Computer Supported Cooperative Work (CSCW) (2017) 26:165–203
DOI 10.1007/s10606-017-9267-z

Infrastructuring for Cross-Disciplinary Synthetic Science: Meta-Study Research in Land System Science

Alyson L. Young[1] & Wayne G. Lutters[2]
[1]*Department of Human-Centered Computing, Indiana University-Purdue University Indianapolis, 535 W Michigan Ave, IT 587, Indianapolis, IN, 46202, USA (Phone: 317-278-9209; E-mail: youngaly@iupui.edu);* [2]*Department of Information Systems, University of Maryland Baltimore County, Baltimore, MD , USA*

Abstract. Traditionally infrastructure studies are post-hoc analyses of emergent phenomena. While acknowledging the contextual complexity of co-evolution, there has been a turn toward exploring these processes from a design perspective. In this paper we examine a new interdiscipline, Land System Science, whose scientific inquiry is predicated on a deep and ongoing integration of radically disparate data from across the natural, physical, and social sciences. We report the results of a three-and-a half year field study of meta-study practice. In doing so, we perform infrastructural inversion to foreground the backstage scientific work practice to identify points of infrastructure. We used these insights regarding breakdowns and workarounds to inform the design of GLOBE, infrastructural tools that support this community's needs for communication, cooperation, and knowledge construction. Our insight comes from being embedded both with domain scientists and software developers. Through four cases, we highlight the scientists' unique challenges, strategies developed to address them, and the system components designed to better support many of these tactics. Specifically, we address the difficulties of *finding, standardizing, interpreting,* and *validating* data. This advances the infrastructuring literature by illustrating how design can be used to engage a scientific community in active self-reflection.

Keywords: Land system science, Meta-study, Infrastructuring, Information infrastructure, Collaborative design, Synthetic science

1. Introduction

The scientific enterprise is currently in a period of rapid evolutionary change. While there are many factors driving this transformation, the foremost is the necessity for science inquiry to address complex systems. Systems-level science privileges attention to the interdependences present in complex systems, such as our terrestrial atmosphere and human micro-biomes. Here more holistic, integrative methods trump traditional reductionism and atomistic thinking. Consequently, this necessitates a greater reliance on multi-disciplinary investigation. Understanding complex systems requires a plurality of perspectives – integrating the theory base, knowledge production practices, and analytic methods developed in our historically siloed scholarly disciplines (Lutters and Winter 2012). The rise of computational science, enables all of this. Major innovations in computing, such as simulation and data mining, have turned traditional physical disciplines into information-centric disciplines, enabling new forms of scientific inquiry (Nielsen 2011). This movement, known by some as the "fourth paradigm", is made possible by the availability of large datasets, which allow patterns to be derived directly rather than through more inferential means (Edwards et al. 2011; Hey et al. 2009). Through four integrated cases, we examine the processes of infrastructuring for science in this new era.

Traditionally, studies of infrastructure have been post-hoc analyses of emergent phenomena, many of which have studied the short-term aspects of workplace information technologies (Monterio et al. 2013). In this paper, we present findings from a three-and-a-half year ethnographic study of meta-study practice in Land System Science (LSS). LSS is knowledge "infrastructure-in-the-making" (Parmiggiani et al. 2015). While this rapidly growing community exhibits some external markers of an established discipline, such as an organizing scientific body with core conferences and journals, they continue to debate what it means to be a LSS researcher and what it means to do LSS research. This can be seen, in part, in their 2016 rebranding as Land System Science from Land Change Science, further emphasizing the dynamic interactions within socio-ecological systems. These at times roiling discussions make visible key tensions, affording a unique opportunity to observe a field in self-reflection and self-definition. Young communities in the active process of infrastructuring foreground phenomena that can be buried in more mature disciplines, such as negotiations around community identity, scientific methods, and knowledge production. These insights, in turn, can provide valuable information to inform the design of infrastructure tools that support the changing dynamics of a community. As Edwards et al. (2007) note: "…it is difficult to alter infrastructures once they have become established, and thus… choices in the early phase of development really make a difference" (p. 8).

We perform what Bowker (1994) calls an infrastructural inversion, in order to examine the on-going processes of infrastructuring in Land System Science. This analytical lens is particularly valuable for targeting the evolution of young infrastructures over time. We focus, in particular, on LSS's adoption of meta-study research as one way to advance an understanding of human-driven changes to the terrestrial surface. This analytic lens is helpful in detailing the truly backstage elements of work practice, such as human and technical processes for locating, transforming, and synthesizing diverse data in meta-study research. In the process, we identify points of infrastructure (Pipek and Wulf 2009) – breakdowns in scientific practice brought on by mismatches in technology and desired work practices. This in turn allowed us to develop new innovations in technology to amplify LSS researchers' current assemblage of tools and social practices. The resulting system is called GLOBE (globe.umbc.edu).

2. Infrastructure perspective

Infrastructure traditionally refers to the large technical systems that enable groups, organizations, or societies to function in certain ways. This includes such things as roads, railways, bridges, power grids, and communication networks that make up the fundamental facilities of a civic society. Studies have focused, in particular, on the standardization, and emergence of these systems from an installed base (Pipek and Wulf 2009). The term has also come to refer to the constellation of technical tools, protocols, and systems that underpin the Internet, often referred to as information infrastructure or cyberinfrastructure (Karasti et al. 2010). Infrastructure in this sense consists of an assemblage of technical infrastructure, social norms, and organizational practices (Edwards et al. 2007). Typically, information infrastructure is characterized as open (multiplicity in user types), interconnected (multiplicity of purposes, agendas, and strategies), constantly evolving (dynamic ecosystem of systems) and (re)shaped by an installed base of existing systems and social norms (Monterio et al. 2013). Information infrastructure also stretches across multiple scales of action: It is a "technological venture", which seeks to support and enable collaborative work activities; it is a "social venture", requiring human work, organization, and maintenance to function and persist; and, it is an "institutional venture", working to provide accessible and stable amenities to organizations and communities at both national and international levels (Ribes and Finholt 2009, pp. 377–378). It is also fundamentally relational, marked by multiple meanings and ambiguity (Star and Ruhleder 1996).

Infrastructure by its very nature is designed to appear as "timeless, un-thought, even natural features of contemporary life" that are invisible to everyday use (Jackson et al. 2007, p. 2). Thus, they can be difficult to examine. One productive way to understand infrastructure is to invert it. This requires going backstage to study infrastructure in-the-making (Star 1999; Star and Bowker 2002). Bowker proposes using infrastructural inversion as a methodological device to foreground infrastructural elements and unpack the invisible work of infrastructures. This means "learning to look closely at technologies and arrangements that, by design and by habit, tend to fade into the woodwork" (Bowker and Star 1999, p. 34).

This process allows the researcher to see those aspects that have become standard, routine, transparent, and invisible (Edwards 2010). One prominent way to invert an infrastructure is when it becomes visible upon breakdown. Jackson (2014) proposes that these moments of breakdown afford a means to see and engage with technologies in new and interesting ways. This can be seen in the study of a distributed sensing network by Mayernik et al. (2013) in which the researchers show how sensing technologies were initially incompatible with field based environmental research because they did not support the social processes that underlay this type of research. Once the researchers inverted ("unearthed") the infrastructure, they were able to reconfigure it in such a way that reintroduced human skill and expertise into data collection.

While an information infrastructure perspective helps us to understand how information is transported and shared, a focus on knowledge infrastructures allows us to understand how data and information flow through systems and are converted into reliable forms of knowledge (Edwards 2010). These infrastructures thus comprise "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds" (ibid, p. 17). From this perspective, knowledge production is a widely distributed sociotechnical system.

Land System Science is knowledge infrastructure-in-the-making. Just as Climate Science "systematically produces knowledge of the climate" (Edwards 2010, p. 8), LSS is working to systematically produce insight into the ways that land systems act as causes and consequences of global change. To produce such knowledge, they are working to establish community norms and standards, to solidify the role of the Global Land Programme (GLP) – the international institutional body supporting LSS research – as the de facto organization for this type of research, to develop LSS-specific theories and frameworks, and to establish accepted means of knowledge production. One of the ways they are doing this is through the adoption of research synthesis techniques.

2.1. Designing infrastructural tools

A primary objective of this paper is to identify when and how infrastructural inversion happens in an emergent knowledge infrastructure and then use this information to inform the co-design of infrastructural tools to transform current knowledge production processes. There has been much debate about whether infrastructure can, in fact, be designed ex nihilo or if designers should focus more on growing an infrastructure (Edwards et al. 2009; Hanseth and Lyytinen 2008). Indeed, research has suggested that it is often impractical to talk of designing infrastructure. Rather, infrastructure design projects should focus more on designing modules, which interoperate with an installed base or pre-existing infrastructure (Edwards et al. 2007; Jackson et al. 2007). As Edwards et al. (2007) suggest: "…the language of building (cyber)infrastructure (in the sense of creating it either from scratch or according to an orderly progression from plan) may be misguided, and seriously overstate the capacities for action and control available to central system-builders" (p. 38).

This is because infrastructure is large, complex, and layered (Hanseth and Lyytinen 2008). It holds different meanings for different groups of people at the local level (Edwards et al. 2007; Karasti et al. 2010), and it is continually evolving in scale, scope, and functionality (Hanseth and Lyytinen 2008). Further, the work of developing infrastructure requires more than just attention to technical fixes, but also human arrangements, institutional resources, and maintenance of technical systems (Ribes and Finholt 2009). In this way, infrastructure is largely an accomplishment of scale. It starts as a locally constructed, centrally controlled collection of systems that are then assembled into globally guided (inter)networks (Jackson et al. 2007). They extend their reach through a complex process of transfer or translation whereby disparate elements are fit together through adaptation and mutual adjustment (ibid). These changes take time and negotiation, requiring modifications to the installed base to enable interoperation (Edwards et al. 2007). Often, this is achieved through gateways: strategic intermediaries, whether technical, social, political, or legal, that permit multiple heterogeneous systems to interoperate (Edwards et al. 2009). An example of such a gateway would be an adaptor that allows appliances designed for one part of the world to work with the plug sizes found in others. It can also be achieved through a recurring process of adjustment in which an infrastructure adopts to, reshapes, or internalizes elements from its environment (ibid).

Infrastructure is intended to operate over long-term scales. The design of infrastructure is thus a "visionary activity," expected to not only support the work of users today, but also what they intend to do in the future (Ribes and Finholt 2009). Consequently, infrastructure development implies working both with short-term and long-term timeframes, addressing the immediate needs of a community while at the same time anticipating their future requirements (Karasti et al. 2010). This becomes particularly challenging with the introduction of a new science. As users change focus or new users arrive on the scene, they bring with them different requirements and infrastructure needs. In many cases, the existing information infrastructure is insufficient to answer new kinds of research questions (Bietz and Lee 2009). For LSS has this meant developing workarounds and designing fixes as the field matures.

A conceptual strategy that can assist in the design of infrastructural tools is that of infrastructuring. Infrastructuring emphasizes the ongoing and processual quality of infrastructuring activities as well as the extended durations under which infrastructuring unfolds (Karasti and Baker 2004; Karasti and Syrjänen 2004). For analytical purposes, Bossen and Markussen (2010) argue that it is more helpful to think of infrastructure as a verb as this enables "the variety of material and non-material components of which it consists, the efforts required for their integration, and the ongoing work required to maintain it" to come to the fore (p. 618).

In particular, Pipek and Wulf (2009) describe infrastructuring as a methodological approach to the successful design and use of work infrastructures. Infrastructuring activities can inform the search for "points of infrastructure" that evoke improvements in work infrastructures. Points of infrastructure are the moments when an infrastructure becomes visible to its users. This visibility may occur for one of two reasons. The first is an infrastructure breakdown in which part of the technological infrastructure ceases to operate as expected. These are temporary breakdowns in which a work infrastructure is unable to provide services, whether perceived or actual. The second is a use innovation in which users successfully appropriate a new infrastructural tool within a local context. These use innovations can either occur when the technological infrastructure changes to afford new work activities or when new work requirements can be supported with pre-existing technologies. Points of infrastructure can thus be seen to catalyze formal and informal in-situ design work that encourages both users and system designers to reconfigure or extend existing work infrastructure in order to repair a breakdown. We use infrastructural inversion to identify these points of infrastructure to aid in the co-design of infrastructural tools to support the evolution of Land System Science.

