August 29, 2014

GEOSHARE: Geospatial Open Source Hosting of Agriculture, Resource & Environmental Data for Discovery and Decision Making

By

Thomas Hertel and Nelson Villoria (Purdue University)[1]

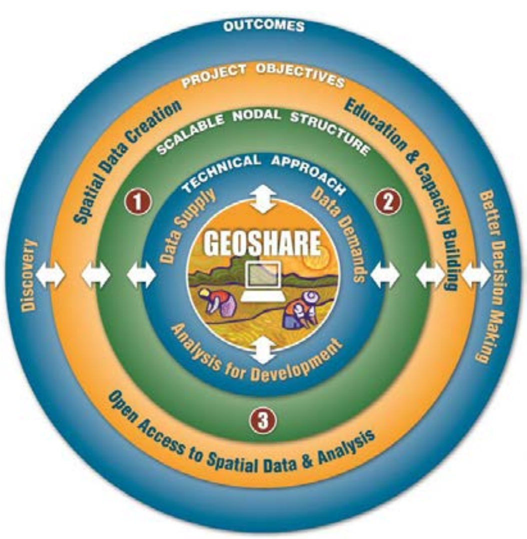
In collaboration with

Noah Diffenbaugh (Stanford University), Glenn Hyman (CIAT), James W. Jones (University of Florida), Jawoo Koo (IFPRI), Andrew Nelson (IRRI), Navin Ramankutty (McGill University), Stefan Siebert (Bonn University) and Carol Song (Purdue)

[1] The authors acknowledge prior support from the UK Foresight Project, UK-DFID, UK-DEFRA and ERS-USDA, ILSI-Research Foundation/CIMSANS, as well as seed funding from the Purdue Global Policy Research Institute and the Purdue Climate Change Research Center. Paul Hendley, Gill Shepherd, Dave Gustafson and Lowell Hardin provided valuable comments on an earlier draft. This document was prepared during Tom Hertel's Residency at the Rockefeller Foundation's Bellagio Center: July, 2013 and he would like to thank his fellow residents for valuable discussions on this topic.

Motivation and Overview

Feeding 9 billion people in the face of a changing climate, while preserving the environment and eliminating extreme poverty, is one of the key development grand challenges facing the world as we look forward to 2050. Households surviving on less than $1.25/day are still predominately rural and remain disproportionately dependent on agriculture and forests for their livelihood. Yet, the data currently available to understand how global and local phenomena affect the agriculture-environment-poverty-trade nexus are insufficient to advance discovery and promote effective decision making (Hertel et al. 2010). Most geospatial datasets for agriculture are regional or national in scope, mutually incompatible, one-time efforts, and, if they are publicly available, specialized knowledge and costly software licenses may significantly limit access. This lack of readily accessible, up-to-date information has greatly inhibited the ability of scientists, practitioners and policy makers to address the socio-economic and environmental impacts of contemporary policy issues related to environmental quality, poverty reduction and the long run sustainability of the world food system. An accurate assessment of these policies, as well as effective resource allocation to solve development problems, requires knowledge of local conditions for a number of factors affecting sustainable agricultural productivity, including: soil, water, topography, access to fertilizer and other agricultural inputs, climate, and current farming practices. At the same time, these local decisions are being made within an international context, for which global analysis is required to capture the drivers of change as well as to avoid misleading conclusions.



GEOSHARE stands for Geospatial Open Source Hosting of Agriculture, Resource and Environmental Data for Discovery and Decision-making. **GEOSHARE's mission is to develop and maintain a freely available, global, spatially explicit database on agriculture, natural resources, and the environment accompanied by analysis tools and training programs for scientists, decision makers, and development practitioners.** GEOSHARE will become the focal point for a network that enhances international cooperation across the agriculture-environment-poverty nexus. Its comprehensive data network will involve geospatial analysis with the most qualified scientists in the world, to meet the demands of data consumers and decision makers. The GEOSHARE concept has been validated by two dozen peer-reviewers (Hertel et al. 2010) and funding for several proof of concept initiatives has been provided by national and international organizations (UK-Foresight, UK-DFID, UK-DEFRA, USDA-ERS and CCAFS), along with two foundations (CIMSANS and Rockefeller), as well as Purdue University's Global Policy Research Institute, the Climate Change Research Center (PCCRC) and the Rosen Center for Advanced Computing.

GEOSHARE is different from existing activities and programs in a number of ways.

- Firstly, it has both a global and a local impact.
- GEOSHARE emphasizes data production, consistency, validation, and interoperability using state-of-the art methods. In addition, unlike other open data portals, GEOSHARE uniquely enables the management of "workflows" (a term adopted from computer science).
- GEOSHARE seeks to generate time-series of geospatial data to support analysis of far-reaching development challenges as well as validation of analysis tools.
- HUBZero technology sets GEOSHARE apart from other data portals in important ways. HUBZero allows non-expert users, domain scientists and students to rapidly develop online applications which they can publish and share with others who can launch computation on state of the art cyber-infrastructure – all done within the environment of a simple web browser, without downloading and installing any software or data bases. Its cyber-infrastructure facilitates technology transfer, management of workflows, and builds capacity for stakeholders in developing and developed country institutions by engaging university students and scientists. HUBZero encourages open-source access to data and associated data processing and analysis tools, thereby offering users analytical and exploratory capabilities beyond spatial zooming or automated summary statistics.
- GEOSHARE, is complementary to other global data efforts relating to the agricultural sector (e.g., *S-World*), which can be linked to GEOSHARE allowing users to combine these data with other relevant information archived in GEOSHARE.
- GEOSHARE makes use of site-specific farm-level data collection and monitoring efforts such as those proposed by Sachs *et al.* (Sachs et al. 2010) in selected sites in the world, as well as crowd-sourced data as documented in Fritz et al. (2009) and those planned under the new partnership between the World Bank and Google (Anstey 2012).
- As it develops in the future, GEOSHARE has the potential to serve the nutrient management and hydrologic modeling communities as well.
- GEOSHARE is an important source of inputs for other modeling exercises including the biophysical and economic models used in the context of the Agricultural Model Inter-comparison and Improvement Project (AgMIP: www.agmip.org) such as well as the family of models based on the Global Trade Analysis Project (GTAP: www.gtap.org) data base.
- GEOSHARE can provide a translational bridge between communities, such as the climate impacts modeling groups and those seeking to analyze the consequences of such impacts (https://geoshareproject.org/tools/cropdatatool/)
- GEOSHARE has been named as the preferred source of open data and preferred workflow environment for an assessment of "sustainable nutrition security" to be performed by a new public-private partnership recently announced by CIMSANS and AgMIP.

This document describes key design features of GEOSHARE, beginning with a description of how it operates, followed by discussion of the proposed governance structure as well as its role in the global institutional landscape.

BOX 1. HUBZERO TECHNOLOGY: NETWORKING IN THE CLOUDS

The cyber-infrastructure for GEOSHARE is built on the HUBzero® Platform for Scientific Collaboration (McLennan and Kennell 2010). A hub combines unique middleware with Web 2.0 functionality, providing a platform that is much more powerful than an ordinary website. Users are not only able to network and share information, but they can also create, publish and access interactive visualization tools powered by a computer cluster built to render computer-generated imagery, as well as facilitate online collaboration supporting research, education and outreach.

HUBzero was created by the NSF-funded Network for Computational Nanotechnology starting in 2002 with the development of the first hub, named nanoHUB.org (Klimeck et al. 2008). Since then, usage of nanoHUB.org has grown exponentially. In 2011, it served 400,000 visitors from 172 countries worldwide. In 2007, HUBzero was spun off from nanoHUB.org as a separate project and software package to power newly created hubs. It was released as open source software in April 2010, and today, HUBzero supports more than 40 hubs with a combined audience of more than 650,000 visitors each year. GEOSHARE is one of these hubs. In the pilot project we have created several workflows on the hub – one which culminates in reconciled, gridded cropping data, and one accessing more than 36,000 global climate impact gridded outputs, selects and aggregates these into a usable format for the economic modeling community. Both of these share a common dependence on geospatial source data. The ongoing development of geospatial data support within the HUBZero environment is a key to the success of GEOSHARE (Zhao et al. 2011), and these advances will be greatly advanced through a recently approved NSF-DIBBS project led by Carol Song at Purdue University, and aimed at improving the GIS capabilities of HUBZero.

