

Legal Agreements Governing Archiving Partnerships: The NGDA Approach

Julie Sweetkind-Singer and Tracey Erwin, Stanford University; Mary L. Larsgaard, University of California, Santa Barbara (USA)

Abstract

The National Geospatial Digital Archive, funded by the Library of Congress, is a partnership between the University of California, Santa Barbara and Stanford University with the purpose of providing long-term preservation of and access to geospatial data and imagery. Geospatial content in the public domain as well as that which is not is being accessioned into our collections. While public domain data may often be accessioned without having to seek explicit permission from the agency or group creating and/or hosting the materials, private data should be captured with a clear understanding of the rights and responsibilities of the stakeholders, be they the content owner or the university. In addition, the two collecting nodes (and any future nodes) must agree upon standards of conduct in order to trust that any content brought into the network will be managed well now and into the future.

Joint agreements have been written by UCSB and Stanford that codify all of these relationships. This paper will lay out those agreements, which include the Content Provider Agreement, the NGDA Content Collection Node Agreement, and the NGDA Content Collection Node Procedure Manual. These three agreements provide the legal and structural framework of the NGDA allowing the stewardship network to grow over time with the understanding that each new partner will meet the standards necessary to provide secure long-term preservation of the content.

Background

In December of 2000, the United States Congress authorized nearly \$100 million to fund a national effort to “set forth a strategy for the Library of Congress in collaboration with other federal and nonfederal entities, to identify a network of libraries and other organizations with responsibilities for collecting digital materials that will provide access to and maintain those materials.” [1] Congress mandated that the money be used to develop policies, protocols, and strategies for the long-term preservation of “at-risk” materials. The Library of Congress formed the National Digital Information Infrastructure & Preservation Program (NDIIPP) to administer and oversee the awards. [2] The first round of funding was announced on September 30, 2004 with awards going to 8 partnerships, including a joint award to the University of California, Santa Barbara (UCSB) and Stanford University (Stanford) for the formation of a national geospatial digital repository dedicated to collecting and preserving geospatial data and imagery across a broad spectrum of geospatial formats. The repository was formally named the National Geospatial Digital Archive (NGDA).

UCSB and Stanford quickly set out to write collection development policies to guide the acquisition of content. Geospatial data and imagery are created at an astoundingly high

rate, requiring judicious selection of content closely aligned with the research needs of the faculty, researchers and students. While much geospatial content, especially in the United States, is in the public domain, it was clear from the start that contracts would have to be written governing the acquisition of content from copyright and/or license holders. The two collecting partners, or nodes, also needed to be bound together to assure compliance with agreed upon principles of conduct when accessioning and, perhaps, sharing data across the network of nodes.

Content Provider Agreement

The Content Provider Agreement (CPA) is a document that governs the rights and responsibilities of the copyright holder or one who has the full authority and right to deposit the materials with a node of the NGDA and of the node that accepts the content. [3] The agreement consists of three parts: the main body, Exhibit A, and Exhibit B. Although UCSB’s and Stanford’s agreements are essentially the same, the document has been written in such a way to provide flexibility at the individual node level. The main body allows each university the ability to customize how they receive their content, how the content provider can get a copy of the materials they deposit, and under what circumstances the contract would be voided. Exhibit A outlines the content itself. Exhibit B details the users, uses, and management of the materials and must be the same across all of the nodes. Content Providers are assured that no matter in which node they deposit their content and whether or not it is shared, they can expect the same policies will govern their materials.

Our work on the document began with a review of other available depositor agreements, including the California Digital Library (CDL) /UC Libraries Digital Assets Agreement. [4] Other NDIIPP partnerships were also creating depositor agreements, which they shared with us, such as the Data-PASS project lead by the Inter-University Consortium for Political and Social Research (ICPSR). [5] We looked for areas of agreement across the documents, for clauses that defined the rights and responsibilities of the different parties, and at the basic structure of the agreements. In draft form the NGDA agreement was vetted by a group of NDIIPP members at the national meeting held in the summer of 2006.

Main Body

The main body of the CPA begins with a standard explanation of intent and definition of terms. It provides for a grant of license with terminology that is broad enough to cover changes in technology over time. The grant of license states, “Content Provider hereby grants to Custodians a paid-up, non-exclusive, world-wide, transferable license to reproduce, prepare derivative works of, distribute, perform publicly, display publicly, digitally transmit and otherwise use the Licensed Materials at no cost in any

media now known or hereinafter created in accordance with the terms of the Agreement.” The NGDA does not intend to function as a dark archive, hence the wording allowing for the public display and performance of the materials.

