Introduction

Agricultural researchers and public users world-wide have ever-increasing access to a plethora of online resources, including "grey literature" not published in commercial or society publications.   In the United States, most of the agricultural grey literature is contributed by land grant university Agricultural Experiment Station (AES) and associated Cooperative Extension Service (CES) units. Their publications and resources are often created in a digital format and presented online even when a parallel distribution option is maintained for a print or hard-copy version. While increasing the amount of free online agricultural information is critically important, many issues must be addressed to ensure that users can both easily locate and retain long-term access to this information. It is also imperative to create a user-friendly process that enables researchers, extension agents, and farmers to more easily contribute valuable digital agricultural content to online venues, while at the same time centralizing access points to enable faster, easier access by all users. Most importantly, literature being born-digital[1] today must be described and archived in a manner that preserves access for searchers ten, fifty, or one hundred years from now, enabling them to learn from today's lessons rather than repeat them.

A 2006 pilot project at the University of Minnesota (UMN) demonstrates how to enhance current access and preserve long-term access to this body of literature. The goal of this project was to investigate issues, policies and processes involved in depositing born-digital Minnesota AES (MAES) and University Extension Service (ES) publications[2] into a recently launched local institutional digital repository, the University of Minnesota's Digital Conservancy (UDC). Many discussions were held locally and with national colleagues to define and discuss policy issues surrounding content scope, metadata standards, access, copyright, and long-term maintenance of the resources. Also discussed are issues surrounding the formation of bridges between this collection and the many similar agricultural information silos around the country and the world, to further the goal of forming a single national or international repository and portal for agricultural information.

The Problem: Here Today, Gone Tomorrow

A major disadvantage to the new online publishing paradigm, one rarely addressed directly in the literature, is the potential loss of access for future researchers to digital content being published online today. Some may ask whether older literature is really that important in this age where it seems any information need can be met with an Internet search. Don't users need and expect the most up-to-date, current information? Certainly the general public normally seeks the most current (and thereby theoretically the most scientifically accurate) information, but one of the founding tenets of scientific research is the importance of building upon and adding to the knowledge gained and documented in the past. Although peer-reviewed society and commercial journals document the progression of high quality basic research, the practical applications of agricultural research are documented primarily in the grey literature. This practical component is a critical knowledge output for the world's agricultural community, yet much of this literature is now in grave danger of being lost to future generations. As individuals and as a society we have long relied on others to selectively retain our intellectual output and ensure that future generations have access to the historical record, in all fields from arts and humanities to the sciences. Those "others" (primarily libraries and museums) are relatively new at addressing selection and preservation issues for born-digital resources.

---

[1] Word Spy (http://www.wordspy.com/words/born-digital.asp) defines "born-digital" as "relating to a document that was created **and exists** only in a digital format. " For the purpose of this project, I have defined "born-digital" as related to those resources whose first incarnation is in electronic format. While these objects may also exist in hard copy or print form, the primary archival copy is the digital file. This may include html web pages, PDF documents, images, etc.
[2] Minnesota Agricultural Experiment Station (MAES) and University of Minnesota (UMN) Extension Service units are administered jointly. The UMN Extension Service will be referred to herein as "Extension" for brevity. Extension publications were the initial focus of this project, although the UDC "Extension" collection will also include MAES publications.

In the meantime we are continually losing content and/or access to resources. "It's a fallacy that if something is on the web, it will stay there," says Kristine Hanna, director of web archiving services for Internet Archive. "It's not like a piece of paper you put in a file folder and it will be there forever." (McClure 2006) Many webpages and associated pdf and image files disappear into the great digital black hole titled, "Error 404: Not Found", indicating that the resource has been deleted, withdrawn, replaced by more current information, or posted by an organization no longer in operation. At land grant institutions, web-site managers for departments or agricultural colleges have a mandate to keep their sites current, and web-only content is regularly withdrawn, updated, or overwritten – typically without being preserved or archived in a way that enables and ensures long-term access. As noted by Dr. Karen Lilley, web services manager for UMN Extension, "Web-only materials can be lost when they're deleted or updated online, in an effort to continually keep the site current. Unlike print materials that are sent to the Libraries for long-term access, no similar process is in place to save and share web-only items for future research."

Some researchers self-post resources on departmental websites, where they are also likely to disappear over time. Researchers abandon or cease to maintain web resources as other priorities take precedence, or as they become bored with the process, see little return on their time investment, move to other research interests, transfer to other institutions, or retire. We have all heard stories about valuable information stored in individual databases or on hard drives destined for obsolescence, or burned on CDs stored in file drawers. That literature is lost to future researchers more surely than if it were a print document in a folder in their home or work office[3]. As noted in a University of Rochester study, "Most faculty members already assume—usually erroneously—that their materials are rigorously backed up and permanently accessible, even though many have lost data and documents due to crashes or software obsolescence" (Foster and Gibbons 2005). Clearly, grey literature is particularly vulnerable, including the vast stores of agricultural information published primarily online by many academic institutions in the United States. At a minimum, the project being completed at the University of Minnesota will address this problem for this one institution, with the hope of also providing a preliminary roadmap for other institutions and organizations to follow as we move into an e-Agriculture future.

From Silos to Cooperatives

It is very difficult to perform a reasonably comprehensive search for extension or experiment station publications about a certain topic (e.g. conservation tillage) even using digital literature that is deemed "current", with resources fragmented into silos that bear no connection or relationship to similar silos produced and maintained elsewhere. Currently CES and AES content users having to visit each institution or repository to gather different content. These silos tend to be established by the Agricultural college at a specific institution in each state, at best, and at worst (and more commonly) authors post to individual websites or to their departmental website. The information is also fragmented in virtual space by the fact that topic categories or descriptive terms applied to the content in each silo tend to be different, thus reducing the ability of the user to easily cross-search multiple websites or silos. Although some states' extension publications, including born-digital, are now so comprehensively well indexed that they could almost be considered "white" literature (see the Colorado Agriculture Bibliography Project at lib.colostate.edu/research/agbib/), most others are very difficult to locate and would be considered more grey literature. Search engines are not the answer for searching across silos, as they do not provide the "one-stop" searching functionality that many believe exists, even for the literature that is considered current.

Search engines in a broader sense have several limitations. First, they cover the Web much less comprehensively than many users might assume. In 2002 the extent of the Internet covered by any one search engine was estimated at 40% (Hutchinson and Greider 2002). A more recent study indicated that by 2005 Google was indexing 68.2% of an estimated 11.5 billion surface websites (Gulli and

---

[3] Print is to this day the most stable archival format, lasting hundreds of years, while digital formats change so often it is difficult to keep hardware and software in place to read these resources.

Signorini 2005). This highlights another limitation of popular search engines, namely that these "surface" web sites they cover comprise only a fraction of the total amount of information actually available in digital form. One study found that, "the sixty known, largest deep Web sites contain data…roughly forty times the size of the known surface Web. These sites appear in a broad array of domains from science to law to images and commerce [and] roughly two-thirds of these sites are public ones, representing about 90% of the content." (Bergman). Search engines are now beginning to utilize new approaches to crawl or search the enormous amount of content buried in the "deep Web" (Wikipedia contributors).

Searchers seeking to navigate the massive amount of information produced online need more help limiting their search by subject and context, e.g. application of the resource need. Having the ability to restrict their search to one body of agricultural information would serve the needs of many of these users without restricting that access for others. This situation is improving all the time with the implementation of portals that enable users to perform searches of many different silos, e.g. the Plant Management Network's new Institutional Search service that enables cross searching of multiple institutional extension sites (http://www.plantmanagementnetwork.org/). However, the PMN is a fee-based (albeit non-profit) service, and these types of initiatives also do not address the issues of archiving and description to ensure access to this information far into the future. Selecting and applying some minimal standardized ontology to this content will maximize the ability to pull or harvest the information from all of the silos into one central cooperative and greatly enhance the effectiveness of searches performed comprehensively across silos.