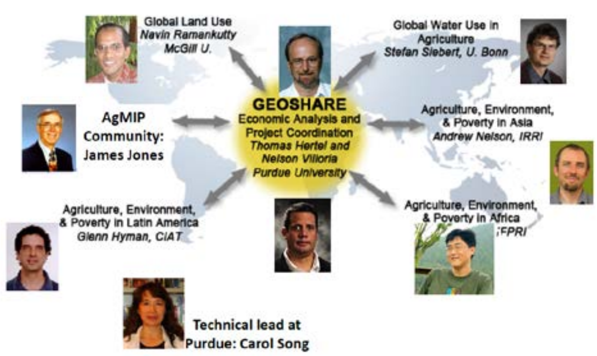
Design of GEOSHARE

The institutional design of GEOSHARE has evolved from its initial conception (Hertel et al. 2010), aided by a pilot project, with funding from UK-DFID, UK-DEFRA, USDA, CIMSANS and Purdue University. This section describes the design of GEOSHARE from several different perspectives.

GEOSHARE as a collection of 'nodes': A key feature of GEOSHARE is its reliance on a nodal structure (Figure 2). Given the fast-paced development of geospatial data and analysis tools, no single 'Center' can manage to be on top of all the aspects of the individual fields pertinent to the assessment of long run food security, environmental sustainability and poverty reduction. Furthermore, individual domain experts are no longer willing to simply send their data into a large 'Data Center' to be re-packaged and sent out as a "unified" data product. And even if they were willing to do so, without serious coordination of the effort up front, the bits and pieces brought together such a Data Center would not be interoperable (e.g., different country boundaries, differing assumptions on land cover, etc.). Instead, a coordinated, decentralized approach will be used -- supported by unified standards, common data bases and set of spatial modeling tools.

An illustration of this approach, based on a modified version of the original GEOSHARE pilot project proposal to UK-DFID, is portrayed in Figure 2. Here, there are six nodes, three of which are global in scope and three of which have a regional emphasis. Two of the global nodes emphasize domain-specific expertise in geospatial data and analysis – in this case, land cover and land use (McGill) and irrigated agriculture (Bonn). Each of these nodes is led by a scientist who is a recognized leader in the field and who is actively publishing geospatial data and related analyses. In each case, the node leader is an individual who firmly believes in the sharing of geospatial information, and who has a track record in this area. For example, Navin Ramankutty, along with his collaborators, has published definitive global data sets on land cover (Ramankutty et al. 2008) and land use (Monfreda, Ramankutty, and Foley 2008) which have formed the basis for dozens of publications in this field. Siebert and collaborators (Portmann, Siebert, and Döll 2010) have published a global data set (MIRCA 2000) which is widely used for analysis of irrigated agriculture, and Siebert is the key resource person for development of the UN-FAO AQUASTAT data base covering the world's irrigated croplands. These two nodes have been partially funded in the GEOSHARE pilot project.

Figure 2. GEOSHARE as a Collection of Nodes



The other global node in Figure 2, AgMIP, represents a user community. This particular community focuses on the inter-comparison and improvement of agricultural models. As such, it is potentially an important consumer of GEOSHARE data. And, as we will see below, it is also a supplier of intermediate model results which may be further processed and made available to the public within the open source environment provided by GEOSHARE. The interplay between AgMIP and the other global nodes will become clearer below when we discuss the workflow.

The Project Coordination node located at Purdue plays several roles. First of all, this is where the cyber-infrastructure is developed (Box 1). This is a critical function in GEOSHARE, as it is through this cyber-infrastructure that the various nodes communicate, exchange information and publish their findings for wider use. The second role played by the Purdue node is that of administration and coordination. In addition, the Purdue node provides a link to the GTAP user community of global economic modelers.

There are three regional nodes portrayed in Figure 2: Asia, Africa and Latin America. The idea of including regional nodes in GEOSHARE was originally proposed by Stanley Wood, then Co-PI of HarvestChoice (a Gates Foundation-funded initiative with a focus on Sub-Saharan Africa) and a founding member of GEOSHARE. He emphasized that regional expertise was critical for 'ground-truthing' the global data bases, as well as supplementing them with additional detail obtained from regional institutions. In the case of OECD countries, we expect that much of the regional detail will have already found its way into the global data sets. Thus, regional nodes for Europe and North America, while important in the longer run, were deemed to be less pressing in the pilot project. However, in the case of many developing countries, this regional information is less comprehensive and is often not yet fully integrated into the global data bases.

Thus there is a more important role for regional nodes. Also, given the focus of the project on poverty impacts in developing countries there is additional justification for such regional nodes.

Thanks to Stanley Wood's early leadership, GEOSHARE was fortunate to engage three regional nodes. They are each led by a prominent geospatial scientist within the Consultative Group for International Agricultural Research (CGIAR). In the case of Africa, the regional node is based at IFPRI, where it is led by Jawoo Koo. In Asia, the regional node is at IRRI and is led by Andrew Nelson. And in Latin America, the regional node is based at CIAT and is led by Glenn Hyman. (In the current GEOSHARE pilot effort, we have only been able to obtain funding for projects with the IFPRI and IRRI nodes.) It is important to note that GEOSHARE does not intend to substitute for the many excellent, regionally focused projects currently underway around the world. Rather, GEOSHARE's unique comparative advantage derives from the interplay amongst these different nodes – the global nodes with one another, and the regional nodes with their additional depth of data access, with the global nodes - and the regional nodes with one another.

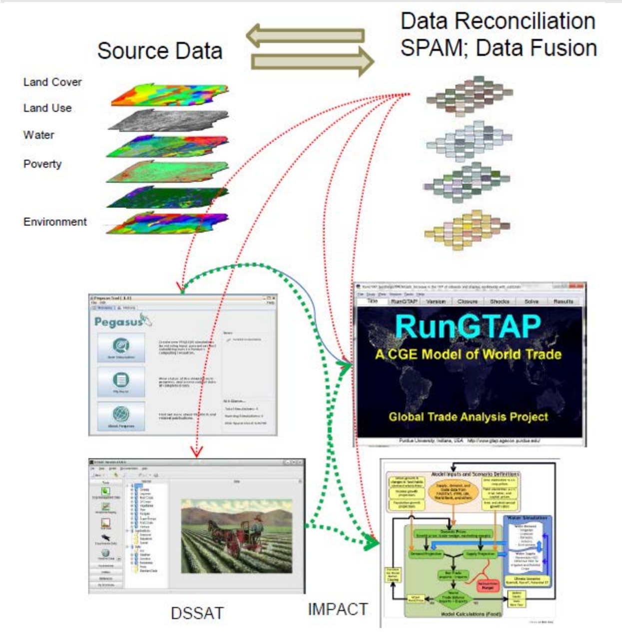
The final set of nodes identified in Figure 2 relate to the user community. Given the workflows being developed in the pilot project, the two most important user communities at this stage include the crop modelers and the global economic modelers. There are many diverse groups representing these interests. However, we have chosen to select representatives from two of the most prominent groups. In the case of the crop modeling community, this node is led by Jim Jones, chief architect of the DSSAT family of crop models and co-PI of the Agriculture Modeling Intercomparison and Improvement Project (AgMIP). In the case of global economic models, the link to the user community is provided by Tom Hertel, founder and executive director of the Global Trade Analysis Project (GTAP). The purpose of these nodes is to communicate to the GEOSHARE project the evolving scientific needs of these user communities. This feature will also permit the addition of user community nodes as the project capabilities expand. (Policy linkages will be provided through the advisory board about which is described in more detail below.)

