of all the issues discussed in this paper, perhaps this one might be the furthest along towards some kind of useful resolution. Key examples of the progress here are provided by Guan and Clarke (2010), Akhter et al. (2010), Huang et al. (2012), Stojanovic and Stojanovic (2013), Wang (2013), Zhang and You (2013), Eldawy and Mokbel (2013), Qin et al. (2014), and Shook et al. (2016). It is also very encouraging to see some recent PhD theses addressing GIScience problems using HPC methods (e.g., Puri 2015; Laura 2016). The CyberGIS initiative

[8] GeoSpark (http://geospark.datasyslab.org/) is an excellent example.

[9] For example, see https://paralleldwarfs.codeplex.com/.

led out of the University of Illinois Urbana-Champaign[10] and the Centre for Intelligent Spatial Computing at George Mason University[11] deserve a specific mention, having both produced some excellent research aimed at moving some of the analytical methods used in GIScience (and GISystems) towards HPC platforms.

Note, however, that moving GIS analysis on to virtualized servers in the Cloud does *not* solve the HPC problem at all, though it may provide more performant data infrastructure or a larger memory footprint than a typical desktop machine, which can in turn alleviate a different class of performance bottleneck. But the Cloud does not provide any kind of parallel acceleration; to get this we need to rework our algorithms. Note too that without a community-wide and permanent infrastructure for these re-expressed algorithms to reside in, the GIScience HPC progress reported above will not become operational in any lasting or relevant way—a major loss.

Issue 3: The world is not static and lifeless

Despite many well-reasoned pleas in the GIScience literature for the better representation of time (e.g., Langran 1992; Peuquet 1994, 2003), we have so far failed to incorporate this vital dimension. A useful summary of the challenges and opportunities is provided by O'Sullivan (2005). A challenging critique of contemporary information systems in general for being "static and lifeless" is given by Sowa (2003).

The research focus for GIS and time so far seems to revolve around providing "temporality" for objects. At a very deep level, our GIScience ontology has favoured the primacy of objects over processes, so we have chosen to model the map and not the processes that give rise to the map. Perhaps a good place to begin is the question: *What would GISystems be like if they had been designed around dynamic processes rather than static maps?* Possibly they would better serve geography (as opposed to cartography) because processes are often more scientifically relevant to substantive research questions, with objects being in fact just a snapshot of some process at a point in time. Even human-made or *fiat* objects can be represented in this way, as the outcome of a social or political process. If we could represent in GISystems the processes that gave rise to objects, or simulate processes acting on objects, then we would have a much more powerful modelling environment. Such processes might be physical (such as water drainage), or socio-political (such as how a voting district was defined), or even personal and interpretational (such as the process of understanding a map), all of which might put us in a better position to address some of the concerns of critical theory too. The current state of supporting a set of coverages representing different times (at best) is a very poor facsimile of a dynamic system.

What GIScience has failed to do, where many other x-informatics domains have succeeded, is to move beyond the static world of representing objects and into the dynamic world of representing processes—as interactions between actors, over time, with feedbacks that can actually change the underlying model. A simple example of this is landscape erosion and deposition—the process *changes* the model as material is redistributed across the landscape over time. As such it requires that the underlying data model (such as a surface) be modified as a result of the outworking of the process, not simply that a few snapshots of the process are taken and stored. Whereas there are many good examples of dynamic models in fields such as cheminformatics and even eco-informatics, this kind of functionality eludes most GIS. A notable and worthy exception is the *PCRaster* software from Utrecht (Karssenberg et al. 2009), that actually represents a landscape as the result of physical processes, with equations defining how the landscape changes as a process such as erosion/deposition plays out. Granted, this is a narrow application domain, but it shows at least that dynamic landscape processes can be amenable to modelling.

