JournalMap: a service for geolocating academic literature and generating inferred metadata for improved discovery

NLG Program Goal and Objective: 3.2

Introduction. The University of Idaho (UI) Library seeks $249,958 from the IMLS NLG program to develop JournalMap over three years. JournalMap is a prototype web service and database that indexes journal literature, parses it for the location of the study area, and then records that location. Once obtained, it uses the study location to generate further metadata using spatial data layers such as soil type, precipitation, and biome. Using this system, we can identify environmentally context-similar papers that enhance the precision and recall of searches in the areas of agriculture and environmental research. Researchers can search for articles by location or environmental similarity either through JournalMap's website or via JournalMap integrations with other websites. This project builds on an FY2021 Planning Grant LG-246411-OLS, which allowed us, with partners at Kansas State University (K-State), to plan out and test a redesigned version of JournalMap with greater functionality and fewer limitations. Funds will be used to develop the core application and service and integrate the system with two partner institutions, K-State and the University of Arizona (UA). In each case, partners will bring different types of document collections and database systems that require integration of the service in different ways. JournalMap will create an "Unpaywall"-like database for geolocation metadata of literature, along with the associated contextual metadata. At the end of this grant, JournalMap will be scalable, rely on automated geolocation, and be openly accessible. Key secondary elements of the project will be designing replicable integrations with partner's systems and a developing a sustainability plan for JournalMap.

Project Justification. While access to scholarly literature has become dramatically easier in recent years, existing bibliographic search tools (e.g., Google Scholar, Web of Science, library catalogs) still focus primarily on the *what* of research while largely ignoring the *where.* This prevents efficient searching based on research location, or on location-related attributes including environmental, climatic, social, and economic features (Karl et al. 2013). Yet much of the scholarly literature is either location-based, its applicability dependent on spatial context, or the results (or even the questions asked) influenced by the place and time in which it was conducted (Livingstone 2003). Thus, the lack of usable location information from literature and the lack of location-based search tools limits knowledge discovery.

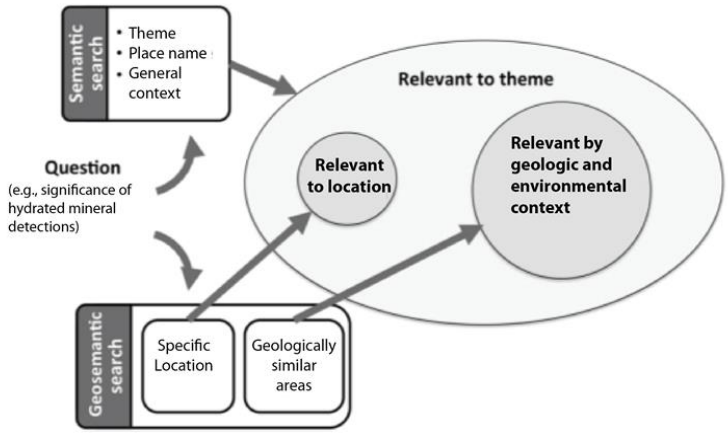

Figure 1. The ability to search by theme and location (i.e., geosemantic search) can improve the relevance of search results. Figure from Karl et al. (2013).

Researchers have documented a problem in agricultural and environmental information-seeking behavior. Resource managers, students, policymakers, landowners, and scientists have difficulty finding information that is salient to the context of their work (McNie, 2007; Wallis, 2011; Schmitt and Butler 2012; Karl et al. 2013). Salience, in this case, means information that is not only topically relevant but possesses spatial or contextual attributes aligned with the user's information need (Figure 1, Karl et al. 2013). For example, if a researcher needs information on practices to control soil erosion in Namibia, a topical literature search will include many sources that are not appropriate to the region or its soil types. Locally-generated research would be difficult to find without well-developed social networks or *a priori* knowledge (see Zimmerman 2007), and discovery of relevant information from other regions with similar soils and climate (e.g., New Mexico) would be nearly impossible without incredibly broad subject knowledge.

Most literature database tools have been constructed without an appreciation of the complexity of spatial searching, relying largely on geographic place names to describe locations. This approach, however, is flawed because there are no publication standards for reporting place names. For example, Karl (2018) described a Web of Science search of ecological literature for studies with "Chihuahuan Desert" in the abstract and associated indexed information. Of the more than 800 articles returned by this search, only one third of them actually occurred within the Chihuahuan Desert (Figure 2a), due largely to the presence of irrelevant place names in articles (see Karl 2018). Additionally, many more studies occur within this region but used different names to describe their study areas (Figure 2b).

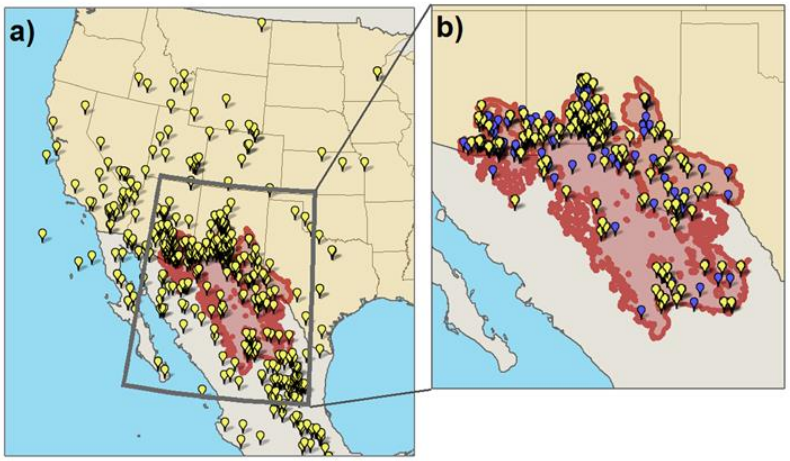

Figure 2. Example of the challenge of searching for scientific literature based on location using existing search tools. Only 33% of over 800 articles returned from a Web of Science search for the location name "Chihuahuan Desert" were within that area (red shaded region, Map a). A search of JournalMap.org returned many additional articles in the area that did not use the term "Chihuahuan Desert" (Map b). From Karl (2018).