GEOSHARE as a collection of standards: In order to ensure inter-operability of the GEOSHARE data bases produced by each node, it is essential to have a common set of standards – both for data creation (e.g., a common definition of countries and country borders) – as well as for data documentation (i.e., metadata). Where possible, these standards will be adopted from existing global organizations (e.g. GEOSS). Where new standards are required, they will be drafted and publicized for comment and peer review by the scientific committee of GEOSHARE (see section on governance, below) which will be comprised of the leaders of each GEOSHARE node. Andrew Nelson of IRRI currently chairs the GEOSHARE sub-committee focusing on standards.

Assessing the added value from GEOSHARE – the role of Workflows: The task of improving global geospatial data is potentially quite onerous. Indeed, ever-increasing temporal and spatial resolutions offer infinite possibilities for extension and refinement. Therefore, a central requirement of GEOSHARE is to assess the added value which might be obtained from specific improvements – both to the source data and to the methods for processing these data and utilizing them to inform decisions. In order to support such assessments, we have introduced the concept of a 'workflow' into the GEOSHARE pilot project. Box 2 develops this idea – adopted from computer science -- in more detail. When used in conjunction with the HUBZero technology (Box 1), this permits users to replicate, and proposed improvements to, existing assessments of issues ranging from the impacts of climate change to the consequences of water shortages for irrigated agriculture, food consumption and trade.

BOX 2. Defining Workflows on HUBZero

At the outset, we envisioned GEOSHARE as a set of inter-operable data bases which could be accessed and improved upon by network participants. However, over the course of the pilot project, it has become clear that to be effective, *GEOSHARE must allow such data to be part of larger workflows that serve the needs of specific communities and effectively connects them*. The adjacent figure illustrates the workflow concept in the context of the pilot project. At the upper left hand corner of this figure are the source data sets which are being developed by GEOSHARE nodes. In the case of the pilot project, we are particularly interested in new data sets on irrigated area (contributed by Stefan Siebert at the University of Bonn) and land cover (contributed by Navin Ramankutty of McGill University). In order to be useful in the context of analysis tools such as the PEGASUS crop model, the DSSAT crop model, or the GTAP and IMPACT models of trade and production, these data sets must be integrated with other information, such as data on yields and harvested area, by crop.

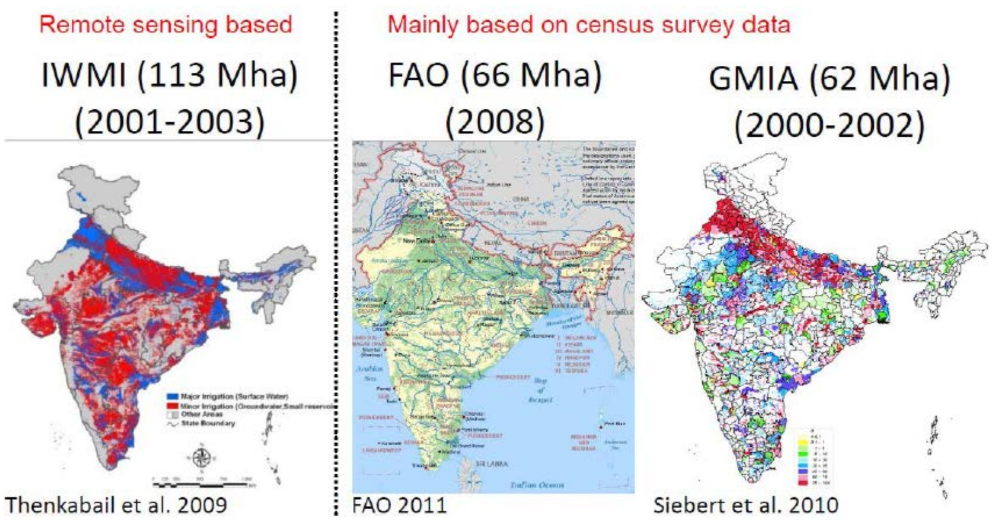


If such integration is to be replicable, it requires a formal framework such as SPAM (see Box 3) which produces global, gridded data for agricultural production at uniform resolution. However, decision makers are primarily interested in the results from the biophysical and economic models shown at the bottom of the figure, and these are run using reconciled data (e.g. coming from SPAM) as an input. Therefore, the results of these models depend critically on the quality of the source data inputs.

The beauty of running such workflows on HUBZero is that each of the inputs employed in generating a particular set of results can be automatically documented and published. Thus, for example, the described workflow might be used to assess the effects of the improved dataset on irrigation in India being developed by Stefan Siebert under the GEOSHARE pilot project by a current irrigation layer and subsequently re-reconciling the data, then running a crop model such as PEGASUS in order to assess the change in climate impact results stemming from the new input data. Alternatively, a new workflow might involve changing the priors in SPAM. This, too, could generate a different set of outcomes. Finally, alternative workflows could arise at the final stage in which a different crop model is used. Once published, each of these workflows can be replicated on the HUB. Results can be compared, and scientific progress can be made.

One example of particular interest in the context of the pilot project, is the role of improved irrigation data for India. Figure 4 shows that there are very significant differences in the estimated total area under irrigation in India, depending on the source of the estimates. Differences in spatial distribution are even more striking. In the case of the International Water Management Institute (IWMI), which relies heavily on remote-sensing, they estimate that there are 113 Mha of irrigated cropland in India. On the other hand, the UN-FAO and the Government of India estimate this figure to be only about half as large as the IWMI figure. This seems like a huge difference. But the question remains -- is it likely to lead to significantly different estimates in support of a particular decision? By running through the entire workflow – from source data to impact assessment -- it is possible to determine how important it is to narrow the range of uncertainty regarding Indian irrigated area. One can imagine that this might be very important in some cases, but less so in others. Based on the policy and investment decisions currently in play, public institutions then can determine which data sets are of highest priority for updating/improvement, and perhaps also whether additional resources should be devoted to the provision of data base infrastructure more generally.

Figure 4. Alternative Estimates of Irrigated Area in India



Source: Produced by Stefan Siebert.

A related advantage of treating GEOSHARE as a set of workflows is that it creates a natural incentive for individuals to contribute new source data. Absent the availability of this workflow ‘in the clouds’, the only incentive for a data producer to upload her/his data is to advance the public good -- while perhaps garnering some additional citations. However, with this workflow in place on the HUB, a contributor of new source data can themselves draw out implications of their improved information. This greatly increases the incentive for contributing new data. Indeed, particularly salient contributions, when run through this workflow, may well change the outcome in influential papers, thereby generating interesting publication possibilities. While the publishing of a new data set alone is often difficult, any paper which builds on new and improved data in order to significantly alter an important published finding is far more likely to garner recognition (and perhaps additional funding) for the author, thereby generating incentives to contribute.

Of course not all new source data added at the top of the workflow in Figure 3 will significantly alter important findings. However, the ability of members of the GEOSHARE community to simply replicate the steps in this workflow will be a major advance over the current state of play in geospatial analysis of agriculture, environment and poverty. After all, replication is the foundation for scientific advance in the face of the massive complexity posed by these problems. Without first being able to replicate the current state-of-play, it is difficult to know whether a given innovation is indeed an improvement. And, over time, the scope for improvement will extend not only to the source data, but also to the spatial allocation methods themselves. Indeed, the HUBZero infrastructure outlined below is ideally suited to facilitating such incremental advances. Because of the open-source environment in which this infrastructure is developed, any of the models run on the HUB can be readily modified. This opens the way to a systematic approach to model testing and improvement.