The scope for the UMN Extension archives project was to recommend roles, responsibilities, and processes for the integration of born-digital extension publications into the Digital Conservancy of the University of Minnesota Libraries (UDC at http://conservancy.umn.edu/). The UDC is charged with "the development of a digital institutional repository to capture and manage a wide range of electronic information that is created daily at all levels of the institution." The outcome of this project is in essence a new silo, but this silo is being designed with the important, critical awareness that the content must adhere to international standards, and readily lend itself to OAI metadata harvesting so it can easily be incorporated into a large cooperative collection to be built out of many other silos of digital agricultural information. Included are recommendations for possible strategies and delineation of the issues that need to be addressed before we can form bridges between this collection and the many similar silos around the country (e.g. Michigan State University's Knowledge Repository). Underlying this effort is the longer-term vision for the eventual inclusion of all metadata available for CSREES, NDLAR/USDA, and other agricultural resources (including older digitized print as well as born-digital resources) into national and/or international digital agricultural repositories.

Background-Literature Review

The importance of a global information network for agricultural information in all formats has been documented for many years by librarians (Gwinn 1994) and the extension community (Heatley and Gardner 2003; Schmitz 2005), although there is a degree of disconnect between faculty, AES administrators, and the library community. In a recent survey of extension faculty and administrators, the author notes the importance of "mak[ing] more agricultural research findings available in new and interpretive way[s] to a broader constituency" but does not mention the need for archival access, or the ways in which institutional repository managers might ease the depository burden for their faculty (Kjelgren 2007). Results of a recent survey conducted as a follow-up to the 2007 World Summit on the Information Society (WSIS) affirmed that "improving information and communication processes should be the top priority in furthering global discussions and related actions in e-Agriculture," noting that "almost half of stakeholders working in agriculture identified e-Agriculture with improvements in processes such as information dissemination, access, and exchange; and communication, participation, and network/community-building activities…"(World Summit on the Information Society (WSIS)).

National Extension Service administrators noted several years ago that digital technologies and distance education offer limitless possibilities to engage multiple audiences, expanding Extension's educational

role as a "brand name" quality source for unbiased, research-based information and education" (Extension Committee on Organization and Policy [ECOP] 2002). Vicky Reich (LOCKSS) contributed a comment to the Chief Collection Development Officers (CCDO) of Large Research libraries Discussion Group at an ALA meeting in 2006 that, "born digital materials are of much greater risk; collection and preservation priority should be placed on materials that are not replicated in print" (Reich 2006). The Digitization Policy Task Force of the American Library Association's Office for Information Technology Policy introduced nine Principles for Digitized Content that echo many of the themes addressed here, including the importance of digital content, the need for collaboration and communication, and preservation and standards issues (Liss 2007). The importance of addressing interoperability, infrastructure development, and coordination of collection development have been noted by individuals closely involved with the development of a National Digital Library for Agriculture (Heatley and Poley 2006)

Many initiatives are under way nationally to build local silos of content that could be incorporated into a collective data repository, e.g. the Texas Cooperative Extension collection in the Digital Repository at Texas A&M University (http://txspace.tamu.edu/handle/1969.1/3683). At the national level, the National Agricultural Library has been working for years with other national organizations to develop the concept of a National Digital Library for Agriculture (National Agricultural Library 2001; National Agricultural Library, ERAL Committee 2006; Young 2004) At the international level, efforts to address standards and collections issues are underway, as exemplified by the Food and Agriculture Organization of the United Nations' AGRIS Open Access model to "promote the availability of online content, such as that of grey literature which is not available through commercial distribution channels but significantly contributes to agricultural research and development" (Subirats et al. 2007). The time is certainly ripe for the University of Minnesota to join these efforts, and develop a silo of UMN Extension content that can be incorporated into these broader efforts, while also providing a roadmap for informing their development.

Methodology and Project Steps
The first and to some degree most critical step for a successful outcome in this project was to secure the confidence and support of UMN Libraries and UMN Extension administrators at the highest level. The next step was to identify appropriate collaborative partners in the local (UMN) and national community. Correctly identifying and including all stakeholders in the process was critical to successfully address and clarify issues including content scope, metadata standards, access, copyright, and long-term content maintenance. Identified stakeholders included Extension Service administrators, researchers, and IT personnel, colleagues in national land-grant libraries, AES, and CES units, the University's Digital Conservancy (UDC) team[4], and additional Libraries administrative, digital collections, and technical unit staff. As each issue was addressed, the project coordinator identified appropriate staff involvement, scheduled, and facilitated meetings.

One unforeseen issue encountered during this phase of the project was communicating with other land-grant universities to assess the state of their extension collections and obtain feedback about potential issues and opportunities. Identifying the appropriate contacts either in extension or in the libraries was surprisingly difficult due to the different existing document routing processes and administrative structures in place (or not) at each institution. The extension and/or library staff contacted sometimes had little or no knowledge of existing archival processes or collection status of extension documents, in any format. Attempting to determine content coverage via perusal of extension and library websites at each institution was also problematic. Factors included:
- Difficulty establishing scope of coverage due to varying cataloging policies. For example, some libraries catalog extension series only by the series title, while others catalog each piece separately;

---

[4] The UDC Management and Working Groups are comprised of representatives from across the Libraries including Co-Directors Philip Herold (Academic Programs-Sciences) and Beth Kaplan (Academic Programs-Archives), and administrators and staff with expertise in repository architecture and metadata, collection development, database/web development, and copyright.

- Varying policies at each institution in terms of what extension content is collected by the libraries, what is included in a fee-based extension "store", and what resources are offered freely to the public on a general extension website.
- Varying levels of website sophistication and structure at each institution's extension and/or AES website (some institutions have combined these two arms administratively), presenting a new and time-consuming learning curve for identifying content at every institution investigated.

In spite of the strong cooperative spirit demonstrated by everyone contacted, these factors greatly reduced the amount of information gathered that could be contributed to a national AES and CES content inventory.

Other project tasks:

- Inventory and evaluation of UMN Extension and MAES websites, including Extension Store (http://shop.extension.umn.edu/) resources, to determine what content should be included in the archives project and to provide recommendations for a systematic process for selecting, reviewing, capturing, archiving, and making accessible this academic content via the UDC.
- Research the websites of the CSREES[5], NASULGC, USDA/ARS, USAIN, AgNIC, the NAL, ADEC, *eXtension,* and the libraries and extension services at land grant institutions (including but not limited to the twelve states in the CSREES North Central region), to assess the current state of affairs with regard to the collection and preservation of born-digital extension publications.
- Perform a literature review[6] to ensure that recommendations provided built upon the significant body of research produced by USAIN, AgNIC, the NAL, ADEC, and institutions around the country that have long worked toward the goal of building a national digital library or portal for agriculture.
- Incorporate information from the activities indicated above to formulate recommendations for policies regarding access, selection or collection development, and copyright, to discuss and address metadata issues, and to recommend specific roles and responsibilities for authors, extension staff, and libraries/UDC staff.
- Articulate lessons learned from the project in a manner the enables national and international colleagues to use this project as a roadmap or springboard for forming information silos into cooperatives.