Several search tools have been developed that have begun to change the thematic-only literature search paradigm through map-based searching. For example, JournalMap (Karl et al., 2013) provides a map-based search interface for georeferenced journal articles and an API for embedding search results and article maps. The USGS Science Base website permits basic map browsing of USGS reports and articles by agency scientists from geographic coordinates or bounding boxes assigned to each source. The data repositories such as Pangaea, Earthworks, and DataONE map locations of datasets, many of which can be tied back to published articles. Article abstracting services like GeoRef and CAB Abstracts assign general geographic regions (i.e., place names) to articles. BioStor extracts geographic locations from historic articles in the Biodiversity Heritage Library and provides article-level maps of specimen locations (Page, 2011). There are also numerous examples of "georeferenced bibliographies" that offer maps of article locations related to specific themes (e.g., van Vliet et al. 2012, Pert et al. 2015, Howell et al. 2019). Of the existing examples, most assign locations to published articles either via their source data (e.g., Pangaea), self-reporting by the authors, or manually (e.g., ScienceBase, CAB Abstracts). Currently, only JournalMap and BioStor employ automated geolocating algorithms to mine locations (from printed geographic coordinates) from article text. However, these approaches will continue to be of limited utility until the total amount of georeferenced literature is greatly increased, and this will hinge on developing new techniques for rapidly and accurately georeferencing existing literature.

Offering searchable maps of literature based on location or geographic names is a step in the right direction, but most of the available services do not provide for identifying literature from contextually-similar but geographically separate regions (see Namibia erosion example above). Of the sites offering geographic searching for literature, only JournalMap (with limited spatial attributes; Karl et al. 2013) and the GLOBE project (focused solely on land-use change studies; Schmill et al. 2014) provide functionality for searching based on location similarity. Enabling similarity-based searching is possible through inclusion of additional and readily available environmental (e.g., elevation, biome), social/political (e.g., population, political regime), or economic (e.g., GDP) spatial layers.

The value of georeferenced literature databases has been established in many fields including ecology and conservation (Page et al. 2011, Martin et al. 2012), land management (Wallis et al. 2011, Karl et al. 2013), environmental science (Schmill et al. 2014), library sciences (Johnson et al. 2009, McKee 2019), and infectious disease (Hendrickx et al. 2010). However, in most cases, assembly of these databases is a laborious process of manually geotagging articles and ultimately these efforts are typically not sustainable beyond their original research

objectives. Additionally, until a very large number of articles are georeferenced, the potential for geographic-based literature searching is limited to narrow domains or specific questions. JournalMap's automated approach to identifying and extracting geographic coordinates from articles (Karl 2018) is an important step toward scaling up geographic-based searching, but articles with coordinates account for only about half of papers published in ecology journals (and less for other knowledge domains; Karl et al. 2018). Thus, in order to scale-up the concept of geographic literature searching, more robust, automated georeferencing approaches need to be developed and implemented. Automation is a logical step towards scalability but has seen limited use due to the complexity of toponyms (Gritta et al, 2017; Gritta et al, 2019). Named-entity recognition software, like Spacy, combined with geocoding resources, like Geonames.org, have introduced the capacity to do automated geographic indexing more effectively.

JournalMap demonstrates the capacity to search for a given topic and use its location to find other sources from areas with similar attributes. But a greater value of JournalMap would be integrating it with other search tools and databases to improve knowledge discovery and infer relationships with other resources (e.g., literature, images, datasets), especially those based in libraries. JournalMap was originally developed independently and not in coordination with other efforts. Thus, integration is limited and challenging. Once an article is georeferenced, however, its accompanying metadata becomes richer over time with no additional labor cost as that location is overlain with new data layers. In fact, adding more data layers and new attributes to an indexed article is a trivial additional cost. A revised JournalMap designed to support existing literature databases or data repositories, in a constellation of different subjects, would allow institutions to stand up a powerful, customizable database solution to provide maximum benefit to their users.

The NLG program has focused objective 3.2 on innovative approaches to collection management and improving cataloging and inventory processes. JournalMap offers an automated approach to classifying articles by geographical location (Karl 2018, Olsen et al. 2021), an important step toward scaling up and making feasible geographic-based searching. JournalMap – once well-integrated with other efforts – has the potential to improve discovery for users of these systems with minimal effort on the part of librarians and curators.

Finally, the FY2021 IMLS Planning Grant positioned this project to immediately begin development and to initiate integration efforts with two partners. During the Planning Grant, we were able to address the original limitations of JournalMap and plan to move beyond them – by three primary enhancements: 1) moving beyond only journal articles, 2) moving beyond a coordinate point-based geographic model to accepting points, lines, and polygons to describe study areas, and 3) introducing the ability for two-way (i.e., read/write) communication between providers through the JournalMap API. Further, we obtained feedback about the value, approach, and system through two groups: 1) our Advisory Board, which included librarians, scientists, and technologists, and 2) a small survey of an international research group. Both groups were invaluable in introducing needs and requirements that are incorporated into the following work plan.

Project Work Plan. Over the three years of the project, we will follow two tracks of objectives – technical objectives focused development of the service and its integration with our partners, and outreach objectives focused on engaging with advisors, reaching out to potential content providers, and working on the sustainability plan for the project. The UI will lead the project and both tracks. K-State and UA personnel will participate in both tracks as detailed below. Specifically, Project Director Jeremy Kenyon and co-Principal Investigators Jason Karl and Livia Olsen will oversee the outreach objectives throughout the project, while the technical objectives will be overseen by other project members, detailed below.