3. Methods and approach

The findings we present here are based on a three-and-a-half year ethnographically-informed field study of situated scientific synthesis practice in Land System Science. Our primary data resulted from observations of international workshops, interviews with lead investigators and their students, and a community-led survey. This resulted in 96 h of recorded material and more than 200 open-ended survey responses.

Observations of community-organized workshops provided insights into the perceived strengths, challenges, and opportunities for scientific synthesis in LSS. We attended four multi-day international workshops in (1) Australia (November 2011), which focused on integrating human behavioral data into studies of land use change; (2) The Netherlands (May 2012), which focused on the different synthesis methods used in LSS and how the process could be improved; (3) the United States (June 2013), which focused on the content of LSS meta-studies, seeking to identify patterns across synthesis studies in order to build conceptual models; and (4) Germany (March 2014), which focused on the ways scientific synthesis could be used to enhance models of land use change.

We supplemented these observations with semi-structured, contextual interviews with investigators to further understand the techniques and processes they use for synthesis. We completed twelve in-depth interviews with LSS researchers who had recently worked on a meta-study. Given that this is an emerging practice finding these individuals was challenging, but we were able to leverage our involvement with the workshops to recruit successfully. Our interview guide asked participants to reflect on the benefits and challenges of conducting meta-studies in LSS. It then dove into practical detail asking them to describe how they selected, coded, and analysed case study data, and what technical tools they used along the way. Each interview was conducted in person or via Skype and lasted approximately one hour. All interviewees were academics and held the following positions: Associate Professor (2), Assistant Professor (5), Postdoctoral Researcher (3), recent PhD graduate (1), and a doctoral researcher (1). The sample was highly international with informants from institutions in Canada, Denmark, Germany, the Netherlands, and the United States to best reflect the existing global centers of LSS research.

This fieldwork was augmented with a community driven survey on the current state and future trajectory of LSS. It was administered to the 650 attendees of their premier conference venue – the triennial Global Land Programme Open Science Meeting (GLP-OSM) in Berlin Germany, in March 2014. This afforded the rare opportunity to survey the majority of the global LSS population at a point of communal self-reflection. The open-ended questions relevant to our project mirrored our interview guide. Through this survey, we were able to obtain a baseline understanding of the community, including their common practices, tools, and perspectives. This helped to situate our field findings within the larger population. The final survey sample consisted of 205 respondents (32% response rate, 44% women) from 31 different countries.

Our iterative cycles of data collection and analysis followed a constructivist grounded theory approach, in which earlier findings were used to shape the on-going investigation (Charmaz 2014; Clarke 2005). The first full cycle occurred in late 2012 with open coding of the field notes and transcripts from the first two workshops resulting in a set of coarse themes. Our preliminary codebook was expanded and refined in a second cycle in 2013, integrating the data from the third workshop. These themes informed the initial design of our interview guide and survey instrument. The next iteration involved importing and coding both the interview transcripts and the free text survey results in Nvivo. The final cycle in 2014 captured complete video of the United States workshop and notes from GLP-OSM conference along with the second round of interviews. Throughout the full project analytic memos were created to record emerging concepts and their interrelationships. Integrative diagrams were produced to assist with sense-making with regards to the emerging theory. The data for this paper are drawn primarily from the workshop observations and investigator interviews as supported by survey responses.

Throughout the project, the authors were embedded as equal partners on the GLP-sanctioned, U.S. National Science Foundation-funded software development team to design new digital meta-study support tools ("GLOBE"). This afforded us the opportunity to iterate rapidly between field study and system design. As insights were gained from our analysis, we were able to incorporate them immediately into our weekly design meetings. Likewise, when the team faced a difficult design choice we were able to direct specific data gathering activities with our participants, including small-scale field experimentation and usability assessments. At key points in the project lifecycle, we engaged users outside the design team to test the system functionality. This included both user interface design, such as how to represent the variable displays, and user workflows, such as the most intuitive way to enter a case study into the system. This highly collaborative process and tight coupling between data collection, analysis, and implications provided us with distinctive insight into the interplay between scientific practice and the processes of infrastructuring.

4. Case background: the land system science community

Land System Science is a community forged by a shared worldview that the changing interactions among human systems, the terrestrial biosphere, atmosphere, and other Earth systems are best understood as mediated by human use of land (Turner et al. 2007). It is an emergent interdiscipline with researchers spanning the natural, physical, and social sciences, many of whom have earned their degrees in disciplines that differ from their departments of employment and current research projects. As it is not yet an institutionalized discipline of study, none of these researchers have been formally trained in LSS.

Institutionally, much of the LSS community is respondent to the Global Land Programme. This multi-year coordination project was the result of a series of scoping meetings that began in 2001, culminated with a Science Plan in 2005, and the establishment of the GLP International Project Office in 2006. In 2016, the GLP transitioned to Future Earth, a new 10-year international research initiative aimed at transforming Earth system science and human dimensions research. In this move to Future Earth, the GLP gained programme status solidifying its role and importance in contributing to knowledge and sustainable solutions to global environmental change. Further, as land is at the heart many of the challenges of climate change, the GLP and LSS are viewed as critical for identifying sustainable solutions for the planet.

Traditionally, the referent disciplines of LSS have generated knowledge through local case studies based on detailed field observations of land use change or regional studies that combine remote sensing observations with socio-economic and biogeophysical data. However, the influence of the GLP drew attention to the need for more integrative assessments of the drivers and impacts of land change as part of global environmental change (Verburg et al. 2013). This led to the adoption of synthetic practices such as integrative remote sensing products (e.g., anthromes: Ellis and Ramankutty 2007) and agent-based models.

The LSS community has particularly benefited from advancements in remote sensing. With the use of a single satellite (e.g., Landsat 8) these researchers are able to see large portions of the Earth, rather than relying on reconstructions from independent instrument readings or field observations (Edwards 2010). In turn, this allows them to see how certain landmasses are changing over time, such as the encroachment of cities or the shrinking of forests. Over the past two decades, remote sensing capabilities have improved dramatically offering even finer spatial, spectral, and temporal resolution data improving their understanding of how land is changing (Brown et al. 2013; Kummerle et al. 2013).

While remote sensing has revolutionized LSS researchers' ability to observe and model global patterns in natural systems, human systems are not directly observable from space. A satellite sensor cannot capture, for example, how political or economic decisions impact land use. This means that while they have a good understanding of what is happening to land over time, they have a much weaker understanding of why and how it is happening.

One way that LSS researchers are trying to address this gap in understanding is through the synthesis of results from the existing body of case literature. Synthesis refers to the process of integrating research questions, methods, theories, and data across dissimilar scales of analysis, study systems, and levels of expertise in an effort to uncover consistencies and account for differences in studies. Popularized within the medical sciences as a way to reason across the case study literature, it has had a significant impact on the health sciences. In epidemiology, for example, synthesis is used as a way to make timely decisions about risk assessment and public health concerns, and in pharmacology it has led to the discovery of new applications of existing compounds. Through synthesis LSS researchers hope that they will be able to better understand how human-environment systems act as agents of global change.

4.1. Information infrastructure in land system science meta-study research

Research synthesis in LSS is supported by an assemblage of tools, social processes, and disciplinary norms appropriated from LSS researchers' referent disciplines. This infrastructural assemblage, however, has largely been insufficient to support the production of rigorous syntheses of land use change. Thus, while meta-studies are highly valued in this community and seen as an important step in advancing the production of systematic knowledge on the global causes and consequences of land changes (Magliocca et al. 2014; Rudel 2008), they are currently difficult to do well. In response, the community has mobilized around the need to improve meta-study processes with new methods and innovative tools. The research we present here describes this process of infrastructuring.

In the survey, respondents were asked to describe the technology they use for finding, organizing, and analyzing data in their meta-study research (summarized in *Table 1*). In addition to the specific types of technical tools they use, many respondents reported on the social processes they undertake to enable this type of research. This additional detail further illustrates that research synthesis, like many research activities, is not merely a technical activity, but rather relies on social processes of negotiation and translation.

Table 1.. Summary of tools and social processes used in meta-study research in Land System Science as reported by survey respondents.

Activity	Technical tools	Social processes
Finding Data	Citation databases (Google Scholar, Web of Science); government and project databases; remote sensing and GIS	Systematic queries; snowball searching; cross-referencing; traditional library search; social search; interfacing with government agencies
Organizing Data	Citation management software (Mendeley, Endnote, Zotero); GIS (ArcGIS, ArcMap); Database tools (Access, SQL, Excel); Cloud storage (Dropbox, Google Drive)	Data normalization (R, SPSS, STATA)
Analyzing Data	End-user programming (Excel, SPSS, SAS, STATA); programming languages (R, FORTRAN, Python, SQL); spatial tools (GIS and remote sensing)	Statistical analysis (regression); qualitative methods (narrative interpretation, ideal types, semantic analysis, participatory research); synthesis (meta-analysis, literature review); data-specific (spatial analysis, landscape analysis, network visualization, commodity computing clusters, systems dynamics)

Survey respondents reported using search engines and citation databases, particularly Google Scholar and Web of Science, to find peer-reviewed published literature; government and project-based databases and data repositories to find relevant datasets; and, remote sensing and geographical information systems (GIS) to locate data with a spatial context. Searching for these sources was done primarily online, with respondents using a number of search strategies that included systematic queries, snowball searching, and cross-referencing. In a few cases, respondents relied on physical library searches to find relevant publications.

An interesting category, described in section 5.1.2, was respondents' use of social ties to obtain relevant data sources. They reported sharing data with collaborators, contacting authors to obtain original datasets, and attending workshops and scientific meetings to gain access to data. They proclaimed that, "colleagues and social networks have been more important for finding difficult to obtain datasets" (Survey Respondent 71) and the "best technology [is to] contact the authors behind the data!" (SR149). In a few cases, respondents reported working directly with government agencies to obtain needed data for their analyses.

Once the desired data were obtained, respondents used a number of technical tools and social processes to organize their data. Published literature was stored in citation management software, including Mendeley, Endnote, and Zotero, while spatial data were housed in GIS-related products, such as ArcGIS and ArcMap. Some also used database tools, such as Access, SQL, and Excel for data management. In a few cases, they mentioned the use of cloud storage services, such as Dropbox and Google Drive, to support distributed collaboration.

For some respondents, organizing data also meant normalizing the different data formats and spatial scales to allow for comparison. Respondents mentioned using programming languages, such as Cran R and Python for this purpose: "I like program languages, in general, because data comes in different formats and have to be organized before being used in modeling and research" (SR 72). Statistical software, like SPSS and STATA, served a similar purpose.

Regarding tools for analysis, three prominent categories emerged: end-user programming tools, such as Excel, SPSS, and SAS; programming languages, such as R and FORTRAN; and, spatial tools, namely GIS and remote sensing software. The first two categories were used primarily to conduct statistical analyses on the data. Indeed, the use of regression was the foremost mentioned analytic technique used by respondents. Spatial tools, on the other hand, were used to organize the data geographically. This particular analytic technique allows researchers to see the spatial distribution of their cases and to attempt to determine the degree to which the collection of cases is representative of the global extent.

While quantitative analysis methods dominate LSS meta-study practice, a few survey respondents referenced qualitative analysis techniques, such as narrative interpretation, ideal types, and semantic analysis. They also referenced specific epistemological orientations, namely participatory research and interpretivist approaches. In a few cases, respondents also reported methods they used to analyze specific classes of data: for images, they referred to such methods as spatial analysis, landscape analysis, and network visualization, and for economic data they reported the use of commodity computing clusters and systems dynamics approaches. Finally, some respondents explicitly mentioned the use of meta-analytic techniques.

Although not explicitly mentioned in the survey, findings from other components of our study point to the importance of participatory approaches to meta-analysis for data interpretation. LSS researchers often engage with original case study authors, study staff, or experts to obtain information on how data were constructed and what the data mean. This is a necessary step to ensure that data are appropriately (re)used. The process by which this is done is described in section 5.3.2.

