Title
A review of the emergent ecosystem of collaborative geospatial tools for addressing environmental challenges

Authors
Palomino, Jenny
Muellerklein, Oliver C
Kelly, Maggi

Publication Date
2017-09-01

DOI
10.1016/j.compenvurbsys.2017.05.003

Peer reviewed

Computers, Environment and Urban Systems 65 (2017) 79–92

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Review

A review of the emergent ecosystem of collaborative geospatial tools for addressing environmental challenges

Jenny Palomino [a,b], Oliver C. Muellerklein [a], Maggi Kelly [a,b,c,*]

[a] Department of Environmental Sciences, Policy and Management, University of California, Berkeley, Berkeley, CA 94720-3114, United States
[b] Geospatial Innovation Facility, University of California, Berkeley, United States
[c] University of California, Division of Agriculture and Natural Resources, United States

ARTICLE INFO

Article history:
Received 30 December 2016
Received in revised form 17 April 2017
Accepted 17 May 2017
Available online xxxx

Keywords:
Spatial Data Science
Collaboration
Tools
Multi-user
Workflows

ABSTRACT

To solve current environmental challenges such as biodiversity loss, climate change, and rapid conversion of natural areas due to urbanization and agricultural expansion, researchers are increasingly leveraging large, multi-scale, multi-temporal, and multi-dimensional geospatial data. In response, a rapidly expanding array of collaborative geospatial tools is being developed to help collaborators share data, code, and results. Successful navigation of these tools requires users to understand their strengths, synergies, and weaknesses. In this paper, we identify the key components of a collaborative Spatial Data Science workflow to develop a framework for evaluating the various functional aspects of collaborative geospatial tools. Using this framework, we then score thirty-one existing collaborative geospatial tools and apply a cluster analysis to create a typology of these tools. We present this typology as a map of the emergent ecosystem and functional niches of collaborative geospatial tools. We identify three primary clusters of tools composed of eight secondary clusters across which divergence is driven by required infrastructure and user involvement. Overall, our results highlight how environmental collaborations have benefitted from the use of these tools and propose key areas of future tool development for continued support of collaborative geospatial efforts.

Contents

* Corresponding author at: Department of Environmental Sciences, Policy and Management, University of California, Berkeley, Berkeley, CA 94720-3114, United States.
E-mail address: maggi@berkeley.edu (M. Kelly).

http://dx.doi.org/10.1016/j.compenvurbsys.2017.05.003
0198-9715/

1. Introduction

Environmental challenges such as biodiversity loss, wildfire management, climate change, and rapid conversion of natural areas due to urbanization and agricultural expansion are recognized as "wicked problems" (Allen & Gould, 1986; Balint, Stewart, & Desai, 2011; Carroll, Blatner, Cohn, & Morgan, 2007; Temby, Sandall, Cooksey, & Hickey, 2016), or "complex social-ecological systems" (Akamani, Holzmueller, & Groninger, 2016). Many of these challenges can be described as global in scale, at the nexus of interdisciplinary approaches, and/or part of coupled processes. Research teams have also become larger, more distributed, and multi-disciplinary (Elwood, Goodchild, & Sui, 2012; MacEachren & Brewer, 2004). To address these challenges, researchers have called for collaboration not only in the environmental management and decision-making processes (Daniels & Walker, 2001; Frame, Gunton, & Day, 2004; Selin & Chevez, 1995), but also in the knowledge production process, including the sharing of data, methods and tools (Cravens, 2014; Head & Alford, 2015; Temby et al., 2016). Consequently, understanding how various technologies, including geospatial tools, can support collaborative efforts for environmental problem-solving is a critical area of ongoing research (Cravens, 2014; Cravens, 2016; MacEachren & Brewer, 2004; Wright, Duncan, & Lach, 2009).

Contemporaneous to the emergence of these complex and large-scale research challenges has been a rapid expansion in the sources of geospatial data from mobile devices, environmental sensors, and Unmanned Aerial Vehicles (Miller & Goodchild, 2015) as well as from increased public access to administrative data through cloud/web-based Application Programming Interfaces (APIs; Anselin, 2015). In addition, Volunteered Geographic Information (VGI; Goodchild, 2007) as well as data captured by citizen scientists continue to increase in volume (Dickinson, Zuckerberg, & Bonter, 2010; Dickinson et al., 2012), both complementing and challenging the anonymity and centralized nature of traditional geospatial data produced by large organizations (i.e. governments and proprietary companies). Available data are now more detailed, with changes in scale from local to global extents, from coarse spatial resolutions in 2D planimetric to fine grain sizes with 3D and 4D options, and from seasonal/monthly temporal scales to daily or real-time capture. As such, researchers working on environmental challenges are increasingly leveraging large, multi-scale, multi-temporal, and multi-dimensional geospatial data in search of solutions (Goodman, Parker, Edmonds, & Zeglin, 2014; Miller & Goodchild, 2015).

Complementing this explosion in data has been the development of diverse array of geospatial analytical tools (i.e., scripting libraries, open source and cloud/web-based mapping options) and increased functionality to support multi-user workflows (i.e. standardized working environments, code-sharing, data exchange, status updates). Through advances in Web 2.0 technologies (Haklay, Singleton, & Parker, 2008) and Free and Open Source Software for Geospatial (FOSS4G; Steiniger & Hunter, 2013), the primary use of geospatial data is evolving from proprietary desktop software and data formats used to create static cartographic products toward the leveraging of open source and cloud/web-based tools, open data format and standards, and APIs to create dynamic web visualizations shared by collaborative teams across technology, science, and the public.

These intertwined evolutions in available geospatial data and tools also highlight the ongoing discussion regarding the role of technology within collaborative projects and how to best leverage technology to support collaborative tasks. Successful collaboration is dependent on many things including dynamics of negotiation, equity in knowledge and power, inclusion and access, and trust, which have been explored by various researchers (Elwood, 2006; Sieber, 2000; Wright et al., 2009). In addition to these social dimensions, collaboration is also dependent on the technology used to complete and achieve the desired tasks and outcomes (Cravens, 2014; Cravens, 2016). In their seminal work on "geocollaboration", MacEachren and Brewer (2004) identify four "stages of group work" as "explore, analyze, synthesize, present" (pg. 7) and explain that these stages represent "collaborative tasks for knowledge construction" (pg. 19) that can be accomplished using technology, especially those for geovisualization.

MacEachren and Brewer (2004) also offer a definition of collaboration that applies well to the context of leveraging geospatial data and technology for environmental problem-solving: "a committed effort … of two or more people to use geospatial information technologies to collectively frame and address a task involving geospatial information" (pg. 2). MacEachren and Brewer (2004) categorize these multi-user collaborations into four types: same place-same time, same place-different time, different place-same time, and different place-different time, stating that these last two (different place) were still primarily in the prototype phase at the time of their publication and were being driven by advances in database and web technology.

Since then, as these technological advances have progressed further, there has been a rise in technologies that support all of these collaborations, most notably for different place-different time collaborations. In particular, the logistics and mechanisms provided for collective work by technology in general, and geospatial ones in particular, have been identified by other researchers in varying descriptions of collaborations between scientists, non-scientists, and the general public: "collaboratories" (or collaboration laboratories; Pedersen, Kearns, & Kelly, 2007; Wulf, 1993) and "geocollaboratories" (specifically "work by geographically distributed scientists about geographic problems" MacEachren et al., 2006, pg. 201), participatory planning and management (Jankowski, 2009; Kelly, Ferranto, Lei, Ueda, & Huntsinger, 2012; Voss et al., 2004; Wright et al., 2009), citizen science efforts (Connors, Lei, & Kelly, 2012; Dickinson et al., 2010; Dickinson et al., 2012), observatory networks such as National Ecological Observatory Network (NEON; Goodman et al., 2014), virtual networks for collaboration such as Geosciences Network (GEON; Gahegan, Luo, Weaver, Pike, & Banchuen, 2009) and Human-Environment Regional Observatory (HERO; MacEachren et al., 2006) and "action ecology" (White et al., 2015). Through these collaborative efforts, researchers highlight how advances in geospatial data and tools provide technical support for collaborations through facilitation of: (i) group use and development of technology (i.e. field data collection at broad and long scales; dispersed responsibility of tasks); (ii) sharing and peer reviewing of data and results (i.e. crowdsourcing of data validation; data editing by multiple users); (iii) communication between stakeholders (i.e. ability for stakeholders to share their different representations of space and project outcomes); and (iv) integration of complementary tools (i.e. combining geospatial and communication-oriented tools; integration of big data tools and open data formats). Hence, the technical capabilities of geospatial tools can provide the practical mechanisms and infrastructure that allow people to successfully work together on tasks and goals, despite their distributions across time and space.

While it is evident that geospatial tools can support collaboration through providing the technological infrastructure needed for collaborative tasks, existing literature does not yet provide a clear framework for evaluating geospatial tools based on how well they support completion of these collaborative tasks. Furthermore, as projects can differ greatly in their requirements, there is no single tool that fulfills all needs and often, multiple tools must be integrated into workflows. As such, in addition to features that support workflows across multiple users, geospatial tools also need to support interoperability between tools (i.e. transfer of data, methods and results between tools). Consequently, successful navigation of the ever-expanding list of collaborative (i.e. multi-user) geospatial tools requires an understanding of their strengths, synergies, and weaknesses, specifically regarding functionality for collaborative tasks and capabilities for tool interoperability.

A typology of geospatial tools can provide a roadmap for these explorations by focusing on technical infrastructure for collaborative tasks such as setting up common working environments and shared data exploration, analysis, and visualization. This typology would also illustrate connections between collaborative geospatial tools as an ecosystem with identifiable niches of functionality. In this paper, we provide such a typology of the emergent ecosystem of collaborative geospatial tools by evaluating how key multi-user tools address technical barriers to collaboration through their varying capabilities and functionality.

The three objectives of this paper are to:

1. Select representative case studies (i.e. collaborative geospatial tools) that have been developed to support multi-user geospatial workflows;
2. Develop a quantitative and reproducible framework to evaluate the tools based on the key components of a collaborative Spatial Data Science workflow; and
3. Apply a cluster analysis to develop a typology of collaborative geospatial tools.

To provide a conceptual understanding of our evaluation framework, we first review the key factors that have led to the evolution of a collaborative Spatial Data Science workflow. Next, we describe how others have previously outlined typologies of geospatial and collaboration tools. Last, we apply our quantitative framework to score and cluster multi-user geospatial tools based on their functionality for collaborative tasks. Overall, we use this typology to present a map of the emergent ecosystem and niches of tools, highlight how environmental research collaborations have benefitted from the strengths of these tools, and propose key areas of future tool development for continued support of collaborative geospatial workflows. We believe that understanding the current ecosystem of collaborative geospatial tools can highlight opportunities for expanded or new functionality, promote stronger interoperability between existing tools, and help stakeholders to leverage the best tools for their needs.