Equation-based dependencies between the process and the representation are needed to allow GIS to expand into a range of dynamic processes that have been marginalized or pushed into other disciplines, such as hydrology in the above example. By failing to include dynamic models with feedbacks, we have created a version of geography as static and lifeless, which is at odds

[10] See http://cybergis.cigi.uiuc.edu/cyberGISwiki/doku.php.
[11] See http://cisc.gmu.edu/.

with so much geographical research, and research in related fields such as ecology and hydrology. Those who need dynamics have of course built their own tools and systems. But it is a real shame that we still cannot integrate these methods back into GISystems.

I believe the aspiration here should be to support the integration of complex, dynamic models: using research progress in systems dynamics and multi-physics (Karnopp 2012; Keyes et al. 2013). As an example, the current state of the art for modelling the functioning of the heart in bio-engineering requires solving different sets of partial differential equations for fluid flow, electro-potential, and biochemistry, at the cellular level. These different physical models obviously affect each other, so must be integrated at every step. In an ideal GIS, we may want to integrate models that are physically based with those that are socially driven. But the same underlying technology and mathematics would allow us to create a model of (say) water catchment and use in a subsiding environment that is subject to a population explosion—as a single, dynamic system with feedbacks. Such models are important now and will be likely more so in the future. They will need to draw heavily from the analytical and representational capabilities of current GISystems. So why not support them within research-oriented GISystems?

Of course, there's a great deal of variety in the kinds of dynamic systems we might wish to model, so it is reasonable to question whether adding dynamics is really a practical goal. One of the problems with the design of current GISystems is that they are based on a fairly ad hoc and restricted set of logical and statistical operators, rather than on a comprehensive library of mathematical functions that can be easily exposed to application developers, and upon which many different kinds of dynamic functionality can be constructed. The library of equation forms needed to cover a very large number of varied applications is surprisingly small, even though at the user interface the applications may appear radically different. Using a standard equation library as the basis for our algorithms also means that the HPC templates described in Issue 2 are more likely to be directly amenable to us.

Issue 4: Too little of geographical research is catered for by GIS

Taking the above issues together, it is perhaps clear to see why GIScience is not leading the development of the tools and methods needed by geographical research fields more broadly. Obviously, though, it is challenging to engage with such a "*multidisciplinary and multiparadigmatic field*" (Blaschke and Merschdorf 2014, 196). There have been many serious attempts to broaden GIScience theory to meet a wider swath of geographical needs, in the areas of time and process, critical theory, feminist theory, visualization, qualitative spatial reasoning, and others; Reitsma (2013) provides a very useful summary of such efforts. But still the majority of geographical problems cannot be solved in GIS and the majority of problems that are addressed by GIScience seem to me to be ones that GIScientists have chosen for themselves—the bigger agenda of enabling geographical research appears to be largely absent or in some cases still just an agenda. A very pertinent and fascinating critique of GISystems is provided by Hacıgüzeller (2012), which has the helpful property that it is from the perspective of another discipline—archaeology.

It is not only the lack of temporal dynamics and process modelling that restricts the utility of GIS to geographical analysis more generally. Existing GISystems are built around two useful but limited representational paradigms—the raster grid and the vector map (usually but not always including point, line, region topology)—and offer analysis methods that fit within these paradigms. There are a lot of potentially useful methods in this space, both for analysis and for visualization, and GIScience has done well here to propose, test, and improve what we have. But there are entirely missing representational paradigms that are (or could be) very useful and possibly better for some tasks, such as Voronoi Spaces, hexagonal grids, spaces defined by field equations (i.e., continuous mathematical functions as opposed to discrete shapes), mereological and qualitative spaces, to name just a few. These are important research topics in GIScience and geographical analysis, and have real value in terms of engaging with a broader set of research questions, but are marginalized because they are not available in mainstream GISystems for researchers to build on and turn into useful platforms. As such, we are stuck with the geographical modelling paradigms that were developed *first*, not necessarily the ones that are *best*. For example, it seems nothing short of a tragedy that the decades of work on qualitative and topological spatial reasoning have yet to find a home in any comprehensive GISystem (the Conference on Spatial Information Theory has been running biennially since 1993).