RECOMMENDATIONS: Policies, Metadata Issues, Roles and Responsibilities

Step 1: Policies (Access, Selection, Copyright)

Access: Open?
One of the first issues to address with any institutional repository content is to determine the level of access, i.e. does any user have access, or is access limited to University-affiliated users or to a specific community of users within the University? The UDC is primarily envisioned as an "open access" repository, although Libraries administration recognizes that some internal communities may need to restrict access to certain confidential or internal administrative content. This issue was discussed early in the process with Extension administrators, who granted approval for this content to be openly accessible from the UDC, i.e. made freely available to all users. This includes content currently sold in the online Extension Store (usually for a relatively minimal fee covering production expenses). Extension administration expressed their belief that the greater exposure to freely available resources from their store would not negatively impact their sales. Some reports indicate that this type of greater exposure can actually increase sales of print publications, including more expensive commercial resources. Steve Potash, president and CEO of Overdrive.com, a provider of inventory and rights management to online booksellers, notes that, "Ten years ago there was [fear among publishers] that digital versions of titles would detract from print sales [however today] everyone knows that e-books add to the ROI. [By] publishing e-books, they introduce their authors and titles to a new audience"(Chiara

---

[5] See CSREES Glossary of terms (Cooperative State Research, Education and Extension Service, USDA)
[6] Bibliography appended.

2007). Google advertises its Google Book Search project with testimonials from publishers claiming the increased exposure has increased book sales (Anonymous b), especially for older titles (Anonymous c).

In any case, information produced and distributed freely by governmental and public institutions is now being collected, repackaged, and sold by the private sector, albeit with the addition of extra metadata enabling search enhancements[1]. Miami University Professor Bettina Fabos warns, "If we want to go beyond a mainstream, commercialized, sponsored online information repository we need to turn to a different structure that offers a more inclusive, democratic information environment" (Fabos 2007).

Access: Revisions

One rather strong concern mentioned by Extension staff throughout this project was the need to protect consumers from erroneously using "old" or outdated information, although librarians know from working with researchers that it is also imperative to retain access to previous editions or versions of a resource. This access to the complete record is important in historical research, where a longitudinal perspective is often critical (e.g. documenting how the recommended use of a specific pesticide changed over time). One solution to this concern is to flag revisions as such and present all versions of a resource to users, while ensuring that the most current content is the first and most obvious resource in a result list. Michigan State University's Extension Knowledge Repository (http://www.msue.msu.edu/portal/) clearly demonstrates this functionality, serving the most current resource at the top of a results page as the most predominant result, followed by any previous versions. Older editions are clearly tagged as such both on the first result page and on a cover page appended to each resource with the following statement: "**MSU Extension Publication Archive.** Archive copy of publication, do not use for current recommendations. Up-to-date information about many topics can be obtained from your local Extension office." This is an excellent solution to a sticky issue, and it demonstrated to the author the value of holding discussions throughout the project with national colleagues who have often resolved similar issues.

Selection: Content Scope & Priorities

One of the most difficult tasks in this type of project is identifying and selecting which resources should be included in the collection over the long term. Although a comprehensive digital collection will include older resources not yet digitized, those are protected for the present by virtue of the longevity inherent in their physical format. Born-digital grey literature in all subject areas is in the gravest danger of being lost unless it is specifically included in the scope of library collections worldwide. Not only are AES and CES resources increasingly created in digital format, in many cases these resources do not have any corresponding "print" equivalent (e.g. pdf version) but exist only in html format, as a webpage. UMN Extension estimates that approximately 40% of their publications exist only in digital format, and that percentage will continue to increase.

Selecting and "collecting" webpages or websites, including description and archival storage, is a relatively new arena for libraries, although many libraries add records and links to web resources to their OPACs. UMN Libraries and Extension staff struggled to define boundaries and arrive at a collection development statement that would result in the collection and preservation of resources deemed valuable on a long-term basis, without casting the net too broadly and gathering resources more programmatic or administrative in nature. While the latter are important, they fall more clearly into the domain of the University Archives unit, which is experimenting with *ArchiveIt* for collecting the record of the University's departments and colleges, processes, and policies. As an enhancement to this project, the UDC recently used *ArchiveIt* to collect all MAES and Extension web content in the Wayback Machine (http://www.archive.org/web/web.php) from 1997 to present, for the purpose of enhanced description and permanent preservation.

Resources recommended for inclusion in the Extension collection of the UDC are all digital resources categorized as publications, products, or newsletters by the University Extension Service and the Minnesota Agricultural Experiment Station, including all e-format resources made available to users via the online Extension Store (http://shop.extension.umn.edu/). While most of this content is currently

in pdf and html format, images embedded in websites are included and it is probable that future formats will include streaming video, podcasts, etc.

Exclusions currently include (but are not limited to) blogs, events/workshops, educational programs, calendars, and program descriptions. It has been noted that blogs are sometimes used to deliver "newsletter"-type content of value to future users, so this category should be revisited. As with all collections, the selection criteria will need to be reviewed and revised regularly, and adapted to recognize changing technology and communication trends.

Selection: Priorities for inclusion

While continuous and comprehensive coverage is critical for any special collection, many of these resources are still not available in digital format. Resources to be added to the extension collection in the UDC were prioritized in the order in which they should be added, based on danger of being lost and immediacy of availability in digital format.

1. Digitized resources served/accessed via the online Extension Store, if available in a print-equivalent (e.g. pdf, Publisher) format.

2. Withdrawn content currently archived on an Extension Service server (estimated at approximately 8,000 documents).

3. Freely available pdf or html content considered a "publication" or "newsletter" on identified, authorized extension websites. This phase captures the MAES resources and includes webpages that can be described individually as unique resources, with content that remains relatively fixed in time until a revision[7] is created. All images associated with selected webpages, typically in jpg or gif format, will be archived with the associated html content.

4. Older microform and print content, which will have to be digitized. These resources are a collection strength at the University of Minnesota, with over 2.7 million government documents and a recommendation from a recent *R2 Consulting* study to digitize in particular extension and agricultural documents, creating a "reborn digital (converted from another format)" (Dempsey 2006) component to the collection.

Extension staff, subject librarians, and UDC staff are currently evaluating websites to determine what content should be archived. The University of Minnesota College of Food, Agriculture, and Natural Resource Sciences (CFANS), the administrative unit for Extension, is moving to a new web content management system, which heavily impacts processes for collection development. The selection scope will be finalized as the college begins their transition to that new system in winter or spring 2007/2008.

Copyright

One of the most difficult issues to address when developing digital collections is intellectual property rights, or determining copyright ownership. A mechanism must be in place to ensure that permission is granted to authorize the deposition of the content into an open access repository, while ensuring that there are not any overlapping or competing copyright holders (e.g. *eXtension*). Clarifying these issues for this collection involved multiple discussions between upper level Extension and Libraries administration, the University's Office of the General Counsel, and the directors of the UDC.

The UDC has a copyright policy that covers all UDC content. At some stage in the submission process content authors grant permission for the deposition of their IP into an open access archive. This permission may be granted at the University level, at a subcollection level, or at an individual author level. Historically, the copyright owner for AES and CES content was the Regents of the University of Minnesota. A change in national copyright law in 2003 resulted in the University shifting their stance

---

[7] Determining when a change can be considered a "revision" and labeled as such is very important, and is more difficult to define with web content than with a physical object. This determination will be made by Extension staff, with the mechanism for tagging and presenting revisions as such to be established by UDC staff. Useful possible mechanisms include scripts that scan the content and automatically create some assessment of, for example, the extent to which content has changed. The script would then produce reports that technical staff can utilize to flag and tag content revisions.

on Intellectual Property (IP) to a default that assumes the author retains copyright unless they specifically assign it to the University Regents. Although Extension staff estimate that 95% of their contributors trust the University to protect the content and ensure its proper use, they had to design a new process to address this policy change at the University level. Extension thus developed a statement that enables contributors (authors) to transfer <u>non-exclusive</u> copyrights (i.e. permission only to use the content, with proper attribution to the creator, with no permission to alter the content, etc.) to the Regents of the University of Minnesota. This is incorporated into a Product Action Form contributors sign when they submit their content, which also includes the following note to inform authors of Extensions' intent to archive the resource:

> *The Board of Regents mandates that University Archives collects a copy of anything the U of MN produces to preserve them for future users and historians, so we will automatically send copy to U Archives.*

Both of these statements need to be retained and incorporated when Extension moves from paper-based submission to online submission of content files directly into the UDC.