2. The evolution of a collaborative Spatial Data Science workflow

In their fundamental work on geocollaboration, MacEachren and Brewer (2004) identify that while many geospatial projects are pursued as group efforts, most geospatial technologies at the time of their writing were developed and evaluated for individual use. To address this discrepancy, the authors propose “geocollaborative environments” that are focused on providing a shared working environment, whether or not the users are in the same physical environment or collaborating in real time. In addition to technical barriers to collaboration in working environments, Steiniger and Hunter (2013) identify barriers to open science stemming from the lack of transparency in analysis methods and programming code. These authors highlight various publications (e.g. Ince, Hatton, & Graham-Cumming, 2012; Morin et al., 2012; Rocchini & Neteler, 2012) arguing against “proprietary- ‘black box’ - programs that hinder scientific advancement and testing” (p. 147). Specifically, Rocchini and Neteler (2012) urge ecologists to embrace Stallman's (1985) “four freedoms” paradigm of FOSS to freely execute, modify, and share programs, while also identifying the need for better mechanisms (i.e. tools) for scientists to share “the backbone of ecological software: its code” (p. 311). Seemingly responding to the call by MacEachren and Brewer (2004) for better “multi-user system interfaces” as well as calls for increased application of FOSS and open science ideas to geospatial research, there are now more multi-user options to share data and code than ever before. As such, new concerns have arisen about how to choose the *right* tool, especially when evaluating newer FOSS4G and cloud/web-based tools (Steiniger & Hunter, 2013).

The proliferation of cloud/web-based and FOSS4G tools also highlights the progression from the traditional desktop model of Geographic Information Science (Goodchild, 1992; abbreviated to GI Science per Hall, 2014) to an advancing geospatial cyberinfrastructure, or CyberGIS (Anselin, 2012; Wang et al., 2013; Wright & Wang, 2011; Yang, Raskin, Goodchild, & Gahegan, 2010). In particular, the CyberGIS community has promoted the integration of existing GI Science and spatial analysis tools with cyberinfrastructure tools that harness cloud and high performance computing technologies (i.e. distributed, parallel, clustered) for scalable geospatial data research. Another strong focus of CyberGIS has been on tool interoperability in order to promote the sharing of data and methods as well as reduce the plethora of narrowly customized tools and “non-sharable stove-piped data systems” (Yang et al., 2010, pg. 272). In the quest to transform the technological infrastructure available for geospatial research, CyberGIS has also recognized the importance of support for shared problem-solving, distribution of geospatial data in flexible and secure ways, and community-driven solutions for wrangling and analyzing large and complex datasets (Wang et al., 2013; Wright & Wang, 2011; Yang et al., 2010).

Supported by CyberGIS technical frameworks for tool integration and interoperability, a complementary field of Spatial Data Science is emerging as an interdisciplinary approach to leveraging the spatial data explosion provided by sensor networks, VGI and mobile technologies, and by the open data and science movements (Anselin, 2015; Jiang & Thill, 2015; Wang, 2016; Yuan, 2016). Identifying Spatial Data Science as a branch of the broader Data Science field, Anselin (2015) describes it as “a combination of… exploratory spatial data analysis, spatial statistics, and spatial econometrics from a statistical disciplinary perspective, with spatial data mining, spatial database manipulation, and machine learning from a computer science disciplinary perspective” (p. 1). Maintaining the connection to GI Science, Yuan (2016) similarly describes Spatial Data Science as the domain of “Geography, Statistics, Computer Science” communities that focus on “Spatial Statistics, Spatial Big Data, Machine learning” (p. 5). While exploring synergies between CyberGIS and Spatial Data Science, Wang (2016) identifies key aims of Spatial Data Science as “scalable spatial data access, analysis and synthesis” (p. 3), with key challenges to these goals being data aggregation and data integration. Similarly, Anselin (2015) identifies related challenges as issues of “scale … endogeneity … [and] computational efficiency to deal with large amounts of very fine grained geographical data in near real-time” (p. 2).

While not explicitly mentioned in these definitions of Spatial Data Science, data wrangling (i.e. harnessing, cleaning, transforming) “often constitutes the most tedious and time-consuming aspect of analysis” (Kandel et al., 2011, pg. 271). As such, effective data wrangling plays a key role within modern geospatial workflows and is integral for overcoming the identified challenges of data aggregation, integration and scalability. Similarly, in these descriptions of Spatial Data Science, little emphasis is placed on data representation and visualization, as compared to analysis and synthesis. In particular, cartographic principles remain central to the display and representation of geospatial data, especially for the web. This is evident through the focus on color palettes (i.e. tools like ColorBrewer), vector line simplification (i.e. algorithm-based tools like MapShaper and Simplify), typological representations (i.e. data formats like TopoJSON), and efficient rendering of basemaps and large datasets (i.e. data formats like Vector Tiles). Similarly, visualization has also been identified as a key component of data analysis, as it is “particularly essential for analysing phenomena and processes unfolding in geographical space” (Andrienko & Andrienko, 2013, pg. 3). For example, visual analytics provides methods and tools for analyses of large spatial datasets through interactive visualization of iteratively mined data and has proven particularly important for movement data such as mobile and VGI (Andrienko & Andrienko, 2013; Beecham, Wood, & Bowerman, 2014; Stange, Liebig, Hecker, Andrienko, & Andrienko, 2011).

In light of these descriptions, Spatial Data Science can be seen as standing at the intersection of the three fields of GI Science, Data Science and CyberGIS (Fig. 1). Through this intersection, Spatial Data Science unites the statistical, data mining, and web-enabled data visualization techniques of Data Science with fundamental GI Science methods and principles of spatial analysis, geoprocessing, and cartography within the computational infrastructure and interoperability potential provided by CyberGIS. With an emphasis on standardized and repeatable workflows, Spatial Data Science promotes the compilation and

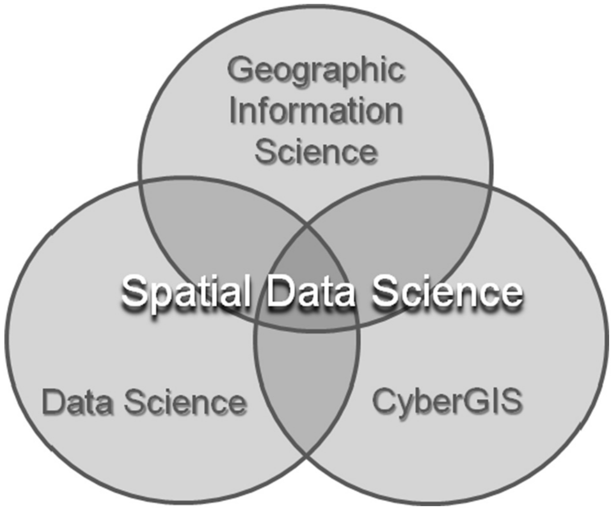

Fig. 1. Spatial Data Science at the intersection of GI Science, Data Science and CyberGIS.

integration of disparate data from multiple sources, the use of open source and cloud/web-based technologies for robust data analysis, and the leveraging of an expanding suite of data visualization and publication tools to support communication between project collaborators, the public, and other stakeholders. Due to the increasing overload of geospatial data available, the harnessing of tools that assist users in data wrangling, management, analysis, visualization, and publication is critical for collaborative geospatial research. To this end, Spatial Data Science provides a path (i.e. workflow) for navigating the rapidly expanding field of data, methods, and multi-user tools for working with and analyzing large and complex geospatial data.

At the heart of Spatial Data Science is a common workflow (Fig. 2) that leverages cloud/web-based and open source geospatial tools to address technical impediments to collaboration such as non-standardized working environments, siloing of data, unreproducible analyses, and static map visuals. While there are many possible routes available when navigating a collaborative Spatial Data Science workflow, these routes generally consist of four key primary tasks through which collaboration can not only be fostered, but are actually fundamental to geospatial problem-solving in the 21st century: (1) setting up the working environment; (2) data wrangling (i.e. harnessing, cleaning, transforming); (3) data analysis; (4) data visualization and publication. Both data management and visualization are deeply embedded within all tasks, particularly data wrangling and analysis. Data visualization is highlighted specifically with publication (Fig. 2) to emphasize its important role in facilitating the dissemination of results and knowledge gained. Facilitating this workflow are (5) the integration and support of FOSS4G and (6) user involvement by the public, scientists, technologists, practitioners, and governments. Given the iterative nature of collaboration, this Spatial Data Science workflow is adaptive; the tools chosen for each task can be modified or replaced as the needs of the projects are further refined or new tools become available. By addressing technical challenges at each step, this collaborative Spatial Data Science workflow allows researchers and stakeholders to more easily share research ideas, analyses, code, results, and conclusions to work toward the integration and synthesis of knowledge.

3. Typologies of geospatial tools and collaboration

Existing typologies (or classifications based on general type) of geospatial tools have been created through qualitative categorization and comparison of tool capabilities. To date, reviews of tool applications have been conducted to identify technical approaches to building tools, domain-specific evolutions in tools, and capabilities of tools for specific applications (Table 1). Although these qualitative typologies provide fundamental understanding of the evolution and landscape of geospatial tools, none provide a framework for evaluating how tools address technical barriers to collaborative tasks.

Though specifically highlighting only FOSS4G options, Steiniger and Hunter (2013) have provided the most comprehensive qualitative typology of GI Systems (GIS) software to date, expanding to nine categories of software from the seven original types identified in Steiniger and Weibel (2009): (1) desktop GIS; (2) spatial database management systems; (3) server GIS; (4) mobile GIS; (5) exploratory spatial data analysis tools; (6) remote sensing software; (7) GIS libraries (i.e. projection and geometry libraries); (8) GIS extensions, plug-ins, and APIs; and (9) Web Mapping Servers and Development Frameworks (p. 136 and 139). In

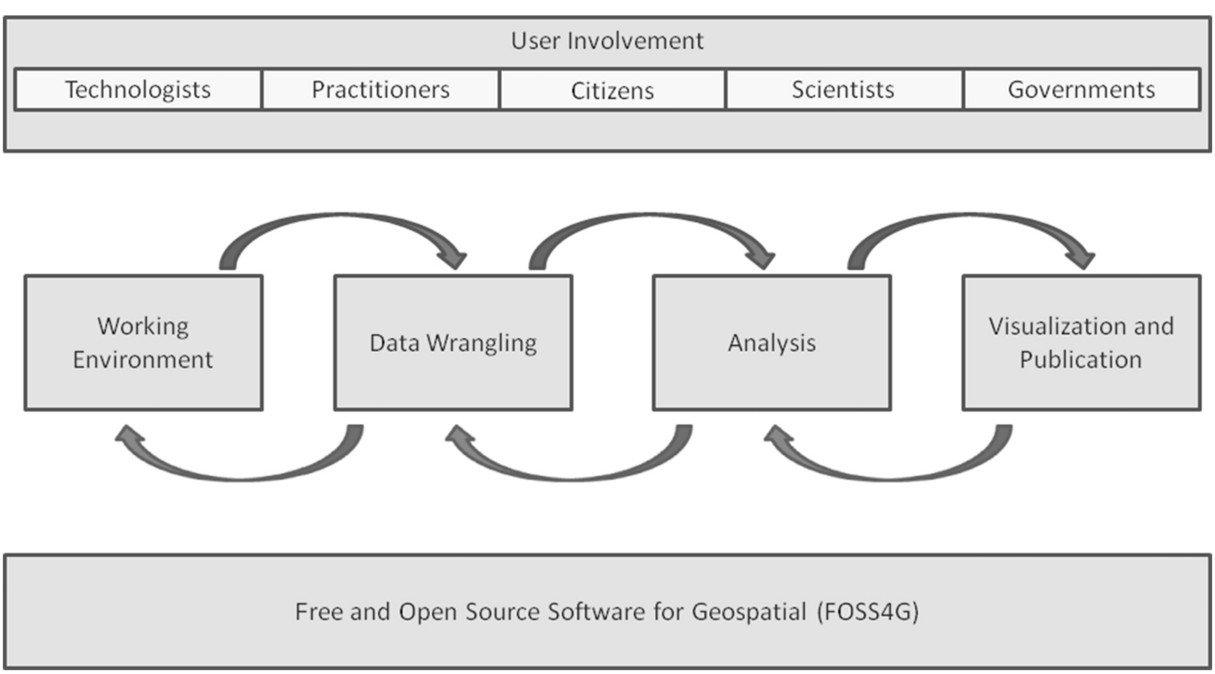

Fig. 2. Collaborative Spatial Data Science workflow.