The same argument can be made for missing analytical functionality, such as cartographic animation, agent-based models, spatial interaction models, and the like. During the quantitative revolution, those involved imagined quantitative geography to be a vast field, with many useful analytical representations that could be explored (Haggett and Chorley 1967; Macmillan 1997), or later that emerging fields such as geocomputation would help to restore some of the missing analytical functionality (Gahegan 1999). What we have settled for in GISystems is far narrower than this initial vision and we are impoverished as a result.

Another useful development here is that it is getting easier for researchers to add in or link missing analytical functionality to existing systems (*ArcGIS* server object extensions for example) and this certainly helps. But such changes are always going to be limited in terms of scope. Adding missing conceptual models and related data structures requires that we have our own codebase. From an even more radical point of view, perhaps we could extend GIScience to embrace different epistemologies and ways of knowing, beyond those used in normal science and towards those used in more *critical* approaches to geographic research (Thatcher et al. 2016)?

This points to two important needs, going forward:

- To design systems so that they can accommodate new methods, modelling paradigms, and theories, i.e., that can evolve. Designs will need to be very modular and provide well-documented interfaces at multiple level of abstraction.
- To recognize the methods, models, and theories we are currently using in GISystems, and their shortcomings, and make plans to address them. We are good at the former (recognition and creation of research agendas), but poor at the latter.

Issue 5: Fostering of a GISystems development community is urgently needed

In the same article cited at the beginning of this paper, Bush (1967, 11) describes the importance of those who turn imagined ideas into real (infra) structure: "There are those who labor to make the utility of the structure real, to cause it to give shelter to the multitude, that they may be better protected, and that they may derive health and well-being because of its presence."

The open source movement extends across the research sector and has done a marvellous job of lowering the financial barrier of entry into many research domains and of democratizing the software development process (Warger 2002). One of the most exciting things to emerge in the GIS universe in the last few years is the Open Source Geospatial Foundation (colloquially known as OSGeo[12]) whose dedicated community is committed to developing and delivering GISystems and technology following open-source principles. The work here is highly laudable and the community is clearly both passionate about their vision and supportive of the idea that GIS needs to be open for it to progress healthily. They have created some strong, alternative GIS platforms based on open codebases (e.g., *OSGeo4W* and *QGIS*[13]), platforms that are owned by the community and thus can be modified as the community sees fit. These examples demonstrate that building and maintaining a GISystem codebase (or an innovation platform for GIScience) is indeed a viable goal for a community of determined people. Note also that these efforts are by and large unfunded, these are volunteers choosing to share their efforts "to make the utility of the structure real" (Bush 1967, 11).

Perhaps equally importantly, these activities are producing a new generation of programmers, with all the knowledge and skills needed to create and maintain their GIS platforms and systems (Elliott and Scacchi 2008). As a community, I believe we should be more engaged with OSGeo, and I note that both the International Cartographic Association and the International Society for Photogrammetry and

[12] For a brief introduction to OSGeo, see http://www.osgeo.org/content/foundation/about.html.

[13] See http://trac.osgeo.org/osgeo4w/ and https://www.qgis.org/en/site/about/index.html.

Remote Sensing are already collaborating with OSGeo in a variety of ways. For our long-term health, we need to join forces with this group, GIScience needs what they can offer and we also have much to give.

The missed opportunity—so far—is that open GIS projects are often based around replicating the functionality of existing commercial GIS, and so are usually based on the same restrictive assumptions about the world, the map, and the importance of structure over process. But we could work with this community to create an open GIScience research platform as a fast track to resolving some of the issues described above, with shared governance.