Objective	Type	Year One	Year Two	Year Three
JournalMap Service (UI)	Technical	Complete the JournalMap primary development	Complete any remaining tasks and work with partners	Any remaining bug fixes and optimizations.
CRDB/BioDIS (K-State)	Technical	Develop data pipelines for moving content to/ from JournalMap	Begin adding JournalMap-enhanced records to CRDB and BioDIS	Complete the visualization and integration into CRDB and BioDIS

Rangelands Gateway/Rangeland Ecology and Management Archives (UA)	Technical	Develop data pipelines for moving content to/ from JournalMap	Begin adding JournalMap-enhanced records to the Rangelands Gateway	Complete the visualization and integration into the Rangelands Gateway
Collections and Community	Outreach	Identify opportunities for indexing new content or providers with content	Index relevant open collections; set up meetings with potential new providers	Curate new content; present at meetings and conferences
Sustainability	Outreach	Explore scenarios, models, and plan for meetings	Evaluate scenarios and meet as possible	Formalize a sustainability strategy and plan
Advisory Board	Outreach	Provide feedback and assist with reaching out on other tasks	Provide feedback and assist with reaching out on other tasks	Provide feedback and evaluate the project's outcomes

Table 1. A high-level overview of the planned objectives and tasks for the proposed 3-year project.

Technical Objective 1. **Develop JournalMap into a production service.** In the first year, we will develop our redesigned JournalMap prototype into a production-level service and application that will replace the existing JournalMap (note: the current website for JournalMap does not feature this redesign). During our FY2021 Planning Grant, we re-engineered the original JournalMap from legacy Ruby-on-Rails to a Python/Flask framework, and redesigned the database to: 1) include most major document types (originally restricted to journal articles) using the Citation Style Language (CSL) schema, and 2) accommodate any type of spatial object (originally restricted to coordinate pairs, or points) using GeoJSON format. A significant improvement in the redesign made possible by the Planning Grant was extending JournalMap's web service API to support all database operations associated with creating and editing document records, not just retrieval (i.e., read) access of the existing API. However, due to the limited nature of that project, we did not completely rebuild the entire application, only the requisite parts to demonstrate its potential and work on the specific advancements mentioned above. Thus, the first year of this project will focus heavily on fully implementing JournalMap's full set of features in the Flask framework, importing all existing content, and making the new site fully operational (see Supporting Document 3 for a draft Gantt chart that details our anticipated JournalMap development effort).

Our goal in year one will be to ensure that the redesigned JournalMap service is ready to receive new collections through its provider API and return geolocated and enhanced metadata about indexed resources (see Supporting Document 4, pg. 13 for the workflow diagram and other details). From our FY2021 Planning Grant, most of the basic functionality and database design is already in place, including an interface for displaying of a range of spatial object types (i.e., points, lines, polygons),a provider dashboard for editing and modifying records, and a moderator dashboard for reviewing, correcting, and approving document metadata and location information. Implementation will require developer effort to bring these elements up to a production-level standard, "harden" the service through rigorous testing, and make it ready for public use. To ensure high-quality and where applicable, the development team will perform regression testing using custom-built automated test harnesses to ensure that functionality is preserved as new features and capabilities are added and defects are identified and fixed. Automated regression testing will be most applicable around enhancements of the bi-directional API and provider integration endpoints. These regression tests will help preserve progress towards milestones and deadlines as code is subsequently added and revised.

One major outcome for this objective is to implement a generalized approach to ingesting new resources (i.e., content) via the API. Currently, the JournalMap prototype's API can receive items in a JSON CSL-formatted payload. Many systems export metadata as CSL (e.g., Zotero, DataCite, and Crossref); however, custom literature databases like K-State's Croplands Research Database (CRDB) and the UA's Rangelands Gateway (RG) do not. To address this challenge, we will create a set of scripts (e.g., XSLT, Python) designed for mapping a generic SQL or XML database output to CSL so that any provider who needs to transform their resources to CSL can do so. The two partners in this

project, K-State and UA provide two different types of databases, allowing us to design an ingest process that works across different systems.

One other redesigned feature we will fully implement is the assignment of attribute values (e.g., elevation, precipitation regime, population density) to JournalMap documents based on their geolocation. These attributes enable searching for articles from similar, but geographically separate areas. The original JournalMap used a limited set of attributes, and two priorities identified by the Advisory Board for our Planning Grant were expanding the suite of attributes and making it easier to add attributes later. Our current JournalMap prototype did not implement this new approach to assigning attributes to articles, so a task in year one will be to fully implement a revised attribute assignment function. Part of this process will include providing clear, transparent information about what data layers are used to generate the attribute information for each JournalMap record to enable sound provenance of all automatically generated metadata. This information will be added to the website.

A final improvement is JournalMap's approach to automated geoparsing. Presently, JournalMap can automatically parse geographic coordinates from documents with a high degree of accuracy (Karl 2018). Depending on the source, this can often geolocate over half of all documents ingested. The remaining documents must currently be geolocated manually based on place names used in the document. The tools to automatically geolocate scientific documents perfectly from place names do not yet exist. However, in our Planning Grant, we implemented the Mordecai geoparsing system, which uses Spacy and machine learning models to identify likely placenames and Geonames to provide geolocation (Halterman 2017). Identification of a study's country of origin was identified by Mordecai reliably enough for us to include this in the geoparser for JournalMap in year one. Work on more accurate placename geoparsing with Mordecai and similar systems is ongoing. We are seeking additional funding (e.g., USDA-NIFA Data Science for Food and Agricultural Systems Program) to advance our use of Mordecai to return more precise and accurate location information and assign a confidence interval that conveys our degree of certainty to users.

We recognize that unforeseen challenges can introduce delays, and our work plan build in an expectation that we might need to extend completion of development of the JournalMap service into year two. In addition, we will move towards a strategy of working on specific issues raised by K-State and UA during their first year, and finalizing the service. If development proceeds ahead of schedule, our goal in year two is to do at least one round of updates and enhancements based on the issues raised by our partners. We expect that the JournalMap service will be functional and production-ready in year three. Further technical work in year three will be addressing issues raised during the integration process with K-State and UA and assisting with their portions of the project.