Table 1
Existing typologies of geospatial tools.

Technologies reviewed	Publication
Web-based spatial decision support tools	Rinner (2003)
"Trends and developments in GIS-based multi-criteria decision analysis" tools	Malczewski (2006), p. 1
"Visually-enabled geocollaboration" tools	MacEachren and Brewer (2004), p. 1
GI Systems for public participation	Sieber (2006)
Historical evolution in web tools for geospatial applications	Haklay et al. (2008)
Historical evolution of participatory GIS	Jankowski (2009)
FOSS4G tools for landscape ecology	Steiniger and Hay (2009)
Marine geospatial ecology tools	Roberts, Best, Dunn, Treml, and Halpin (2010)
"Enabling technologies" for CyberGIS	Yang et al. (2010), p. 266
Evolution of software for spatial analysis	Anselin (2012)
"Domains of VGI"	Elwood et al. (2012), p. 573
"Capabilities and interfaces of existing tools" for GIS and spatial analysis tools integrated within CyberGIS	Wang et al. (2013), p. 2026
FOSS4G landscape in 2012	Steiniger and Hunter (2013)
"Major classes of technology tools and needs they might meet", among them being GI Systems, decision support tools, visualization tools, and "distance collaboration" tools	Cravens (2014), p. 23
"Map-based web tools supporting climate change adaptation"	Neset, Opach, Lion, Lilja, and Johansson (2016), p. 1

addition to this fundamental qualitative typology, the authors also identify benefits of FOSS4G, key factors to consider for evaluations of options, and the primary barriers to FOSS4G adoption (referencing others such as Cruz, Wieland & Ziegler, 2006 and Nagy, Yassin & Bhattacherjee, 2010). Not identified as key functionality, the potential to support collaboration is not addressed in the criteria for evaluation or adoption.

The literature focusing on categorizations of tools with an explicit focus on collaborative work have been broader in scope and not specific to geospatial options. In support of geocollaboration tools, MacEachren and Brewer (2004) provided a summary of Computer-Supported Cooperative Work tools, or "CSCW technologies, often called groupware…characterized as information technology that allows people to work together… with an emphasis on sharing tasks and decision-making" (p. 10). In the listed categories of CSCW tools, multi-criteria evaluation tools integrated with GIS are the only geospatially enabled options. Similarly, in discussing the evolution of web mapping technologies, Haklay et al. (2008) presented a general "series of 'technologies of collaboration'" from Saveri, Rheingold, and Vian (2005), including "Self-organising mesh networks… Community computing grids… Peer production networks… Social mobile computing… Group-forming networks… Social software… Social accounting tools… Knowledge collectives" (p. 2025–2026).

While some geospatial tools can be embedded within these broader typologies of collaborative technologies (i.e. OpenStreetMap and other FOSS4G tools are products of peer production networks), we expand on these works by providing a new typology of geospatial tools that is specifically centered on collaboration. We ask the following specific questions: what are the common types of collaborative geospatial tools, and what functional niches do they fill? To answer these questions, we develop a quantitative and reproducible framework to evaluate multi-user geospatial tools based on their functionality for supporting common tasks in collaborative projects (i.e. wrangling, analyzing, visualizing, and publishing geospatial data) and present a typological map of the emergent ecosystem of collaborative geospatial tools.

4. Methods

4.1. Selection of tools

We evaluated thirty-one multi-user geospatial tools based on their functionality to support collaborative tasks (Table 2). The tools represent a variety of platform types: cloud-based (i.e. hosted on the cloud by the tool provider), web-based (i.e. hosted by user on a web server), local installation (i.e. installed locally on an individual computer or cluster of computers), and mobile (i.e. application installed on mobile device). The tools also vary in their FOSS status and in the industry type of their primary creators and contributors. Specifically, the included tools express a mission of supporting collaboration and/or offer functionality for supporting collaboration (e.g. sharing of data and code, asynchronous tasks, status updates). This requirement excludes tools focused on big data processing such as distributed computation engines (e.g. Spark) or scenario modeling such as ecosystem valuation tools (e.g. Integrated Valuation of Ecosystem Services and Tradeoff, or InVEST). The included tools also provide a set of analytical and/or data collection functionality within a multi-user environment (e.g. beyond basic online data providers or interactive web maps such as Cal-Adapt, WorldMap, etc). This requirement also excludes tools focused primarily on workflows by individuals (e.g. desktop GIS tools such as GeoDa). In addition, included tools provide an out-of-the-box user interface and do not require the creation of a custom user interface or the use of a third party user interface. This definition excludes tool extensions such as widgets and plug-ins, which are not considered to be distinct from the platform onto which they are installed. This requirement also excludes geostack components whether open source or not (e.g. ArcGIS Server, OpenLayers, PostGIS, Leaflet). Tools currently in Beta mode were also excluded (e.g. GeoGig, a promising versioning tool). Finally, multi-user tools not exclusively limited to geospatial tasks were also included, if the stated criteria were met and the tool was able to integrate geospatial tasks (e.g. Jupyter Hub, RShiny). Though not an exhaustive list, the thirty-one tools evaluated in this paper are representative of the wide range of available platforms that support multi-user workflows for geospatial data.

4.2. A workflow-based evaluation of functionality

The included tools were evaluated on twenty-nine different features that support multi-user workflows for geospatial data (Appendix A). Based on a collaborative Spatial Data Science workflow (Fig. 2), these features represent functionality provided to address traditional technical impediments to collaboration and are organized into groups that represent the key components of the workflow: (1) setting up the working environment; (2) data wrangling; (3) data analysis; (4) data visualization and publication; (5) the integration and support of FOSS4G; and (6) user involvement.

A standardized scoring rubric was used to assign a value of 1–3 for each feature, with 1 indicating little to no application of that feature within the tool to 3 indicating that the feature is critical to the functionality of the tool (Appendix A). As no single tool can provide functionality for all features, we chose to treat the features as categorical variables referred to as factors, instead of as continuous variables. As such, tools with scores of 1 for particular features are not automatically clustered more closely with tools scoring 2, than with tools scoring 3; each score is simply considered to represent a different level of functionality. For example, all tools were scored on their reliance on cloud/web-based functionality. Tools that rely primarily on local installations (i.e. Desktop, Server, and Mobile) were given a score of 1, while tools that are completely cloud/web-based (i.e. no local installations of any kind) were given a score of 3. Tools that have both local and cloud/web-based components were given a score of 2. As factor variables, this scoring does not promote a score of 3 as more desirable for collaboration than a score of 1 or 2; these scores simply indicate different functionality based on the level of cloud/web-based integration. Last, the scores for all features were based on the mission statement or stated capacity provided on the tool website as well as professional experience by the authors. All of the included tools that are available for download or online access were tested by the authors; the exceptions being Seasketch, NASA NEX, and ExchangeCore, as these tools require granted permissions to

Table 2
List of multi-user geospatial tools included in analysis.

Label	Name	Platform type	Creators/contributors	FOSS status[a]
T1	CARTO	Cloud-based	CARTO (private sector)	Limited free, not open source
T2	MapGuide	Web-based	OSGeo (non-profit)	FOSS
T3	XchangeCore	Web-based	National Institute for Hometown Security (non-profit)	Free (restricted access), not open source
T4	Jupyter Hub	Web-based	NumFOCUS Foundation (non-profit)	FOSS
T5	NASA NEX	Local install	NASA (public sector)	Free (restricted access), not open source
T6	OS Geo Live	Local install	OSGeo (non-profit)	FOSS
T7	ROpenSci	Local install	Project of the NumFOCUS Foundation (non-profit)	FOSS
T8	Rshiny	Local install or cloud-based	RStudio (private sector)	Limited free, limited open source
T9	Global Forest Watch	Cloud-based	World Resources Institute (non-profit)	FOSS
T10	NextGIS	Local installation or cloud-based	NextGIS (private sector)	Limited free, limited open source
T11	QGIS Cloud	Local installation or cloud-based	Sourcepole (private sector)	Limited free, limited open source
T12	FME	Local installation or web-based	Safe Software (private sector)	Neither free nor open source
T13	Google Earth Engine	Cloud-based	Google (private sector)	Free, not open source
T14	Madrona	Local installation or web-based	Ecotrust (non-profit)	FOSS
T15	MapBox Studio	Cloud-based	MapBox (private sector)	Limited free, limited open source
T16	Field Papers	Cloud-based	Stamen Design (private sector)	FOSS
T17	iNaturalist	Mobile	California Academy of Sciences (non-profit)	FOSS
T18	OpenDataKit_GeoODK	Mobile	University of Washington, Seattle (academia)	FOSS
T19	OpenStreetMap	Cloud-based	OpenStreetMap Foundation (non-profit)	FOSS
T20	eBird	Cloud-based	Partnership between Audubon (non-profit) and Cornell University (academia)	Free, not open source
T21	GeoLocate	Local installation or cloud-based	Tulane University (academia)	Free, not open source
T22	HOLOS	Local installation or cloud-based	University of California, Berkeley (academia)	FOSS
T23	Data Basin	Cloud-based	Conservation Biology Institute (non-profit)	Free, not open source
T24	ESRI Collector for ArcGIS	Mobile	ESRI (private sector)	Limited free, not open source
T25	Geopaparazzi	Mobile	HydroloGIS (private sector)	FOSS
T26	Locus Map	Mobile	Asamm Software (private sector)	Limited free, not open source
T27	Orux Maps	Mobile	OruxMaps (private individuals)	Free, not open source
T28	ArcGIS Online	Cloud-based	ESRI (private sector)	Limited free, not open source
T29	Seasketch	Web-based	University of California, Santa Barbara (academia)	Neither free nor open source
T30	AmigoCloud	Cloud-based	AmigoCloud (private sector)	Limited free, not open source
T31	ArcGIS Open Data	Cloud-based	ESRI (private sector)	Limited free, not open source

[a] See Appendix A for more information on FOSS status.

download or access. When a feature could not easily be scored using online references or professional experience, questions regarding those features were sent to the tool provider, with a 100% response rate. The individual tool scores were also used to calculate an average score across all tools (i.e. an average tool score) for each of the twenty-nine features (Appendix B).