5. Cases of infrastructuring

In the following section, we describe four cases of infrastructuring as they relate specifically to meta-study practice in Land System Science. In each case, we start with a general problem confronted by these researchers as they attempt to synthesize case data from disparate disciplinary sources using the current assemblage of tools and social processes appropriated from their own home disciplines. We then contextualize the specific problem to the LSS context and describe workarounds developed by these researchers to overcome or alleviate these impediments. We use these insights to guide the design of technical tools to help amplify the evolution of the existing knowledge infrastructure to better support meta-study practice in LSS. We focus in particular on technical solutions that address issues of scale and technical skill, while acknowledging the limitations of technology for supporting the more social aspects of meta-study practice.

In order to interpret these cases, the reader needs to understand the concept of *geocontextualization*. As a community focused on studying specific Earth systems and their human dimensions, spatial context is paramount. Rather than abstract processes such as in atmospheric chemistry, in LSS the location where something happens matters. LSS scientists maintain a geographic gestalt of Earth systems to which they constantly orient their reasoning. Spatiality is a fundamental component of all LSS research and is one of the ways that the community is defining what it means to be a LSS researcher: all data and findings must be reported relative to specific locations on the planet. As Edwards notes: “world maps undergird our ability to conceive global space. They are an infrastructural technology, a principal material support for ‘thinking globally’” (Edwards 2010, p. 27). For researchers from other disciplines interested in doing LSS research, they must therefore learn to situate their data in space and to appreciate the interconnectedness of the variables constrained by that geography.

This emphasis on geocontextualization manifests itself in how LSS researchers assess the validity of their and other researchers' results. We continuously observed in the workshops and scientific meetings researchers using maps to illustrate the degree to which their findings reflected an appropriate distribution of a phenomenon across the globe. LSS researchers have implicit, tacit spatial knowledge; they can look at a map and know, for instance, exactly where cities are located or what areas of the world have minimal tree coverage. Geographic representations therefore are central to any kind of data analysis done in LSS.

The need to understand the spatial context of their data is also expressed in the tools they use to conduct their meta-studies. In the survey, participants reported using geographic information systems to support nearly all stages of the meta-study process, including organizing spatial data and using GIS layers to explore the interrelatedness of their variables. In the interviews and workshops, participants further described their use of data visualisation tools, such as Google Fusion Tables, to understand the spatial distribution of their cases. One meta-study researcher even described his development of a custom visualization tool to try to assess the representativeness of his case collection.

5.1. Case 1: finding studies

5.1.1. *Online searching*

Most meta-studies in LSS start with a search to identify published case study results relevant to the purpose of their meta-study. How LSS researchers define a case study depends largely on the scope or scale of their intended synthesis. For some, a case is equivalent to a publication, for others it can refer to a single study site, and for still others it refers to the unit of observation used within one study (Margulies et al. 2016). Despite these differences, two common criteria exist for a publication to be accepted as LSS relevant: it must be grounded in physical geography and it must focus on how humans are changing land.

Like all researchers, LSS scholars uniformly begin their search for cases with a keyword search via an established information broker, such as Web of Knowledge, Google Scholar, or Science Direct. A common query structure is:

> keywords (swidden or shifting cultivation (slash and burn)) and (change or driver* or impact) for studies published in 2000-2010).
>
> (Amsterdam workshop, presenter slides)

In the process of searching for cases, many participants uncovered a core challenge of LSS research: determining the correct keywords or keyword combinations is difficult when dealing with cross-disciplinary cases. For one postdoctoral researcher, this meant that his searches routinely yielded either too many or too few results.

These researchers also use these databases to locate data with a spatial context. This is where traditional searching becomes particularly problematic for this community.

As keywords map tenuously to geography, it is often difficult for them to determine the correct terms to find cases in a particular area of the world. For instance, one of the interviewed postdoctoral researchers described the difficulties she had finding cases in the Sahel because each author used a different term to represent the region. While some cases used a particular city name, such as Timbuktu, others used the name of the region, namely West Africa. This complicated her search for relevant studies.

Different disciplines have found multiple ways to cope with the limitations of keywords to represent space. The creation of standardized ontologies, such as administrative boundaries or World Protected Areas, is one such solution. These ontologies parse areas of interest into clearly defined geographic spatial entities, representing the different ways that disciplines understand and talk about space. For example, political scientists care about provinces and precincts compared to economists who care about postal code based marketing clusters. Although these standards have proved effective for the individual disciplines that created them, they do not translate well across disciplinary boundaries.

One solution to this is the use of standardized study location data and environmental attributes to search for scholarly research geographically. While location-based search engines are starting to exist (e.g., Journalmap: Karl et al. 2013) all of them are based on coordinates, typically latitude by longitude. For LSS researchers, these systems are of limited utility because they lack the contextual details needed to understand the variables likely at play in a particular site: “If you actually need to know what’s going on in a location, point data is crap” (comment by domain scientist at GLOBE design meeting).

The limitations of point data become particularly apparent when considering the three real-world examples below (*Figure 1*). In the first example (top left), the point is located just outside of the actual study area. This means that what is represented by the point is not the intended study area but rather a neighboring site. In the second example (top right), the point is located within one of two study sites and therefore only accounts for the conditions of one of the case locations. In the third example (bottom), the point is located in a very large geographic region. This would be the equivalent of placing a pin in the middle of Canada, which is extremely geographically diverse. Thus, it is impossible to figure out what the pin actually represents.

Recognizing these limitations, a handful of journals in the environmental sciences (e.g., *Journal of Environmental Management*, *Journal of Archaeological Science*, and *Engineering Geology*), now allow authors to upload GIS files as part of their papers’ metadata. This allows researchers to search for case studies within an exact geographic area. This can be as specific as a town in Brazil or as generic as an entire country. For LSS researchers, this level of detail is necessary to fully understand the contextual factors that are likely at play in a particular area. In other words, by knowing the exact physical boundaries of a study site, LSS researchers can visually check whether or not variables of interest will be represented there.

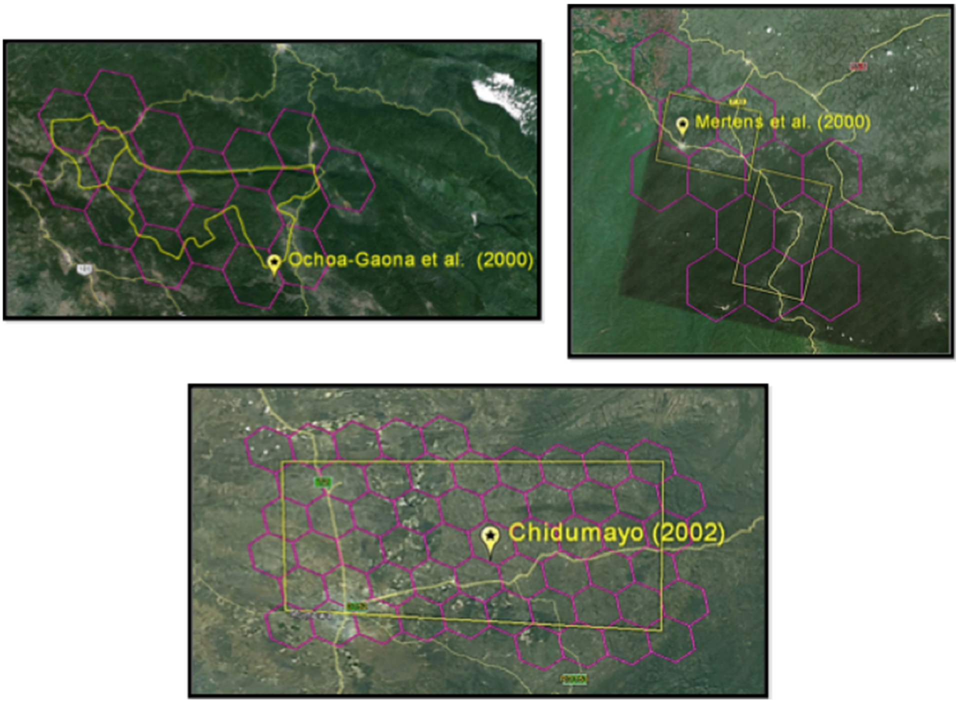

Figure 1. Examples of problematic point geometries.

5.1.2. *Social searching*

To overcome some of the logistical challenges imposed by the limitations of citation databases, LSS researchers are turning to their social networks and original case study authors to obtain difficult to find information, including full case details. This process of social search helps meta-study authors to more easily identify relevant cross-disciplinary literature than is possible using a keyword search in a citation database. One postdoctoral researcher noted:

> I mean without their input this analysis would not be possible because it's often data that is not published yet or is data that is even if it would be published… it's information that is not provided in one publication (interview, doctoral researcher).

Enlisting the help of colleagues and original authors then is often the only way to obtain the provenance information required for successful synthesis. However, this is not a fail-proof solution; original case study authors are not always so willing to help meta-study authors: "we asked people who we knew had expertise in certain underrepresented regions… It was tough while we were doing the study to get much help though. We used listservs, email, professional groups [but received] very little feedback at all" (interview, assistant professor). For this participant, harnessing the power of his social network did not prove to be a fruitful activity.

5.2. Case 2: standardizing for reuse

A challenge of all interdisciplinary research is that data come in multiple formats, which are not readily comparable. Researchers must first transform the data to a standardized format before it can be synthesized: "Each paper talks differently about the same topic. They don't use the same words. But in the end if you want to summarize you need to put them in an overarching category" (interview, assistant professor). For data to be commensurable, it must therefore be removed from its original context. Reduction is necessary for comparability (Latour 1999).

The processes by which cases are conducted and findings are reported impact researchers' ability to standardize data for synthesis. Disciplinary traditions, for instance, impact the techniques of measurement used to create data, imbuing them with judgments and values about what constitutes data, what is considered the best unit of measurement, and the ways in which different entities are grouped to make a measurable whole (Pine and Liboiron 2015). Differences in methods, orientations, and analytic tools ultimately result in heterogeneous data that are incommensurate (Baker et al. 2005b). For LSS, these differences often mean that case studies frequently differ in how they report and weight variables, and important factors in one study are often overlooked in another. It also often reflects how scientists view their own data: for some, observations may be research findings, while for others they may merely provide background context (Borgman et al. 2007).

The predominant approach to reconciling data heterogeneity has been the development and enforcement of metadata standards (Baker et al. 2002). "Standards act as lubricants. They reduce friction by reducing variation, and hence complexity, in sociotechnical processes, and they 'black box' decisions that would otherwise have to be made over and over again" (Edwards 2010, p. 251). Standardization through classification thus facilitates interoperability (Lee et al. 2009; Ribes and Lee 2010) and comparability (Bowker and Star 1999; Latour 1999). Standards have been proposed, designed, and developed to better facilitate data sharing (Karasti et al. 2010) and data reuse (Faniel and Jacobsen 2010) across many scientific disciplines. This is one reason that research synthesis has been successful in highly structured and mature fields such as the health sciences and, more recently, ecology. Standardized study procedures, data reporting, and metadata practices make data reuse and integration more tractable (Baker et al. 2005a; Blake and Pratt 2006a, 2006b; Rolland and Lee 2013; Yarmey and Baker 2013).

While standardization has proved useful in these contexts, the nature of LSS meta-study research is such that a universal classification scheme does not currently exist and may never be a feasible solution. Three prominent issues limit its utility. First, most LSS analyses are second hand syntheses of primary data not typically collected by the scientist herself. These data are thus already imbued with their own disciplinary standards. In addition, the original data may have been previously reused by another discipline, causing it to be wrapped in secondary layer of disciplinary-specific metadata. Second, LSS researchers come from different disciplinary backgrounds and therefore bring their own standards with them that they apply to the transformation or reuse of data. Finally, as a loose confederation of researchers, no authority or governing body exists imposing uniquely LSS standards. Lacking such standards, LSS researchers have developed a number of strategies to deal with data incomparability.

5.2.1. *Transforming cases to binaries*
How the data are transformed is often a consequence of how they were originally described in the individual case studies. The level of precision at which a case study author measures inputs, for instance, affects the degree to which the meta-study author can talk about the strength of a phenomenon:

> The problem is that you see a land tenure security variable and you see a deforestation or forest change variable, but you don't exactly know the direction of that effect, you don't know if changes in forest cover are maybe actually creating different security scenarios or whether the security itself is influencing the change in forest cover (interview, assistant professor).