The JournalMap development will be directed through the UI's Research Computing and Data Services (RCDS) unit, led by project team members Luke Sheneman (Director) and Jennifer Hinds (Research Applications Architect). Both were instrumental in developing the JournalMap prototype for our Planning Grant and have extensive experience in both web and server-side application development and system architecture. One of the benefits of working with RCDS will be a flat hourly rate for development activity, allowing the team to allocate time and responsibility to whomever within the unit is best for a particular task, while maintaining an overall commitment to the project. RCDS will host the JournalMap website, API service, and database and will continue to do so after the project; the current cost for application hosting internally for UI is absorbed by institutional funds. Development for JournalMap will be managed between teams at UI, K-State, and UA in a private GitHub repository. Application bugs and feature requests will be tracked with GitHub Issues. Once the new implementation is stable and ready for release, the GitHub repository will be made public.

Technical Objective 2. **Integrate K-State's databases with JournalMap.** K-State Libraries will integrate JournalMap with the CRDB as well as their natural history database, BioDIS (currently not available online, but will be updated and online by the end of summer 2022). CRDB contains primarily traditional research literature about greenhouse gases and cropping systems worldwide, but BioDIS introduces collections from K-State's herbarium and other biodiversity specimen information, a different kind of resource, but one in which geolocation is essential.

In year one, K-State will begin to develop a pipeline for pushing content to JournalMap for indexing and geolocating (increasing the number of records in JournalMap) and retrieving the geolocation-enhanced records using the API. During the FY2021 Planning Grant, K-State developed an approach for submitting resources into JournalMap using a Zotero library, but we plan to develop a workflow to push content into JournalMap directly from the canonical CRDB and BioDIS databases. Additionally, K-State will implement the use of CSL natively in the CRDB to simplify metadata transfer to JournalMap and to improve the database itself.

Working with BioDIS is a foray into importing and capturing location information for images and non-textual content with JournalMap. The database presents an opportunity to demonstrate the utility of JournalMap with a fundamental resource in environmental and agricultural research, the natural history collection. We are please that Dr. Ferguson is equally excited to utilize the JournalMap framework (see Supporting Document 5). Mapping the metadata in BioDIS into CSL will require more effort than for publications like articles, but CSL is a flexible schema with customization options. In fact, the JournalMap prototype uses CSL's "custom" fields currently to capture geolocation information along with the standard citation details. The image metadata in BioDIS typically does not contain geographic coordinates but does contain other location information like county names or research area which would require using JournalMap's functionality to locate items on the map. We plan to use these customizable fields in CSL to capture relevant and unique metadata from the specimen database.

In year two, with the pipeline in place, the K-State will focus on building out its user interface integration with JournalMap. The goal of these integrations is to provide researchers with map-based search tools (enabled by JournalMap's location metadata and search API) for more efficiently identifying relevant resources within the CRDB and BioDIS applications. Based on the FY2021 Planning Grant, results from JournalMap can be filtered for specific collection providers, such as the CRDB collection. K-State will create a searchable and browsable map interface for their applications using this provider filter. We have identified two possible approaches to integration with CRDB and BioDIS, one that uses an iframe to import a map from JournalMap directly, and another that involves creating a native map in the CRDB/BioDIS websites that makes calls to JournalMap through the API. In both cases, users are intended to remain on the CRDB or BioDIS websites to search those collections. Any user that wishes to see other collections, or the entire JournalMap database, will be directed to the primary JournalMap website. These options will be evaluated as we proceed, in discussions among the team members and with the Advisory Board to gain feedback on preferences and practicality.

A final unique contribution of the partnership with K-State is the inclusion of new attributes and data layers to JournalMap. For example, currently the CRDB uses the Köppen Climate Classification System designations, and a cropping system classification as additional metadata. Including these layers in JournalMap will eliminate the labor spent adding this metadata to CRDB records in the future. In fact, this is one of the promises of JournalMap's value enhancing approach – to eliminate time spent assigning standardized concepts to records when they can be automatically generated from reasonably accurate layers (and re-indexed over time as those layers improve or change). In year two, we plan to explore the addition of these layers.

By year three, we will complete any remaining tasks necessary to complete integration with partner institutions. Specifically, K-State will have a system that uses the API from JournalMap and widgets to integrate the mapping features and layers into the CRDB and BioDIS. We plan to make the widgets available as reproducible software for future providers.

The K-State team will be led by Livia Olsen and Julie Bell, both participants in the FY2021 Planning Grant, and leads for the CRDB and BioDIS projects at K-State Libraries. Olsen oversees agricultural research services and collections, including the content of the databases, while Bell oversees systems and digital projects for K-State Libraries. They will work with Tom Misilo, a systems administrator and developer at K-State Libraries.

Technical Objective 3. **Integrate UA's databases with JournalMap.** UA will focus on indexing content from and integrating JournalMap with their discovery tool, the *Rangelands Gateway* (RG), a repository of more than 25,000 peer-reviewed journal articles, websites, images, databases, videos, maps, reports, and decisions making tools.

Specifically, UA will initially index the journal archives for *Rangeland Ecology and Management (REM)*, a core collection it maintains for the Society for Range Management. Aside from US-centric collections, the RG also includes content indexed from groups including the Australian Rangelands Society and the Grasslands Society of Southern Africa. Currently, the Gateway can only be searched by Collection, Title, Author, and free-text search of the abstract. The team at UA is in the process of extending the RG's search interface to use keywords facets and filters. The timing is right for introducing extended map-based search capabilities afforded by JournalMap into the RG.

In the first year, UA will develop a pipeline for loading content from the REM archives (hosted in the RG's Dspace instance at the UA) into JournalMap. Our goal is that this pipeline will also take shape as a general tool that others with Dspace repositories can use to push content to JournalMap. The UI team also operates a Dspace instance for other projects and so both teams are familiar with its related metadata standards and protocols such as OAI-PMH and Dublin Core and will devise a generalized schema-mapping approach. The RG uses a customized "extended Dublin Core" schema which offers a similar challenge for schema mapping. UA will extend their data model to include locational and ecological database fields that will eventually be populated by the JournalMap auto-tagging algorithms. The locational and ecological information provided by JournalMap will improve the RG's search interface, search results, and resource details page. Ultimately, the differentiation between the K-State and UA systems also will provide an opportunity to better develop our generalized approach to ingestion. No matter our efforts, some work will always be required by external providers of content to JournalMap, depending on their database schema and unique systems, but our hope is to reduce the effort by working with multiple systems and developing best practices.