4.3. Cluster analysis

Next, we applied a statistical clustering method on the individual tool scores to determine the common typologies among tools and identify existing niches. Our clustering method uses a custom R package called Threshold Smoothing Ensemble Clustering (TSEC or TSEClustering, developed by Oliver C. Muellerklein), which incorporates the Weighted K-Means Clustering method from the R package *wskm* (Zhao, Salloum, Cai, & Huang, 2015). The TSEC model clusters the observations (i.e. tools) and variables (i.e. features) based on an ensemble of co-occurrences across pre-defined subsets of the features using a smoothing threshold function. Conceptually, the final cluster memberships are the result of a threshold approximated ensemble of similarities between the tools across subsets of the features (see Appendix A for list of features and subsets). This method is targeted at datasets with low sample size yet a high number of variables, building an ensemble of similarity measurements used to assess intra-cluster and inter-cluster variance for optimized information gain. Further, cluster membership weights are generated for observations at local (i.e. for subsets of features) and global (i.e. complete list of features) levels of variable importance. In sum, this unsupervised approach generates similarities among the observations based on various subsets of the features (i.e. subsets of predictor space) to assign clusters, concluding with a final ensemble of all cluster assignments.

The cluster analysis workflow is shown in Fig. 3. The workflow begins with (a) input data of n × p (i.e. number of observations x number of features, or predictors). From the input data, six subsets of predictor space (i.e. g_i, predefined subsets of the twenty-nine features) are created manually by splitting the features into groups that represent one of the six components of the collaborative Spatial Data Science workflow (i.e. from the first subgroup describing the working environment to the sixth subgroup outlining user involvement, see Appendix A). Six additional subsets of predictor space are created by iteratively grouping all subsets except the initial g_i subset (i.e. for all g_x not g_i, the leave one-out method). A final subset of predictor space is created by grouping all g_i subsets (i.e. the complete list of twenty-nine features).

Intra-observational similarity matrices (b) are generated through correlation matrices for each of the thirteen subsets of predictor space (i.e. g_i). Then, using entropy-weighted K-means clustering (c), cluster assignments are produced for each of the thirteen similarity matrices. The Elbow Method is used to obtain the optimal number of clusters by calculating the relative percentage of variance captured by the clusters versus the total number of clusters (Tibshirani, Walther, & Hastie, 2001). Next, observational co-occurrences (d) are generated for each

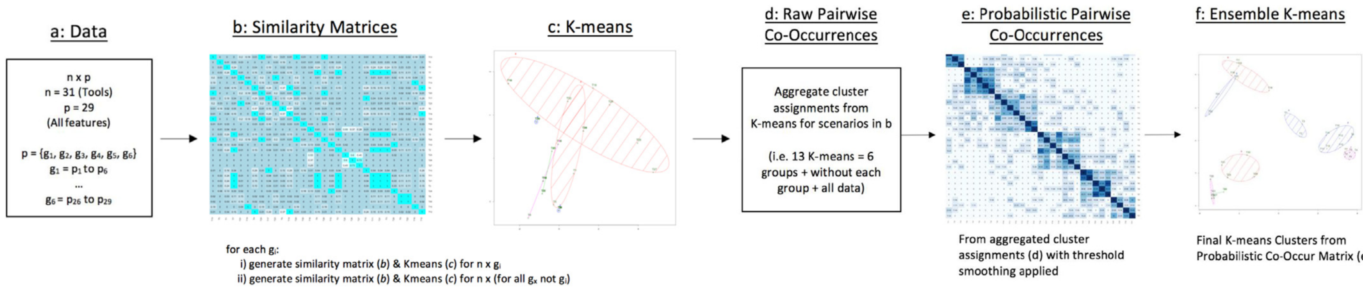

Fig. 3. Cluster analysis workflow. Steps a-c generate a similarity matrix and K-means cluster assignment for thirteen subsets of features (i.e. each of the six g_i, iterative grouping of all g_x except the initial g_i, and for all g_i as a complete dataset). The resulting thirteen similarity matrices and K-means cluster assignments are aggregated in steps d-f with a threshold smoothing function to produce a final ensemble K-means cluster assignment.

of the thirteen K-means cluster memberships to obtain an ensemble of probabilistic assignments (e) (i.e. a correlational matrix of co-assignment among observations). These co-assignments are displayed as a dendrogram, a visualization of the partitions from hierarchical clustering (Fig. 4). Threshold approximation is used to dropout low pairwise observational assignments (i.e. force to zero). Finally, a last K-means clustering (f) is run on the resulting Smooth Ensemble, an $n \times n$ correlational matrix of observational assignments, to generate the final cluster memberships. These final cluster memberships are visualized in a bivariate cluster plot, which uses Principal Components to make a two-dimensional representation of the clusters (Fig. 5). The final cluster assignment for each of the thirty-one tools is listed in Table 3.

5. Results

Three primary clusters composed of eight secondary clusters were revealed from the cluster analysis (Table 3). The divergences between the clusters are visualized in complementary ways in a dendrogram (Fig. 4) and a map of the K-means Two-Dimensional Space (Fig. 5). The first primary cluster A, composed of subclusters 1–3, contains

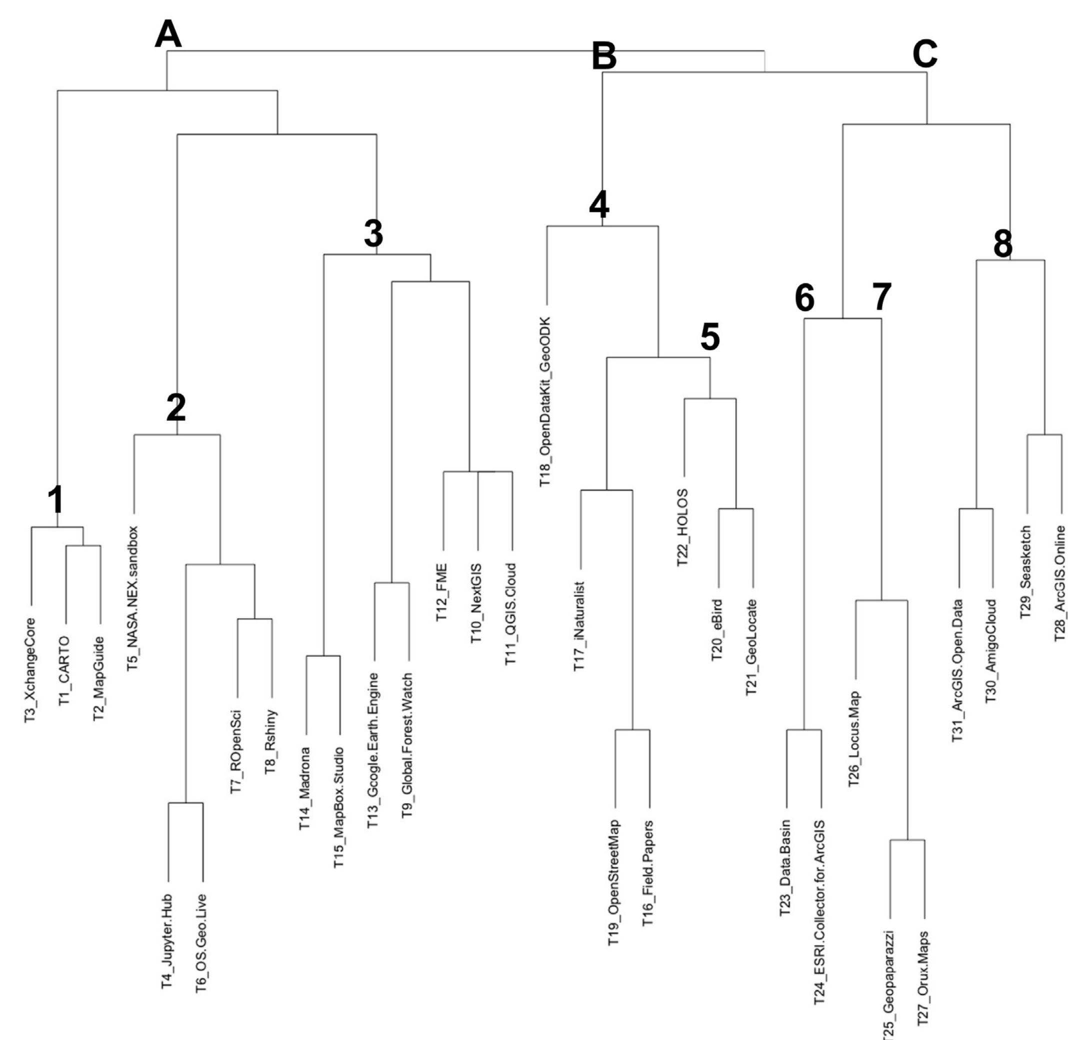

Fig. 4. Dendrogram of collaborative geospatial tools. Primary clusters are designated as A–C, with secondary clusters labeled as 1–8.