For this participant this meant limiting his data transformation to an indication of whether or not there was an effect of land tenure insecurity on deforestation. The data did not allow him to defensively extrapolate to the strength or direction of the effect.

It is very common for LSS meta-study researchers to quantify the qualitative case data in their meta-studies. In interviews, participants consistently mentioned the processes they used to transform narrative data to allow for quantitative comparison. This is in line with conventional thinking of natural sciences that only through quantification can valid results be obtained (Knorr Cetina 1999). A postdoctoral researcher, for instance, showed how he had transformed qualitative descriptions of significant drivers of agricultural change from a narrative account to a numeric representation that would allow him to run statistical analyses on the data. If the case study authors had reported that a driver had a significant relation to land use change, he coded this as a "1" if present and "0" if not.

In the process of transforming qualitative data, contextual information about the data is lost. This loss is problematic because understanding human impact is critical to their research practice. Consequently, outliers in the data provide needed context to understand this impact. In an effort to preserve context, some LSS researchers have adopted coding practices that not only quantify qualitative data but that also provide qualifiers for why the data were transformed in a particular way. In describing his data transformation practices, a doctoral researcher indicated that he tries to retain contextual information by attaching qualifiers to his codes: “If I put a 1, I note or give a quote from the report qualifying why there is a 1” (interview, doctoral researcher).

One way that LSS researchers balance the need for large-scale data transformation with the preservation of context is by using a visual programming environment like an Excel spreadsheet:

> I was thinking of using a database manager like Access or others and developing a form to put in a structured way the data, but I found that is sometimes too restrictive and that it’s more easy if you have one [Excel] spreadsheet where you can easily add a column or skip through columns and compare and so on. Of course … it can get a lot of columns, but somehow this makes it more transparent. And you are better able to understand your data (interview, doctoral researcher).

While this participant identifies the potential benefits of using a database manager to create forms to transform, organize, and search across his data, he also realizes the limitations it would impose, in particular the loss of data transparency afforded by the more visible and interactive spreadsheet.

5.2.2. *Amplifying with global data*

Even more challenging is that case studies often do not report certain contextual details, such as the direction of a change. One of the ways that LSS researchers are trying to get at some of these missing contextual details is by using trusted external data as a means of validation and explanation. The use of commonly accepted global datasets (e.g., OECD, IMF, WorldBank) is seen as a particularly promising approach.

Global data are made up of heterogeneous, irregularly spaced instrument readings that have been converted into complete, consistent, gridded global datasets for areas of the world in which observations do and do not exist (Edwards 2010). Thus, they provide a complete and standardized global record that, when overlaid onto the heterogeneous multidisciplinary cases, help to fill in gaps in understanding. This is reported in one workshop participant’s published meta-study as:

> A common problem in the meta-analysis of qualitative and narrative case studies is the comparability of the driving factors mentioned, and the limitations on extrapolating the results due to their qualitative character. We have, therefore,

used georeferenced datasets, in addition to the commonly-used qualitative meta-analysis, to describe in a comparative manner, the location conditions of the case-studies (citation removed for anonymity).

Instead of inferring what the case study authors meant by certain variables and then organizing similar variables into an overarching category, this researcher chose to use global datasets to help interpret and describe the conditions of the locations in the case studies used for her meta-analysis. This takes an alternative, top-down approach to research synthesis, focusing more on which cases can be used to represent certain parts of the world and using global datasets to provide ground support to explain the local results.

The use of global data, however, is not a panacea. Most LSS researchers do not have an adequate statistical understanding of global data nor do they have the GIS skills required to apply them. Nearly half of the participants in the US workshop felt that this would involve a steep learning curve. Because global data are reported at different scales and resolutions compared to case study data they are inherently incommensurable. Even when researchers have the skills needed to transform the global data, finding the best fit dataset is difficult. One meta-study author described how he tried this approach to control for the unobservable effects in his analysis:

> One way to try to get around that was to think about using this World Bank data, which at the country level they have these measures of governance and they measure and define governance in several ways… What I did was I attempted to use these variables about governance to try to help control for that relationship that we can't really assign causality to without something like that (interview, assistant professor).

While he was able to transform the data, ultimately the external variables were not appropriate for his particular study:

> …there's like certain statistical criteria that those external variables need to meet to actually do the things that we want them to do and they didn't really meet those so they weren't really valid for trying to help us control for the unobservable effects that we couldn't control for in our analysis (interview, assistant professor).

Ultimately, despite all of his hard work, he was unable to strengthen his analysis given the statistical limitations of this dataset. While frustrating, he still sees promise in using global data to understand what drives outcomes in a more comprehensive way.

5.3. Case 3: interpreting data from distal disciplines

A related problem is interpreting data from dissimilar disciplines. Like others who have studied reuse (e.g., Birnholtz and Bietz 2003; Faniel and Jacobsen 2010; Zimmerman 2007), we observed our scholars struggling with concerns about data interpretation and quality. Participants likened it to "comparing apples to oranges" since land use changes have been studied from so many different disciplinary perspectives (SR65).

Recall that in the process of moving from data collection through analysis to publication important details about study procedures are often lost. The inevitable post hoc rationalization and deletion of process is necessary to make data comparable and communicate study findings (Gooding 1982; Latour 1999; Latour and Woolgar 1979; Star 1985); however, this leaves the meta-study researcher at a loss as contextual information needed to evaluate data for reuse is often missing (Faniel et al. 2013; Rolland and Lee 2013).

For those researchers from proximal disciplines, similarities in background and experience have been found to aid in data interpretation and validation (Zimmerman 2007). For instance, Baker et al. (2005a) describe how although the interdisciplinary field of ocean informatics draws from such vast domains as the physical, chemical, biological, geographical, and atmospheric sciences, these all have overlapping techniques for data handling, modeling, and visualization. These commonalities consequently enable them to develop strategies for dealing with their distinctive knowledge interests. Study sampling grids, for example, serve as boundary objects (Star and Griesemer 1989); they represent a shared framework of language and experience that the interdisciplinary oceanographic researchers can use as a reference and basis of comparison.

As we move towards greater multidisciplinarity in scientific practice, with researchers reusing data from increasingly distal disciplines, this becomes much more difficult due in part to differences in practice and instrumentation (Borgman et al. 2009). In these instances, it is no longer a case of finding ways for researchers studying ocean currents to interoperate with marine biologists; rather, it becomes a question of how to facilitate multidisciplinary activities between, for example, landscape ecologists, cultural anthropologists, geoinformaticians, and environmental economists – disciplines that differ vastly in their epistemologies, ontologies, and approaches to knowledge construction. LSS researchers have thus developed workarounds, with varying degrees of success, to deal with this added complexity in data interpretation.

5.3.1. *Using authors' definitions*

The first technique was to adhere to the case study authors' definition: "we try to take the interpretation of the original researchers rather than make the interpretation ourselves" (comment at US workshop). While data may be coded under a different label to match the meta-study authors' categories, the reported observation is not reinterpreted. A downside of this approach is that the meta-study author's framework often does not align with that of the case study author. This is evidenced in the following discussion in which a geographer questions a modeller's chosen classification scheme:

> Geographer: It looks like although you define urbanization as external but urbanization could be a global driver as well or global driver could be external drivers as well. And subsidies could come from the government and could be an external driver too.

> Modeler: It is at least partly influenced by our background as land use modelers and as a modeler the question is always what is 'endogenous in your system and what is exogenous in your system?' When we simulate agricultural change then I take external drivers as those that are not driven by farmers or agriculture (comments from US workshop).

Here we see that the modeler's process of data interpretation is influenced by his disciplinary worldview. This means categorizing variables into binaries of whether the variable is external or internal to the model. This logic, however, does not map to the geographer's worldview. From his perspective many of the variables the modeler has classified as endogenous could in fact be exogenous. Edwards et al. (2011) report a similar challenge in which members of a working group expressed confusion over terminology used to label elements of their models. This resulted in the need for continual explanation throughout the meeting. This highlights how one's disciplinary perspective can unconsciously influence how data are understood and treated.

5.3.2. *Engaging original authors*

A second technique was to enlist the help of original case study authors in the meta-study process. This kind of engagement has been found to be one of the only ways to obtain contextual information needed to assess data relevance (Hackett et al. 2008; Rolland and Lee 2013; Zimmerman 2007). Through a participatory meta-analysis approach, LSS researchers are better able to ensure the validity of their codes and resulting models (Young and Lutters 2015). This is because "the first thing people in a meta-analytic team who did some of the case studies are going to do is they're going to look at how their case was coded" (comment at US workshop). This, in turn, helps ensure the codes reflect the original authors' intentions.

Our participants described two types of participatory meta-analysis. The first was an interpretative approach. This involved "bringing in primary authors working in these systems and sitting around the table and thinking about drivers of change as well as consequences of change" (comment at US workshop). Original authors were not asked to code their cases, but rather to describe the land change processes they had witnessed in the region. These discussions provided the needed contextual information to enable the meta-study authors to more accurately code their cases for analysis.

The second was a partnered approach, in which case study authors were actively involved in each stage of the meta-study process. This often consisted of asking case study authors clarifying questions about their study sites, coding cases based on this information, and then asking them to confirm whether or not their cases had been interpreted correctly. The original authors were then able to amend codes where necessary. These case study authors were often given co-authorship in exchange for their assistance.

While this process is relatively simple for the reuse of small data datasets (Rolland and Lee 2013), it becomes much less practical when scaled up. For LSS, where there are multiple cases from multiple authors, each with its own unique set of data, the process of engaging original authors is much more complex. Consequently, those researchers who engage in this type of participatory meta-analysis prioritize their inclusion of case study authors to those who are experts on the topic and can speak across multiple studies simultaneously.

5.4. Case 4: determining representativeness

A core challenge for LSS researchers is determining if the collection of cases in their meta-study is actually representative. In the optimal meta-study, cases are selected at random across the globe. For LSS, this is often not feasible because the universe of cases has not been conducted at random locations across the Earth. There is an overrepresentation of certain areas of the world. The majority of all case studies are found in four types of locations: (a) adjacent to a university, (b) near an airport, (c) in warm climates, and (d) in developed countries. Currently, there is no easy way to assess a collection of cases for representativeness. The present strategy is to find all cases on a particular topic and hope that this approach is viewed as systematic. One way LSS researchers are trying to determine the representativeness of their case collection is by loading their cases onto a map. One interviewed associate professor went as far as to create his own global visualization system to attempt to determine representativeness – the result of his effort is shown in *Figure 2*. A visual representation is the best way LSS researchers are able to make sense of their case distribution.

Even when case studies are available, often they have been conducted experimentally rather than observationally. This is problematic because LSS researchers are interested in understanding overarching principles based on the routine natural behaviours of the global populace, which cannot be adequately captured in interventionist or experimental study designs. One of our workshop participants used an example from the journal

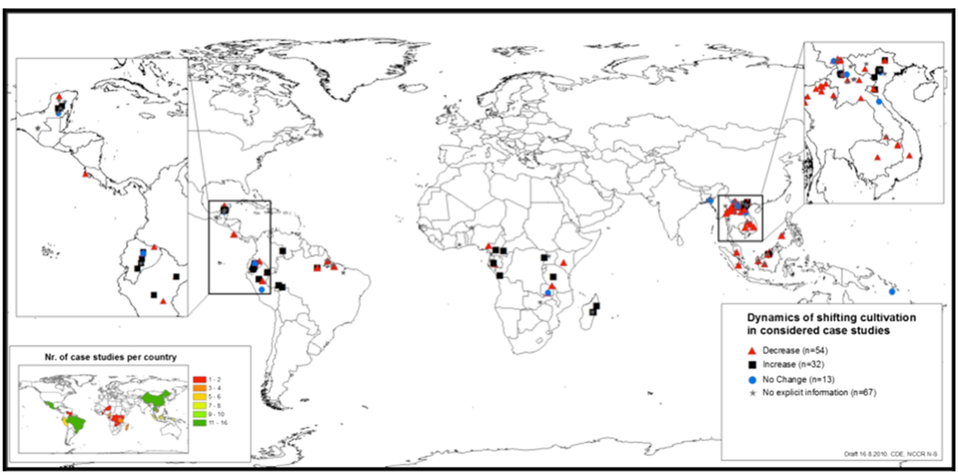

Figure 2. Spatial map of case studies presented at Amsterdam workshop.