A related issue for this specific collection arose when **eXtension**, a national CSREES and NASULG online extension content initiative, drafted a "business agreement" that appeared initially to prohibit even land-grant institutions from offering their content to users in an open access repository, if that content was also used by **eXtension**. This issue was resolved when **eXtension** revised their business agreement in 2007, granting contributors the *non-exclusive* right to use their content as desired, including depositing it into an institutional repository.

Step 2: Metadata Issues

UDC staff members are responsible for ensuring that the metadata applied to repository content is OAI compliant and that the content is exposed for harvesting, to enable the incorporation of these records into other repositIt is critical to consult and involve staff who are engaged with national groups articulating and defining standards for digital collections, and who possess the requisite expertise and awareness to address these interoperability issues at the outset.

Providing users with open access to content is only effective if they can locate it, which involves applying appropriate metadata to the content. The centralization of metadata into one cooperative repository is possible if interested participants use the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), which essentially enables cross-searching multiple subject gateways. The increased standardization of descriptive metadata will greatly enhance access and "findability"[8] for these resources. UDC content is OAI compliant and will rank relatively high in Google users search results due to Google's prioritization of repository content in their algorithm. However, research indicates that even OAI-PMH repository content is not comprehensively indexed by search engines, with Yahoo indexing approximately 65% and Google indexing only 44% (McCown et al. 2006). One mechanism for increasing exposure for this particular UDC collection is to catalog and include it in the UMN Libraries OPAC, which will ensure its eventual reflection in WorldCat holdings. As a final mechanism for maximizing exposure to these resources, incorporate the records for this content in an internationally utilized index. The clear indexing choice is AGRICOLA, with NAL's recent User Needs Assessment Survey demonstrating high brand recognition of this index (see summary at http://www.nal.usda.gov/news/ossummary.shtml)

Given the issues presented with access mechanisms and search engine coverage, the most critical aspect for this particular project is for this collection to be interoperable with others, enabling in essence the conversion of silos to cooperatives. The following specific questions helped to articulate issues, and define processes and policies for metadata application.

---

[8] **Findability** refers to the quality of being locatable or navigable. At the item level, we can evaluate to what degree a particular object is easy to discover or locate. At the system level, we can analyze how well a physical or digital environment supports navigation and retrieval (Wikipedia contributors).

1.   What level and types of descriptive metadata need to be applied to every extension resource deposited in the UDC?
Does the addition of a formal ontology or similar descriptive metadata truly enhance the findability of information? Some have suggested that today social "tagging" might suffice in place of more rigorous assignation of subject headings or descriptors and other formal metadata. UMN metadata librarian John Chapman disagreed, noting that tagging is a way for people to organize their information and while "most people like <u>organizing</u> their stuff…very few really enjoy <u>describing</u> it" (personal communication). In other words, tagging can help us organize information once we find it, but it does not work well at grouping information into subject or topical sets unless applied across a very large group of users on the same site (e.g. http://del.icio.us/).

2.   Should a formal thesaurus or subject taxonomy be applied and if so, which one? ?
"[P]recision searching means knowing the vocabulary of the subject being searched" (Hutchinson and Greider 2002, 533). But whose vocabulary should be used, and from what perspective? A farmer will use different terminology than an extension agent, agency scientist, or graduate student. The National Agricultural Library's Agricultural Thesaurus (NALT) is becoming standard for agricultural indexing in the U.S., but future initiatives might focus more on identifying and empowering the users with the ability to apply some of their own keywords or "tags". As noted earlier, tagging is no substitute for the use of a formal taxonomy, but it should be considered as an enhancement to increase users' full-text searching capabilities.

Current UMN Libraries practice is to apply Library of Congress (LC) subject headings to content cataloged for inclusion in the OPAC, including extension service content. However, the NALT is the national standard thesaurus now used to describe content in the most important collections of agricultural information nationwide, including all USDA publications, as well as many related collections of digitized content (AgNIC). In order to facilitate the inclusion of this content into national or international repositories of agricultural information, and ensure that the content is correctly indexed in AGRICOLA, NALT subject headings should be applied to all UMN AES resources incorporated into the UDC. AGRIS, an international agricultural index, uses AGROVOC, a multilingual agricultural thesaurus (http://www.fao.org/agrovoc/) developed by the Food and Agricultural Organization (FAO) of the United Nations. The FAO is developing a cross-walk between AGROVOC and the NALT, which will greatly enhance the facility for including U.S. resources in international indexes and repositories.

3.   Who applies different types or levels of metadata?
The overarching philosophy is to integrate metadata assignation processes into existing workflows to minimize the impact of the project on any one office or staff member, both in the Libraries and in Extension. It is important to apply as much descriptive metadata as possible at the local level, partly to allow for local idiosyncrasies and also to minimize bottlenecks incorporating the metadata into a national repository due to possible staffing shortages at that level. Reliance on local expertise and oversight as much as possible is the most effective and efficient way to ensure that content is described in a timely manner and that the collection meets local needs.

4.   What metadata schema should be used to enable the harvesting of the metadata for inclusion in other repositories?
The UDC is using the DSpace platform for the first collections, including Extension resources, and will be modifying their Dublin Core metadata template to meet specific characteristics unique to this collection[9]. Persistent Uniform Resource Identifiers (PIs) will be assigned, enabling indexing agencies to link directly to content and thus increasing the probability that those agencies will include this content in their scope.

5.   What specific fields or types of metadata should be applied?

---

[9] For example, one of the issues not mentioned often in the literature about a national repository for agricultural information is the enormous importance of geographical context. It is recommended that a field be included in the metadata schema that addresses the issue of spatial context.

The exact metadata template is not yet finalized for this collection, although some specific fields have been identified as critical (e.g. Extension's Topic headings). In addition to descriptive terms, preservation metadata should be applied to ensure long-term access to digital collections. Joy Davidson states that, "Metadata can play a vital role in enabling the effective management, discovery, and reusability of digital information. Preservation metadata provides provenance information, documents preservation activity, identifies technical features, and aids in verifying the authenticity of a digital object. The PREMIS Working Group released its Data Dictionary for Preservation metadata in June 2005 (http://www.oclc.org/research/projects/pmwg/premis-final.pdf). This is a core set metadata elements recommended for use in all preservation repositories of the type of materials archived and the strategies employed."(Davidson 2006) She also highlights the importance of developing national standards for different data types.

Technical specifications are not completed, with the UDC and the University's new CMS in the very early stages of development. Preliminary metadata field definitions have been developed for the CMS, and preliminary data fields are now being selected and mapped for existing (as opposed to newly created) digital Extension content to be batch loaded into the UDC.

Step 3: Roles and Responsibilities

The final step after content has been selected and described is to design new work flows to ensure that each resource deposited into the UDC as part of the Extension Service collection follows a linear process without loopbacks from creator to Extension, through the Libraries technical services area, and to Libraries IT/UDC staff. In the print world, the print object acts as a process driver in that the person with a resource in hand knows that they are the responsible party at that point in time. Lacking the existence of a physical object, we can use print models as a template or guide for building the process for routing electronic objects. The goal is to break the process into manageable pieces, ensure that everyone knows their role, and automate deposition and description to the greatest extent possible while avoiding loopbacks and/or duplication of effort. As engineer H.M. Gladney noted in his assessment of progress on the National Digital Information Infrastructure Preservation Program (NDIIPP) (Gladney 2007):

> *Until archiving is partitioned into almost independent components, tasks that can be defined by objectively described rules are hopelessly confounded with subjective tasks that require human creativity, judgment, and taste. Many engineers favor "divide and conquer", designing no mechanism or process until it is segregated from the overall task sufficiently so that it can be treated with minimal, well-defined interfaces with other components.*

An important component of this type of process is the automated routing of content to the next appropriate processing point at the completion of each step, concurrent with a notification mechanism at points requiring manual intervention. Roles and responsibilities are defined more specifically below, beginning with the critical role of project coordinator.