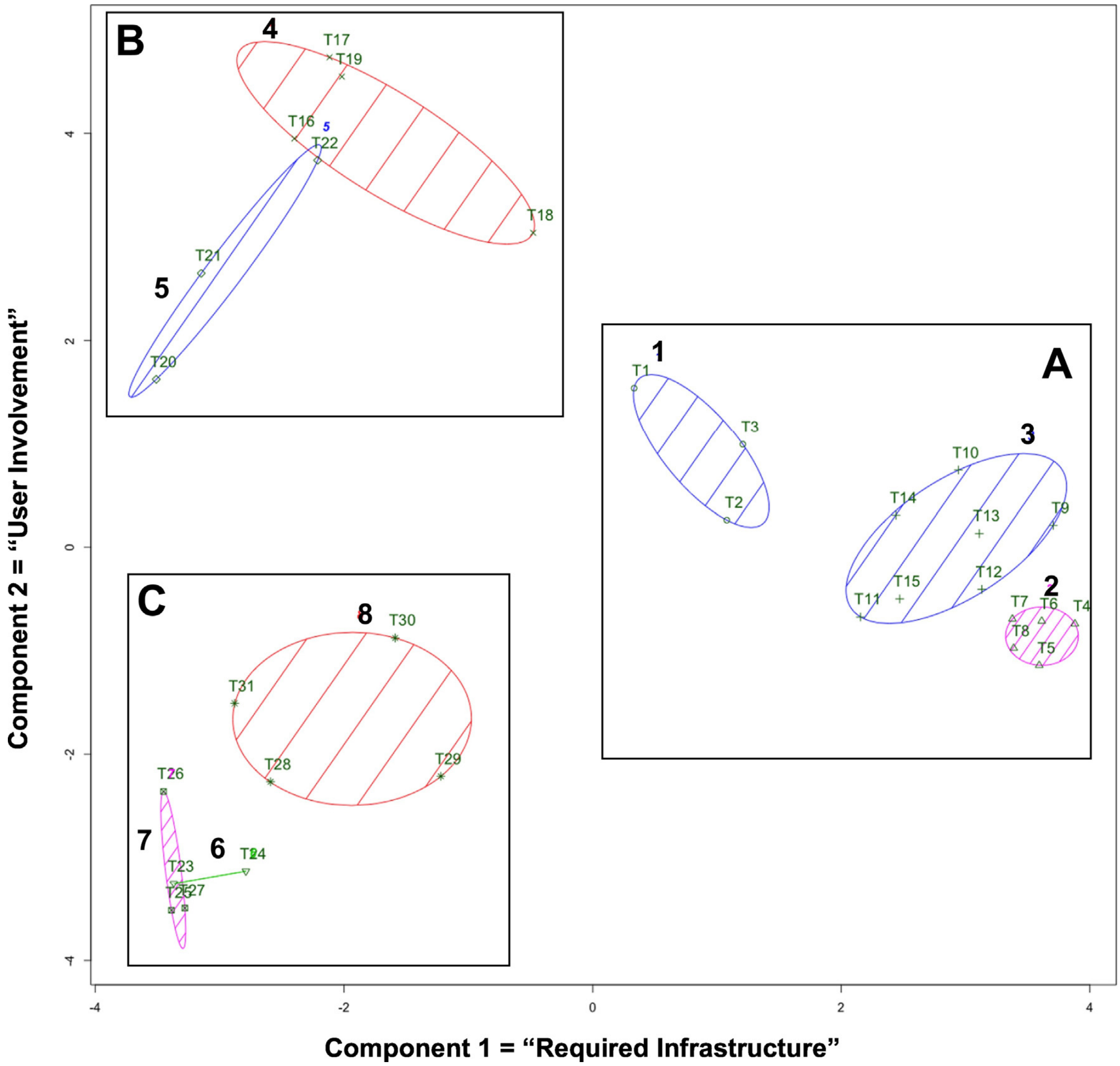

Fig. 5. Typological map of collaborative geospatial tools, based on the K-means Two-Dimensional Space. Bifurcation driven by required infrastructure is reflected along Component 1, from low (left) to high (right). Divergence driven by user involvement (i.e. optimal number of users and public accessibility) is reflected along Component 2, from project-based content managers (bottom) to participatory data aggregators (top). The identified clusters account for 76.3% of the total variance.

tools that are highly scalable and customizable, allowing for the highest integration of advanced spatial analysis and visualization techniques as well as interoperability with other tools. The second primary cluster B, composed of subclusters 4–5, demarcates participatory data aggregators that have inherently large project scopes (i.e. functionality is optimally leveraged with high numbers of public users). The third primary cluster C, composed of subclusters 6–8, identifies content managers, or project-based tools focused on managing access to data and tasks for a predefined set of users.

5.1. Primary drivers of divergence

The first bifurcation between the clusters emerges from the required level of infrastructure needed to best leverage the tool (i.e. required user setup, use of high performance computing and cloud services, support for multi-tier users, user knowledge needed to extend functionality). This bifurcation is reflected along Component 1 of the K-means Two-Dimensional Space (Fig. 5), wherein tools with the heaviest infrastructure needs are clustered on the right-hand side (primary cluster A of highly scalable and customizable tools), while tools with lighter infrastructure requirements are clustered on the left-hand side (primary cluster B of the participatory data aggregators and primary cluster C of the content managers).

A second key divergence between the clusters is driven by user involvement, a key determinant of project scope (i.e. optimal number of users and public accessibility). The K-means Two Dimensional Space reflects this divergence along Component 2 (Fig. 5). Tools with inherently larger scopes are clustered toward the top (i.e. primary cluster B of the participatory data aggregators). The functionality of these tools is best leveraged with high number of public users engaging in data collection. Tools with smaller scopes due to a focus on managing user access to data and tasks are clustered toward the bottom (i.e. primary cluster C of the content managers). The functionality of these tools does not vary with a change in the number of users, and access to these data is controlled by a project manager. Along the center of Component 2 are the highly scalable and customizable tools of primary cluster A. For these tools, an increase in the number of users leads to a leveraging of expandable functionality that is not necessary for small user groups, such as differential access to datasets and workflows facilitated by custom web visualizations and APIs (i.e. multi-tier versions of the tools with differing functionality and access based on the user type).

5.2. Highly scalable and customizable tools

Primary cluster A (subclusters 1–3) is characterized by tools that are the most extendable for integrating advanced analysis and data

Table 3
Summary of cluster results.

Label	Name	Primary cluster	Secondary cluster
T1	CARTO	A	1
T2	MapGuide	A	1
T3	XchangeCore	A	1
T4	Jupyter Hub	A	2
T5	NASA NEX	A	2
T6	OS Geo Live	A	2
T7	ROpenSci	A	2
T8	Rshiny	A	2
T9	Global Forest Watch	A	3
T10	NextGIS	A	3
T11	QGIS Cloud	A	3
T12	FME	A	3
T13	Google Earth Engine	A	3
T14	Madrona	A	3
T15	MapBox Studio	A	3
T16	Field Papers	B	4
T17	iNaturalist	B	4
T18	OpenDataKit_GeoODK	B	4
T19	OpenStreetMap	B	4
T20	eBird	B	5
T21	GeoLocate	B	5
T22	HOLOS	B	5
T23	Data Basin	C	6
T24	ESRI Collector for ArcGIS	C	6
T25	Geopaparazzi	C	7
T26	Locus Map	C	7
T27	Orux Maps	C	7
T28	ArcGIS Online	C	8
T29	Seasketch	C	8
T30	AmigoCloud	C	8
T31	ArcGIS Open Data	C	8

visualization and for supporting reproducible code and interoperability through APIs and tool integration. Due to this flexibility, these highly scalable and customizable tools also require the most infrastructure to best leverage the full range of functionality offered. The primary products of these tools are datasets and workflows resulting from advanced analytical, data visualization, or querying methods. As the user base increases, the functionality provided (particularly by APIs and tool integration) can be leveraged to create multiple versions of the tools based on user needs (i.e. from low to high interactivity and privileges). The divergences between subclusters 1–3 are driven by differences in out-of-the-box functionality. Subcluster 1 highlights cloud/web-based tools (with APIs available) that have more built-in functionality for data exploration, visualization, and publication than for spatial analysis and querying (CARTO, MapGuide, XchangeCore). Subcluster 2 identifies tools with a stronger focus on facilitating reproducible workflows or standardized working environments (Jupyter Hub, NASA NEX, OS Geo Live, ROpenSci, RShiny). These tools require either server or desktop installed components and are best leveraged by integration with additional tools and packages for spatial data analysis and visualization. Subcluster 3 differentiates tools with the highest built-in functionality for geoprocessing or spatial analysis (Global Forest Watch, NextGIS, QGIS Cloud, FME, Google Earth Engine, Madrona, MapBox Studio). These tools provide APIs or expose open source capabilities to users, and thus, easily integrate custom scripting or can be expanded to build new tools in a multi-tier user environment.

5.3. Participatory data aggregators

Primary cluster B (subclusters 4 and 5) represents crowdsourcing tools that have inherently large and public scopes. The functionality of these tools is optimally leveraged with a high number of users. Often with a specifically defined focus (i.e. crowdsourcing data for a particular ecological phenomenon), the primary products of these tools are aggregated datasets compiled from many public contributors. The divergence between subclusters 4 and 5 is primarily driven by the differing roles of citizen scientists. Subcluster 4 delineates FOSS4G tools focused on crowdsourced data collection by the public either in real-time or asynchronously from the field or based on lived experience (Field Papers, iNaturalist, OpenDataKit/GeoODK, OpenStreetMap). Subcluster 5 represents research-driven participatory tools that are more focused on expert data curation by scientists (i.e. no mobile applications or syncing of field data) and provide APIs to engage and exchange data with the public (eBird, GeoLocate, HOLOS).

5.4. Content managers

Primary cluster C (subclusters 6–8) delineates project-based tools that focus on the management of users and their access to data and tasks. The primary products of these tools are content management systems (some with supporting APIs) controlled by a project manager. The divergences between subclusters 6–8 are driven by differences in data management functionality and infrastructure. Subcluster 6 characterizes tools with functionality for managing projects and tasks (i.e. organization of workspaces, group communication tools, assignment of tasks), in addition to managing user access to data (Data Basin, ESRI Collector for ArcGIS). Subcluster 7 designates Android-based mobile data collectors that provide functionality for navigation and surveying (Geopaparazzi, Locus Map, Orux Maps). These tools allow a predefined set of users to collect geospatial data asynchronously and are not restricted to cloud-based databases. Finally, subcluster 8 identifies tools with light spatial analysis or querying capabilities that rely on "live" databases (i.e. cloud/web-based database services) for managing the exchange of data (ArcGIS Online, Seasketch, ArcGIS OpenData, AmigoCloud).

6. Discussion

Based on previously cited calls for more collaboration in geospatial research and technologies, it is clear that evaluation of geospatial tools must also include how they facilitate collaboration in the wrangling, analysis, visualization, and publication of geospatial data. Previous typologies have qualitatively categorized and compared geospatial tools without explicit consideration of the functionality provided to support collaborative tasks (see Table 1). By providing an essential assessment of geospatial tools specifically centered on functionality for collaboration, this typology can help geospatial researchers and stakeholders of collaborative geospatial projects evaluate and choose the best tools for their needs. By following the Spatial Data Science tenet of standardized and reproducible workflows, this typological map can evolve and expand over time, as more collaborative geospatial tools continue to be developed and adopted. In this paper, we use this typology to highlight the strengths of existing collaborative tools, identify key areas of future technical development, and elucidate ongoing challenges for collaborative geospatial tool development.

6.1. Strengths of collaborative geospatial tools

Even as the ecosystem of collaborative geospatial tools continues to expand, research focused on global environmental change are already benefiting from the existing technical strengths of these tools. Key benefits stem from increased integration of open source technologies as well as from an increased focus on interoperability through APIs and integration across tools. Users benefit not only from the cost-effectiveness of additional functionality from open source integration and interoperability with other tools, but also from being able modify and expand on these built-in open source capabilities and APIs.

Overall, the scored tools range from a moderate to high level of open source integration on the backend, whether a mix of open and closed source technologies to completely built on open source technologies (average tool score = 2.61, along a gradient in which 1.0 indicates no integration of open source technology and 3.0 indicates completely open source). Similarly, the scored tools support a range of moderate to high level of interoperability, from being able to push or pull data from other tools to providing a fully open API (average tool score = 2.68).