Nature to show how despite rigorous selection criteria and state-of-the-art statistical analyses, a study can be weakened by poor sampling procedures:

> The problem is the population of studies that they've got from the developing world, most of the studies were done in experiment study settings… This is a real no-no because you're not talking about a population, you're not talking about a setting that at all is representative of the universe of cultivators you want to generalize (comment at US workshop).

As a result, the study's conclusions were viewed as suspect. This same participant goes on to suggests that since meta-studies are exercises in abstraction, researchers need to be cognizant of the "garbage in, garbage out" problem otherwise their study results will not be worth "the electronic space it occupies". A core group of LSS researchers are currently attempting to address this issue of 'comprehensiveness' by creating a protocol for conducting meta-study research in LSS, similar to that of the PRISMA statement used by health professionals (Liberati et al. 2009).

Next, we illustrate how we used the findings from our field study to design a system, called GLOBE, to address many of the technical limitations impacting meta-study research in LSS.

6. The GLOBE system

The GLOBE system resulted from a need identified by the LSS community for innovative tools to better support meta-study practice. Commissioned by the Global Land Programme and endorsed as one of its core products, GLOBE depends strongly on pre-existing LSS infrastructure, combining existing services, such as global datasets, local published case studies, and map-based visualizations, in new and innovative ways. Our objective was not to replace those technical tools that work, or to try to technically reduce the social aspects that underlay meta-study practice, but rather to amplify the current evolving LSS information infrastructure to better support this type of research. We focus, in particular, on those enhancements that help to address issues of scale and technical skill, providing LSS researchers with new ways to locate, visualize, and determine the representativeness of a collection of cases. Thus, GLOBE does not attempt to solve all the challenges of conducting a meta-study in LSS, nor does it attempt to replace all of the workarounds developed by the community. Rather, it provides a means for LSS researchers to share, compare, and integrate local and regional case studies with global data to better assess the global relevance of their work. We are currently in the process of transferring ownership of GLOBE to the GLP.

GLOBE uses paradigms from collaborative social computing and advanced statistical modeling to aggregate and analyze local and regional case studies to enable the synthesis and integration of globally-relevant information for land system scientists. In GLOBE, a case consists of three main parts: source (i.e., bibliographic information), geometry (i.e., study site), and case notes (i.e., provenance information). While this information is extracted from the original source, PDF copies of published case studies are not stored in GLOBE. Rather, links are provided to download the publication directly from the publisher. Case information is entered into the system by community members who have an account. Accounts are approved by members of the GLOBE research team and can be requested by filling out an online form. Our decision to make GLOBE a controlled access environment was to safeguard it against bots filling it with garbage cases, which would significantly reduce the quality of the analyses.

GLOBE also does not support micro-scale data analysis. While the system can be used to identify geographically relevant local and regional case studies and package them with global data, researchers still need to manually review the published literature for relevancy and code the cases based on their particular research topic. In other words, GLOBE cannot determine if the published literature actually describes the variables of interest. It can only pull those cases from areas of the world where the variables are present. GLOBE helps to support these manual processes by enabling users to export the packaged case information and global data in an Excel spreadsheet for further analysis outside the system. Users can also attach notes to the cases, serving a similar function to the qualifiers used by the doctoral student to describe the contextual factors in his cases. Recall that Excel is the preferred tool for this type of micro-scale data analysis. GLOBE thus acts as a gateway, permitting multiple systems to interoperate

Next, we describe how our four cases of infrastructuring informed the design of the GLOBE system functionality. We illustrate in particular how we built on the current LSS infrastructure to make access to global data, local and regional case studies, and social contacts easier, as well as how we nudged the infrastructure to enable new forms of analysis.

6.1. Enhancing spatial search capabilities

We leveraged this community's emphasis on geographic sensemaking by designing GLOBE as a geographic search engine (*Figure 3*). Remember that while place name keyword searches can pull cases that reference a specific geographic location, they cannot confirm if a case is actually located in the specified area. Representing searches on a map assists both with information seeking and with sensemaking across the results: LSS researchers can better assess data relevance when they understand where in the world the data are located. Visualizing case searches in space is therefore a powerful tool for this community.

We included global data filters to help streamline the search process. Users can limit the extent of their search to specific global variables of interest, such as mean annual temperature, or areas of the world, such as tropical regions. This allows them to instantly eliminate seemingly irrelevant cases from their sample, significantly reducing meta-study production time. Current search parameterization processes require users to manually inspect each potential case for significance – a time-consuming and laborious activity.

Figure 3. GLOBE case search interface with global variable filters.

GLOBE supports this process by encouraging case contributors to upload geometries (KML, shapefiles) with their cases. *Figure 4* provides an example of how a case with multiple geographies is treated in GLOBE. For those researchers who do not have such files, we offer tutorials on how to create them. Alternatively, case contributors can select an existing geographic unit, for example, from the World Protected Areas index, or they can trace the boundaries of their study site using

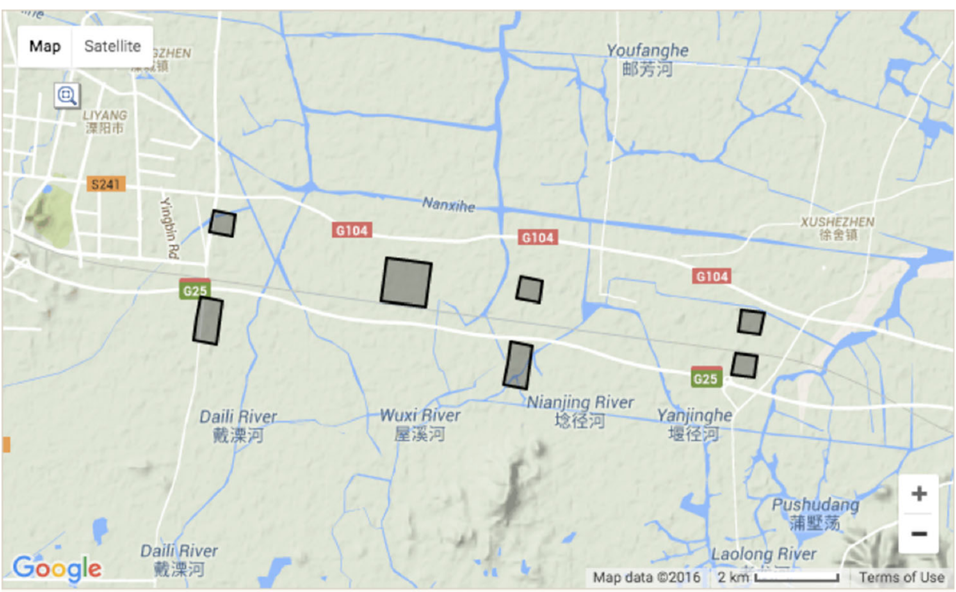

Figure 4. Representation of case geometries in GLOBE illustrating the treatment of one case consisting of multiple study sites.

GLOBE supported drawing tools. Recent efforts are examining the effectiveness of machine learning to deduce geographies automatically. This use of full case geometries thus enables a much more accurate analysis. Since the boundaries of the study site are clearly defined, the GLOBE system can better determine the actual set of global variables at play in that area.

6.2. Providing access to normalized global data

Recall that one of the ways that LSS researchers are dealing with data heterogeneity, incommensurability, and issues of data precision is the use of global datasets to fill in contextual gaps. In particular, GLOBE was designed to address the challenge of locating relevant global datasets by acting as a single repository for a suite of common global variables from disparate and often difficult to navigate sources. Each dataset is broken down into its individual variable component, such as population density or market access. Users can rapidly select and try out over 100 variables of interest to see which ones are relevant for their particular assessment (*Figure 5*). They can also download the variables to an Excel spreadsheet for external use and recommend additional datasets to be added to the repository.

GLOBE was also designed to address the challenges associated with mismatches in scale and resolution between global data and case study data. Indeed, a major innovation of GLOBE is that we normalize all global variables to a universal geographic form for consistent analysis across datasets. We call these GLOBE Land Units (GLU), which are 96 km^2 equal area hexagons distributed across the Earth's land surface (*Figure 6*). As human systems are typically less well quantified than biophysical systems, this surface area is the finest spatial resolution that offers direct use of high-quality global land use area estimates (Ellis et al. 2011).

This approach enables rapid computation and integration of global data with case study data, an otherwise complicated and often time-consuming process that requires advanced GIS skills and expertise, which many LSS researchers do not have. A disadvantage of this approach is that finer-resolution global datasets are also transformed into these more coarse universal units, lowering the quality and precision of the original data. It is this process of data normalization, however, that affords nearly instantaneous and seamless integration of local and global data compared to other GIS-enabled methods. This computational benefit also supports real-time analysis and interactive visualization significantly enhancing the user experience.

6.3. Supporting participatory meta-analysis

A prominent approach to data access and interpretation is engagement with case study authors or topic experts. When looking for individuals who can help authors

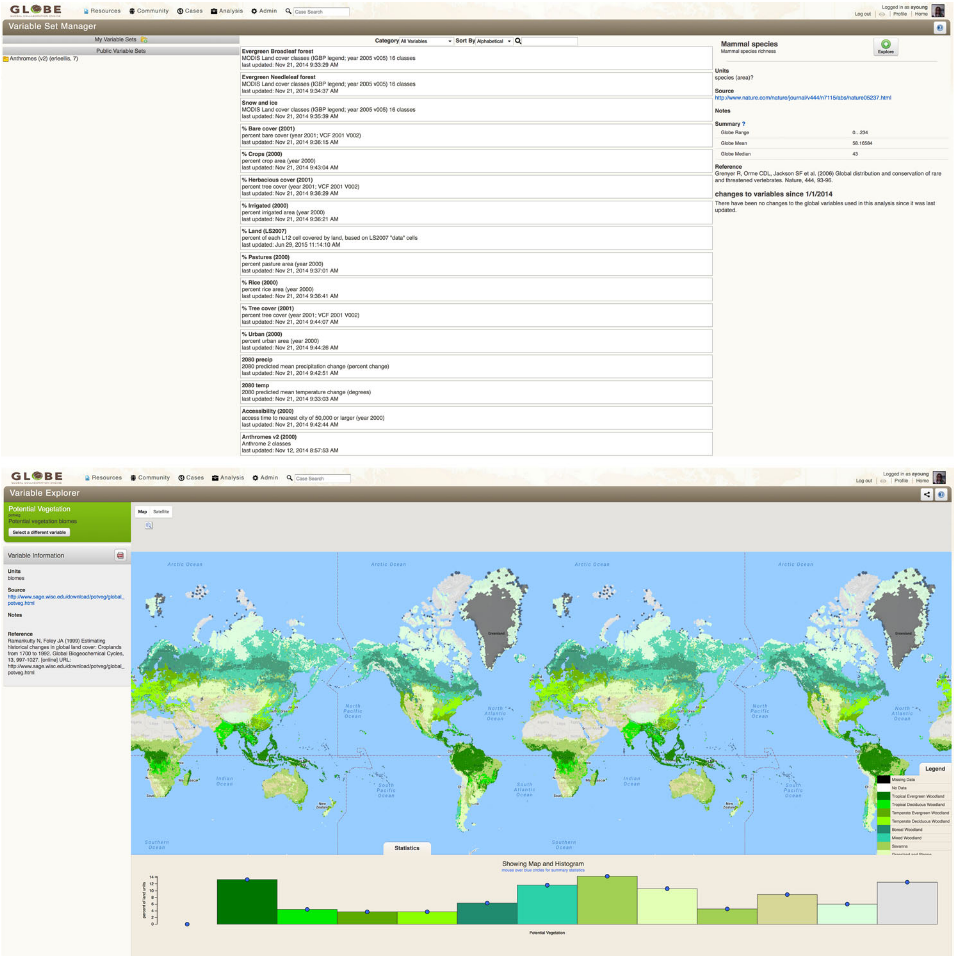

Figure 5. GLOBE variable repository (top) and GLOBE variable explorer (bottom).

understand and interpret cross-disciplinary cases on a particular topic or within a particular area of the world, the most common way is to find similar case studies. This approach allows meta-study authors to indirectly identify researchers who are studying the same phenomenon that they are studying. Recall, however, the challenges of searching for cross-disciplinary cases using traditional keyword-based approaches. As different disciplines use different keywords or ontologies to describe variables, finding relevant or similar cases is not always an easy process.