Coordinator
Throughout this process it became clear that assigning a project coordinator with a "big picture" perspective is critical for an initiative of this type. This individual can identify and help work through various issues as they arise, even though they may have no particular depth of knowledge or special expertise in the issue being addressed. The coordinator identifies stakeholders who need to be involved in resolving specific issues (e.g. Libraries technical services, extension IT, Libraries web services), and communicates and coordinates the necessary meetings to resolve those issues. Finally, a coordinator could serve as the primary institutional contact in the future, as national or international initiatives expand to include content in these local repositories.

Author/Creator
- Enters Level 1 metadata (Title, Author, Copyright/Permissions choices).

- Uploads the electronic resource and associated images (including a print-ready version, if the resource is also to be produced in print). "Strongly recommend" to creators that they submit content as a pdf file, to expedite the inclusion of a resource into the UDC.

Extension Service staff

General responsibilities for extension staff capitalize on their intimate knowledge of their users, contributors, and resources, and should not differ significantly from those currently performed for web and Extension Store content. The primary difference is that content will be deposited via a relatively automated mechanism (e.g. John Chapman's DropBox concept[10]) into the Libraries UDC, rather than being deposited on an Extension server. Metadata will then be harvested and content linked from the appropriate site(s) via the Content Management System (CMS) being implemented at the University of Minnesota. Extension staff will:

- Review content to ensure quality is of the high level commensurate with Extension's mission.
- Assign Extension Service topic terms (Appendix A) to ensure the resource is categorized correctly in the Extension Store and on the Extension website.
- Enter basic metadata (as specified by Extension staff) required for ES tracking and administrative purposes.
- Mark content to indicate desired posting site(s) (e.g. Extension Store, ES website)
- Flag "revisions" and other unique characteristics of the resource.
- Upload content into UDC.
- Manage internal links: Add and remove link(s) to Extension Store or ES website (e.g. when content is replaced with a revision or with more "current" content).
- Ensure that a hard copy of each item, if available, is forwarded to the University Libraries for cataloging and addition to the print collection.

Finally, extension should assign a liaison to meet regularly with a small group comprised of representatives from the Libraries (e.g. metadata librarian, the agricultural or government documents subject librarian, and Libraries Technical Services staff) to review scope and process, and to update processes and/or systems as required.

Libraries: Subject Liaison (Agriculture and/or Government Documents Librarian)

Libraries' contributions to the process capitalize on the added-value services that libraries have long provided for information in all formats. For the subject liaison, this involves preliminary and ongoing collaboration with Extension staff to identify and select content appropriate for inclusion in the collection. Content should meet the selection criteria outlined in a collection development statement/guidelines developed by the selector(s). The liaison may also be called upon in the future to serve as local (UMN) coordinator with larger efforts, working to ensure that the collection continues to be developed in a manner that ensures the integration of content into national or international collections.

Libraries: Cataloging and Technical Services

Cataloging staff have long developed and refined MARC records and interested staff will have no difficulty learning or adapting to a different scheme, e.g. Dublin Core or taxonomy, e.g. NALT. Utilizing cataloging staff to describe content being deposited into the UDC will require additional training, but involves no new responsibilities given that these staff members have been describing resources and adding metadata to records for many years - in the form of MARC records. Advanced cataloging staff members are trained not only in ensuring that the particulars of each field of a record are correct, but are also skilled at utilizing formal taxonomies to describe coStaff members have expressed a strong interest in acquiring expertise and developing new skills desperately needed for the success of digital libraries world wide.

---

[10] Web forms could be created to automate many of the routing processes, presenting each individual with only those metadata fields appropriate to their role in the process, offering options for flagging content for inclusion into the Extension Store, an Extension web site, and/or the UDC, etc.

Eells, Linda
Born-digital Agricultural Resources: Archives and Issues

Staff contributions to the process capitalize on staff expertise in providing descriptive metadata to enhance access to information resources in all formats. The application of descriptive and preservation metadata enables access to resources today and into the future by fostering the development of targeted, focused access. Specifically, Libraries cataloging staff will be responsible for the:
- Review of metadata applied by Extension staff for adherence to cataloging standards defined by the UDC for this collection.
- Assignation of controlled vocabulary (NAL Thesaurus).
- Completion of additional descriptive metadata fields to enable full access to records via the Libraries OPAC.

Libraries: University Digital Conservancy (UDC)/Digital Initiatives
Libraries IT and UDC staff, including web and metadata librarians, provide unique expertise and services to expedite the submission, description, access, and long-term preservation of resources, including the development of routing tools and processes. These staff members work with content providers, in this case Extension staff, to:
- Establish appropriate access mechanisms and policies (e.g. revisions, open vs. limited access).
- Ensure interoperability with other digital agricultural initiatives and collections via the application of and adherence to national and international standards.
- Establish metadata field definitions that provide the functionality required to meet local and national needs.
- Ensure the application of metadata appropriate to the long-term preservation and ability to migrate content to future platforms/technologies.
- Participate in national technology infrastructure groups working on the integration of this collection into a regional, national, or international digital agricultural library.

National Agricultural Library (NAL)
Institutions implementing local archival initiatives rely on national and international partners to include in their larger efforts those resources selected, described, and made accessible at the local level. NAL staff could provide the same supportive services currently provided for physical resources that document the work of agriculture researchers and extension agents in the U.S. This would include the routine, systematic harvesting of applicable ES collection metadata from U of MN's Digital Conservancy for inclusion in the National Digital Library for Agriculture, and the application of additional metadata or mapping as required for inclusion of record(s) in the AGRICOLA index.

Summary and Lessons Learned
After defining roles and responsibilities, the project as initially designed was complete. The environmental scan validated the initial impression that born digital agricultural literature is in many institutions an endangered species, or one whose status is unknown, and that much of this literature is fragmented in institutional silos lacking interoperability features and functionality. It also brought to light the need for the application of standards and infrastructure to support the collection of metadata from these silos into one cooperative access point, a need that several national groups (AgNIC, NAL, and ADEC) are working on resolving. Those efforts informed the recommendations presented for the UMN's extension archive, including access, selection, and copyright policies, clarification of metadata issues to be addressed, and delineation of roles and responsibilities for moving the resources from author to Extension and into the UDC, with appropriate description applied along the way. As I worked through the project, the following issues surfaced as being very important to the success of this type of project and others that may be planned on a larger scale, e.g. a National Digital Library for Agriculture (NDLA).

Lessons Learned: Critical Issues for Collaboration
The long-term vision for a national collection of digital agricultural resources is that it will include (but not be limited to) all content, born-digital, or retrospectively digitized microform and print, that has been published by public land grant (including 1890s) institutions from their inception to the present

(see National Digital Library for Agriculture design concept website at
http://www.nal.usda.gov/ndla/about.shtml). At the international level, the vision requires
collaboration between the NDLA and international organizations (e.g. FAO) to designate and maintain
an international archive that enables users to access from a single point of access all metadata and/or
content currently held in a multitude of independent "silos", e.g. Cornell's Core Historical Literature of
Agriculture (http://chla.library.cornell.edu/), AgNIC (http://www.agnic.org/), MSU's Knowledge
Repository http://www.msue.msu.edu/portal/), and *eXtension*
(http://about.extension.org/category/info/), to name just a few.   The following issues were
identified as very important to the success of this project, and are shared here for others interested in
establishing collaborative repository initiatives and relationships.