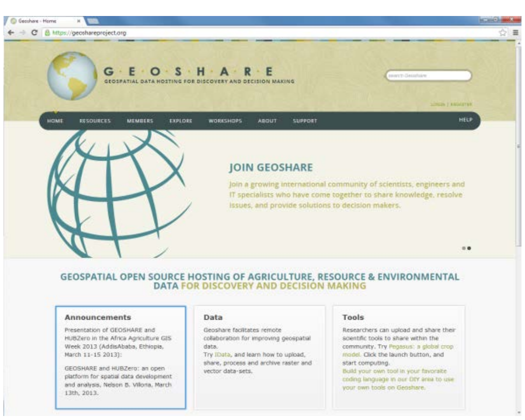
Consider, by way of example, the instances for which independent observations are available at the 10km grid level for a given region. *Here 'ground-truthing' becomes possible*. Performance of the SPAM approach to downscaling of more coarsely defined data down to the 10km level could be compared to the actual observations in this case. Where it performs well, no modifications may be required. However, where it performs poorly, this should lead model developers to look for reasons why it fails, leading to potential improvements which, in turn, could be tested against new observational data. With time, we expect there to be alternative implementations of SPAM, or altogether different spatial allocation tools, running on the HUB. This will facilitate significant improvements in global, gridded data for the analysis of agriculture, environment and poverty issues.

Estimation of the likely impact of climate change on Indian agriculture, and investments needed to facilitate adaptation to such impacts, represents an important area of decision making. We know that the response of rainfed and irrigated agriculture to elevated temperatures will be quite different, due to induced water stress. Might decisions on new investments in adaptation be misleading as a consequence? Our suspicion is that introducing these differing irrigated source data into a workflow like that shown in Figure 3 could lead to a significant difference in climate impacts analyses. However, without running through the entire workflow, we cannot know for sure how much difference this makes in the final result. Source data improvements in some areas may be relatively cheap, yet impactful, while other improvements may be costly and making little difference in the final decision. Distinguishing between these two types of investments would be enormously useful for policy makers.

BOX 3. SPAM: A Tool for the Spatial Allocation of Agricultural Activity

One of the greatest challenges in working with geospatial data derives from the underlying inconsistencies in the different data sets being combined for use in a given workflow. In many cases, the variables are reported at different resolutions – e.g., national, regional, county levels or pixel. In order to be useful in geospatial modeling, these need to be brought to a common resolution. In short, they need to be placed 'on a common grid' so that they can be aggregated to agro-ecological zones, watersheds, etc. This is fundamentally an exercise in 'downscaling' – something which is common in the climate community. In addition to the challenge of down-scaling, the data are often inconsistent – as would be the case, for example, if the area equipped for irrigation of crops in a given grid cell exceeded the total area in agriculture recorded in the crop land cover data base. And perhaps the harvested area data base shows production in a region where the land cover data base shows no cropping. All of these differences must be resolved in order to obtain an analytical data base which can be used in the subsequent biophysical and economic models described in Figure 3. One approach to achieving such consistency is the Spatial Allocation Model (SPAM) (You and Wood 2006), developed at IFPRI.

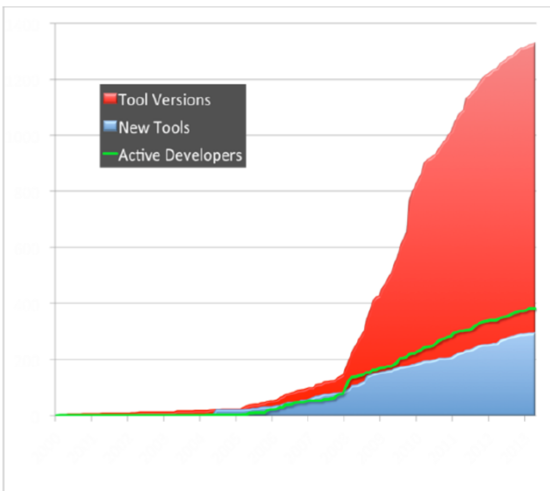
SPAM begins with roughly 500 layers of geospatial source data which have been pre-processed and placed in a common format. These include: crop production statistics, farming system characteristics, satellite-derived data, crop-suitability assessments and population statistics. Before running SPAM some of the irreconcilable differences between data sets much be resolved at a pre-processing stage. After that, SPAM uses the diverse sources of information to build 'priors', which are simply advance conjectures of how the disaggregated data might actually be distributed. For example, in the case of commercial agriculture, these priors are based on a 'revenue-maximization' model of behavior within the grid cell. After forming these priors, SPAM seeks to minimize an objective function which measures the departure of the down-scaled data from the priors, all while respecting the observed totals for variables like harvested area, yields, etc. SPAM was originally developed and tested on Brazil, where it produced promising results, when compared to observed data at the municipality level (You and Wood 2006). It has subsequently been extended to cover the globe, although it is still run one country at a time. In the GEOSHARE pilot project, we have put SPAM 'in the clouds', focusing on Ghana as a case study.



GEOSHARE as Cyber-infrastructure: The workflows outlined above are enabled through the HUBZero infrastructure, based at Purdue University (Box 1) which allows individual researchers to interact, form sub-groups and collaborations, upload data and analysis tools, and create interoperable geospatial data sets for analysis of agriculture and the environment (www.geoshareproject.org). HUBZero allows non-expert users, domain scientists and students to rapidly develop online applications which they can publish and share with others who can launch computation on state of the art cyber-infrastructure – all done within the environment of a simple web browser, *without downloading and installing any software or data bases*. HUBZero also bundles social networking features that support scientific communication and collaborations (WIKIs, tagging, reviews, citations, Q&A, forums, project groups).

HUBZero technology originated with nanoHUB (Box 1) a NSF-funded research and teaching community focused on the sharing and development of tools for nano-technology simulation, analysis and capacity building. One of the most impressive aspects of nanoHUB is its ability to foster innovation. This is illustrated in Figure 5, which plots, over the past decade, the number of individuals supplying tools on nanoHUB (developers), the number of new tools, and the number of new versions of existing tools. While the number of suppliers and new models grows linearly, experience over the first ten years of nanoHUB shows that innovations on existing tools can grow exponentially. In the case of the GEOSHARE workflow described above, these innovations might relate to new source data, new versions of SPAM, or even new versions of PEGASUS or GTAP. The main point is that *it is the ability to modify the data and tools available on the HUB that makes this such a powerful vehicle for building a community of practice* around geospatial data for analysis of agriculture, environment and poverty.

Figure 5. Growth of contributors, new tools and modified tool versions on nanoHUB



Source: Compiled by Gerhard Klimeck, nanoHUB Director

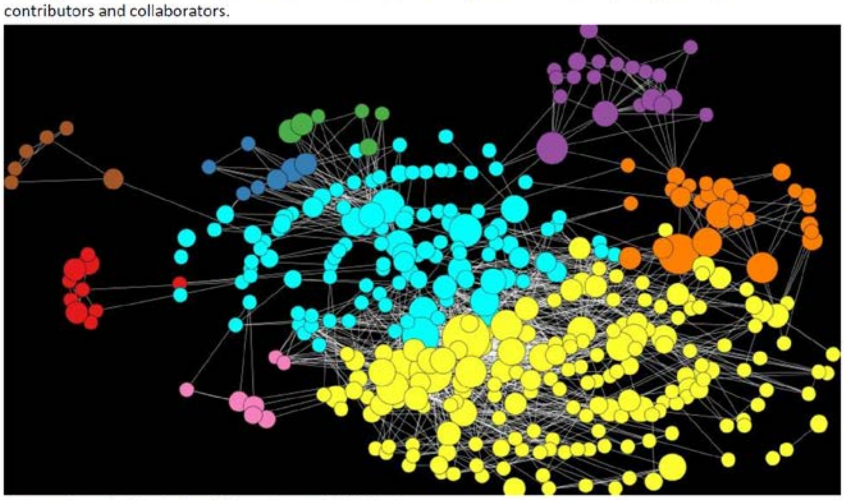
A very important feature underpinning the success of nanoHUB in fostering innovation is the ability to document workflows in order to permit replication and proper scientific citation. Indeed, *failure to document the processing of geospatial data bases is one of the biggest obstacles to advancement in this field.* Most authors want to document their work, but they do not have the time or the tools/infrastructure to do so properly. The beauty of operating these workflows on the HUB is that all of the critical steps in the workflow shown in Figure 3 are automatically documented each time it is run – i.e., metadata are produced which document which source data were used, which spatial allocation model was applied, and which version of the end-of-pipeline model was run to get the results. Furthermore, if the resulting product is to be used by others, it can be "published" on the HUB by assigning a digital object identifier (*doi*) and archived for others to access or reference in the future. By making documentation and citation easy, such an environment can facilitate the adoption of best practices and thereby facilitate to scientific advance.