GLOBE is designed to support and simplify this process by enabling researchers to conduct a similarity analysis. This is an assessment that examines the global context of a selected case based on one or more specified global variables, such as population density, percentage crops, or potential vegetation. Users can also specify the global extent of their analysis to areas of the globe that they want to consider in their assessment, such as tropical biomes or ice-free land. Results are then presented

Figure 6. Globe Land Units covering metro Washington, DC USA.

as statistics, histograms, and maps. Areas most similar to the original case study site are highlighted in purple and areas least similar in pale blue. A rank-ordered list of similar cases is also presented to users (*Figure 7*).

Users can then use this list of similar cases to connect with case study authors. GLOBE facilitates this by providing direct links to the site expert's verified email

Figure 7. An example of a similarity search in GLOBE.

address and personal website (*Figure 8*). Each case entered into the GLOBE system is attached to an originating author with clear provenance metadata. System-generated quality metrics also help to identify cases with weak links that will likely be difficult to interpret and integrate. In these cases original authors are invited by the system to work collaboratively with case contributors to verify, correct, or re-enter case details and geographic information. Cases vetted in collaboration with original case study authors earn higher quality scores.

These features have been expanded to include visualizations of GLOBE users who share similar interests in topics or geographic locales. Accessible from the GLOBE user's profile page, these features provide a means for meta-study researchers to quickly learn about and potentially connect with other relevant GLOBE users. A linked top keyword list is also generated for all GLOBE users who have created cases or case collections (*Figure 9*). When a user selects one of these keywords, a list of cases containing that keyword of interest is returned and visualized on a map. Users can then use the previously described method to contact case study authors for more

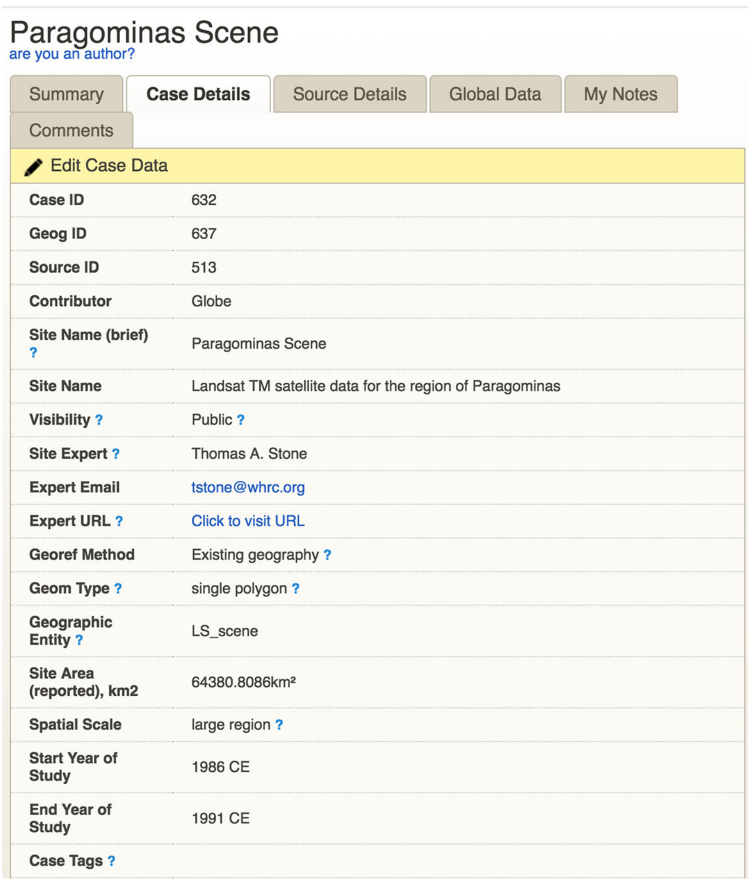

Figure 8. Site expert contact information as represented in GLOBE.

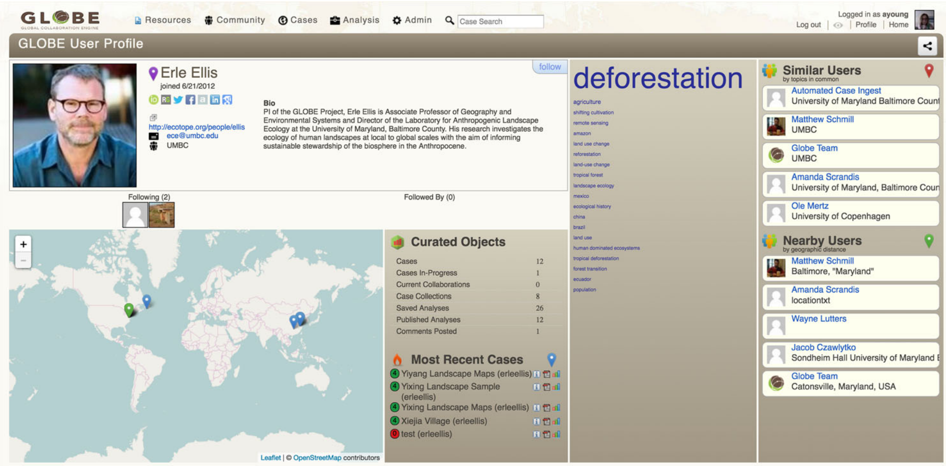

Figure 9. GLOBE user profile page illustrating new social network features.

information about their cases. Near-term evolution of the system will include a means to search for users, not just cases, based on keywords or specific global variables. Longer-term plans consist of developing greater social networking capabilities that will allow GLOBE users to identify and connect with friends-of-friends-of-friends rather than cold-calling case study authors based on their journal articles.

6.4. Enabling the exploration of representativeness

Although LSS researchers frequently use data visualisation tools to assess the spatial distribution of their cases, an inherent geographical bias in case study data means that the sample of available cases on any given phenomenon may not have been conducted at random (Martin et al. 2012) and therefore, despite their best efforts, it may not be possible to achieve representativeness with only published case study data. GLOBE was developed in part to address this issue. The system uses global datasets matched against case study geographies to help researchers determine which variables (e.g., population density or percentage tree coverage) are relevant for their study. For each variable of interest, a user can use the system to visually check how well their cases fit within the selected variable space. Once the user has determined the set of variables relevant for her study, she can run a representativeness analysis in GLOBE to assess whether her collection of cases accurately captures the variation observed globally (Magliocca et al. 2013). A binning visualization illustrates which variable ranges are over or underrepresented. For areas that are underrepresented, possible cases to include are presented to the user (*Figure 10*). This process also helps researchers identify overrepresented or underrepresented areas of the world, which could help remediate case study production biases and direct future data collection efforts.

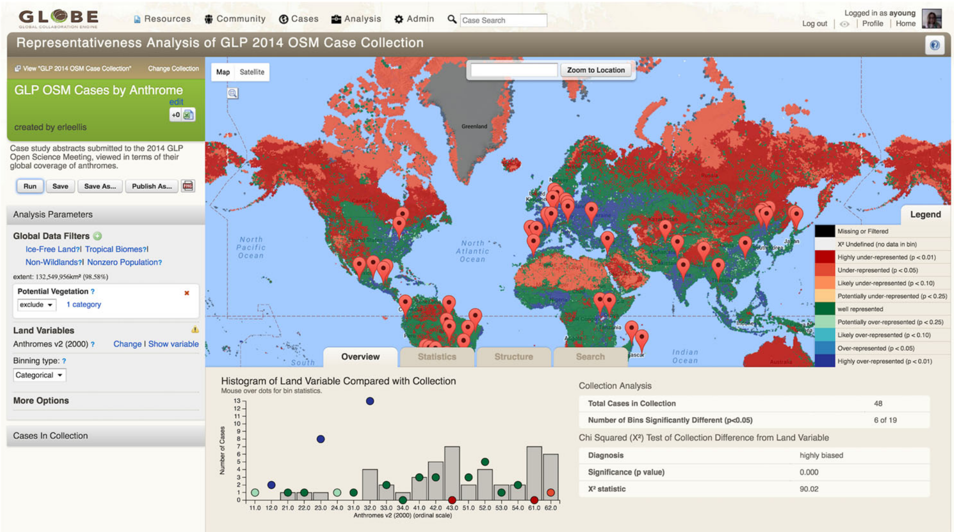

Figure 10. An example of a representativeness analysis in GLOBE.

7. Discussion

Studies of infrastructure have traditionally been post-hoc analyses of emergent phenomena. In this paper, we examined a young interdiscipline, Land System Science, in the active process of infrastructuring. We were particularly fortunate to happen upon an open and reflective scientific community at a key inflection point in its development. We focused specifically on their co-option of meta-study research practice as one way to advance their understanding of the human impact of land use on global environmental change. By studying a still evolving knowledge infrastructure, we have been able to uncover the infrastructural inversion that goes into (re)appropriating an assemblage of tools, techniques, and practices to enable this type of research. In the process, we identified points of infrastructure – breakdowns in scientific practice – brought on by their attempt to use legacy infrastructure to answer new kinds of globally-representative research questions (Bietz and Lee, 2012; Pipek and Wulf 2009). We then used this information to inform system design.

While some of our findings confirm those identified in previous research, we have been able to extend these discussions in interesting and significant ways. Like others, we found that lack of contextual information on how and where data are produced (Faniel and Jacobsen 2010) leads to difficulties in data interpretation (Birnholtz and Bietz 2003) and concerns about data quality (Zimmerman 2007). Understanding how data were initially constructed is a highly collaborative and time-consuming process (Rolland and Lee 2013), which often requires engagement of the original study staff (Lee et al. 2009). Beyond this, the nature of LSS research is such that these reseaachers are forced to take a radically multidisciplinary approach, which involves treating their phenomena of interest as a coupled human and natural system. This means finding ways to mix biogeophysical data with socio-economic data: a much more complex form of cross-disciplinary knowledge production than dealing with each type in isolation. As the gulf between disciplinary traditions widens, it becomes much more difficult for researchers to establish common ground. This added complexity means, for instance, that they are unable to draw on their informal knowledge to try to interpret cases (Zimmerman 2007). This not only affects data interpretation; it also complicates data analysis and integration. For our research community, this has necessitated innovations in traditional practice. These include the use of global data (instead of metadata schemas) to fill in gaps in understanding, geocontextualization to understand the spatial distribution and context of case data, and representativeness to calculate the extent to which a collection of cases constitutes an unbiased, globally-representative sample.

A primary objective of this paper was to identify when and how infrastructural inversion happens and to use this information to inform the co-design of infrastructural tools to scale up current knowledge production processes. To unpack the work involved in meta-study research, we adopted the concept of infrastructuring with its concomitant methods to understand the current knowledge infrastructure and identify avenues for design. This approach was particularly useful for pinpointing the technical and social aspects of meta-study practice that were problematic for this particular community. While we had a general sense that meta-study research was difficult for LSS researchers, we did not know the explicit reasons for this difficulty. By attending to points of infrastructure, we were able to focus our attention on those specific aspects of work practice that did not function properly (i.e., breakdowns) and to observe how these individuals worked to address these challenges through either social or technical means. For instance, this approach helped us to identify the challenges these researchers face in quantifying qualitative data that lacks precision (such as the direction or strength of a change) and how they are trying to use global datasets (to varying degrees of success) to fill in missing contextual details. We were then able to use this information to design infrastructural tools that allow them to more easily find and try out different global variables of interest.