Coordination: A "shepherd" or coordinator can identify all stakeholders and participants, verify or work
to ensure buy-in, identify issues and work with appropriate offices or individuals to resolve those (e.g.
copyright, access), and bring stakeholders together who are not traditionally be aware of the work of
others whose feedback is important to the process. The coordinator clearly delineates the roles,
responsibility, and value of all contributors, content owners, and possible participants including
extension staff, library staff, systems and metadata experts, and leaders at the national level, and
maintains an awareness of initiatives and a strong connection to colleagues nationally, to ensure that
selection policies, interoperability issues (e.g. metadata schema, OAI compliance, etc.) are compatible
with and complementary to other efforts.

Stakeholders: In 2004-2005 several surveys were done to document collaborative initiatives between
libraries and extension units (including Agricultural Experiment Station and Cooperative Extension
services) at land-grant institutions. The report on the results of these surveys (Hutchinson et al. 2005)
would be an excellent resource for identifying possible collaborative partners, as it is clear from this
report which institutions have already formed strong connections between their libraries and their
extension offices that is critical for the success of this type of project.

Financial and Staff Resources: Governance, maintenance, and sustainability of the repositories cannot
depend on an intense and sustained infusion of resources over the long-term, although significant
resources will be required in the 3-5 year start-up phase for coordination and the development of
infrastructure. Resources committed to the UMN project included a three-month paid research leave
for the author/coordinator, and a special University appropriation for staffing and infrastructure
development for the UDC, without which this project could not move forward. Similar (albeit scaled
up) resources would be required for coordination and infrastructure development on a larger scale.
For long-term maintenance, processes and tools must be established to integrate into normal
workflows in the same way the acquisition, cataloging, and preservation of print resources is an
integrated part of current workflows.

Selection Criteria and Content identification: Achieving a vision of a comprehensive collection of
extension resources can become a reality if participants contribute to the incremental development of
a detailed inventory of extension resources, digitized and print, of all institutions nation-wide.
Developing this inventory from the top-down has proven extremely difficult due to the complexities
and vastly different levels of awareness about these resources at each institution. The most effective
method for compiling this inventory would be for each institution to inventory their local collection and
enter the specifics into a database coordinated at the national level. This will require the
participation of both extension and library staff at each institution. The inventory would include for
example specific series titles, years covered, format, and current and/or archival location.
Coordination at the national level and cooperation at the local level would be greatly enhanced by the
development of a web-based form or template in which institutions could enter their collection data.
Content redundancy can be minimized by defining collections of distinction and formulating a
mechanism to minimize content overlap or duplication. Recent trial reports from OCLC's Collection
Analysis Tool indicate that the University of Minnesota holds the largest collection of agricultural

material in the CIC, although much of it is not "unique"[11]. In a recent analysis of UMN serials and government documents holdings by the libraries consulting firm R2 (http://www.ebookmap.net/), data indicated that UMN holds the second largest depository library collection in the U.S., with particular strength in agricultural documents. Building on these strengths without overlapping or duplicating effort in collecting content should be a core goal of any collaborative CD efforts.

Users: It is important to remember and define the audience for whom this repository is being created, and to design the national library to meet their needs, rather than allowing staffing limitations or libraries, extension, or IT staff exigencies to define the architecture. In spring 2006 the USDA National Agricultural Library conducted a survey that can be used as a starting point for creating a picture of the potential users of a digital agricultural library. The goal of this survey was "to identify the library's current and potential customers; to determine those customers' information needs, sources, and behaviors; and to assess their view of NAL's existing and potential services."(National Agricultural Library 2006) Follow-up surveys or data harvesting efforts should be initiated to identify characteristics of additional end user populations, defining to the extent possible the needs of those end-users, and gathering information about their search terms and patterns. Data collected about visitors to extension websites, and local and national agricultural repositories (gathered and distributed following federally mandated methods that ensure the protection of individuals' privacy) should inform the development of resources and services that actually meet their needs. While USDA, university researchers, and extension staff are very important (primary, currently?) users of this information, the scope of the collection would be severely limited if the needs of public users outside of that scope – who are the traditional end-users of extension resources – are not addressed.

Standards, Metadata, and Taxonomies. A repository should adhere to national or international metadata standards yet remain flexible enough to accommodate local idiosyncrasies. Interoperability is of paramount importance, but contributors also need to retain a high level of local autonomy. Different stakeholders will want to use different taxonomies, from tagging by the end-user, to category assignation by extension staff, to the addition of access (e.g. NAL terms) and preservation metadata by libraries and NAL staff. It is important to select a metadata schema that accommodates metadata enhancement at various points in the process and minimizes duplicate entry of data (e.g. MARCXML, MODS). Different silos use different schema, have different peer review or access requirements, etc. Thus at a higher level it may be most effective to push the design and control of each component of a repository down to the local level to the greatest degree possible, while insisting on adherence to a core set of standards, e.g. requisite metadata has to be Open Archives Initiatives (OAI) compliant, exposed and harvestable. Local content providers and managers will only engage and contribute to a repository if they can retain control over processes that were developed in and for their local environment and users.

Bridges to the Future: Connections and Next Steps
Building the legacy for tomorrow's researchers is the responsibility of today's libraries, unless they are comfortable relying on commercial interests to manage the stewardship function. Bettina Fabos once stated that, "ongoing work toward subject gateway development—all of it developed as free, open source software—provides a small but growing countervailing force to the commercialization of "the universe of knowledge" (Fabos 2007). Libraries must not abandon their role as stewards of society's knowledge regardless of the format in which information is produced (Farb 2006). Librarians have long collected and preserved the intellectual output of civilizations, and society will assume libraries retain that role in the absence of any stated and widely acknowledged shift of that responsibility to some other entity.

As noted early and often, the long-term goal is to form local collections (silos) of digital content with a concurrent vision of the eventual (or immediate) incorporation of that content into larger or

---

[11] The tool indicates that U of MN has the 2nd largest "unique" collection of agricultural resources in the CIC. The OCLC Collection Tool is limited by varying comprehensiveness of coverage and accuracy of individual library holdings, although this data would be a useful supplement to an inventory process.

collections (cooperatives) of subject content gathered and served to users from centralized access points. Several steps could be taken move the community forward toward that goal.

- A project could be undertaken to perform a centrally coordinated, extensive follow-up survey to that done in 2004-2005 (Hutchinson et al. 2005), to more specifically <u>inventory and define the scope of extension content</u> held by a limited (pilot) group of institutions (e.g. U.S. Land Grant Universities in the North Central Region) including *exact* coverage dates, series titles if applicable, and current location of print and digitized resources. Connie Britton at Ohio State University compiled an excellent spreadsheet of Agricultural Experiment Station content for all 50 states in the U.S. This spreadsheet could serve as the basis for designing a small national web-accessible database that institutions could access to easily update the status of their collections, including Extension Service publications.

- At the national level a reconfigured AgNIC could be repurposed to serve as the cooperative home, i.e. the metadata repository for extension content served through a National Digital Library for Agriculture portal. Alternately, the extension community could work with other stakeholders to form an entirely new infrastructure that incorporates the strengths of existing repositories. At the international level, organizational partners led by the FAO are working on this issue, with the new e-Agriculture initiative offering exciting possibilities. In the recent discussion around the theme "Responding to Demand: the Focus of e-agriculture" on e-Agriculture.com (Anonymous2007), many organizations around the world mentioned their work on developing regional and/or national online repositories for agricultural information, and several posts noted the need for collaborative searching across those repositories.