Finally, since the most important impact metrics for scientists worldwide now involve citation counts, any environment that increases the likelihood of citations will increase author contributions. In the case of nanoHUB over the past four years, primary citations to work on the hub doubled to more than 1,000, and secondary citations tripled to nearly 9,000. Furthermore, the H-Index – commonly used as a metric of quality for published work, rose from 27 in 2009 to 48 in 2013, suggesting that not only was the overall number of citations rising, but they were also having greater individual impact.

GEOSHARE as a network: GEOSHARE aims to foster a network of individuals and institutions working on problems of agricultural development, global environmental quality and poverty reduction, as the previous discussion has shown. There are many different ways to develop such a network, and there are many different types of global networks: some focus on research, some emphasize decision support, and others exist mainly to facilitate dialogue and communication. In the case of GEOSHARE, we draw inspiration from the Global Trade Analysis Project (GTAP) network (www.gtap.org), which Tom Hertel founded 20 years ago, and which now comprises more than 10,000 members in more than 150 countries around the world. Of particular interest is the fact that this network serves a variety of different communities – all of which share a common interest in high quality, global economic data on bilateral trade, production, consumption, environmental metrics, and associated policies. Similar to GTAP, GEOSHARE is envisioned as a vibrant, global network which will bring together world class researchers, as well as local, regional and international policy makers, private sector stakeholders, students, and data contributors, all sharing a common interest in geospatial analysis of agriculture, environment and poverty issues.

A critical part of fostering such a network will entail providing an appropriate platform for interaction on geospatial data and analysis. Here again, we believe that HUBZero can play an integral role. Figure 6 shows the constellation of tool developers in the nanoHUB environment. Each dot refers to an individual developer, with the area denoting the number of tools developed. These are color-coded by institution, but importantly the light blue dots in the middle all represent developers who were not at institutions formally engaged in the underlying NSF project. These are 'volunteers' – individuals who have found nanoHUB on their own and begun uploading their own tools, as well improving existing ones. The lines in Figure 6 connect the dots, denoting tools which have been successfully transferred to, and modified by, other developers. These lines show that the connections are not just within institutions, but also across institutions and between original project members and newcomers. In short, Figure 6 gives inspiration to the idea that the HUBZero environment can indeed facilitate development of a community of practice.

Figure 6. Network links between developers (dots). Areas reflect usage and lines connecting developers represent analysis tools linking developers. Colors represent institutions. Light blue dots are new developers who were not part of the formal NSF network, but have subsequently emerged as contributors and collaborators.



Source: Compiled by Gerhard Klimeck, nanoHUB Director

GEOSHARE as Capacity-Building: In order to participate in GEOSHARE, some level of training is required. And such training can be difficult to obtain – particularly for members working in developing countries. Yet this is precisely where the need is greatest for geospatial data and analysis tools to support decision making on issues surrounding agriculture, environment and poverty. For example, hundreds of thousands of hectares of land in Africa are being leased each year, yet the governments writing these leases often have little idea of the potential productivity of these lands in commercial use, or the tenure status of the land. They need access to the data and analysis tools being developed on HUBZero in order to simulate potential yields and determine an appropriate range of lease prices. Decisions about public investments in research and development as well as crop suitability or climate adaptation are also frequently made in the absence of information about potential returns to such investments. Even when data and tools are available they are often insufficiently utilized due to a lack of training. Therefore, one of the GEOSHARE pilot project case studies – led by IFPRI – has focused on the evaluation of irrigation technology in Tanzania. As designed, the GEOSHARE framework is capable of engaging a broad, diverse learning community ranging from seasoned practitioners and policy makers, researchers undertaking cutting-edge empirical work, to undergraduate students taking coursework related to geospatial analysis of agriculture, environment and poverty.

GEOSHARE plans to build capacity through a variety of methods, including on-line tutorials, in person and on-line workshops and webinars, sponsored sessions at conferences coupled with travel grants for participants from developing countries, travel scholarships to facilitate inter-disciplinary experiences, graduate research fellowships and undergraduate internships. The precise form of these activities will be shaped by input from the user/stakeholder community. GEOSHARE also plans to offer an annual *global development challenge program* for undergraduate students. Through the program, student teams will be given the opportunity to work with GEOSHARE mentors to apply their creativity in solving a development problem or meeting a development need by competing for funding that will allow them to cultivate their idea, field-test their methodology, and measure the impact of their work. GEOSHARE's open-source data platform lends itself well to innovation and will actively encourage new ideas to improve development objectives by establishing a second challenge program topically aligned with stakeholder needs and focused on accelerating innovation through public- private sector partnerships – a topic to which we now turn.

GEOSHARE as a Public-Private Partnership: GEOSHARE offers considerable potential for public-private partnership. As interest in the global environmental impacts of national policies increases, the need for reliable data and analysis tools with which to assess such impacts has become critical. Perhaps the most striking example over the past five years has been the challenge posed by the requirement that biofuel production in the EU and US be accompanied by an environmental impact assessment which includes the prospective indirect land use change (iLUC) induced in the rest of the world. The economic modeling challenges associated with such assessments are daunting (Hertel and Tyner 2013), however, without a reliable data platform, such assessments are virtually impossible. Both private sector investors and public sector policy makers have a strong incentive to ensure that the data and analysis tools available for such assessments be replicable, publicly available, and reflective of the state of the art in geospatial analysis. Development of the GEOSHARE HUB will facilitate these outcomes, thereby leading to more robust and informed policy decisions, as well as a more stable and predictable environment for private sector decision making.

GEOSHARE can benefit from private sector engagement in a variety of ways. Increasingly, private organizations are willing to share some aspects of their internal data, and a forum such as GEOSHARE will offer a platform whereby private partnerships can flourish such that spatial data can be shared with the global community. These data could also be used for 'ground-truthing' GEOSHARE results – assessing the validity (e.g.) of the spatially allocated data available on the HUB. Private sector actors can also offer important insights into emerging issues which may be appropriate for GEOSHARE-based analysis. By involving these stakeholders in the governance structure of GEOSHARE, these important insights can be incorporated into the ongoing process of priority-setting within the project.