Another challenge in the process of ingestion is merging records of the same publications across the different, independent systems. The RG is a non-custodial catalog of primarily open-access content with some recognized record duplication with JournalMap. In discussions with UA, we have identified several solvable concerns, such as identifying the version of record, verifying details against systems like Crossref, dealing with records with different levels of description or enhancement (e.g., a record with author initials vs. a record with full names), and then the challenge of re-integrating those records back into the provider's databases so that they can use the JournalMap enhancements (location, attributes, etc.). In short, we estimate developing a series of processes to ensure that records from the source database are verified and enhanced.

Finally, like K-State, the UA partnership offers an opportunity to explore the use of new layers and attributes. The RG contains literature on social and economic dimensions of agriculture, not only its biophysical dimensions. Adding relevant data layers (e.g., population density, demography, per capita income) for these subjects will improve not only the relevance of the tool for those partners, but for other future users in other domains as well. There are many clear, well-known, and heavily used data layers that could expand the usefulness of JournalMap and give users a new way of thinking about spatial content of the items for which they are searching. We will consult with our Advisory Board on deciding which layers to include and to gain some perspective on them or where those layers may be problematic.

By year three, the search interface on the RG will include an interactive web map and attribute filters, such as average yearly precipitation, soil-type, elevation, state/province, country. Users will be able to see their search results tabularly or spatially, then continue to filter/reduce their search results using the facets. Clicking on a specific search result will display more details about the resource but will also include a preview map and ecological keywords derived by JournalMap's indexing.

The UA team will be led by Matt Rahr, Director of Cyber & Information Technologies for the UA College of Agriculture and Life Sciences (CALS). Rahr has led numerous technology projects at UA, including the development of the RG. Rahr will assign the tasks as appropriate to his development team, which includes application architects and web designers.

Outreach Objective 1. **Grow JournalMap's content and community.** One of our objectives for the project will be to continue indexing new content and collections in order to increase the JournalMap database. As mentioned earlier, the service grows in value as more content is added. While some collections will be obtained through our project partnerships, we will also continue to ingest documents from other sources, focusing first on open-access literature for

the sake of expediency. Our task in year one will be to identify the best and most appropriate repositories available and begin discussions with the providers, as necessary. One example of a repository we may pursue is African Journals Online, which has relevant open access journals with content which will enrich the global scope of the database.

In year two, we will discuss potential indexing of these collections with the collection owners. If appropriate, we may seek to recruit new members of our Advisory Board based on these conversations. We will work towards an equitable and agreeable approach to indexing of this content. Building collections and community is a goal for this project, and while we could simply index open-access material without discussions, we feel that the opportunity to build positive partnerships, grow awareness of JournalMap, and create an equitable relationship with providers makes it worthwhile.

In year three, we have two goals. First, we aim to aggressively begin presenting the work of the past two years at conferences and meetings, building awareness about the JournalMap service further, and demonstrating its capabilities through the integrations with K-State and UA. Project team members will submit proposals to conferences such as ACRL, USAIN, DLF, CNI, and others as appropriate. Some project team members are already planning to attend these conferences and will aim to discuss JournalMap at them.

Our second year three goal is curation. For those items that we have begun indexing and continue to index, we will work on cleaning up any metadata and ensure that the geolocation process has been checked, and that requisite fields and metadata are appropriate. The current JournalMap interface makes this process simple, with a moderator dashboard that provides a clear window on what was processed successfully and where errors or incomplete information may exist. A successful service at the end of the project's performance will require not only new content, but also accurate and high-quality content. As other tasks wind down and we move towards outreach, we will commit time to curation and cleaning up records as needed.

Outreach Objective 2. **Develop a sustainability plan for JournalMap.** A major objective of this project is to plan for the maintenance and long-term viability of the JournalMap service beyond an NLG Implementation Grant. Recognizing the wide range of options and the time involved, we will begin the sustainability planning process in year one. We plan to use the Lyrasis *It Takes a Village* Guidebook (Arp and Forbes, 2018) and associated toolkit as a starting point, and as a framework for defining our future structure and planning scenarios. To begin developing a plan, the project team will work through the toolkit exercises, and as appropriate, work with the Advisory Board to develop different scenarios for a sustainable service and plan to meet with and discuss those options over the subsequent two years. Our board, bringing a range of different perspectives and experiences, will be able to offer different ways of thinking about sustainability, from how to continue the technical development and improvement of the service, to managing financial viability, to building community and creating enthusiasm for JournalMap. We also hope to leverage the board's connections to others who may provide valuable insight and perspective on how to make an open-source project sustainable in the library/publishing context.

In year two, we will evaluate the potential scenarios. For example, there are models involving non-profit organizations, such as OurResearch and Duraspace, which demonstrate the viability of open-source projects supported outside a university structure. We will plan meetings with representatives from these groups to discuss their thoughts on how viable such a model is. We will also pursue meetings with foundations to obtain insight into the potential for those organizations to provide funding, such as the Arcadia Foundation or the Mellon Foundation, or at the very least, discussions with groups that have obtained funding from these organizations. We will also discuss commercial publishers as possible partners in growing and sustaining the system. Regardless, *our goal is that the automated geolocation of literature, the assignment of metadata, and the capacity to use the API to add and return data will always be available to non-commercial entities for exploration or re-use.* Lastly, we will also consider maintaining the JournalMap service as a function of the UI. Currently, we anticipate the hosting and basic administration of the system to continue without fees, based on the university's current internal service model, and the data curation and management might be incorporated into the work of librarians and staff at the UI Library. Both K-State and UA currently have models in which their databases are generally sustainable by being incorporated into their organizations' regular work.

Our goal for year three will be to formalize a document or series of them that expresses a mission, vision, strategy, and direction for the project, as well as for maintaining the architecture and continuing to grow the content and collections represented in the JournalMap index. The formal plan will be a deliverable from the project and will be posted on the website for reference.