The content provider agrees to provide the node with one copy of the content, which may, but does not have to be, shared with other nodes in the NGDA for preservation purposes. A clause in the Agreement allows the collecting node to remove the materials for any reason from its archive given it provides written notice to the depositor. Materials that are found to be in violation of copyright must be removed within 48 hours from any hosting node and returned or destroyed within 15 days of the receipt of the request.

Every node must agree to be bound by the terms in Exhibit B, explained in detail below, or to use terms that are substantially similar. It is agreed that the provisions in Exhibit B will change over time. The question arose as to whether or not the content provider had the right to remove the content if the provisions in this section changed. Both nodes agreed that notice had to be given to the content providers when “substantial” changes were suggested. They would be given the opportunity to withdraw their content only if the proposed changes were unacceptable. “Substantial” is purposely left undefined in the Agreement, allowing the nodes the flexibility to decide this on a case-by-case basis with the other nodes and the content providers.

Three reasons are given as to why the Agreement would be terminated: the copyright term or proprietary licensing rights expire; changes are made to Exhibit B that are unacceptable to the content provider (say, we decide we want to sell reproductions of their content); or if there is a breach of copyright or license where it is discovered that the content provider was not legally able to deposit the content with a node in the first place.

Exhibit A

Exhibit A is a form with blanks to be filled in created jointly with the content provider when the materials are identified for deposit. It should include information about the exact nature of the collection with a detailed description of the content’s scope, the number of files to be received, the file formats to expect, the metadata for each file, and the rights governing the collection. A section on transmission of the data is also included. Will the materials be delivered on hard drives or via file transfers? How often will materials be deposited, once or as an ongoing stream of submissions? How should the owner of the content be identified and does the collection have a formal name? What are the rights issues surrounding the collection? Consideration is also given to the wishes of the depositor as to the end use of the collection. For example, the depositor may want a node to mediate all contact between those wanting to use the data and/or imagery for publication, handling the permissions and access to the materials. The Content Provider may also stipulate the methods by which communication takes place and the nature of that communication. They may wish to know how often the collection is accessed, when it is being shared with other nodes, and if access has been interrupted due to routine maintenance of the archive. By not providing a strict structure for Exhibit A, the node and the content provider are given the maximum flexibility to create a document that meets the needs of both.

Exhibit B

Exhibit B includes three sections: Authorized Users of Licensed Materials, Authorized Uses of Licensed Materials and Management of Licensed Materials by Custodians. This section was the most difficult to craft and took the most negotiation between the UCSB and Stanford library staff and lawyers. All wanted this section to be acceptable to both a public and a private university. As we were writing this agreement, anything the group considered essential for all nodes had to go in this section. Only if all existing nodes and future nodes agreed to these provisions could content be shared across the collecting network.

Four groups are included as authorized users of the collection. The first class of users is students and employees (including faculty, staff, affiliated researchers, and independent contractors) of any holder of the NGDA content regardless of the physical location of the person. The second group is any walk-in patron to one of our libraries. The third segment of users is the Library of Congress. A provision of the contract signed with the Library of Congress states that they may request a copy of the content be delivered to them at which point it would become the physical property of the United States Government for the collections of the Library of Congress. This provision has been exercised by the Library of Congress with both UCSB and Stanford delivering archived content for, at this point, dark archiving and redundancy rather than immediate access. The Library of Congress retains the right to make the content accessible on their premises, but has yet to do so. Finally, future access to the general public is allowed. Preservation environments have been created at both nodes with a future goal of providing access to all materials accessioned into their respective collections. At this point, access to any of the user classes is limited and often mediated by library personnel.