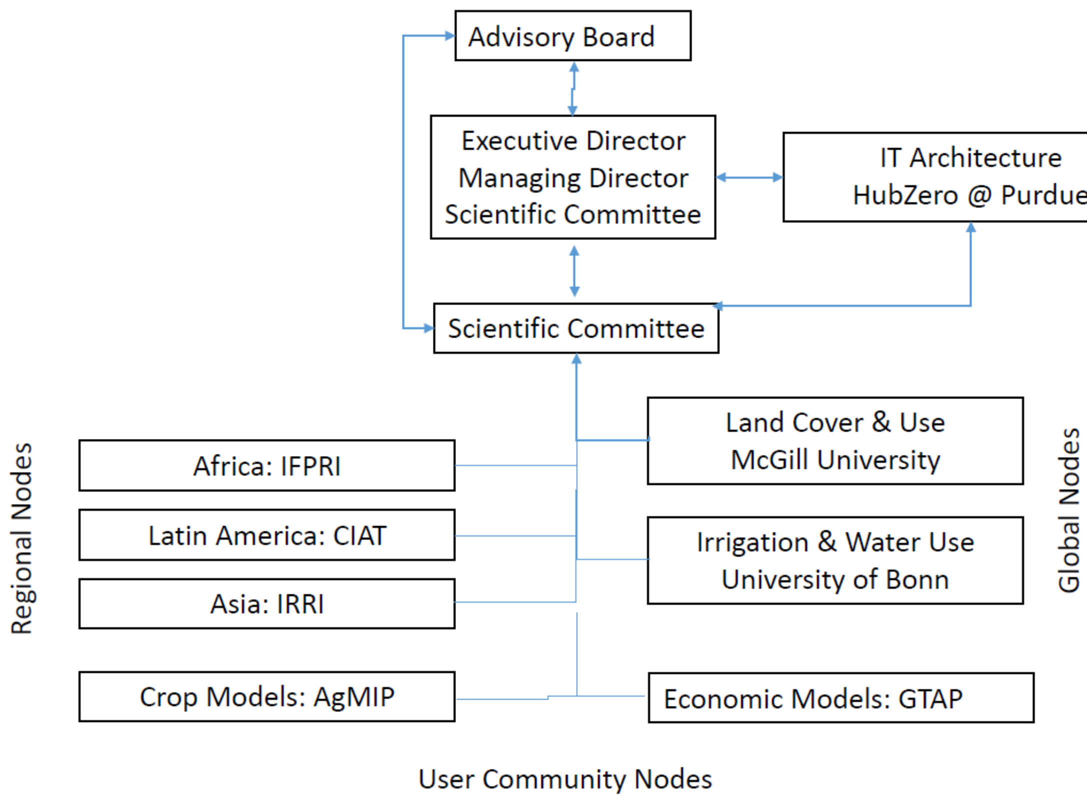
In an era of tightening federal budgets, the private sector can also be an important source of funding for GEOSHARE activities. However, when it comes to information about land, its productivity, location, tenure, and other characteristics, making these data available to the general public will be critical to maintaining the credibility of GEOSHARE. It is essential that any funding flowing from the private sector into the project be clearly designated for the provision of public goods, available to all network members, not just the sponsor organization. In fact, the public sector can benefit from the integration of such data with other publicly available information from the global community contributing to GEOSHARE. An example of how this partnership might work out is provided by the CIMSANS Open Data Working Group of the ILSI Research Foundation, which is funded in part by the private sector. They stepped forward to support development of two prototype workflows. This is a pure public good, the creation of which would not have been possible at this stage without their input. In short, the global community will benefit from the provision of these workflows.

Finally, from the viewpoint of potential network members, the private sector is an important source of employment opportunities. Critical to creating a vibrant network, and attracting young professionals and students to GEOSHARE, will be the prospect of a first job – or, in the case of those seeking mobility – a new job. By involving private sector stakeholders in the network, we will greatly facilitate the identification and matching of individuals with such employment opportunities – both in the private and public sectors. Prospective employers will have a good understanding of 'what they are getting', given the capacity building and networking activities in which a given individual has participated, as well as tools and analysis that they have posted on the HUB. And new entrants to the labor market will have a much better idea of the interests of potential employers as well as the opportunities which they might offer.

Proposed Governance Structure for GEOSHARE

An important output of the GEOSHARE pilot project is the development of a proposed governance structure for GEOSHARE. At present, this is expected to involve three key elements: an Advisory Board, a Scientific Committee, and an Executive arm comprising the Executive Director (20% time), the Managing Director (half time), a Scientific Coordinator (25% time), as well as IT support staff who will maintain the GEOSHARE cyber-HUB discussed previously. Figure 7 offers a schematic overview of the governance structure.

Figure 7. Illustrative schematic of the GEOSHARE governance structure



The Advisory Board:

Responsibilities: The Advisory Board will assume fiscal responsibility for the project as well as setting the overall policy direction for GEOSHARE. Perhaps the most important way in which this will be accomplished is through the approval of new nodes, a process through which the Board will determine the depth, breadth and overall orientation of the project. The Board will also initiate requests of the Scientific Committee for periodic assessment of the quality of various GEOSHARE data and analysis tools. With Board approval, these can become official GEOSHARE releases. In this sense the Board will share responsibility with the Scientific Committee for the integrity of the official data base and associated tools.

Membership: The Board is comprised primarily of individual representatives of stakeholder agencies, each contributing financially to GEOSHARE's core funding. Depending on the interests and membership of the sponsor organization, these individuals will serve up to a 5 year term, renewable once, for a maximum of 10 years. In addition, the Executive Director will serve on the Advisory Board, as will the Chair of the Scientific Committee. The Advisory Board will select a new Board Chair, by vote, every two years, with possible re-election for two more terms, and a maximum of six years total.

The Scientific Committee**:**
Responsibilities: The Scientific Committee will be responsible for all technical aspects of the project. In particular, it will be responsible for recommendations on database content, meta-data, standards and interoperability, as well as analysis tools and methods of delivery. Most of the work of the Scientific Committee will take place on sub-committees which will be formed to address particular technical issues. One of these sub-committees will oversee the cyber-infrastructure, ensuring that it is developed in a manner which supports the work of the nodes. The Scientific Committee will periodically evaluate the workflows available on the HUB and the Executive Sub-Committee will recommend to the Board which of these should be endorsed by the Board for use by decision makers. New GEOSHARE nodes, and hence new members of the Scientific Committee, may be proposed to the Board at any time. Approval of new nodes will be at the discretion of the Advisory Board. Existing nodes will be periodically reviewed by the Advisory Board.

Membership: The Scientific Committee will be staffed by the leaders of each of the GEOSHARE nodes, including also the Scientific Coordinator and the IT lead. Leadership of the Scientific Committee will reside in the elected Chairperson and the three-person Executive Sub-Committee consisting of the Chair, the Chair-elect, and the past-Chair. Terms for each position will be for two years, so that any given individual will spend a maximum of 6 years on the Executive Sub-Committee. The Scientific Committee Chair will also serve on the Advisory Board. In the initial start-up phase, the runner up in the election for Chair will assume the role of Chair-elect so that the Scientific Executive Committee will begin with two members, before expanding to three after the first two years.

The Research Nodes: At the heart of GEOSHARE are the research nodes representing the research and user communities with a stake in the project. The node leaders are selected based on a variety of factors, including: (a) they are recognized leaders in their field, (b) they are committed to the GEOSHARE mission of providing improved geospatial data to the global community, (c) they have a track record of providing public goods to their respective communities, and (d) they are willing to contribute their data and/or analysis tools via HUBZero, adhering to the GEOSHARE standards established by the Scientific Committee. Research node leaders will be given a modest honorarium each year in recognition of their service to the community. GEOSHARE will also cover their travel costs to the annual meeting of the Scientific Committee which each member is expected to attend. Beyond that, funding of the nodes will be pursued on a bilateral basis. In particular, the Advisory Board and the Executive Director will help to facilitate funding initiatives designed to support thrusts at each research node which are in keeping with the vision and needs of the broader project. Membership on the GEOSHARE Scientific Committee will be viewed as a professional honor and an assignment which makes that research node more attractive to funding agencies – particularly those represented on the Advisory Board.

The Executive Director:

Responsibilities: The Executive Director of GEOSHARE (20% FTE) will be the Chief Executive Officer of the organization and will be responsible for official GEOSHARE communications with the public, the network, the Scientific Committee and the Board. In consultation with the Chair of the Advisory Board, the Executive Director will prepare the agenda for the annual Board meeting. Together with the Board Chair, the Executive Director will be responsible for recruiting new Board members. The Executive Director will work closely with the Executive Sub-Committee of the Scientific Committee in order to ensure smooth functioning of that group, as well as assisting in the development of proposals for new nodes. With the support of the Board and the Scientific Committee, the Executive Director may also take the lead on development of new grant proposals, as well as fund-raising activities in support of the Research Nodes. The Executive Director will also work closely with the Managing Director to ensure high quality capacity-building activities, research and dissemination events. The Executive Director will also appoint the Managing Director and the Scientific Coordinator.