These high scores for open source integration and interoperability (as compared to all other scored features) provide a clear understanding of the role of cost and accessibility, as the resulting clusters do not represent groupings based on the price of the tools. Each cluster is a mix of free and/or completely open access tools to "freemium" (i.e. cost applied to access higher levels of functionality) and/or restricted access tools (i.e. domain specific applications that are free once access is granted, such as XChangeCore and NASA NEX from primary cluster A). Focusing solely on the cost of the tools as the key barrier to collaborative projects would result in a very different typology of collaborative geospatial tools, one that would not fully account for the niches of technical functionality that tools provide for collaborative tasks.

In our analysis, the individual clusters reflect differences in the level of open source integration, the level of modification allowed by users, and overall interoperability (see Appendix B for individual tool scores). Primary cluster A of the highly scalable and customizable tools demarcate tools that are primarily built on open source technologies, provide strong support for users to modify built-in open source capabilities, and generally, are highly interoperable either through the ability to integrate with other tools or through APIs. On the other hand, primary cluster C composed of content managers represent tools that are primarily a mix of open and closed source technologies, provide less support for users to modify the built-in open source capabilities, and as such, are generally less interoperable (with a few exceptions of tools that provide API access such as AmigoCloud, ArcOpenData and Locus Map). This is perhaps unsurprising as the strength of content managers are built-in functionalities that do not require much modification or technical knowledge by users (i.e. support for asynchronous tasks - such as offline capture of data, task assignment, and status updates - and user/content management - such as access/workspace control and group definitions). Primary cluster B of the participatory data aggregators is more evenly split; about half of the tools are mixed open and closed source that provide less support for modification by users (similar to primary cluster C), while the other half are primarily built on open source technologies and provide mechanisms for users to modify open source functionality (similar to primary cluster A). As compared to primary clusters A and C, the distinguishing characteristic of primary cluster B is that all of its tools scored the highest value (3.0) for interoperability (i.e. primary clusters A and B had wider ranges of scores). In fact, all participatory data aggregators included in this analysis provide access to APIs and/or Software Development Kits (SDKs).

For research centered on global environmental change, increased open source integration and support for interoperability in collaborative geospatial tools are allowing for unprecedented cross-disciplinary integration of data and methods, beyond simply powerful data processing or spatial analysis capabilities. Citizens with varying levels of technological skillsets (from non-scientists to practitioners) are commonly leveraging these strengths through access to source code on Github and public API access to data and analytical methods. On the data side, citizen science projects that leverage participatory data collection tools are becoming more cost effective due to the availability of low-to-no-cost, easy-to-launch tools that require little infrastructure investment or technical knowledge by users. Through these citizen science efforts, researchers are granted a mechanism for "dovetailing research with conservation and management" (Dickinson et al., 2012, p. 294). On the methods side, repeatability of complex workflows is facilitated by increased availability of APIs that allow for seamless exchange of geospatial data and by the ability to integrate functionality from other specialized tools.

A key example of open source integration benefitting citizen science efforts is the application of the mobile geospatial data collector iNaturalist from primary cluster B of the participatory data aggregators. This lightweight mobile application is increasingly being used in BioBlitzes, which are short-duration field collection efforts to inventory biodiversity or to monitor a particular species within a specified area, typically parks and protected areas (Dickinson et al., 2012; Francis, Easterday, Scheckel, & Beissinger, 2017). Citizen scientists simply download the free mobile application and capture photos and notes that automatically sync to the iNaturalist database. All data collected with iNaturalist are available for public exploration and use through their web mapping application and API and are also shared with free and open access scientific databases such as the Global Biodiversity Information Facility. In a unique global collaboration, National Geographic, the iNaturalist team, and citizens in 100 countries participated in The Great Nature Project between 2013 and 2015 to collect "over half a million images of over 20,000 different species of plants, animals, and fungi" (Francis et al., 2017; National Geographic, 2016).

Key examples of the benefits from increased interoperability are Global Forest Watch (of primary cluster A of the highly scalable and customizable tools) and Seasketch (of primary cluster C of the content managers). Global Forest Watch leverages the Google Earth Engine API and the CARTO platform (both tools also in primary cluster A of the highly scalable and customizable tools) to create interactive web maps that can analyze forest change on-the-fly for an area of interest. Building off of the Google Earth Engine API, Global Forest Watch freely provides its own customized APIs as well as templates for ArcGIS Online (in primary cluster C of the content managers) to facilitate additional tool building and data sharing by others. Leveraging the benefits of tool integration, Seasketch is a key example of a collaborative environmental planning tool that has benefited from integration with widely used spatial decision support tools such as Marxan and InVEST, as well as from integration with ArcGIS Online for content management. Focused on marine area protection, Seasketch is currently being used "around the globe by 4441 users in 229 active projects" to provide stakeholders with the capability to explore scenarios and propose their own plans for new marine protected areas (Seasketch, 2016). For example, through collaboration between Parque Nacional Galapagos, Conservation International, and World Wildlife fund, user-sketched plans from Seasketch are being integrated with the InVEST toolkit to allow public stakeholders to evaluate habitat risk and explore outcomes of proposed zoning scenarios for marine protection around the Galapagos Islands (Seasketch, 2016).

6.2. Key areas of future technical development

In addition to highlighting the strengths of collaborative geospatial tools, our typology can help identify key areas of future technical development. One such area of needed development is the continued integration of cloud and high performance computing (average tool score = 1.7, along a gradient in which 1.0 indicates no integration of cloud and high performance computing and 3.0 indicates full integration). For collaborations centered around global environmental change, the leveraging of cloud and high performance computing can shift the cost-benefit structure, such that research questions that previously would have been very difficult or even possible to answer (due to computing time and resources) can now be addressed. Ongoing support of CyberGIS research as well as collaborations between scientists and technologists are key for continued integration of cloud and high performance computing into geospatial tools.

Our analysis indicates that primary cluster A of the highly scalable and customizable tools has the highest overall application or potential for cloud and high performance computing (i.e. all tools scored at least 2.0, with the majority scoring 3.0). An exemplar of this cluster is Google Earth Engine, which was successfully leveraged to create the Hansen Global Forest Change dataset by a team consisting of fifteen collaborators, including technologists from Google, scientists from the USGS and Woods Hole Research Center, and researchers from the University of Maryland-College Park, SUNY-Syracuse, and South Dakota State University. Hansen et al. (2013) applied the distributed computing power of Google Earth Engine to map global forest loss and gain for 2000–2014 at the finest combined spatial and temporal resolution of any global product to date (yearly data at a 30 m pixel resolution). A BBC News article quoted lead author Matt Hansen: "This is the first map of forest change that is globally consistent and locally relevant. What would have taken a single computer 15 years to perform was completed in a matter of days using Google Earth Engine computing" (BBC, November 14, 2013).

Another key area of future development is increased support for non-traditional raster and vector data formats (i.e. open data options, cloud-based tile services). Although these non-traditional data formats are becoming critical for environmental collaborations investigating questions of larger extents and finer resolutions, existing functionality to support these formats varies greatly depending on the data type and the task. For example, across all features scored, the highest average tool score is for data download of non-traditional vector formats (average tool score = 2.81), while the lowest average tool score is for data editing of non-traditional raster formats (average tool score = 1.39). Overall, average tool scores for non-traditional vector formats are higher than for non-traditional raster formats across all data tasks (i.e. creation, editing, upload, download), and for both non-traditional vector and raster formats, average tool scores for data uploads and downloads are higher than for data creation and editing.

Regarding discrepancies between non-traditional vector and raster formats, there are two primary contributing factors. First, increased integration of open source technologies and development of APIs have both lead to and been reinforced by stronger support and wider use of non-traditional vector formats such as GeoJSON, Vector Tiles, and MBTiles (average tool score = 2.55 for data uploads to 2.81 for data downloads). This dual reinforcement is not as strongly reflected within collaborative geospatial tools for non-traditional raster formats such as HDF5 and Tile Mapping Services or for older raster formats that are seeing a resurgence such as NetCDF (average tool score = 2.32 for data uploads and 1.68 for data downloads). This could be driven by the fact that non-traditional raster formats are increasingly being used to cover larger extents and/or finer resolutions, resulting in larger datasets and storage needs, which are ongoing challenges for geospatial tools in general.

Second, the typical process for creating and editing raster data often differs greatly from that of vector data. Most raster data are still expert-curated in single user environments, and unlike editing of individual features in vector data, editing of raster data typically involves global re-calculations of pixels for which GUI-based editing tools are not as useful. These differences in the curation of vector and raster data are reflected in the average tool scores. For both editing and creation, the average tool scores for non-traditional vector data are higher than for non-traditional raster (for editing, average tool score = 2.13 for vector compared to 1.39 for raster; for creation, average tool score = 2.42 for vector compared to 1.52 for raster).

Regarding the higher average tool scores for data uploads and downloads of both non-traditional vector and raster formats (as compared to editing and creation), these scores are reflective of the unique strengths of each primary cluster which focus on a different aspect of data management. For example, the highly scalable and customizable tools of primary cluster A provide flexibility and expandability for data integration, while the participatory data aggregators of primary cluster B provide infrastructure for data aggregation. Similarly, the content managers of primary cluster C provide strong built-in functionality for managing access by users to data, projects, and workflows.

Other key areas of needed tool development include the wider adoption of functionality to support reproducibility of workflows (i.e. sharing of code or steps of workflow, average tool score = 1.9), custom scripting for analysis (average tool score = 2.03) and for data visualizations (average tool score = 1.9), and the integration of time (average tool score = 2.16). For tools that best support these options at present (i.e. primary cluster A of the highly scalable and customizable tools), users are able to modify open source capabilities or harness APIs to create tailored analyses and applications for a second tier of users. However, an intermediate to high level skillset in programming is often needed for leveraging these functionalities. In addition, stronger support for integrating time into analyses is an outstanding need in Spatial Data Science beyond that of collaborative geospatial tools, particularly for visualization and analysis across continuous timelines (i.e. dynamic modeling approaches). While of all geospatial technologies, remote sensing analytical tools have most successfully addressed time, these same tools are not structured to provide multi-user support (with few exceptions such as Google Earth Engine), and typically function within discrete timelines. Similarly, support for reproducibility of workflows and results is a key component not only for collaborative geospatial workflows, but for Spatial Data Science as a field of study focused on repeatability and transparency of workflows.

A final key area of needed functionality is user controlled versioning of data and workflows; this feature was not scored in our evaluation, as so few of the representative tools offer this functionality. While many collaborative geospatial tools provide some light versioning capabilities (i.e. revision history of code in Google Earth Engine, the ability to create and compare different runs of a model in Seasketch, user contribution history for participatory data aggregators), what is not yet available is true distributed versioning of data, workflows and code that allows users to track changes at the object level (i.e. a data attribute or function), to reconcile conflicts that arise in competing edits (i.e. multi-user versioning), and to roll back changes as needed (i.e. adaptive management of data and workflows). In a fully versioned environment, all of these tasks are documented and available for review. For environmental planning and management projects, these kinds of native versioning capabilities would provide stakeholders with a structured and transparent mechanism for examining the trajectory of data and models and for actively contributing to their construction. Stakeholders could move from being primarily users of scenario exploration tools to active developers of them, as constructors of alternative stories and models beyond just offering their version of a controlled output map.