Discussion

To revisit the opening quote: it is easy to throw rocks, far harder to be constructive and suggest how they might instead be fashioned into a useful structure. The problem is NOT commercial GIS, because commercial GIS are responsive to the commercial market. The problem is creating and maintaining an effective platform for research, and by researchers. The major issues to address here are:

1. *Market failure*: The market has not provided what we need, a research-focused software platform, nor will it. We need to take responsibility for this ourselves, as many other x-informatics communities have before us.
2. *Funding failure*: Lack of focus on sustaining research infrastructure, as opposed to the next round of new ideas. Or conversely, lack of incentives for the community to resolve this problem in the absence of funding.
3. *Education failure*: Lack of coding and software engineering skills among GIScientists. GIScientists emerge with many kinds of backgrounds, but we need to equip more of them with the tools they need to build. Software Carpentry[14] is a useful place to start for practical skills needed for coding.
4. *Cultural failure*: An academic culture that fails to reward those who build or maintain tools and software and encourages a short-sighted and individualistic approach to research. Clearly this problem is much more complex to address since it involves challenging the norms of research productivity. But these are indeed being challenged across a range of research fields (e.g., Barnes 2010; Baker 2016).

GIScience sometimes reminds me of a smart college graduate: passionate, articulate, and capable; brought up by loving but rather self-absorbed parents; still living at home, in the basement, playing video games and afraid to take on any real responsibility. If we want GIScience to remain vibrant and relevant, I believe we urgently need to rethink how we meet some of the challenges described above. We need to set our sights higher, think bigger, take a deep breath and imagine what a shared, open, dynamic, geographic analysis platform could do for us, if it integrated some of those complex models and simulations that now live outside of GIScience. Add to that the layers of societal, infrastructural, demographic, and health-related data that we DO analyze well; now that would be an impactful and highly relevant x-informatics, one perhaps worthy of being called a science.

In “The Builders”, Bush (1967) goes on to describe the importance of those who unearth new ideas, cautions about those who take credit when they did not really do the work or spend their time arguing about less important details. He underscores the need for a functioning community of many roles, all working together, including: (a) visionaries who can see in advance where pieces may fit, (b) those who give meaning and context to the edifice, (c) those who construct the emerging pieces into a working structure and make it useful, and (d) even those who encourage and facilitate behind the scenes. All these roles are needed in a healthy community. I’d humbly suggest that GIScience has done rather well on (a), (b), and perhaps also (d), but is currently failing with (c). This is fixable: we need to nurture and reward more “builders.”

Some self-reflection is needed by each of us: *What is our contribution to the building of GIScience and how does it help*? In the GIScience literature we seem to have a lot of papers on research agendas, conceptual frameworks, reviews of progress in the field, and papers defining what the field is or critiquing it. We also have a large pile of interesting

[14] A description of the Software Carpentry community and their teaching materials can be found at https://software-carpentry.org/.

ideas, most of which are still lying on the floor of the quarry, in a heap, waiting for somebody to incorporate them into a working system and thereby establish their worth.

It seems appropriate, by way of encouragement, to let Vannevar Bush have the last word.

> Finally, the edifice itself has a remarkable property, for its form is predestined by the laws of logic and the nature of human reasoning. It is almost as though it had once existed, and its building blocks had then been scattered, hidden, and buried, each with its unique form retained so that it would fit only in its own peculiar position, and with the concomitant limitation that the blocks cannot be found or recognized until the building of the structure has progressed to the point where their position and form reveals itself to the discerning eye of the talented worker in the quarry. (Bush 1967, 11)

Acknowledgements

Many thanks to the two anonymous reviewers, whose thoughtful comments helped to improve this manuscript.

References

Akhter, S., K. Aida, and Y. Chemin. 2010. GRASS GIS on high performance computing with MPI, OpenMP and Ninf-G programming framework. In *Proceedings of the ISPRS Technical Commission VIII Symposium*. International Society for Photogrammetry and Remote Sensing, 580–585.

Armstrong, M. P. 1995. Is there a role for high performance computing in GIS? *Journal of the Urban and Regional Information Systems Association* 7(1): 7–10.

Asanovic, K., R. Bodik, B. C. Catanzaro, J. J. Gebis, P. Husbands, K. Keutzer, D. A. Patterson, et al. 2006. The landscape of parallel computing research: A view from Berkeley. Technical Report UCB/EECS-183. Berkeley, CA: Electrical Engineering and Computer Sciences, University of California at Berkeley.