Selection: The Executive Director will be selected by the Advisory Board. This individual will serve for five years, renewable for a second term, for a maximum of ten years total. The Executive Director will serve on the Advisory Board. The Executive Director, with support from the Managing Director, will be responsible for preparing the agenda for the annual Board meeting, as well as making decisions with regard to staff hiring. (In launching the Project, Thomas Hertel will assume the role of Executive Director at the outset of the project. After 5 years, the Board will review his performance and consider appointing a new Executive Director.)

The Managing Director:

Responsibilities: The Managing Director (50% FTE) will be hired by the Executive Director with approval by the Advisory Board. This individual will be responsible for ensuring the smooth functioning and operation of GEOSHARE and will oversee administrative activities and support personnel. The Managing Director will work with the Executive Director and Board Chair to prepare the annual report. The Managing Director will also ensure that background materials for the annual Board meeting are prepared and distributed in advance and will organize and follow-up on any deliverables emerging from that meeting. The Managing Director will organize meetings and tele-conferences on behalf of the Scientific Committee and its various Sub-Committees and will track deliverables resulting from those meetings. This individual will also coordinate capacity building activities.

The Scientific Coordinator:

Responsibilities: The Scientific Coordinator (25% FTE) will be appointed by the Executive Director with approval of the Advisory Board. This individual will be work closely with the IT team to ensure the successful implementation of the technical aspects of the project as defined by the Scientific Committee. As needed, the Scientific Coordinator will participate on the different sub-committees defined by the

Scientific Committee. The Scientific Coordinator will contribute to development of new grant proposals and other fund-raising activities, capacity-building activities, project outreach, and joint conferences.

Administration of GEOSHARE: This proposal recommends that, for the foreseeable future, the administrative functions of GEOSHARE will reside at Purdue University. In theory, GEOSHARE could be located anywhere in the world, as it is a global project, with operations taking place 'in cyber-space'. However, in practice, there are great advantages to basing GEOSHARE in a University setting. As has been demonstrated with the Global Trade Analysis Project (GTAP), a University home can provide state of the art computer support, institutional continuity, and a neutral forum in which national and international institutions can find common ground – outside of the usual political constraints. Based on the GTAP experience, Purdue University has proven to be a very reliable home for this type of global project. In addition, as the home of HUBZero, Purdue is in an excellent position to host GEOSHARE's cyber-infrastructure. Indeed Purdue's Rosen Center for Advanced Computing has shown significant commitment to GEOSHARE by donating two years of support for the HUB and through its recent NSF proposals aimed at bringing geospatial capabilities into the HUBZero environment. In short, there is a compelling case for locating the administration of GEOSHARE at Purdue.

Sustainability: We envision GEOSHARE becoming financially sustainable in the medium run. We draw inspiration from the World Climate Research Programme which oversees the Coupled Model Intercomparison Project (CMIP), as well as the Purdue-housed Global Trade Analysis Project (GTAP), a successful database infrastructure effort initiated by Professor Hertel 22 years ago, which now supports a network of 12,000 individuals in more than 160 countries undertaking global economic analysis. The CMIP and GTAP experiences show that, with appropriate crafting of incentives, it is possible to fund public goods through consortia of public and private agencies which spread the costs of data and analysis production and delivery. Finally, the emergence of the HUBZero community at Purdue University, and the continuing investments being made in these activities by the National Science Foundation – most recently the NSF-DIBBS project which includes GEOSHARE as an important use case, also bode well for long run sustainability of GEOSHARE.

Placing GEOSHARE within the Institutional Landscape: A Stakeholder Map

GEOSHARE is part of a rapidly growing geospatial landscape focused on global food security and environmental issues. It is important to understand how GEOSHARE fits into this landscape and how it connects to other institutions working in this area. Its distinguishing features are the emphasis on global, inter-operable data bases and the HUBZero technology which allows users to go beyond spatial zooming and summary statistics, thereby facilitating the use of analysis tools 'in the clouds', sharing of data and implementation, documentation and publication of complex workflows. The ability of users to achieve this without downloading data and software to their own computer makes this environment particularly friendly to users in developing countries with limited bandwidth and minimal access to software licenses.

GEOSHARE is complementary to many important existing projects and initiatives. *HarvestChoice* is a Gates Foundation-funded project with a strong regional focus on Africa. Given the level of resources involved and the sharp regional focus, this project goes into far greater detail and engages more deeply with regional stakeholders than will be possible under GEOSHARE. This is why key members of the *HarvestChoice* team are involved in GEOSHARE. Africa regional node leader Jawoo Koo ensures that there is close communication between the two projects. Similar synergies are achieved through geospatial projects taking place in Asia (through IRRI-based regional node leader Andrew Nelson) and Latin America (through CIAT-based regional node leader Glen Hyman).

GEOSHARE is also complementary to many of the important global geospatial biophysical data bases maintained and offered online, such as *S-World* (soils), *CRU* (climate data) *CMIP* (climate projections). These data can be linked to GEOSHARE, and they can be integrated into the data base and workflows (recall Figure 3). The UN-FAO maintains many key data bases which can be used to provide national control totals for GEOSHARE data base construction and comparison. The FAO also maintains an important geo-spatial portal for distributing processed and down-scaled data – most notably the GAEZ data base. The GAEZ crop suitability indexes are used in SPAM, which is part of the workflow in Figure 3. However, unlike the GEOSHARE workflow, users cannot run the GAEZ spatial allocation models which distribute area, production and irrigation themselves. And they cannot change the source data and the downscaling algorithms used in GAEZ. In that sense, these represent very different approaches to supplying geospatial data for wider use.

GEOSHARE adheres to the broad guidelines for geospatial data developed by the Group on Earth Observations which seeks to knit together the many different geospatial activities underway in different parts of the world. However, given its sharp focus on agricultural land use, environment and poverty, GEOSHARE does not attempt to link into all of these diverse efforts. One important GEOSS activity which does have immediate relevance for GEOSHARE is GEO-GLAM, a project run out of the University of Maryland which focuses in remote sensing of cropland activity and land use change. GEOSHARE is currently in discussions with GEO-GLAM to identify areas of overlap and complementarity between the two projects.

GEOSHARE is complementary to new initiatives on agricultural monitoring such as the Gates Foundation-funded initiative focusing on Africa (Sachs et al. 2010). First of all, since GEOSHARE is generating global scale data, with relatively fine resolution (10km grids in the case of SPAM output), it is important to 'ground-truth' these estimates wherever possible. By providing detailed local observations, the monitoring projects offer data against which the performance of the spatial allocation models can be compared, evaluated and improved upon. A second area of complementarity rests in the need of the monitoring studies to 'scale up' site-specific findings. By offering globally consistent data, comparisons can be made between characteristics, stresses and responses in the monitored sites with other regions which have not been monitored. This comparative information will also be important when it comes to selecting new sites for agricultural monitoring. The third area of complementarity relates to the analysis tools (e.g., crop models and economic models) available on HUBZero. By applying the same tool, with different data, to both monitored and non-monitored sites, it should be possible to extrapolate some findings from the observed to unobserved regions.

Finally, and perhaps most importantly, GEOSHARE will offer a source of input data for other projects. AgMIP is a large scale project, funded by UK-DFID, USAID and USDA which has the goal of improving models used for analysis of climate impacts on agriculture. As such, it is a big consumer of geospatial data. There are many elements to this project and GEOSHARE aims to contribute in several of these areas. The most obvious one relates to the gridded crop models which produce climate impact results at the level of a geospatial grid (e.g., 10km), as opposed to a simple point. These crop models rely on gridded data of the sort produced by SPAM in the workflow identified in Figure 3. By improving the source data used in SPAM, making its workflow and processes more generally available, and potentially improving on the spatial allocation algorithms, GEOSHARE will contribute to unifying and improving the operation of AgMIP gridded model analyses.