Of the key areas of future tool development presented in this paper, support for distributed versioning in collaborative geospatial tools is clearly in the earliest phase of its evolution. In general, collaborative geospatial tools have focused on other asynchronous tasks (i.e. off-line capture of data, task assignment, status updates; average tool score = 2.26) and user and content management (i.e. access and workspace control, group definitions; average tool score = 2.32). As the integration of "live" databases through web and cloud-based data services continues to become more standard in collaborative geospatial tools, particularly for multi-user collecting of data (average tool score = 2.39), versioning capabilities can also continue to be expanded. Mechanisms for supporting further integration of versioning can be adopted from existing spatial database engines (i.e. ESRI ArcSDE, PostGIS) which offer versioning of geospatial data or allow it to be programmed, and from existing version control frameworks such as Git/Github, which has played a key role in the FOSS4G movement, allowing any user to contribute to and modify the source code to fix bugs and extend functionality. Tools such as GeoGig, a Git-like versioning tool for geospatial data (currently in Beta testing), can serve as a preview of collaborative geospatial tool functionality that will likely become standard in the near future.

6.3. Challenges and future directions

The technical challenges for continued development of collaborative geospatial tools parallel existing research areas within Spatial Data Science centered around issues inherent to large and complex geospatial datasets. While data mining techniques integrated from Data Science have provided ways to turn massive data into usable information, analysis and visualization of large geospatial datasets remain difficult, as not all approaches scale appropriately (Anselin, 2012; Li et al., 2016). Visual analytics for spatial-temporal data is one area of research that aims to provide scalable methods for analyzing datasets that are too large to be contained within working memory (or random access memory, RAM). For example, Andrienko, Andrienko, Bak, Keim, and Wrobel (2013) outline a methodology for clustering of large movement datasets that begins with sub-setting the data and creating an iterative identification list of each event's neighbors that are "stored in the database, to be later retrieved on demand" (pg. 214). Similarly, Stange et al. (2011) employ various spatial and temporal filters and aggregations to prepare large movement datasets for clustering of trajectories and flows using data mining methods such as self-organized maps (SOM) and algorithms specific to mobile data. Additional solutions to the challenges posed by large datasets are being explored through the use of high performance computing environments for data wrangling and analysis (Leonard & Duffy, 2014; Li et al., 2016) and through geovisualization techniques integrated from visual analytics, or "geovisual analytics" (Anselin, 2012). These techniques include the use of multiple-linked views that allow users to work with multiple visualizations at once and human "vision-inspired" techniques such as foveation that aim to reduce information overload by varying detail depending on area of focus (Li et al., 2016, pg. 124). However, though advances in rendering have been made with emerging data formats (i.e. Vector Tiles, MBTiles, Tile Mapping Services), issues of optimizing geospatial data storage and querying remain. Tiles still require producer-side storage of raw data, and in general, spatial indexing techniques need to evolve for larger geospatial datasets, particularly in real-time applications (Li et al., 2016).

Even as computational techniques to extract and render information from data are improving, collaborative geospatial tools are limited by ongoing conceptual challenges to synthesizing information derived from large amounts of geospatial data. In particular, Miller and Goodchild (2015) point out that key issues resulting from the progression from a "data scarce to data-rich environment" are also longstanding challenges in geographic research: accuracy; uncertainty; representations of data and features; "populations (not samples), messy (not clean) data, and correlations (not causality)" (p. 450). While it is clear that these issues will continue to be ongoing challenges for both theory and technology, collaborative geospatial tools can serve as exploratory testing grounds of proposed solutions. For example, participatory data aggregators have already begun to integrate approaches to addressing issues of quality in VGI data such as biases in geographic coverage, user motivations, and knowledge levels (Quinn, 2015) through crowdsourced-based approaches (i.e. validation, repetition), social-based approaches (i.e. trusted users), and geographic knowledge-based approaches (i.e. spatial dependence and topological rules) (Goodchild & Li, 2012).

Developers of collaborative geospatial tools should also note ongoing concerns regarding the centralized production of technology and knowledge. It is clear that while collaborative geospatial tools are indeed becoming more interoperable and sophisticated, the development of these tools require knowledge that is not equally shared, which serves as a barrier to including stakeholders in the tool development process. For example, Wright et al. (2009) explore how geospatial tools used in collaborative natural resource management projects can either reinforce the technical knowledge divide between scientists and the public or provide alternative ways for the citizens to engage in the storytelling process. In addition, through presenting a "hierarchy of hacking", Haklay (2013) identifies a key barrier to democratization within neogeography as the technical knowledge and skillsets needed for citizens to be empowered to create their own tools, in light of "the current corporatisation of the web" (p. 63). The author concludes that new geospatial tools have increased the access and use of geographic information only at the lower hacking levels; "the higher levels, where deep democratisation of technology is possible… require skills and aptitude that are in short supply and are usually beyond the reach of marginalised and excluded groups in society" (p. 67). Similarly concerned about corporate and top-down control of geospatial tool development, Miller and Goodchild (2015) argue: "We must be cognizant about where this research is occurring— in the open light of scholarly research where peer review and reproducibility is possible, or behind the closed doors of private-sector companies and government agencies, as proprietary products without peer review and without full reproducibility" (p. 460). Consistent with the concerns expressed in the literature, our analysis also indicates an overall high level of user knowledge needed to fully leverage the functionality offered by collaborative geospatial tools (average tool score = 2.55, along a gradient in which 1.0 indicates none needed and 3.0 indicates a high level needed). This score reflects the fact that many tools in primary cluster A of highly scalable and customizable tools and primary cluster B of participatory data aggregators provide both basic functionalities as well as capabilities for expansion of the tools by advanced users.

Looking into the future, continued development of collaborative geospatial tools requires a sustained focus on the eight dimensions of Open GIS proposed by Sui (2014): "Open Data, Open Software, Open Hardware, Open Standards, Open Research, Open Publication, Open Funding, and Open Education" (p. 4). In particular, Open Software and Open Standards have been critical for the previously highlighted strengths of collaborative geospatial tools: integration of open source technology and support of interoperability through tool integration and APIs. These aims are supported by ongoing evolution of Open Geospatial Consortium standards and other open data standards, combined with a renewed focus on standardized and queryable metadata (Sui, 2014). Similarly, Elwood et al. (2012) highlight that the required integration of data across differing formats and media can be a major challenge to data synthesis, which often "can only be achieved if systems are to a large degree interoperable" (p. 582). Steiniger and Hunter (2013) further argue for more open source APIs, as many popular "web-mapping tools work as black boxes and do not give users the freedom to study and modify them" (p. 145).

Finally, tools are but one component in the collaborative process, an iterative exercise in communication between people to "generate (ideas and options), negotiate, choose, and execute" solutions to community and global challenges (MacEachren & Brewer, 2004, p. 7). As such, the process of stakeholders evaluating, implementing, and troubleshooting tools as a group may be more fundamental to the success of collaborative efforts than the functionality provided by the tools themselves. One likely reason is that while many environmental management and planning projects aspire to incorporate collaborative tasks (Cravens, 2016; Wright et al., 2009), tools are often chosen *before* project needs are understood, or are not evaluated until *after* projects are completed (Cravens, 2014). In addition, group discussion regarding the applicability and functionality tools can also serve a strong mechanism of stakeholder engagement, as the negotiation process can allow individuals to feel acknowledged and heard. While we have argued that tool functionality can be leveraged to provide technical support for collaborative tasks, future research can expand on this collaborative geospatial typology to focus on identifying which technical improvements are most critical for strengthening public engagement of non-scientists, particularly in the contexts of citizen science and collaborative environmental planning. It remains "a challenge for future research… how to combine computer technology with facilitation without stifling the creativity of participants" (Jankowski, 2009, p. 1971). In addition to focusing on expanding functionality, research can continue to explore additional ways that tools can empower stakeholders (i.e. further incorporation of theories of communication and decision-making, tool design, and user-computer interactions). As stakeholders become more involved in the applied process of technology design and creation, they can also highlight previously unrecognized barriers and impediments to collaboration (both social and technical) as well as help to redefine both conceptual frameworks and best practices for collaboration.

7. Conclusions

Spatial Data Science, which combines aspects of GI Science, Data Science, and CyberGIS, has emerged as an interdisciplinary field that supports collaborative geospatial research through an emphasis on leveraging cloud/web-based and open source geospatial tools that foster reproducible workflows and address long-standing technical barriers to collaboration. Here, we used a quantitative and repeatable approach to create an adaptable typology of collaborative geospatial tools based on their functionality for collaborative tasks. The resulting typological map reveals three key clusters composed of eight subclusters, across which divergence is driven by required infrastructure and user involvement. These clusters represent three primary types of collaborative geospatial tools: (1) highly scalable and customizable tools with heavier infrastructure needs, (2) participatory data aggregators and (3) content managers, the latter two with lighter infrastructure needs. As the process of collaboration is complex, one way (i.e. one cluster) is not better than another; these clusters represent discrete types of functionality that support communication and collaborative tasks for different needs and purposes. Overall, the development of a typology of collaborative geospatial tools can suggest key areas of future tool development and Spatial Data Science research, as well as help stakeholders evaluate tools by providing an understanding of the strengths of existing tools and highlighting areas of needed development. Thus, our example exploration of the emergent ecosystem of collaborative geospatial tools is not only about tools per se; this work highlights the ongoing need to facilitate communication between scientists and stakeholders in order to support fruitful collaborations that address community and global challenges.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.compenvurbsys.2017.05.003.

References

Akamani, K., Holzmueller, E. J., & Groninger, J. W. (2016). Managing wicked environmental problems as complex social-ecological systems: The promise of adaptive governance. *Landscape dynamics, soils and hydrological processes in varied climates* (pp. 741–762). Springer International Publishing.

Allen, G. M., & Gould, E. M., Jr. (1986). *ComRlexi.* (Wickedness).

Andrienko, G., Andrienko, N., Bak, P., Keim, D., & Wrobel, S. (2013). Visual analytics focusing on spatial events. *Visual analytics of movement* (pp. 209–251). Berlin Heidelberg: Springer.

Andrienko, N., & Andrienko, G. (2013). Visual analytics of movement: An overview of methods, tools and procedures. *Information Visualization*, *12*(1), 3–24.

Anselin, L. (2012). From SpaceStat to CyberGIS twenty years of spatial data analysis software. *International Regional Science Review*, *35*(2), 131–157.

Anselin, L. (2015). Spatial data science for an enhanced understanding of urban dynamics. http://citiespapers.ssrc.org/spatial-data-science-for-an-enhanced-understanding-of-urban-dynamics/ (Accessed on October 3, 2016).