Outreach Objective 3. **Convene and work with an Advisory Board.** The Advisory Board is the primary external feedback mechanism for this project. In our first year, we will convene an advisory group designed to evaluate our progress, provide critique and advice, assist with developing contacts with potential future partners, and help in crafting a long-term sustainability plan for JournalMap. Advisors that we have confirmed include Dr. Jeff Herrick, a soil scientist with the USDA Agricultural Research Service, early advisor to JournalMap and creator of the LandPKS app suite; Meagan Duever, GIS Librarian at the University of Georgia; Dr. Rod Page, Professor at the University of Glasgow and creator of BioStor, a service that harvests location information from the Biodiversity Heritage Library; Andrew Antaya, a researcher at the UA; Dr. Charles Rice, Distinguished Professor at K-State and representative from the Global Research Alliance on Agricultural Greenhouse Gases Croplands Research Group; and Dr. Carolyn Ferguson, Professor at K-State and creator of BioDIS. We will recruit additional members to the Advisory Board with a focus on representing underserved communities whenever possible. We are also seeking to improve representation from individuals with experience managing open-source development projects (e.g., a representative from the team at Washington State University, who manage the Mukurtu CMS project).

The Advisory Board will meet quarterly throughout the project (with the option to visit or discuss more or less frequently as needed or preferred). Agendas will include presentations on technical progress of the project's objectives, discussions regarding problems that may arise, potential outreach and collaboration opportunities, and efforts to grow JournalMap to include additional repositories, including those from the Global South. Discussions regarding the sustainability of JournalMap and the portability of the integrations with K-State and UA will start in the first year but accelerate in the subsequent two years.

Finally, we will provide regular opportunities for our advisory group to rigorously evaluate the service. Based on our goals and outcomes, we will evaluate: (1) the quality of the application/service, including the accuracy of the automation in selecting the right location, (2) the range and diversity of the collections, (3) the precision of attribute data derived from the locations, and (4) the successfulness of the integrations with both K-State and UA, including their portability to other potential partners. The evaluation will include questionnaires designed to elicit their expert opinion as well as our success in relaying the state of the project's performance to them. The project team will develop a report that details the performance of the service on quantitative dimensions (for example, time from submission to indexing, raw numbers of records ingested, accuracy rates of the natural language processing, accuracy of the attribute assignment) to assist in the evaluation.

Diversity Plan. One of the original goals of JournalMap was to aid the transfer of knowledge from well-studied areas to environmentally similar, but less-studied areas of the world (Karl et al. 2013). For many knowledge domains, these less-studied regions occur in the Global South and developing countries in Asia (Wilson et al. 2007, Chan et al. 2011, Martin et al. 2012). In its user interface, JournalMap equalizes representation globally by focusing on the contextual similarity of study sites in its results, not the source or affiliation of the document. JournalMap's goal fits well with the globally-focused missions of the Rangelands Gateway to increase understanding of the importance of global rangelands to economic, environmental, and social health and of the CRDB to support the Global Research Alliance's (GRA) efforts to reduce greenhouse gas intensity and improve crop production for GRA member countries. To this end, we have included a letter of support from Dr. Ladislau Martin-Neto, a Brazilian researcher, who articulates, from a non-US perspective, the value of JournalMap integration with a database like CRDB (see Supporting Document 5).

Through this project we will work to ensure a nationally and globally inclusive approach to engaging contributors to JournalMap and providing access to its content in three ways. First, JournalMap is an open-source web service that offers its metadata for re-use, enabling read access without restriction for all. Second, we will work through the RG and CRDB projects to recruit non-US members to our project Advisory Board. These members will help the project team ensure that design and implementation decisions are accessible and equitable to a global community of users.

Third, we will explicitly build out representation of content from non-Western journals and repositories, and our partners, K-State and UA, have content collections heavily representing studies done in the rural parts of the United States, including Western rangelands and in the landscapes of indigenous communities.

Project Results. The project will produce the following results:

- The redesigned JournalMap will be available as a production-level service through the www.journalmap.org website and as a read/write API for adding content to JournalMap and retrieving it. This will provide researchers will have a website on which to conduct context-similar searching on a large literature set.
- Complete documentation of JournalMap's updated API with code examples to facilitate integration in other applications. We will have created a useful API and geolocation gateway which enables other discovery tools to incorporate this kind of metadata into their own records and systems.
- Web page "widgets" to make simple JournalMap integrations (e.g., map showing an article collection, simple search functionality) with websites easy.
- For the CRDB, K-State's system will push new content into JournalMap directly from the CRDB through the API, increasing the number of records in JournalMap. The CRDB will also feature map-based searching that draws content from JournalMap through the API.
- For BioDIS, K-State will demonstrate the utility of JournalMap for non-literature metadata. They will have a reproducible set of mappings and widgets that can be adapted for other systems.
- The Rangelands Gateway will have an extended data model that includes location and ecological database fields derived from JournalMap. Users will experience an interactive web map with ecological filters, which will enhance the means to which they can interrogate the available literature.
- A sustainability plan for JournalMap developed with the project Advisory Board. While many geolocation bibliographic projects fail because they are manually curated and one time project-based, JournalMap will have a strategy for persistence and will hopefully reduce the need for generalized projects in the future.
- A roadmap for future development of JournalMap developed with the project Advisory Board. We intend for this project to be a starting point for the future of JournalMap, rather than an end in itself. The roadmap will provide us with goals co-produced with representatives of our user community.
- Outreach to potential JournalMap users, integrators, and information science professionals through presentations at professional meetings, publications in scholarly journals, and video tours.

Future Directions. At the conclusion of this project, there will still be opportunities for future development. For example, we perceive that a multi-lingual interface, and the capacity to handle multiple languages in content would be a beneficial addition to the system. However, doing so would be a major effort and falls outside the scope of this proposal. Also, we see opportunities for greater integration with existing tools and systems in the scholarly communications interface, such as ORCID, Unpaywall, data repositories (e.g., Pangea, Dryad), which can positively impact the user experience. However, those integrations would require more time or resources than we are able to devote in this project. Lastly, as mentioned above, we are pursuing separate funding for improving the natural language processing aspects of our application, as we feel that we have identified an appropriate opportunity. Funding from IMLS to develop the core JournalMap services and initial integrations will give us a strong foundation to explore these options.