The authorized uses of the materials section states that the nodes “may make use of the Licensed Materials in any manner consistent with the United States and international law including, but not limited to, any fair use exceptions set forth in copyright law.” The section then goes on to state that the materials may be used for any research, educational, or other non-commercial purpose. The nodes are allowed to make multiple “ephemeral” copies of the entire set of materials to be maintained as backup or archival copies (rather than the three set out in the Digital Millennium Copyright Act). End users are allowed to download and digitally copy or print a reasonable portion of the materials. The nodes are allowed to charge a fee to end users to cover copying or printing costs. A reasonable amount of the content may be used in course packs and be put on electronic reserve in connection with specific courses of instruction. The content holders have the right to incorporate metadata into a publicly accessible catalog. Insubstantial amounts of the content may be transmitted to third party colleagues for personal, scholarly, educational, or scientific use, but in no case for re-sale. Finally, requests for access via interlibrary loan will be granted as long as they are in compliance with current United States copyright law.

The custodians of the content, or nodes, are granted a series of rights that make management of materials transparent to the depositor. If the materials have been provided to another custodian, the original place of deposit can request its return for any reason. If there is copyright infringement involved, the materials must be removed from any public site within 48 hours.

Custodians agree to use reasonable practices and standards for the long-term preservation of the content. They will do their best, as they would with any other content, to prevent unauthorized access to the materials. The depositor will be notified in a timely manner if a security breach occurs. Access to the content by authorized users will be continuous except for periodic maintenance or unanticipated hosting difficulties. The custodians will notify the users as to the terms of use and will credit the copyright holder. Finally, the primary custodian will identify for the depositor the names of the other nodes which hold their content.

Taken together, these three parts (the main body, Exhibit A and Exhibit B) allow each node to craft a document that meets the different needs of the institutions and the depositors, while providing a structure that allows for the federated network of repositories to work in conjunction to ensure the redundancy of the collected content.

NGDA Content Collection Node Agreement and Procedure Manual

One of the main goals of the NDIIPP program is to accession “at-risk” digital content for long-term retention and use. The likelihood that any content will survive into the future is increased if it is replicated numerous times and then distributed geographically across a number of secure repositories, be they all within one institution or across numerous institutions. In order to facilitate this distribution, the nodes within the network must be in agreement as to the standards by which the materials will be managed. The NGDA Content Collection Node Agreement serves this purpose for the NGDA. [6]

From the start, the Node Agreement was conceived as a two-part document, the agreement itself and an accompanying procedure manual. The goal was to create a general structured agreement that would bind the nodes together with an understanding of the obligations of the participating parties. In tandem with this agreement would be a fluid and changeable procedure manual that would allow for flexibility in the way the NGDA chooses to carry out the functions of the federation described in the Node Agreement. The Node Agreement covers the responsibilities, governance structure, removal of content, and discontinuance of a node. The procedure manual lays out for the group the communication mechanisms, meeting frequency, identification of potential new nodes, policies for joining and leaving the group, and information about how content is acquired and removed from the archives.

Node Agreement

The Node Agreement specifies the expectations and obligations for participation as a content collection node of the NGDA and describes the features of the cooperative venture. It requires that each node construct a depositor agreement that must be used when taking in new collections. The new nodes may choose to follow the form adopted by UCSB and Stanford or create their own, but must include the use terms or terms that are substantially similar to those used in Exhibit B of the Content Provider Agreement. It is understood that these use provisions only apply to copyrighted or licensed data acquired by the repository, not to public domain materials, which are free of use restrictions.

Four main features define a content collection node of the NGDA. The node must have a collection development policy for geospatial data and imagery. Three policies have already been developed by the NGDA that serve as examples for other institutions – a general policy for the federation as a whole [7] and two policy documents that have been tailored to meet the research and teaching needs of UCSB and Stanford. [8,9] Each node must have an institutional mandate to collect. It is critical for long-term preservation that the repository have backing from its parent institution to accept, store and preserve content now and into the future. The node must agree to archive the content it collects. Finally, the node agrees in writing to be part of the NGDA collecting network.

The governance structure of the federation will go through at least three distinct phases. The first phase occurs during the life of the Cooperative Agreement with the Library of Congress. During this phase, which is slated to last until December 2009, UCSB and Stanford operate as equal partners and the sole decision makers as to policies and procedures. Their decisions will be made by consensus unless a decision violates an existing university policy. If the two cannot reach agreement, the dispute will be resolved by the Associate University Librarians or University Librarians at their respective schools. The second phase of governance begins when the award period has ended and more nodes join. During this phase, decisions will be made by consensus of the group. The third phase begins when more than five nodes are members of the federation. At this point, the governance structure will be revised to meet the needs of a larger group.