Balint, P. J., Stewart, R. E., & Desai, A. (2011). *Wicked environmental problems: Managing uncertainty and conflict.* Island Press.

BBC News (2013). Forest change mapped by Google Earth. http://www.bbc.com/news/science-environment-24934790 (Accessed on October 26, 2016).

Beecham, R., Wood, J., & Bowerman, A. (2014). Studying commuting behaviours using collaborative visual analytics. *Computers, Environment and Urban Systems*, *47*, 5–15.

Carroll, M. S., Blatner, K. A., Cohn, P. J., & Morgan, T. (2007). Managing fire danger in the forests of the US Inland Northwest: A classic "wicked problem" in public land policy. *Journal of Forestry*, *105*(5), 239–244.

Connors, J. P., Lei, S., & Kelly, M. (2012). Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring. *Annals of the Association of American Geographers*, *102*(6), 1267–1289.

Cravens, A. E. (2014). Needs before tools: Using technology in environmental conflict resolution. *Conflict Resolution Quarterly*, *32*(1), 3–32.

Cravens, A. E. (2016). Negotiation and decision making with collaborative software: How marine map 'changed the game' in California's marine life protected act initiative. *Environmental Management*, *57*(2), 474–497.

Cruz, D., Wieland, T., & Ziegler, A. (2006). Evaluation criteria for free/open source software products based on project analysis. *Software Process: Improvement and Practice*, *11*(2), 107–122.

Daniels, S. E., & Walker, G. B. (2001). *Working through environmental conflict: The collaborative learning approach.*

Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., ... Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, *10*(6), 291–297.

Dickinson, J. L., Zuckerberg, B., & Bonter, D. N. (2010). Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, *41*, 149–172.

Elwood, S. (2006). Beyond cooptation or resistance: Urban spatial politics, community organizations, and GIS-based spatial narratives. *Annals of the Association of American Geographers*, *96*(2), 323–341.

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the Association of American Geographers*, *102*(3), 571–590.

Frame, T. M., Gunton, T., & Day, J. C. (2004). The role of collaboration in environmental management: An evaluation of land and resource planning in British Columbia. *Journal of Environmental Planning and Management*, *47*(1), 59–82.

Francis, J., Easterday, K., Scheckel, K., & Beissinger, S. R. (2017). The world is a park: Using citizen science to engage people in parks and build the next century of global stewards. *Science, conservation, and national parks* (pp. 275–293). The University of Chicago Press.

Gahegan, M., Luo, J., Weaver, S. D., Pike, W., & Banchuen, T. (2009). Connecting GEON: Making sense of the myriad resources, researchers and concepts that comprise a geoscience cyberinfrastructure. *Computers & Geosciences*, *35*(4), 836–854.

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

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, *69*(4), 211–221.

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, *1*, 110–120.

Goodman, K. J., Parker, S. M., Edmonds, J. W., & Zeglin, L. H. (2014). Expanding the scale of aquatic sciences: The role of the National Ecological Observatory Network (NEON). *Freshwater Science*, *34*(1), 377–385.

Haklay, M. (2013). Neogeography and the delusion of democratisation. *Environment & Planning A*, *45*(1), 55–69.

Haklay, M., Singleton, A., & Parker, C. (2008). Web mapping 2.0: The neogeography of the GeoWeb. *Geography Compass*, *2*(6), 2011–2039.

Hall, A. C. (2014). GI science, not GIScience. *Journal of Spatial Information Science*, *2014*(9), 129–131.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... Kommareddy, A. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*(6160), 850–853.

Head, B. W., & Alford, J. (2015). Wicked problems implications for public policy and management. *Administration and Society*, *47*(6), 711–739.

Ince, D. C., Hatton, L., & Graham-Cumming, J. (2012). The case for open computer programs. *Nature*, *482*(7386), 485–488.

Jankowski, P. (2009). Towards participatory geographic information systems for community-based environmental decision making. *Journal of Environmental Management*, *90*(6), 1966–1971.

Jiang, B., & Thill, J. C. (2015). Volunteered geographic information: Towards the establishment of a new paradigm. *Computers, Environment and Urban Systems*, *53*, 1–3.

Kandel, S., Heer, J., Plaisant, C., Kennedy, J., van Ham, F., Riche, N. H., ... Buono, P. (2011). Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, *10*(4), 271–288.

Kelly, M., Ferranto, S., Lei, S., Ueda, K. I., & Huntsinger, L. (2012). Expanding the table: The web as a tool for participatory adaptive management in California forests. *Journal of Environmental Management*, *109*, 1–11.

Leonard, L., & Duffy, C. J. (2014). Automating data-model workflows at a level 12 HUC scale: Watershed modeling in a distributed computing environment. *Environmental Modelling & Software*, *61*, 174–190.

Li, S., Dragicevic, S., Castro, F. A., Sester, M., Winter, S., Coltekin, A., ... Cheng, T. (2016). Geospatial big data handling theory and methods: A review and research challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 119–133.

MacEachren, A. M., & Brewer, I. (2004). Developing a conceptual framework for visually-enabled geocollaboration. *International Journal of Geographical Information Science*, *18*(1), 1–34.

MacEachren, A. M., Pike, W., Yu, C., Brewer, I., Gahegan, M., Weaver, S. D., & Yarnal, B. (2006). Building a geocollaboratory: Supporting Human–Environment Regional Observatory (HERO) collaborative science activities. *Computers, Environment and Urban Systems*, *30*(2), 201–225.

Malczewski, J. (2006). GIS-based multicriteria decision analysis: A survey of the literature. *International Journal of Geographical Information Science*, *20*(7), 703–726.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, *80*(4), 449–461.

Morin, A., Urban, J., Adams, P. D., Foster, I., Sali, A., Baker, D., & Sliz, P. (2012). Shining light into black boxes. *Science*, *336*(6078), 159–160.

Nagy, D., Yassin, A. M., & Bhattacherjee, A. (2010). Organizational adoption of open source software: Barriers and remedies. *Communications of the ACM*, *53*(3), 148–151.

National Geographic (2016). The great nature project. http://nationalgeographic.org/projects/great-nature-project/ (Accessed on October 26, 2016).

Neset, T. S., Opach, T., Lion, P., Lilja, A., & Johansson, J. (2016). Map-based web tools supporting climate change adaptation. *The Professional Geographer*, *68*(1), 103–114.

Pedersen, B., Kearns, F., & Kelly, M. (2007). Methods for facilitating web-based participatory research informatics. *Ecological Informatics*, *2*(1), 33–42.

Quinn, S. (2015). Using small cities to understand the crowd behind OpenStreetMap. *GeoJournal*, 1–19.

Rinner, C. (2003). Web-based spatial decision support: Status and research directions. *Journal of Geographic Information and Decision Analysis*, *7*(1), 14–31.

Roberts, J. J., Best, B. D., Dunn, D. C., Treml, E. A., & Halpin, P. N. (2010). Marine geospatial ecology tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modelling & Software*, *25*(10), 1197–1207.

Rocchini, D., & Neteler, M. (2012). Let the four freedoms paradigm apply to ecology. *Trends in Ecology & Evolution*, *27*(6), 310–311.
Saveri, A., Rheingold, H., & Vian, K. (2005). *Technologies of cooperation*. Institute for the Future (Retrieved on July 9, 2005).
Seasketch (2016). http://www.seasketch.org (Accessed on October 26, 2016).
Selin, S., & Chevez, D. (1995). Developing a collaborative model for environmental planning and management. *Environmental Management*, *19*(2), 189–195.
Sieber, R. (2006). Public participation geographic information systems: A literature review and framework. *Annals of the Association of American Geographers*, *96*(3), 491–507.
Sieber, R. E. (2000). Conforming (to) the opposition: The social construction of geographical information systems in social movements. *International Journal of Geographical Information Science*, *14*(8), 775–793.
Stallman, R. (1985). *The GNU manifesto*.
Stange, H., Liebig, T., Hecker, D., Andrienko, G., & Andrienko, N. (2011, November). Analytical workflow of monitoring human mobility in big event settings using bluetooth. *Proceedings of the 3rd ACM SIGSPATIAL international workshop on indoor spatial awareness* (pp. 51–58). ACM.
Steiniger, S., & Hay, G. J. (2009). Free and open source geographic information tools for landscape ecology. *Ecological Informatics*, *4*(4), 183–195.
Steiniger, S., & Hunter, A. J. (2013). The 2012 free and open source GIS software map–A guide to facilitate research, development, and adoption. *Computers, Environment and Urban Systems*, *39*, 136–150.
Steiniger, S., & Weibel, R. (2009). GIS software–a description in 1000 words. *Encyclopeadia of Geography*, 1–2.
Sui, D. (2014). Opportunities and impediments for open GIS. *Transactions in GIS*, *18*(1), 1–24.
Temby, O., Sandall, J., Cooksey, R., & Hickey, G. M. (2016). How do civil servants view the importance of collaboration and scientific knowledge for climate change adaptation? *Australasian Journal of Environmental Management*, *23*(1), 5–20.
Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *63*(2), 411–423.
Voss, A., Denisovich, I., Gatalsky, P., Gavouchidis, K., Klotz, A., Roeder, S., & Voss, H. (2004). Evolution of a participatory GIS. *Computers, Environment and Urban Systems*, *28*(6), 635–651.
Wang, S. (2016). CyberGIS and spatial data science. *GeoJournal* (Online issue) 10.1007/s10708-016-9740-0.
Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M. F., Liu, Y., & Nyerges, T. L. (2013). CyberGIS software: A synthetic review and integration roadmap. *International Journal of Geographical Information Science*, *27*(11), 2122–2145.
White, R. L., Sutton, A. E., Salguero-Gomez, R., Bray, T. C., Campbell, H., Cieraad, E., ... Poisot, T. (2015). The next generation of action ecology: Novel approaches towards global ecological research. *Ecosphere*, *6*(8), 1–16.
Wright, D. J., Duncan, S. L., & Lach, D. (2009). Social power and GIS technology: A review and assessment of approaches for natural resource management. *Annals of the Association of American Geographers*, *99*(2), 254–272.
Wright, D. J., & Wang, S. (2011). The emergence of spatial cyberinfrastructure. *Proceedings of the National Academy of Sciences*, *108*(14), 5488–5491.
Wulf, W. A. (1993). The collaboratory opportunity. *Science*, *261*, 854–855.
Yang, C., Raskin, R., Goodchild, M., & Gahegan, M. (2010). Geospatial cyberinfrastructure: Past, present and future. *Computers, Environment and Urban Systems*, *34*(4), 264–277.
Yuan, M. (2016). 30 years of IJGIS: the changing landscape of geographical information science and the road ahead. *International Journal of Geographical Information Science*, 1–10.
Zhao, H., Salloum, S., Cai, Y., & Huang, J. Z. (2015). Ensemble subspace clustering of text data using two-level features. *International Journal of Machine Learning and Cybernetics*, 1–16.