The "perfect storm" is forming both nationally and locally, with several projects and initiatives coming together today at the University of Minnesota, where the timing of this AES Digital Archives project dovetails nicely with the development of the University's Digital Conservancy and with the new Extension Service public website going live via the University's new Content Management System (Stellant). The possibilities are exciting, and Extension and Libraries staff are very enthusiastic about the probability that greater access to more resources over the long-term will result from our efforts. This project demonstrates a grounded instance of collaboration that worked and is scalable, but as a community we need to leverage the lessons learned and build on individual institutional strengths. We look forward to future opportunities for participation at the national level in projects ensuring user-centered, long-term access to current and retrospective agricultural information from an authoritative, central site that is accessible from anywhere in the world.

References

Anonymous (a). 2007. *Responding to Demand: The Focus of E-Agriculture.* 2007. Available from http://www.e-agriculture.org/98.html ed.

Anonymous (b), *IDRC increases visibility and sales by making its titles 100% viewable*. Google Book Search: Partner Success Stories. [accessed 8/10/2007]. Available from http://books.google.com/googlebooks/idrc.html.

Anonymous (c), *Springer Sees Record Backlist Sales with Google Book Search*. Google Book Search: Partner Success Stories. [accessed 8/10/2007]. Available from http://books.google.com/googlebooks/springer.html.

Bergman, Michael K. 2007. *The Deep Web: Surfacing Hidden Value.* Available from http://www.press.umich.edu/jep/07-01/bergman.html. University of Michigan Press.

Chiara, Warren. 2007. *E-books' Impact on ROI.* Available from http://www.booktechmag.com/btm_premium/291069522178336.bsp.

Eells, Linda
Born-digital Agricultural Resources: Archives and Issues

Cooperative State Research, Education and Extension Service, USDA. CSREES Glossary. [accessed 8/26/2007]. Available from http://www.csrees.usda.gov/about/glossary.html.

Davidson, Joy. 2006. *[Asis-l] Data Dictionary for Preservation Metadata: PREMIS Tutorial.* Available from http://mail.asis.org/pipermail/asis-l/2006-July/002650.html.

Dempsey, Lorcan. 2006. *The (Digital) Library Environment: Ten Years After.* Available from http://www.ariadne.ac.uk/issue46/dempsey/. UKOLN.

Extension Committee on Organization and Policy (ECOP). 2002. *The Extension System: A Vision for the 21st Century.* Washington, D.C.: National Association of State Universities and Land-Grant Colleges (NASULGC).

Fabos, Bettina. 2007. The Commercial Search Engine Industry and Alternatives to the Oligopoly — Media oktato es kutato kozpont. [accessed 8/10/2007]. Available from http://mokk.bme.hu/centre/conferences/reactivism/submissions/fabos.

Farb, Sharon E. 2006. Libraries, licensing and the challenge of stewardship. *First Monday* 11(7), p.1-14.

Foster, Nancy F., and Susan Gibbons. 2005. Understanding Faculty to Improve Content Recruitment for Institutional Repositories. *D-Lib Magazine*, 11(1).

Gladney, H. M. 2007. Digital Preservation in a National Context: Questions and Views of an Outsider. *D-Lib Magazine,* 13(1/2).

Gulli, A., and Signorini, A. 2005. *The indexable web is more than 11.5 billion pages.* Chiba, Japan: ACM Press.

Gwinn, N. E. 1994. A national preservation program for agricultural literature. *Journal of Agricultural & Food Information,* 2(2), p.25-49.

Heatley, Randall, and Melanie Gardner. 2003. Opportunities for collaboration to develop a digital Land-Grant information system. *Journal of Agricultural & Food Information.Haworth Press Inc., Binghamton, USA,* 5(1), p.35-42.

Heatley, Randall, and Poley, Janet. 2006. *National Digital Library of Agriculture: Call to Action.* Working paper.

Hutchinson, Barbara S., and Greider, Antoinette P., eds. 2002. *Using the agricultural, environmental, and food literature.* Vol. 61. New York: Marcel Dekker.

Hutchinson, Barbara S., et al. 2005. *Collaboration and Cooperation Among Libraries, Cooperative Extension, and Agricultural Experiment Stations in Land-Grant Universities: The Results of a 2004-2005 Survey.* Arizona: U.S. Agricultural Information Network (USAIN).

Kjelgren, R. 2007. Agricultural Experiment Station faculty and administrator attitudes and perceptions regarding placing research results on the World Wide Web. *HortTechnology. American Society for Horticultural Science, Alexandria, USA,* 17(1), p.95-101.

Liss, Ken. 2007. *Principles for Digitized Content.* Available from http://scholcommbc.blogspot.com/2007/03/principles-for-digitized-content.html. Boston, MA: Boston College, University Libraries.

McClure, Marji. 2006. *Archive-It 2: Internet Archive Strives to Ensure Preservation and Accessibility.* Available from

Eells, Linda
Born-digital Agricultural Resources: Archives and Issues

http://www.econtentmag.com/Articles/ArticleReader.aspx?ArticleID=18132&Query=internet%20archive.

McCown, F., et al. 2006. Search engine coverage of the OAI-PMH corpus. *IEEE Internet Computing*, 10(2), p.66-73.

National Agricultural Library. 2001. *National Agricultural Library Assessment Report.* Maryland: U.S. Department of Agriculture.

National Agricultural Library. 2006. Summary from the User Needs Assessment, National Agricultural Library. [accessed 8/26/2007]. Available from http://www.nal.usda.gov/news/ossummary.shtml.

National Agricultural Library, ERAL Committee. 2006. *NAL Electronic Repository for Agricultural Literature (ERAL) Committee Report.* Unpublished committee report ed. Beltsville, MD.

Reich, Vicky. News and Views, Blog Archive, CCDO ALA Midwinter discussion, January 21, 2006. in Chief Collection Development Officers (CCDO) of Large Research libraries Discussion Group ALA Midwinter Meeting, San Antonio, Saturday, January 21, 2006 [database online]. [accessed 8/26/2007]. Available from http://janusconference.library.cornell.edu/?p=83.

Schmitz, John G. 2005. Agricultural Extension on the Web. *Digital Divide* Network, 1-9.

Subirats, Imma, et al. 2007. *Towards an architecture for open archive networks in Agricultural Sciences and Technology.* Rome, Italy: Food and Agriculture Organization of the United Nations.

Wikipedia contributors. Cite. in Wikipedia, The Free Encyclopedia [database online]. [accessed 7/29/2007]. Available from http://en.wikipedia.org/w/index.php?title=special:cite&page=findability&id=111578029&oldid=111578029.

Wikipedia contributors. Deep Web. in Wikipedia, The Free Encyclopedia [database online]. [accessed 7/30/2007]. Available from http://en.wikipedia.org/wiki/deep_web?oldid=147936953.

World Summit on the Information Society (WSIS). World Summit on the Information Society: E-agriculture. in United Nations International Telecommuncations Union [database online]. Rome, Italy, [accessed 7/25/2007]. Available from http://www.itu.int/wsis/c7/e-agriculture/index.html.

Young, Peter R. 2004. National Digital Library for Agriculture. *Quarterly Bulletin of the International Association of Agricultural Information Specialists*, 49(3/4), p.114-120.

Bibliography

Conference issue: USAIN/IAALD joint conference - The information frontier: linking people and resources in a changing world, April 3-5, 1997, Tucson, Arizona. 1997. *Quarterly Bulletin of IAALD,* 42(3/4), p. 114-305.

Akmon, Dharma, and Elisabeth Kaplan. 2006. Developing Electronic REcords Capacity in the Small Collecting Repository: the Documenting Internet2 Project. *RLG DigiNews,* 10(4).