Augen, J. 2002. The evolving role of information technology in the drug discovery process. *Drug Discovery Today* 7(5): 315–323.

Baker, D. N., C. E. Barton, W. K. Peterson, and P. Fox. 2008. Informatics and the 2007–2008 electronic geophysical year. *Transactions of the American Geophysical Union* 89(48): 485–486.

Baker, M. 2016. Is there a reproducibility crisis? *Nature* 533(7604): 452–454.

Barnes, N. 2010. Publish your computer code: It is good enough. *Nature* 467(7317): 753.

Blaschke, T., and H. Merschdorf. 2014. Geographic information science as a multidisciplinary and multiparadigmatic field. *Cartography and Geographic Information Science* 41(3): 196–213.

Borne, K. D. 2010. Astroinformatics: Data-oriented astronomy research and education. *Earth Science Informatics* 3(1–2): 5–17.

Bush, V. 1967. The builders. In *Science is not enough*, ed. V. Bush. New York, NY: William Morrow & Company, 11–13.

Crampton, J. W. 2010. Chapter 8: GIS after critique: What next? In *Mapping: A critical introduction to cartography and GIS*, J. Crampton. London, UK Wiley-Blackwell, 98–111.

Eldawy, A., and M. Mokbel. 2013. A demonstration of SpatialHadoop: An efficient map-reduce framework for spatial data. *Proceedings of the VLDB Endowment* 6(12): 1230–1233.

Elliott, M. S., and W. Scacchi, 2008. Mobilization of software developers: The free software movement. *Information Technology & People* 21(1): 4–33.

Evans, M. R., D. Oliver, K. Yang, X. Zhou, and S. Shekhar. 2013. Enabling spatial big data via CyberGIS: Challenges and opportunities. In *CyberGIS: Fostering a new wave of geospatial innovation and discovery*, ed. S. Wang and M. F. Goodchild. Berlin: Springer, 1–22.

Fox, P. 2016. *Xinformatics 2016*. Course website. Rensselaer Polytechnic Institute, https://tw.rpi.edu/web/Courses/Xinformatics/2016.

Gahegan, M. 1999. What is GeoComputation? *Transactions in GIS* 3(3): 203–206.

Gauch, H. G., Jr. 2003. *Scientific method in practice*. Cambridge, UK: Cambridge University Press.

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

——. 2004. The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers* 94(2): 300–303.

——. 2015. Two decades on: Critical GIScience since 1993. *The Canadian Geographer* 59(1): 3–11.

Guan, Q., and K. Clarke. 2010. A general-purpose parallel raster processing programming library test application using a geographic cellular automata model. *International Journal of Geographical Information Science* 24(5): 695–722.

Hacıgüzeller, P. 2012. GIS, critique, representation and beyond. *Journal of Social Archaeology* 12(2): 245–263.

Haggett, P., and R. J. Chorley, eds. 1967. *Models in geography*. London, UK: Methuen.

Hogeweg, P. 2011. The roots of bioinformatics in theoretical biology. *PLoS Computational Biology* 7(3). https://doi.org/10.1371/journal.pcbi.1002021.

Huang, Q., C., Yang, K. Benedict, A. Rezgui, J. Xie, J. Xia, and S. Chen. 2012. Using adaptively coupled models and high-performance computing for enabling the computability of dust storm forecasting. *International Journal of Geographical Information Science* 27(4): 765–784.

Karnopp, D. C., D. L. Margolis, and R.C. Rosenberg. 2012. *System dynamics: modeling, simulation, and control of mechatronic systems*. 5th ed. New York, NY: Wiley.

Karssenberg, D., O. Schmitz, P. Salamon, K. de Jong, and M. F. P. Bierkens. 2009. A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. *Environmental Modelling and Software* 25(4): 489–502.

Keyes, D. E., L. C. McInnes, C. Woodward, W. Gropp, and E. Myra. 2013. Multiphysics simulations: Challenges and opportunities. *International Journal of High Performance Computing Applications* 27(1): 4–83.