The nodes will convene to discuss topics of mutual interest including reviewing responsibilities and operating procedures, acquisition and removal of new content, adding new or removing exiting nodes, procedure manual updates, and any other pertinent matters. The members are responsible for keeping the procedure manual up-to-date and for meeting once a year to discuss relevant issues to the NGDA whether in person or virtually.

Two indemnification clauses are included in the Node Agreement. The first stipulates that a node will not obligate another node to terms outside of or in addition to those detailed in Exhibit B of the Content Provider Agreement. The second states that if a node accepts copyrighted or licensed materials that it represents and warrants to the other nodes that it has a signed agreement which states that the depositor has the right/authority to deposit the content.

It is acknowledged that content may have to be removed from a node. In such a case, the node removing the content agrees to contact the other nodes in hopes that someone else will be able to accession the content. For example, a node may choose to host only the most current data, say elevation data from the United States Geological Survey, due to the fact that the current data are more correct or complete. The data set may be so large that only one iteration of it can be stored and the most correct or complete set is desired. If the removal of the content is due to copyright or license violations, all of the nodes hosting and serving out the content agree to remove access to the materials within 48 hours, but may keep an archival copy for as long as the dispute remains active.

A node may choose to terminate their participation in the NGDA or they may be removed because of violations to the use

clauses in Exhibit B, the terms of the Content Provider Agreement, or the terms of the Node Agreement.

Procedure Manual

As noted, the procedure manual [10] has been designed as a working document giving the group the flexibility to adapt to the changing needs of the members as standards, methods of communication, and institutional mandates and responsibilities change over time. The manual will be amended as necessary and reviewed on at least a yearly basis when the group members meet. It is not intended that any change to the manual would require a node to depart the federation. Any changes to that effect would be carefully considered. At this early stage, communication will take place through the use of an email reflector with documentation housed on a private wiki and a public Web page. [11]

The meeting specifics were modeled on the successful and simple structure used for over thirty years by the University of California / Stanford Map Libraries Group. The NGDA group agrees to meet at least once a year either in person or virtually depending upon funding. The nodes take turns hosting the meeting with the hosting node bearing the responsibility for organizing the event and setting the agenda. The previous host takes and distributes the minutes of the meeting.

Identification of new nodes is critical if “at-risk” content truly is to be retained for the near and far-term future. Geospatial data and imagery exhibit certain characteristics that make it challenging to collect. While some data are small in terms of file size, such as information about roads, waterways, or administrative boundaries, data in a grid format such as satellite imagery tend to be quite large. The issues around archiving large datasets are complex enough that it may force many nodes to collect content that is specific to the needs of their users. Because of this tendency, broad geographic coverage will be more likely to happen if nodes are recruited from geographically dispersed regions. Both UCSB and Stanford have decided to collect content at the greatest depth for the cities and counties in their immediate area, but allowing for broad coverage of California. It is expected that the existing nodes will use their known network of contacts to suggest other nodes for participation as well as solicit participation through relevant professional organizations and email lists.

New nodes will be vetted by the node that has invited them to consider participation or that they contact about the network. The vetting will include providing the potential node with all relevant documents created by the NGDA, staffing requirements, and hardware and software implications of being a node. The nodes must write their own Collection Development Policy and Content Provider Agreement, and sign the Content Node Agreement. The new nodes agree to follow current best practices for trusted repositories for hardware, software, staffing and self-auditing. The new node is then approved by the governing body of the NGDA.

The mandate of the NGDA is to offer stable, long-term preservation for geospatial data. As such, stewardship of content is of paramount concern. While the expectation is that nodes will continue to be members of the NGDA for the long term, it is recognized that in some circumstances a node may no longer be able to participate or that it may not be able to fulfill its obligations. In either case, the node must notify the other nodes in writing. The departing node is strongly encouraged to offer its

content to another node in the network. If the content is copyrighted or licensed and therefore a Content Provider Agreement has been signed with that specific node, a new document would have to be signed with the node accepting the materials.

When new content is collected, the nodes agree to keep an up-to-date list by collection name, or layers, of what has been selected and served out. Acquisition of collections has been at such a pace that keeping an accounting of the content has not been an unreasonable task. This may change and the policy amended if a node decides to harvest content from the Web or take in large amounts at a steady rate. All other nodes are notified when major new collections are accessioned, such as when UCSB acquired the Citipix Collection of 500,000 aerial images of the United States spanning 65 metropolitan areas.