Jesse Shera (1961) claimed that the purpose of librarianship is to "maximize the social utility of the graphic record." Ranganathan (1931) preceded him by imploring those in libraries to "save the time of the reader." We strongly believe the deliverables of this project – automating geotagging and metadata assignment, providing context-similar searching, an open API to re-use the tagged literature, models for integrating with other bibliographic resources – can produce a positive economy of scale in terms of cost and efficiency for librarians and of access, precision, and recall for researchers that use it. Libraries are the information organizations that often seem to be heavily constrained by institutional budgets and bureaucracy that slow systemic innovation, reduce risk-taking and experimentation, which hamper efforts to change the way their users access information. We feel that this project has an ability to contribute to the library community in a manner that lowers costs, enhances access to knowledge, improves discoverability of spatially-oriented literature, data, and other digital resources in ways that follow Shera and Ranganathan's direction: maximizing the utility of the agricultural and environmental record and saving the time of everyone involved.

Year One	Aug '22	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul '23
Technical Objective 1: JournalMap Service												
Complete the JournalMap primary development (implementation of original logic in new framework, addition of planned enhancements, new resource workflow, testing, security)												
Technical Objective 2: K-State integration of JournalMap with CRDB and BioDIS												
Develop the data pipelines for pushing content to and pulling from JournalMap												
Schema mapping for CRDB/BioDIS												
Technial Objective 3: UA integration of JournalMap with Rangelands Gateway/REM Archives												
Develop the data pipelines for pushing content to and pulling from JournalMap												
Construct ingest scripts for Dspace to JournalMap workflow												
Schema mapping for Rangelands Gateway												
Outreach Objective 1: Grow content and community												
Identify opportunities for indexing new content or partners with content and begin reaching out to them												
Outreach Objective 2: Develop a sustainability plan												
Explore scenarios, models, and establish contacts with experienced individuals												
Project team engage in the "It Takes a Village" guidebook sustainability activities												
Outreach Objective 3: Convene and work with an Advisory Board												
Convene AB meetings, begin discussing Outreach Objectives 1 and 2, present progress on Technical Objectives												
Evaluate progress on year one objectives												

Year Two	Aug '23	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul '24
Technical Objective 1: JournalMap Service												
Complete primary development on JournalMap (as/if needed)												
Work with partners to implement and test integrations												
Bug fixes, refinements, optimizations, cybersecurity improvements												
Technical Objective 2: K-State integration of JournalMap with CRDB and BioDIS												
Begin adding JournalMap-enhanced records to CRDB and BioDIS												
Testing for resource updates, fidelity, consistency and deduplication												
Design web map and functionality for user interface												
Technial Objective 3: UA integration of JournalMap with Rangelands Gateway/REM Archives												
Begin adding JournalMap-enhanced records to the Rangelands Gateway												
Testing for resource updates, fidelity, consistency and deduplication												
Design web map and functionality for user interface												
Outreach Objective 1: Grow content and community												
Index relevant open metadata collections												
Meeting with potential new content providers												
Outreach Objective 2: Develop a sustainability plan												
Evaluate scenarios and meet as necessary with advisors and other contacts developed during first year												
Project team engage in the "It Takes a Village" guidebook sustainability activities												
Outreach Objective 3: Convene and work with an Advisory Board												
Provide feedback and assist with reaching out on other tasks												
Evaluate progress on year two objectives												

Year Three	Aug '24	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul '25
Technical Objective 1: JournalMap Service												
General bug fixes, refinements, optimizations, cybersecurity improvements as needed (ongoing)												
Technical Objective 2: K-State integration of JournalMap with CRDB and BioDIS												
Implement the user interface integration into CRDB												
Implement the user interface integration into BioDIS												
Testing, bug fixes, optimizations (ongoing)												
Technial Objective 3: UA integration of JournalMap with Rangelands Gateway/REM Archives												
Implement the user interface integration into the Rangelands Gateway												
Testing, bug fixes, optimizations (ongoing)												
Outreach Objective 1: Grow content and community												
Curate new content (ongoing)												
Present at meetings and conferences as possible												
Outreach Objective 2: Develop a sustainability plan												
Formalize a sustainability strategy and plan												
Outreach Objective 3: Convene and work with an Advisory Board												
Continue to provide feedback on progress & evaluate sustainability plan												
Evaluate progress on year three objectives and the entire project												

Digital Products Plan

Type

We will produce a JournalMap web application and service, including its website and API. The JournalMap application is written in Python and uses the Flask framework for its web application. The website itself is written using HTML and Angular Javascript and uses Leaflet for the map interface. The underlying database is MySQL. Geoprocessing is done by Mordecai, an open-source application for processing locations from text, and spatial data processing is done through the R library 'rgdal'. All code is stored and shared in a Github repository for the JournalMap project.

In addition, we will produce an improved website for the K-State Croplands Research Database (CRDB) and BioDIS Herbarium database. BioDIS is not currently live but will be migrating to Specify7, an open source, web implementation of K-State Libraries' biological collection data management system. BioDIS is currently running Specify6 and the current data will be transferred to Specify7 by the end of August 2022 and be publicly available again. The CRDB is an application with a SQL database, written in PHP, and uses Apache Solr to provide a faceted search interface.

Further, we will produce a new feature for the Rangelands Gateway (RG) website and an improved data model for the underlying RG database. The RG is a Drupal 9 and PHP application. It relies on a MySQL database. We expect that the web integration will be a Javascript tool using Leaflet for the map interface.

In each case, the technology was selected for several reasons: 1) it fits the skills of the developers at the institution, and 2) it is open source, or openly available for use, which in turn, enables others to work with the projects.

