Planning the Creation of a Multi-Purpose, Public Facing Academic Research Database Platform for Agricultural Data with Geolocation Data Correlation

PI: Jason Bengtson (Kansas State University Libraries)
Co-PI: Jeremy Kenyon (University of Idaho)
Co-PI: Jason Karl (University of Idaho)
Co-PI: Livia Olsen (Kansas State University Libraries)

Statement of National Need

The central research question we intend to address is how academic institutions can apply new technologies in data search, storage, retrieval, and cross-referencing to make scientific data easier to find and re-use. We are asking for $95,578 for this phase of our project. This project will be carried out by the PI's institution (Kansas State University) and the University of Idaho, with no cost sharing anticipated.

This proposal describes planning for a web application and service that will be designed to be modular, portable, flexible, multi-modal, and capable of providing an open source extension to improve search and discovery of agricultural and environmental information. This service will be built as an evolution of two already successful applications: Kansas State University's Croplands Research Database (https://www.lib.k-state.edu/gracroplands/) and the JournalMap tool (https://www.journalmap.org). The Croplands Research Database provides researchers with a curated database of literature about greenhouse gases and croplands, while the JournalMap tool provides a powerful mechanism for extracting geolocation data from academic papers so that it can be applied to the articles as a metadata field for location-based search.

Researchers have documented a problem in agricultural and environmental information-seeking behavior. Resource managers, students, policy-makers, landowners, and other scientists all need information that is salient to the context of their work (McNie, 2007; Wallis, 2011). Salience, in this case, means information that possesses the relevant spatial or temporal attributes to the user's information need. If a student or researcher wants to find literature or data on the environmental stressors of a certain plant species in a particular location, for example, they can do so by searching the species name in any number of useful databases along with the appropriate geographic name. However, if they want to know the stressors of any *similar* species, anywhere in the world, they cannot, not without searching for every possible individual species, which requires them to have incredibly deep and broad knowledge. Many literature database tools are also typically constructed without an appreciation of spatial characteristics. Those that attempt mapping based on location, or geographic names, take a step in this direction, but even then, they are entirely at a loss for ecological parameters – such as elevation, aridity level, number of growing days – that surround that location. This is unnecessary given that geographic researchers have constructed many reasonable and constantly improving spatial data layers that offer up this information. It need only be mined and added to the metadata of the relevant information resources.

The indexing approach demonstrated through JournalMap is the capacity to search for a given topic (such as a plant) and use its location to find other plants with similar environmental attributes. One may not know that certain plants in southern New Mexico encounter many similar ecological stressors as plants in eastern Namibia, but with JournalMap powering the search, one will see both sets of results, each relevant to the ecological characteristics of interest to the user. To some extent, this introduces a basic opportunity to pursue discovery using a improved approach: to use the parameters of the location represented by an information resource to dynamically generate relationships with other resources, be they literature, images, or datasets. It moves us away from a pattern of primarily using text strings and natural language processing for information retrieval and moves us toward also using scientifically agreed-upon earth system models as ways of inferring the relationships between information objects.

But the primary value of this tool is to power and improve others' search tools, thus why we call it a service and not just an application. One only needs to index an object using this method **once** in order to use it. So imagine that one has metadata that gets richer over time, without the labor cost of adding new metadata and a metadata store designed to allow authorized experts to add new facets to data. This application would be something that could be used for many different kinds of literature databases or data repositories, in a constellation of different subjects, allowing institutions to stand up a powerful, customizable database solution to provide maximum benefit to their users.

Project Design

Such a service is exactly what we propose. The team will work with the Croplands Research Database of the Kansas State University Library and the GRA as an example of combining JournalMap with a specific database, to demonstrate that is extensible and portable to any curated environmental and agricultural search system. We intend to create a platform which will be extremely capable out of the box, but which, more importantly, will provide the basis for new modular development by others so that our initial work supplies the core of a continually evolving product.

This grant will fund the planning phase for the development of this application. It will allow the research team the salary support needed to produce a detailed application architecture and plan of work to serve as the basis for a successful research grant which will result in the completion of the digital application. We will also use this time to identify one to two other University teams we can partner with for the next phase of the project with the goal of a final stage database application that provides broad utility.

National Impact

This planning project is calibrated to produce two deliverables (application architecture and detailed plan of action) which will be used to request funds for the next phase of this project (building the application). The deliverables from this phase, like the application to result from the successive phase, will be made publicly available. The availability of these products has national implications for libraries, Universities, and other institutions which will be free to make use of them.

Budget and justification

We are requesting a total budget of $ 95,578. This reflects institutional indirect costs, as well as salary support for four senior personnel: Jason Bengtson, Jeremy Kenyon, Jason Karl, and Livia Olsen. The salary support will be divided evenly with $12,000 supplied per person. Fringe benefits as calculated by Kansas State University total $ 14,880 for a salary support total of $ 62,880. Total Kansas State University indirect costs for a submitted grant of this size are $ 32,698, for an overall total of $ 95,578. This salary support will allow the senior personnel the time to engage in a series of frequent and in-depth planning activities to result in the deliverables described above and outreach to other national teams which may wish to collaborate.

References:

McNie, E.C. (2007). Reconciling the supply of scientific information with user demands: an analysis of the problem and review of the literature. *Environmental Science and Policy*, 10: 17-38. doi:10.1016/j.envsci.2006.10.004

Wallis, P.J. et al. (2011) Mapping local-scale ecological research to aid management at landscape scales. *Geographical Research*, 49: 203–216. doi:10.1111/j.1745-5871.2011.00691.x