The analytic lens of infrastructural inversion further proved particularly useful in detailing the ways in which the evolving LSS knowledge infrastructure, including GLOBE tools, could be aligned with the significant installed base of existing technical tools and social processes (Parmiggiani et al. 2015). That is, rather than design infrastructure from scratch, which has been deemed an impractical task (Edwards et al. 2009; Hanseth and Lyytinen 2008), infrastructural inversion allowed us to identify aspects of the scientific process that could be best augmented and amplified with new technical tools. It also helped us to identify those manual processes that underlay meta-study practice, such as micro-scale data analysis (i.e., interpreting and coding cases), that cannot easily be technically reduced, and consequently were not directly addressed in the system design. Rather, we built tools to provide pathways to access to help researchers identify those individuals who may be able to assist, for example, with data access or interpretation.

At its most basic level, GLOBE is designed to fit within existing scientific workflows, operate in concert with current methods, and link to the installed base of existing information infrastructure. Our goal was to amplify LSS researchers' existing work practices by providing easier access to resources, such as local and regional case studies and global data sets; enabling new types of analyses and search capabilities; and, reducing meta-study production time. (Recall that the core challenge for LSS research conducting meta-study research was the inability to *scale up* to global comparisons.) Rather than require users to adopt completely new work practices, much of the functionality links to pre-existing information infrastructure, thus serving as a gateway permitting multiple systems to interoperate (Edwards et al. 2009; Edwards et al. 2007). For instance, much of work done in GLOBE could easily be exported for use in other software tools more familiar to the scientist's routine meta-study workflow, such as Excel. GLOBE also does not store copies of the published case studies or datasets, but rather links to their location on the publisher's website.

The evolving nature of the LSS community also meant that we needed to design a tool that was specific enough for the community at large, but flexible enough to support individual needs and changing community dynamics. Recall, that as users change focus or new users arrive on the scene, they bring with them different requirements and infrastructural needs. To persist, new infrastructural tools must meet both contemporary needs and unanticipated future uses (Bietz and Lee 2009; Ribes and Finholt 2009). With this in mind, we looked for commonalities across the different disciplines that could be used in a way similar to a boundary object. This led us to identify their emphasis on geographic sensemaking. That is, these researchers understand data and findings best when it is represented within its geographic context. To support this emphasis on geocontextualization, we therefore designed GLOBE around map-based visualizations.

However, while maps tie this community together, not all LSS researchers visualize data in the same way geographically. We therefore did not lock GLOBE down to one specific type of data representation, but rather enable users to modify how they view their data in the system. Users can examine their data across different views, they can enter their own data into the system to examine its spatial pattern, or they can request different global variable sets be uploaded to the system to see how their data fits within different variable distributions. This better ensures that LSS researchers who are drawn to GLOBE because of its the map-based visualizations do not walk away thinking "I can't use this system; it doesn't support my particular approach."

Fundamental to the concept of infrastructuring as a methodology is an understanding of how users appropriate infrastructure to fit their work activities. Appropriation consists of those activities that users enact to fit a specific technology into a particular work setting (Pipek 2005). In terms of infrastructuring, this involves a process of negotiating work practices so that either the "new innovation becomes manifest in a technology usage or so that the obstruction caused by the breakdown can be eliminated" (Pipek and Syrjänen 2006). While we are still in the active process of system evaluation, we have observed early adopters using GLOBE in unexpected ways. Prominent among these is the desire to upload pre-compiled collections of cases. Recall that we designed GLOBE as an alternative to traditional keyword searching. Our expectation was that LSS researchers would use the system to compile these case collections based on geographic relevance. However, what we have observed thus far is that many researchers who are interested in doing meta-study research tend to place more trust in the collections they have spent their careers compiling and thus are skeptical about abandoning these in favor of new collections assembled in GLOBE. Instead, they are using GLOBE to identify any geographic gaps in their case collections and to gain access to the global data sets. To accommodate this need, GLOBE had to be modified to allow for bulk uploads, rather than just single case entry. These preliminary use cases provide direction for future evaluation of how the system is being appropriated by the community to match their current and evolving needs.

In the event that GLOBE itself does not persist, but rather becomes "tomorrow's quaint and inflexible legacy system" replaced by new, more powerful tools (Edwards et al., 2009, p. 371), the hope is that the concepts and analyses it introduces constructively challenge the norms, beliefs, and practices of the community. That is, the Global Land Programme hopes that the project will help promote new standards for producing knowledge within LSS. From the onset, one of the long-term visions of GLOBE is that it will inspire systemic change in LSS by encouraging the community to consider the global relevance of their work. This includes establishing the expectation that all meta-studies should be evaluated in terms of their global representativeness; that is, the extent to which a collection of cases represents an unbiased sample. Additionally, it should encourage all case study researchers to consider the global significance of their fieldwork by placing their study within its global context. For this to happen, representativeness must become an "obligatory passage point" for all LSS meta-study researchers (Callon 1986). The role and importance of representativeness is currently being debated within the community, with many suggesting that it should become a new disciplinary standard. As an example, a team of LSS researchers recently conducted a representativeness analysis to discredit a major scientific paper for overreaching its claims about local biodiversity change (Gonzalez et al. 2016). The GLP is also hosting the lead GLOBE architect at its international program office in Bern, Switzerland to oversee this process. We are currently undertaking a multiyear investigation of the GLOBE system-in-use to see if and how GLOBE inspires this fundamental shift in LSS practice, including the ways in which knowledge is produced, results are evaluated, and evidence is presented.

It is important to recognize that designing infrastructural tools not only creates new functionality, but also affords possibilities for and constraints on action, ways of looking at the world, modes through which we can relate, and the types of knowledges we can construct. For instance, while cyberinfrastructure enhances the scientific process by improving data collection and management, automating data quality control, and expediting communication (Newman et al. 2012), it also requires scientists and information managers to change their data collection and curation practices to match institutional standards (Karasti et al. 2006; Ribes and Lee 2010; Moore and Karvonen, 2008). As a consequence, our infrastructure choices inevitably advance the interests of some while actively harming the prospects of others (Edwards et al. 2013; Jackson et al. 2007). These causalities are the "orphans of infrastructure"; individuals, groups, and processes that fit uneasily, if at all, within the new infrastructure paradigm (Star and Ruhleder, 1996).

Our design choices can also impact the shape and possibilities of knowledge in general. By enabling new ways of knowing and new forms of knowledge production, cyberinfrastructure often does more than just enhance current practices; it often causes a reworking and reordering of traditional knowledge practices, resulting in the loss or devaluing of certain classes of questions, phenomena and ways of knowing (Edwards et al. 2013). The introduction of sensor networks, for example, had an impact on ecology with fewer researchers engaging in actual fieldwork (Mayernik et al. 2013). In light of this, we are also interested in understanding the impact of these changes on the larger LSS community, particularly in terms of those research practices that are pushed to the margins and ultimately devalued.

8. Conclusions

In this paper, we have highlighted critical moments in the evolving Land System Science knowledge infrastructure. These have focused on systematically understanding the ways in which land systems act as causes and consequences of global change and how this can foster solutions for sustainability. We have foregrounded their adoption of meta studies to drive new discoveries through the synthesis of cases across radically different sources. We traced the infrastructural inversion that goes into the appropriation of legacy information technology to address new research questions and methods. Through the unpacking of four cases we have identified points of infrastructure - breakdowns in scientific practice due to mismatches in technical functionality and desired work practices. These insights helped drive the development of new tools to support these practices.

Recall the unique value of examining emerging scientific disciplines such as LSS. It is precisely during these formative years that key tensions and resolutions become visible. Through our study of infrastructure-in-the-making, we demonstrate how design can play a critical role in engaging a scientific community in active reflection on its infrastructural breakdowns. In particular, we illustrate how design enhances these discussions by instantiating key ideas in computational artefacts for uptake into current scientific practice and debate. In this process of actively creating, rather than relying on historical accounts of infrastructural development, we also argue that researchers may be able to gain new insights into the processes by which knowledge infrastructures come into being and stabilize.

Acknowledgements

We would like to thank the LSS community and the Global Land Programme for their support of this research. We would also like to thank our fellow GLOBE team members. This research has been supported by NSF award #115210 "GLOBE: Evolving New Global Workflows for Land Change Science."

References

Baker, Karen S., Geoffrey C. Bowker, and Helena Karasti (2002). Designing an infrastructure for heterogeneity in ecosystem data, collaborators and organizations. In *Proceedings of the annual National Conference on digital government research, Los Angeles, California, USA, 12-22 May 2002*. Digital Government Society of North America, pp. 1–4.

Baker, Karen S., Steven J. Jackson, and Jerome R. Wanetick (2005a). Strategies supporting heterogeneous data and interdisciplinary collaboration: towards an ocean informatics environment. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS), 03–06 Jan 2005*. New York: ACM press, p. 219.2.

Baker, Karen S., David Ribes, Florence Millerand, and Geoffrey Bowker (2005b). Interoperability strategies for scientific cyberinfrastructure: research and practice. In *Proceedings of the American Society for Information Science and Technology,* vol. 42, no. 1.

Bietz, Matthew and Charlotte P. Lee (2012) Adapting cyberinfrastructure to new science: Tensions and strategies. Proceedings of the 2012 iConference. New York: ACM press, pp. 183-190.

Bietz, Matthew J., and Charlotte P. Lee (2009). Collaboration in Metagenomics: sequence databases and the Organization of Scientific Work. In I. Wagner, H. Tellioğlu, E. Balka, C. Simone, & L. Ciolfi (Eds.): *ESCSW 2009. Proceedings of the European Conference on Computer Supported Cooperative Work, Vienna, 7–11 Sept 2009*. London: Springer. pp. 243–262.

Birnholtz, Jeremy P., and Matthew J. Bietz (2003). Data at work: Supporting sharing in science and engineering. In *GROUP 2003. Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work*. New York: ACM Press. pp. 339–348.

Blake, Catherine, and Wanda Pratt (2006a). Collaborative information synthesis I: a model of information behavior of scientists in medicine and public health. *Journal of the American Society for Information Science and Technology*, vol. 57, no. 13, pp. 1740–1749.

Blake, Catherine, and Wanda Pratt (2006b). Collaborative information synthesis II: recommendations for information systems to support synthesis activities. *Journal of the American Society for Information Science and Technology*, vol. 57, no. 14, pp. 1888–1895.

Borgman, Christine L., Jillian C. Wallis, and Noel Enyedy (2007). Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*, vol. 7, no. 1–2, pp. 17–30.

Borgman, Christine L., Geoffrey C. Bowker, Thomas A. Finholt, and Jillian C. Wallis (2009). Towards a virtual Organization for Data Cyberinfrastructure. In *Proceedings of the 9th ACM/IEEE CS Joint Conference on Digital Libraries*. New York: ACM Press. pp. 353–356.

Bowker, Geoffrey C (1994). *Science on the run: Information management and industrial geophysics at Schlumberger 1920–1940*. Cambridge, Mass.: MIT Press.

Bowker, Geoffrey C., and Susan Leigh Star (1999). *Sorting things out: Classification and its consequences*. Cambridge, Mass.: MIT Press.

Brown, Daniel G., Peter H. Verburg, Pontius Robert G. Jr., and Mark D. Lange (2013). Opportunities to improve impact, integration, and evaluation of land change models. *Current Opinion in Environmental Sustainability*, vol. 5, no. 5, pp. 452–457.

Callon, Michel (1986). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fisherman of St. Brieuc Bay. In J. Law (Ed.): *Power, Action and Belief. A new Sociology of Knowledge?* London: Routledge Sociological Review Monograph, pp. 196–229.

Charmaz, Kathy (2014). *Constructing grounded theory (Second Edition)*. Thousand Oaks: SAGE Publications.

Clarke, Adele E. (2005). *Situational analysis: Grounded theory after the postmodern turn.* Thousand Oaks: SAGE Publications.

Claus Bossen, and Randi Markussen (2010). Infrastructuring and ordering devices in health care: medication plans and practices on a hospital Ward. *Computer Supported Cooperative Work (CSCW)*, vol. 19, no. 6, pp. 615–637.

Edwards, Paul N.; Geoffrey Bowker; Steven Jackson; and Robin Williams (2009). Introduction: An Agenda for Infrastructure Studies. *Journal of the Association for Information Systems*, vol. 10, no. 5, 364-374.