We will also produce a simple project website, where the products will be referenced (for example, the Github repository will be linked to on this page), as well as other kinds of outreach materials that may be generated in the course of the project, such as one-page summaries and video tours/training material produced near the end of the project.

Internal documents, meeting notes, meeting recordings will be stored in the University of Idaho's OneDrive using a shared folder structure for the project team. These documents may be used in the course of producing reports, including reports to IMLS. They will take the form of Word, PDF, text, Excel, and other general office documents.

Finally, we anticipate producing conference or meeting presentations, and possibly, publications. These will take the form of PDF documents or a Powerpoint presentation or similar type of file, depending on the presenters. Publications will take the form of PDFs documents.

Availability

All websites and services produced or modified through this project (JournalMap, CRDB, BioDis, RG) will be publicly available during or by the end of the project. The JournalMap API is meant to be permanently open access for read-only activities. Providing content to JournalMap will require an API key, where anyone can request one, but for security purposes, the ability to be a provider will be managed.

All code produced during this project in Github will be freely available by the end of the project. When the JournalMap service is up and running (anticipated by year two), the code will be available publicly. The code for the K-state and UA integrations will be available at the end of the project.

Papers and presentations will be either published as open access or will be placed within an open repository, e.g., FigShare.

Access

JournalMap and CRDB will be licensed as CC BY-SA-4.0 (no restrictions, give attribution, share alike). The BioDIS and RG systems do have some intellectual property restrictions due to the nature of their original development and that original licensing will not be affected by this project. However, the code produced by this project for web integrations from the JournalMap API will be developed as portable, flexible widgets to be re-used by other institutions and available under CC BY-SA-4.0. The contents of the databases are not expected to possess cultural sensitivities or private or confidential information. Any information related to user accounts will be strictly protected and made secure within the host institution's network security.

Sustainability

One of the major objectives of the project is to identify a sustainability plan for the JournalMap service, to enable its growth, while maintaining its quality. However, it has run since 2014 with minimal maintenance and can continue to do so. At the end of the project, the system will continue to be hosted by the UI RCDS, which subsidizes applications and hosting by University of Idaho faculty. Members of the project team will continue to be able to access the application and modify or work with it indefinitely. If, in time, the application begins to become out of date and lacks resources, the project team will assess the future of the service commensurate with the sustainability plan developed as an outcome of the project.

The CRDB, BioDIS, and RG applications are persistently hosted and supported by their respective institutions, Kansas State University and the University of Arizona. Similarly, they depend on occasional opportunities for innovation and development, although the Rangelands Gateway has recently undergone a major overhaul in its design using internal funding. In each case, robust backup and security support is provided by their respective units, the KSU Libraries, and the UA College of Agriculture and Life Sciences.

Organizational Profile

Mission. "The University of Idaho is the state's land-grant research university. From this distinctive origin and identity, we will enhance the scientific, economic, social, legal, and cultural assets of our state and develop solutions for complex problems facing our society. We will continue to deliver focused excellence in teaching, research, outreach, and engagement in a collaborative environment at our residential main campus in Moscow, regional centers, extension offices and research facilities across Idaho. Consistent with the land-grant ideal, we will ensure that our outreach activities serve the state and strengthen our teaching, scholarly and creative capacities statewide."[1] The University of Idaho is authorized by its inclusion in the Idaho State Constitution, Article 9, Section 10.[2] It has been continually accredited by the Northwest Commission on Colleges and Universities, since 1918.[3]

Organization. The University of Idaho Library is organized with its Dean and operations reporting the Provost and Executive Vice-President (EVP) of the University.[4] The EVP reports directly to the President, who in turn reports to the Board of Regents and State Board of Education.

Service Area. The University of Idaho serves the state of Idaho, and its student body comes from all 50 states and from 59 countries. It has an enrollment of 10,791 students. The U of I's residential campus (Moscow, ID), educational centers in four other Idaho locations (Coeur d'Alene, Boise, Idaho Falls, and Twin Falls), as well as the research/extension facilities and programs in 42 of 44 Idaho counties employ over 2,252 faculty and staff in support of students and community members throughout the state. Through U of I's outreach efforts, we build partnerships with stakeholders across Idaho and beyond to create opportunities that foster knowledge and education for youth, college students, and adults. The U of I's community includes our students, faculty, and staff; those who live in Idaho; and individuals throughout the U.S. and the world. As a land-grant university, we recognize that our work cannot be limited to our campus boundaries and contexts; as such, we seek to address the local issues that are relevant to our own communities as well as the societal needs and global issues that affect us all. Through our teaching, research, and outreach, the U of I "brings Idaho to the world" by building contextually competent partnerships with local, national, and international collaborators. These partnerships help transform the educational opportunities of our students and allow researchers to leverage the unique experiences and knowledge of all collaborators to seek innovative and sustainable solutions to these complex challenges.

History. Founded in 1889, the U of I's land-grant mission is visible in our engagement with the community, transformative educational experiences and opportunities, and our drive to "shape the future through innovative thinking." As the largest library in the state of Idaho, the U of I Library reflects this land-grant mission by "bringing the world to Idaho and Idaho to the world." In the past 5 years, the library has developed a culture of, and dedication to, innovative library practices, adding a Data and Digital Services Department, a makerspace, opening a digital humanities center, and both data services center and collaborating across campus to provide services in the areas of GIS, data curation, and data sciences. Members of this grant team are associated with the Library's Data & Digital Services Department, the College of Natural Resources, and the Research Computing and Data Services unit in the of the Office of Research and Economic Development.

---

[1] University of Idaho website. *Our Mission, Vision, and Values.* https://www.uidaho.edu/about/mission-vision-values

[2] Idaho State Constitution, Article 9, Section 10. *State University — Location, Regents, Tuition, Fees, and Lands*. https://legislature.idaho.gov/statutesrules/idconst/ArtIX/Sect10/

[3] Northwest Commission on Colleges and Universities. *Member Directory*. https://nwccu.org/member-institutions/directory/

[4] University of Idaho. *University Organization Chart.* https://www.uidaho.edu/about/org-chart