GEOSHARE will also provide gridded input data for the global economic modeling community seeking to include sub-national detail to their models. Many of the global economic models currently used to analyze global agricultural land use built on either the GTAP or the IMPACT data bases. By improving the geospatial source data on land, water use, and agricultural production, GEOSHARE will improve the data feeding into the GTAP-AEZ and IMPACT models, which include many of those models involved in AgMIP as well, thereby leading to more accurate and informative results from those economic models.

In addition to GEOSHARE supplying improved data to AgMIP, it is also a user of AgMIP outputs. AgMIP has generated long run climate impact results from a suite of different crop models, each of which are run on outputs from a suite of climate models, the climate models having themselves been run on a suite of different economic baseline (forcing) scenarios (Rosenzweig et al. 2013; Elliott et al. 2013). This results in a massive proliferation of climate impacts outcomes which have been largely inaccessible to economic modelers outside of the large research groups. Therefore, the AgMIP tool now provided on the GEOSHARE HUB has proven to be a popular vehicle by which economic modelers are able to access, and aggregate, climate impact results (Villoria et al. 2014).

Summary and Next Steps

It has been four years since we first conceived the idea of GEOSHARE (T. W Hertel et al. 2010), and we are now at the conclusion of a two-year pilot project designed to test and refine our ideas. We will be holding a meeting in Discovery Park on the Purdue University campus, September 10 – 11, 2014, to report on what we have learned over the course of this project and to chart the way forward, including a discussion of the institutional design of GEOSHARE. Many of the original ideas included in the UK Foresight-funded concept note remain present in the current proposal. We still believe strongly in the need for decentralization, with specialized tasks being driven by individual research nodes who adhere to common standards in order to ensure inter-operability of the resulting products. However, our ideas for funding those nodes, and the incentive structures needed to make GEOSHARE functional, have evolved dramatically. We no longer envision a centralized budget disbursed by the administrative offices of the Project. The administrative burdens associated with this institutional model are astronomical. Indeed, the level of frustration experienced by all parties involved has led us to a very different vision of funding. Core activities, including project administration, HUBZero maintenance and development, facilitation of Advisory Board activities and coordination of the research nodes will remain centralized, but all of the major funding for data base development, analysis tools and case studies will flow directly from sponsor agencies to the research nodes in the context of projects which the nodes propose and execute. In this sense, the new institutional model is much closer to GTAP, in which a relatively small Center, with a budget of a bit more than $1 million annually, facilitates the activities of a very large global network (+12,000 members), which taken together, mobilizes tens of millions of dollars each year to undertake individual (and often collaborative) projects. For any group seriously engaged in global economic analysis, the annual consortium membership fee is a small portion of their overall budget, and an excellent investment in order to remain at the top of their field and policy relevant. And, in the case of GEOSHARE, the Advisory Board's ability to approve new nodes gives it great leverage over the future direction of this project.

We are looking forward to feedback on this proposal from the global community, as well as nomination of candidates for the inaugural GEOSHARE Advisory Board. Provision of such global public goods is notably difficult, but the potential gains from doing so are enormous. Success in this endeavor will greatly enhance the quality of decision making surrounding a wide range of issues bearing on global food and environmental security.

References

Anstey, Caroline. 2012. "Empowering Citizen Cartographers." *The New York Times*, January 13, sec. Opinion. http://www.nytimes.com/2012/01/14/opinion/empowering-citizen-cartographers.html.

Elliott, Joshua, Delphine Deryng, Christoph Müller, Katja Frieler, Markus Konzmann, Dieter Gerten, Michael Glotter, et al. 2013. "Constraints and Potentials of Future Irrigation Water Availability on Agricultural Production under Climate Change." *Proceedings of the National Academy of Sciences*, December, 201222474. doi:10.1073/pnas.1222474110.

Fritz, Steffen, Ian McCallum, Christian Schill, Christoph Perger, Roland Grillmayer, Frédéric Achard, Florian Kraxner, and Michael Obersteiner. 2009. "Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover." *Remote Sensing* 1 (3): 345–54. doi:10.3390/rs1030345.

Hertel, T. W, W. Britz, N. S Diffenbaugh, N. Ramankutty, and N.B. Villoria. 2010. *A Global, Spatially Explicit, Open Source Data Base for Analysis of Agriculture, Forestry, and the Environment: Proposal and Institutional Considerations*. Department of Agricultural Economics, Purdue University. http://www.agecon.purdue.edu/foresight/proposal_spatial_database10-15-2010.pdf.

Hertel, Thomas W., and Wallace E. Tyner. 2013. "Market-Mediated Environmental Impacts of Biofuels." *Global Food Security*. Accessed July 18. doi:10.1016/j.gfs.2013.05.003.

Klimeck, G., M. McLennan, S.P. Brophy, G.B. Adams, and M.S. Lundstrom. 2008. "nanoHUB.org: Advancing Education and Research in Nanotechnology." *Computing in Science Engineering* 10 (5): 17 –23. doi:10.1109/MCSE.2008.120.

McLennan, M., and R. Kennell. 2010. "HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering." *Computing in Science Engineering* 12 (2): 48 –53. doi:10.1109/MCSE.2010.41.

Monfreda, Chad, Navin Ramankutty, and Jonathan A. Foley. 2008. "Farming the Planet: 2. Geographic Distribution of Crop Areas, Yields, Physiological Types, and Net Primary Production in the Year 2000." *Global Biogeochemical Cycles* 22 (March): 19 PP. doi:200810.1029/2007GB002947.

Portmann, Felix T., Stefan Siebert, and Petra Döll. 2010. "MIRCA2000—Global Monthly Irrigated and Rainfed Crop Areas around the Year 2000: A New High-Resolution Data Set for Agricultural and Hydrological Modeling." *Global Biogeochemical Cycles* 24 (March): 24 PP. doi:201010.1029/2008GB003435.

Ramankutty, Navin, Amato T. Evan, Chad Monfreda, and Jonathan A. Foley. 2008. "Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000", January. http://www.agu.org/pubs/crossref/2008/2007GB002952.shtml.

Rosenzweig, Cynthia, Joshua Elliott, Delphine Deryng, Alex C. Ruane, Christoph Müller, Almut Arneth, Kenneth J. Boote, et al. 2013. "Assessing Agricultural Risks of Climate Change in the 21st Century in a Global Gridded Crop Model Intercomparison." *Proceedings of the National Academy of Sciences*, December, 201222463. doi:10.1073/pnas.1222463110.

Sachs, Jeffrey, Roseline Remans, Sean Smukler, Leigh Winowiecki, Sandy J. Andelman, Kenneth G. Cassman, David Castle, et al. 2010. "Monitoring the World's Agriculture." *Nature* 466 (7306): 558–60. doi:10.1038/466558a.

Villoria, Nelson, Joshua Elliott, H. Choi, and Lan Zhao. 2014. *The AgMIP Tool: A GEOSHARE Tool for Aggregating Outputs from the AgMIPs Global Gridded Crop Model Intercomparison Project*. Purdue University. https://geoshareproject.org/tools/cropdatatool/.

You, Liangzhi, and Stanley Wood. 2006. "An Entropy Approach to Spatial Disaggregation of Agricultural Production." *Agricultural Systems* 90 (1–3): 329–47. doi:10.1016/j.agsy.2006.01.008.

Zhao, L.*; Song, C.; Lee, J.; Kim, J.R.; Feng, W.; Merwade, V.; and Villoria, N.B. "Bringing integrated GIS data and modeling capabilities into HUBzero platform". Second International Workshop on High Performance and Distributed Geographic Information Systems (HPDGIS 2011), Conference Proceedings, Chicago, IL. November 1, 2011.