Beecher, J. W. 1991. The United States agricultural information network (USAIN): providing information to the end user. *Quarterly Bulletin of IAALD,* 36(1/2), p. 59-61.

Bishop, Ann P., Van House, Nancy A., and Buttenfield, Barbara P. 2003. *Digital library use : social practice in design and evaluation.* Cambridge, Mass: MIT Press.

Borgman, Christine L. 2003. *From Gutenberg to the global information infrastructure : access to information in the networked world.*1st MIT Press pbk. ed. Cambridge, Mass: MIT Press.

Connors, Christine. 2006. Metadata: Practical, Painless, Profitable. *Bulletin of the American Society for Information Science and Technology,* 32(6).

Cotton, Dan. 2007. *Discussion of Contributor and Business Agreements.* Edited by Dan Cotton. RealMedia stream. University of Nebraska.

Crook, Edgar. 2006. For the Record: Assessing the Impact of Archiving on the Archived. *RLG DigiNews,* 10(4).

Ferguson, Julie E. 2005. Bridging the gap between research and practice. *Knowledge Management for Development Journal,* 1(3). p.46-54.

Goldsmith, Beth, and Frances Knudson. 2006. Repository Librarian and the Next Crusade: The Search for a Common Standard for Digital Repository Metadata. *D-Lib Magazine,* 12(9), p.1-8.

Heatley, Randall. 2006. *Thoughts on Institutional Repositories.* Unpublished manuscript. Michigan.

Heatley, Randall, Poley, Janet, and Young, Peter R. 2005. *Plan for Agricultural Information Management and Outreach DRAFT.* Unpublished working paper. Michigan.

Hilton, Chris, and Dave Thompson. 2007. Collecting Born Digital Archives at the Wellcome Library. *Ariadne,* 50.

Hutchinson, Barbara S., Jerry Henzel, and Anne Thwaits. 2006. Using web services to promote library-extension collaboration. *Library Hi Tech,* 24(1), p.126-141.

Ke, Yiping, Lin Deng, Wilfred Ng, and Dik-Lun Lee. 2006. Web dynamics and their ramifications for the development of Web search engines. *Computer Networks,* 50(10), p.1430-1447.

LaFleur, Robbie. 2003. *Minnesota Legislative Reference Library 2003 Pilot Project: Keeping Mandated Documents in Electronic Format.* St. Paul, Minnesota: Minnesota Legislative Reference Library (LRL).

Lynch, Clifford A. 2003. *Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age.*226th ed. Vol. 226. Washington, D.C.: Association of Research Libraries (ARL).

National Agricultural Library. 2006. *AgNIC Specific Cooperative Agreements: Full-text and content-building projects.* Edited by Melanie Gardner. Beltsville, MD: National Agricultural Library (NAL)/USDA.

Parker, Elliott. 2006. *Iowa State University's Agriculture Extension Online Resources to Iowa Farmers.*San Antonio, TX ed. Vol. 26. Association for Education in Journalism and Mass Communication (AEJMC).

PREMIS Working Group, OCLC and RLG (Organization). Data dictionary for preservation metadata final report of the PREMIS Working Group. in OCLC; RLG [database online]. Dublin, Ohio; Mountain View, Calif, Available from http://www.oclc.org/research/projects/pmwg/premis-final.pdf; http://www.oclc.org/research/projects/pmwg/premis-final.pdf Note: View report.

Uhlir, Paul. 1997. *Framework for the Preservation of and Permanent Public Access to USDA Digital Publications.* Baltimore, MD: National Agricultural Library (NAL)/USDA.

Weintraub, I. 1999. U.S. agriculture libraries at land grant colleges: meeting the information needs of a changing industry. *Quarterly Bulletin of IAALD,* 44(3/4), p.177-183.

White, Janet. 2006. *California Agriculture* delivers access to peer-reviewed research. *California Agriculture,*(Jan-Mar).

Willinsky, John. 2006. *The access principle : the case for open access to research and scholarship.* Cambridge, Mass: MIT Press.

Appendix A

Topic List – UMN Extension

Topics	Topic #
Community Economics	
Diversity & Immigration	518
Economic Development	521
Groups & Organizations	537
Internet Education	548
Tourism	569
Workforce Development	578
Leadership & Civic Engagement	
Diversity & Immigration	518
Groups & Organizations	537
Leadership	552
Public Policy	559

Family

Family Relations	
Change & Stress	512
Divorce	519
Early Childhood	520
Infants	547
Parenting	555
Youth & Teens	579
Youth Research	580
Family Resource Management	
Aging	507
Clothes & Stains	513
Housing	544
Parenting	555
Personal Finances	556
Nutrition & Health	
Food & Nutrition	527
Food Preservation	528
Food Safety	529
Health	539

Farm & Agriculture

Ag & Food Safety	
Ag & Environment	506
Food Preservation	528
Food Safety	529
Health	539
Pest Management	557
Safety & Security	561
Ag Business Management	
Agricultural Land Use	508
Farm Business	525

(Farm & Agriculture continued)

MASTER Topics by Category
Community & Economic Development

Crops	
Ag & Environment	506
Alternative Agriculture	509
Corn	515
Fertilizer	582
Forages	530
Irrigation & Drainage	549
Pest Management	557
Small Grains	563
Soil	564
Soybeans	565
Sugar Beets	566
Water Quality	576
Livestock	
Ag & Environment	506
Beef	511
Dairy	517
Horses	542
Livestock Production	583
Manure Management	554
Poultry	558
Sheep	584
Swine	567

Garden

Commercial Horticulture	
Commercial Horticulture	514
Fruit	532
Lawns & Groundcovers	551
Pest Management	557
Tree Professionals	570
Trees & Shrubs	571
Vegetables	573
Home Gardening	
Alternative Agriculture	509
Flowers	526
Fruit	532
Garden Diseases	534
Garden Insects	535
Garden Weeds	536
House Plants	543
Landscaping	550
Lawns & Groundcovers	551
Trees & Shrubs	571
Vegetables	573
Waste & Sewage	575
Wildlife	577

Housing

House & Home	
Home Pests	540
Home Safety	541
Housing	544
Housing Technology	545
Personal Finances	556

Money & Finances

Ag Business Management	
Agricultural Land Use	508
Farm Business	525
Community Economics	
Diversity & Immigration	518
Economic Development	521
Groups & Organizations	537
Internet Education	548
Tourism	569
Workforce Development	578
Family Resource Management	
Aging	507
Clothes & Stains	513
Housing	544
Parenting	555
Personal Finances	556

Natural Resources & Environment

Environmental Science Education	
Environmental Science	523
Housing	
Home Pests	540
Home Safety	541
Housing	544
Housing Technology	545
Natural Resource Mgmt. & Utilization	
Agricultural Land Use	508
Environmental Management	522
Environmental Science	523
Forest Management	531

Forest Products	586
Tree Professionals	570
Trees & Shrubs	571
Wildlife	577
Water Resource Mgmt. & Policy	
Environmental Science	523
Shoreland Management	562
Waste & Sewage	575
Water Quality	576

Nutrition, Food & Health

Food Safety	
Food Preservation	528
Food Safety	529
Safety & Security	561
Nutrition & Health	
Food & Nutrition	527
Food Preservation	528
Food Safety	529
Health	539

Youth Development

4-H Youth Development	
4-H Animal Science	590
4-H Communication & Arts	591
4-H Consumer Science	592
4-H Mechanical Sciences	593
4-H Natural Resources	594
4-H Personal Growth	595
4-H Plant Sciences	596
4-H Promotion	502
4-H Staff	500
4-H Volunteers & Leaders	503
4-H Youth Leaders	501
Community Youth Development	
Early Childhood	520
Youth Research	580