Langran, G. E. 1992. *Time in geographical information systems.* London, UK: Taylor and Francis.

Laura, J. 2016. A taxonomy of parallel vector spatial analysis algorithms. PhD Dissertation, School of Geographical Sciences & Urban Planning, Arizona State University.

Leipzig, J. 2016. A review of bioinformatic pipeline frameworks. *Briefings in Bioinformatics* 18(3): 530–536.

Macmillan, W. D. 1997. Computing and the science of geography: The postmodern turn and the geocomputational twist. In *Proceedings of the 2nd International Conference on GeoComputation*, ed. R. T. Pascoe. Leeds, UK: Centre for Computational Geography, 1–11.

Michener, W. K., and M. B. Jones. 2012. Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in Ecology and Evolution* 27(2): 85–93.

O'Sullivan, D. 2005. Geographical information science: Time changes everything. *Progress in Human Geography* 29(6): 749–756.

——. 2006. Geographical information science: Critical GIS. *Progress in Human Geography* 30(6): 783–791.

Oreskes, N., K. Shrader-Frechette, and K. Belitz. 1994. Verification, validation, and confirmation of numerical models in the Earth Sciences. *Science* 263(5147): 641–646.

Peuquet, D. J. 1994. It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. *Annals of the Association of American Geographers* 84(3): 441–461.

——. 2003. *Representations of space and time.* New York, NY: Guilford Press.

Pickles, J., ed. 1995. *Ground truth: The social implications of Geographic Information Systems.* New York, NY: Guilford.

Puri, S. 2015. Efficient parallel and distributed algorithms for GIS polygon overlay processing. PhD Dissertation, Department of Computer Science, Georgia State University.

Qin, C., L. Zhan, and A. Zhu. 2014. How to apply the Geospatial Data Abstraction Library (GDAL) properly to parallel geospatial raster I/O. *Transactions in GIS* 18(6): 950–957.

Reitsma, F. 2013. Revisiting the 'Is GIScience a science?' debate (or quite possibly scientific gerrymandering). *International Journal of Geographical Information Science* 27(2): 211–221.

Schuurman, N. 2000. Trouble in the heartland: GIS and its critics in the 1990s. *Progress in Human Geography* 24(4): 569–590.

Shook, E., M. E. Hodgson, S. Wang, B. Behzad, K. Soltani, A. Hiscox, and J. Ajayakumar. 2016. Parallel cartographic modeling: A methodology for parallelizing spatial data processing. *International Journal of Geographical Information Science* 30(12): 2355–2376.

Sowa, J. F. 2003. *Signs, processes, and language games. Foundations for ontology.* http://www.jfsowa.com/pubs/signproc.htm.

Stojanovic, N., and D. Stojanovic. 2013. High-performance computing in GIS: Techniques and applications. *International Journal of Reasoning-based Intelligent Systems* 5(1): 42–48.

Thatcher, J., L. Bergmann, B. Ricker, R. Rose-Redwood, D. O'Sullivan, T. J. Barnes, L. R. Barnesmoore, et al. 2016. Revisiting critical GIS. *Environment and Planning A* 48(5): 815–824.

Tolliver, J. 2008. *Musings on 'informatics' found on the Web.* Link no longer active; last accessed June 17, 2013. https://rhino.ncsa.illinois.edu/display/extra/How+others+define+informatics.

Wang, S., ed. 2013. Special issue on CyberGIS. *International Journal of Geographical Information Science* 27(11).

Warger, T. 2002. The Open Source movement. *Educause Quarterly* 3: 18–20

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

Yang, C., Q. Huang, Z. Li, K. Liu, and F. Hu. 2017. Big Data and cloud computing: Innovation opportunities and challenges. *International Journal of Digital Earth* 10(1): 13–53.

Zhang, J. S., and S. You. 2013. High-performance quadtree constructions on large-scale geospatial rasters using GPGPU parallel primitives. *International Journal of Geographical Information Science* 27(11): 2207–2226.