Edward J. Hackett, John N. Parker, David Conz, Diana Rhoten, and Andrew Parker (2008). Ecology Transformed: The National Center for Ecological Analysis and Synthesis and the Changing Patterns of Ecological Research. In G.M. Olson, A. Zimmerman, and N. Bos (Eds): *Scientific Collaboration on the Internet*. Cambridge, Mass.: MIT Press, pp. 277–296.

Edwards, Paul N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. Cambridge, Mass.: MIT Press.

Edwards, Paul N.; Steven J. Jackson; Geoffrey C. Bowker; and Cory P. Knobel (2007). *Understanding Infrastructure: Dynamics, Tensions, and Design.* Retrieved from https://deepblue.lib.umich.edu/handle/2027.42/49353 .

Edwards, Paul N.; Matthew D. Mayernik; Archer Batcheller; Geoffrey Bowker; and Christine Borgman (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*. vol. 41, no. 5, pp. 1–24.

Edwards, Paul N.; Steven J. Jackson; Melissa K. Chalmers; Geoffrey C. Bowker; Christine L. Borgman; David Ribes; Matthew Burton; and Scout Calvert (2013). *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges (Working Paper).* Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/97552 .

Ellis, Erle C., and Navin Ramankutty (2007). Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, vol. 6, no. 8, pp. 439–447.

Ellis, Erle C., Timothy Finin, Wayne G. Lutters, Timothy Oates, and Penny Rheingans (2011). GLOBE: evolving new global workflows for land change science. NSF Grant 115210.

Faniel, Ixchel M., and Trond E. Jacobsen (2010). Reusing scientific data: how earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work (CSCW)*, vol. 19, no. 3–4, pp. 355–375.

Faniel Ixchel M., Eric Kansa, Sarah Whitcher Kansa, Julianna Barrera-Gomez, and Elizabeth Yakel (2013). The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse. In *JCDL 2013: Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, Indianapolis, 22–26 July 2013.* New York: ACM Press, pp. 295–304.

Gonzalez, Andrew, Bradley J. Cardinale, Ginger R.H. Allington, Jarrett Byrnes, K. Arthur Endsley, Daniel G. Brown, David U. Hooper, Forest Isbell, Mary I. O'Connor, and Michel Loreau (2016). Estimating local biodiversity change: a critique of papers claiming no loss of local diversity. *Ecology*, vol. 97, no. 8, pp. 1949–1960.

Gooding, David (1982). Putting agency back into experiment. In A. Pickering (Ed.): *Science as practice and culture*. Chicago: University of Chicago Press, pp. 65–112.

Hey, Tony, Stuart Tansley, and Kristen Tolle (Eds.) (2009). *The fourth paradigm: Data-intensive scientific discovery.* Richmond: Microsoft Research.

Hanseth, Ole, and Kalle Lyytinen (2008). Theorizing about the Design of Information Infrastructures: Design Kernel Theories and Principles. *Working Papers on Information Systems*, vol. 4, no. 12, pp. 207-241.

Jackson, Steven J. (2014). Rethinking repair. In T. Gillespie, P. Boczkowski, & K. Foot (Eds): *Media technologies: Essays on communication, materiality, and society*. Cambridge, Mass.: MIT Press.

Jackson, Steven J., Paul N. Edwards, Geoffrey C. Bowker, and Cory P. Knobel (2007). Understanding Infrastructure: History, heuristics, and cyberinfrastructure policy. *First Monday*, vol. 12, no. 6.

Karasti, Helana, Karen S. Baker, and Florence Millerand (2010). Infrastructure time: long-term matters in collaborative development. *Computer Supported Cooperative Work (CSCW)*, vol. 19, no, 304, pp. 377–415.

Karasti, Helena; and Karen S. Baker (2004). Infrastructuring for the long-term: ecological information management. In *HICSS. Proceedings for the 37th Annual Hawaii International Conference on System Sciences,* p. 10.

Karasti, Helena, and Anna-Liisa Syrjänen (2004). Artful infrastructuring in two cases of community PD. In *PDC 2004: Proceedings of the eighth conference on participatory design: Artful integration: Interweaving Media, Materials and Practices*. New York: ACM Press. pp. 20–30.

Karasti, Helena, Karen S. Baker, and Eija Halkola (2006). Enriching the notion of data curation in E-science: data managing and information infrastructuring in the long term ecological research (LTER) network. *Computer Supported Cooperative Work (CSCW)*, vol. 15, no. 4, pp. 321–358.

Karl, Jason W., Jeffrey E. Herrick, Robert S. Unnasch, Jeffrey K. Gillan, Erle C. Ellis, Wayne G. Lutters, and Laura J. Martin (2013). Discovering ecologically relevant knowledge from published studies through Geosemantic searching. *BioScience*, vol. 63, no. 8, pp. 674–682.

Knorr Cetina, Karen (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge: Harvard University Press.

Kummerle, Tobias, Karlheinz Erb, Patrick Meyfroidt, Daniel Muller, Peter H. Verburg, Stephan Estel, Helmut Habel, Patrick Hostert, Martin R. Jepsen, Thomas Kastner, Christian Levers, Marcus Lindner, Christoph Plutzer, Pieter Johannes Verkerk, Emma H van der Zanden, and Anette Reenberg (2013). Challenges and opportunities in mapping land use intensity globally. *Current Opinion in Environmental Sustainability*, vol 5, no. 5, pp. 484–493.

Latour, Bruno (1999). *Pandora's hope: Essays on the reality of science studies*. Cambridge, Mass.: Harvard University Press.

Latour Bruno, and Steve Woolgar (1979). *Laboratory life: the construction of scientific facts*. Princeton: Princeton University Press.

Lee, Jae W., Jianting Zhang, Anne S. Zimmerman, and Angelo Lucia (2009). DataNet: an emerging cyberinfrastructure for sharing, reusing and preserving digital data for scientific discovery and learning. *AIChE Journal*, vol. 55, no. 11, pp. 2757–2764.

Liberati, Alessandro, Douglas G. Altman, Jennifer Tetzlaff, Cynthia Mulrow, Peter C. Gøtzsche, John P. A. Ioannidis, Mike Clark, P.J. Devereaux, Jos Kleijnen, and David Moher (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Medicine*, vol. 6, no. 7.

Lutters, Wayne G., and Susan Winter (2012). Virtual Organizations. In W. Bainbridge (Ed.): *Leadership in science and technology: A reference handbook*. Thousand Oaks: Sage Publications, pp. 334–343.

Magliocca, Nicholas R., Thomas K. Rudel, Peter H. Verberg, William J. McConnell, Ole Mertz, Katharina Gerstner, Andreas Heinimann, and Erle C. Ellis (2014). Synthesis in land change science: methodological patterns, challenges, and guidelines. *Regional Environmental Change*, vol. 15, no. 2, pp. 1–16.

Margulies, Jared N., Nicholas Magliocca, Matthew Schmill, and Erle C. Ellis (2016). Ambiguous geographies: connecting case study knowledge with global change science. *Annals of the Association of American Geographers*, vol. 106, no. 3, pp. 572–596.

Martin, Laura J., Bernd Blossey, and Erle C. Ellis (2012). Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*, vol. 10, no. 4, pp. 195–201.

Mayernik, Matthew S., Jillian C. Wallis, and Christine L. Borgman (2013). Unearthing the infrastructure: humans and sensors in field-based scientific research. *Computer Supported Cooperative Work (CSCW)*, vol. 22, no. 1, pp. 65–101.

Monterio, Eric, Neal Pollock, Ole Hanseth, and Robin Williams (2013). From artefacts to infrastructures. *Computer Supported Cooperative Work (CSCW)*, vol. 22, no. 4–6, pp. 575–607.

Moore, Steven A. and Andrew Karvonen (2008) Sustainable Architecture in Context: STS and Design Thinking. Science & Technology Studies, vol. 21, no.1, pp. 29–46.

Newman, Greg, Andrea Wiggins, Alycia Crall, Eric Graham, Sarah Newman, and Kevin Crowston (2012). The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, vol. 10, no. 6, pp. 298–304.

Nicholas R. Magliocca, Erle C. Ellis, Timothy Oates, and Matt Schmill (2013). Contextualizing the Global Relevance of Local Land Change Observations. *arXiv:1307.6889 [physics, Stat]*. Retrieved from http://arxiv.org/abs/1307.6889.

Nielsen, Michael (2011). *Reinventing discovery: The new era of networked science*. Princeton: Princeton University Press.

Parmiggiani, Elena, Eric Monteiro, and Vidar Hepsø (2015). The digital coral: infrastructuring environmental monitoring. *Computer Supported Cooperative Work (CSCW)*, vol. 24, no. 5, pp. 423–460.

Pine, Kathleen H., and Max Liboiron (2015). The politics of measurement and action. In *CHI 2015: Proceedings of the Conference on Human Factors in Computing Systems*. New York: ACM Press, pp. 3147–3156.

Pipek, Volkmar (2005). Negotiating Infrastructure: Supporting the Appropriation of Collaborative Software. *International Reports of Socio-Informatics (IRSI)*, vol. 2, no. 1, p. 44.

Pipek, Volkmar; and Anna-Liisa Syrjänen (2006). Infrastructuring as Capturing In-Situ Design. In *Proceedings of the 7th Mediterranean Conference on Information Systems, Venice*.

Pipek, Volkmar and Volker Wulf (2009). Infrastructuring: Toward an Integrated Perspective on the Design and Use of Information Technology. *Journal of the Association for Information Systems*, vol. 10, no. 5, pp. 447-473.

Ribes, David, and Thomas Finholt (2009). The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems*, vol. 10, no. 5, pp. 375-398.

Ribes, David, and Charlotte P. Lee (2010). Sociotechnical studies of cyberinfrastructure and e-research: current themes and future trajectories. *Computer Supported Cooperative Work (CSCW)*, vol. 19, no. 3–4, pp. 231–244.

Rolland, Betsy, and Charlotte P. Lee (2013). Beyond trust and reliability: reusing data in collaborative cancer epidemiology research. In *CSCW 2013. Proceedings of the Conference on Computer Supported Cooperative Work*. New York: ACM Press, pp. 235–444.

Rudel, Thomas K. (2008). Meta-analyses of case studies: a method for studying regional and global environmental change. *Global Environmental Change*, vol. 18, no. 1, pp. 18–25.

Star, Susan Leigh (1985). Scientific work and uncertainty. *Social Studies of Science*, vol. 15, no. 3, pp. 391–427.

Star, Susan Leigh (1999). The ethnography of infrastructure. *American Behavioral Scientist*, vol. 43, no. 3, pp. 377–391.

Star, Susan Leigh and Geoffrey C. Bowker (2002). How to Infrastructure? In L. A. Lievrouw & S. Livingstone (Eds.): *Handbook of New Media: Social Shaping and Social Consequences of ICTs*. SAGE, pp. 151–162.

Star, Susan Leigh, and Griesemer James R. (1989). Institutional ecology, 'translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, vol. 19, no. 3, pp. 387–420.

Star, Susan Leigh, and Karen Ruhleder (1996). Steps toward an ecology of infrastructure: design and access for large information spaces. *Information Systems Research*, vol. 7, no. 1, pp. 111–134.

Turner, Billy Lee II., Eric F. Lambin, and Anette Reenberg (2007). The emergence of land change science for global environmental change and sustainability. In *Proceedings of the National Academy of Science*, vol. 104, no. 52, pp. 20666–20671.

Verburg, Peter H., Karl-Heinz Erb, Ole Mertz, and Giovana Espindola (2013). Land system science: between global challenges and local realities. *Current Opinion in Environmental Sustainability*, vol. 5, no. 5, pp. 433–437.

Yarmey, Lynn, and Karen S. Baker (2013). Towards standardization. A participatory framework for scientific standard-making. *International Journal of Digital Curation*, vol. 8, no. 1, pp. 157–172.

Young, Alyson L., and Wayne G. Lutters (2015). (Re)defining land change science through synthetic research practices. In *CSCW 2015. Proceedings of the Conference on Computer Supported Cooperative Work and Social Computing*. New York: ACM Press, pp. 431–442.

Zimmerman, Ann (2007). Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries*, vol. 7, no. 1, pp. 5–16.