Conclusion

Over the four and a half years of the NGDA partnership between USCB and Stanford, it has become apparent that the stewardship of digital content requires broad thinking across all of the domains of library expertise. During that period of time, libraries have become much more sophisticated in their understanding of the legal framework that is required to collect content that is not in the public domain. This knowledge of digital acquisition and rights management helped to inform the work done on the Content Provider Agreement and the Content Collection Node Agreement.

The NGDA partnership is now entering the final months of its contractual obligations to the Library of Congress and will set out on its own to create a robust and active network of collaborators intent on rigorous acquisition and preservation of geospatial content. The agreements described in this paper will most certainly be tested and adapted to the needs of the members of the federation changing as needs and technologies change.

References

- [1] House Report 106-1033 - Making Omnibus Consolidated and Emergency Supplemental Appropriations for Fiscal Year 2001. Accessed online at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=106_cong_reports&docid=f:hr1033.106.pdf.
- [2] Digital Preservation: National Digital Information Infrastructure & Preservation Program: A Collaborative Initiative of the Library of Congress. Accessed online at: <http://www.digitalpreservation.gov/>.
- [3] NGDA Content Provider Agreement. Accessed online at http://www.ngda.org/research/Rights/Stanford_NGDA_Contentprovider_102307final-1.pdf.
- [4] California Digital Library (CDL) / UC Libraries Digital Assets Agreement, CDL/UC DASA 01-17-06. Accessed online at <http://www.cdlib.org/inside/diglib/submission.agreement.pdf>.
- [5] Data-PASS Data Deposit Agreement. As approved by the Operations Committee, May 18, 2006. Accessed online at <http://www.icpsr.umich.edu/DATAPASS/pdf/deposit-agreement.pdf>.
- [6] NGDA Content Collection Node Agreement. Accessed online at <http://www.ngda.org/research.php>.

- [7] NGDA Collection Development Policy, November 1, 2006. Accessed online at http://www.ngda.org/research/Collections/NGDA_Collection_Development_Policy_11_06_final.doc.
- [8] Collection Development Policy for the National Geospatial Digital Archive: Node, Map and Imagery Laboratory (MIL), Davidson Library, University of California, Santa Barbara, May 2007. Accessed online at: http://www.ngda.org/research/Collections/UCSB_CDP_05_07.pdf.
- [9] Collection Development Policy for the National Geospatial Digital Archive: Node, Branner Earth Sciences Library and Map Collections (Branner), Stanford University, May 2007. Accessed online at http://www.ngda.org/research/Collections/SU_CDP_5_07.pdf
- [10] NGDA Content Collection Node Agreement Procedure Manual. Accessed online at <http://www.ngda.org/research.php>.
- [11] The National Geospatial Digital Archive: A Cooperative Initiative. Accessed online at: <http://www.ngda.org/index.php>.

Author Biographies

Julie Sweetkind-Singer is the Head Librarian at the Branner Earth Sciences Library and Map Collections at Stanford University. Her subject specializations are maps and GIS. She is currently Stanford's project lead on an NDIIPP grant from the Library of Congress focusing on long-term preservation of geospatial data. She has worked at Stanford since May 2000. She received her MLIS from San Jose State University and her MBA from the University of Colorado, Boulder.

Tracey Erwin is the Geospatial Librarian for the National Geospatial Digital Archive (NGDA), a grant-funded project of the University of California at Santa Barbara and Stanford University. Erwin has worked on collection development, selection and acquisition of data and imagery, and legal agreements for the Archive since 2005. She received her MLIS from San Jose State University.

*Mary Lynette Larsgaard is the Head of the Map and Imagery Lab, Davidson Library, at the University of California at Santa Barbara. The Map and Imagery Lab has a collection of remote-sensing imagery and maps of approximately 7 million items and is the largest of its kind in any university library in North America. Larsgaard has published extensively in the field of geospatial data in libraries, most notably with a widely used text, *Map Librarianship: An Introduction* (now in a third edition, published in 1998 by Libraries Unlimited); she is also a coeditor (with Paige Andrew) of the *Journal of Map and Geography Libraries/Geoscapes*. Her specialties are cataloging/metadata creation and twentieth-century and more recent topographic and geologic maps.*