GKR: Consortial Repository Case Study

Marlee Givens
Georgia Tech Library

Cooperative Curation: Inter-institutional Approaches to Supporting Scholarly Communication
August 8, 2012
About me
Although I am originally from Augusta, Georgia (home of Georgia Health Sciences University), this is not me. This is an image from GHSU’s historical images collection in their Scholarly Commons repository, one of the repositories we established as part of the GKR project.

I have been the GKR Manager since February 2010. Prior to that I worked at Lyrasis for eight years, first as a cataloging and metadata trainer and then as manager of preservation and digital services.
Why am I here?

I am here today to talk about the GKR – the GALILEO Knowledge Repository – a project funded by a National Leadership Grant from the Institute of Museum and Library Services, coming to the close of its third year.
Specifically, I will be addressing these aspect of the GKR project: its governance, its services, the outreach and advocacy we have done, how we are evaluating the project, our plans for sustainability of services, and some of the lessons we’ve learned.

We will be covering these aspects in more depth tomorrow, at the Workshop on Georgia's Approach to Statewide Repository Services, with sessions presented by the people who were involved in these efforts.

I’ll be giving you the Cliffs Notes version today.
The GKR is a comprehensive statewide repository, funded by a 3-year IMLS National Leadership Grant from 2009-2012. It is both a repository (a meta-repository that harvests from IRs around the state) and a network of repositories hosted at University System of Georgia (or USG) institutions. It is a grant-funded project and a concept that has existed for eight years. It is a group of very dedicated people.

Leads for the grant-funded project are Georgia Institute of Technology and University of Georgia, and there are several other partners involved as well.

My office is at Georgia Tech, and I report to a PI there, but I work with all of the GKR partners.

This image is from Georgia Southern’s Eagle Scholar, one of the GKR partner repositories.

The grant has a set of five goals and deliverables.
A statewide survey and focus groups, to find out the University System of Georgia faculty’s use and knowledge of institutional repositories (IRs)

There were two parts to the survey: a version sent to librarians in 2009, and a version sent to all of the faculty earlier this year. Results from the most recent survey are published in the Guide to Statewide and Consortial Repositories.

In the coming year, we will be using the results of this survey to build materials for conducting faculty focus groups, and if our grant extension is approved, we will be conducting those focus groups on site around the state.
GKR grant deliverables

2. GKR metadata repository and collection mapping tool

The GKR metadata repository and collection mapping tool, which allows users to search and browse across the 10 IRs that are harvested through this project.

The mapping tool is software, written by Brad Baxter at University of Georgia, which will be released on the open source market as a product of the grant project. This tool allows participants to map items from their home repositories to the subject disciplines around which the central metadata repository is organized.

The central metadata repository is an instance of DSpace repository software that is configured to display content harvested from the GKR partners. It is arranged by subject discipline and searchable by title, author and institution as well.
An IR hosting service, where four independent DSpace instances are hosted on Georgia Tech Library servers and maintained by Georgia Tech Library IT staff, but whose content is managed by staff at the partner site.

We use a unique server configuration that allows multiple instances of DSpace to run independently of each other, supporting multiple developers and allowing each instance to be discretely tuned.

By the way, I think this is actually a picture of a heart-lung machine. It’s definitely not an IR server.
IR related services which help reduce barriers to recruiting scholarly content: rights assistance, digitization, content submission, digital preservation.

These related services were selected based on a stakeholders meeting in 2007.

They are provided in partnership with the Digital Library of Georgia, Valdosta State University, and the MetaArchive Cooperative.
This is what’s happening in this room, right now, as well as tomorrow.
We are publishing a Guide to Statewide and Consortial Repositories as part of this symposium and workshop. You have all received a copy of this on a USB drive. What’s on the drive is version one, based on the knowledge that we have at present, but we plan to update the Guide as we finish up the grant and create our sustainability plan. The final version will be added to our project web site and archived in Georgia Tech’s SMARTech repository.
These are the institutions whose IRs are being harvested into the central site.

4 research universities:
- Georgia State University
- University of Georgia
- Georgia Tech
- Georgia Health Sciences University

2 regional universities:
- Georgia Southern University
- Valdosta State University

2 state universities:
- Kennesaw State University
- Albany State University

2 state colleges:
- College of Coastal Georgia
- Georgia Gwinnett College
These are the institutions whose IRs are being hosted at Georgia Tech.

Albany State University – 4 year state college, HBCU
College of Coastal Georgia – until recently a 2 year state college, offering some 4-year programs since 2010
Georgia Gwinnett College – 4 year state college, one of the fastest growing in the USG
Georgia Health Sciences University – research university specializing in healthcare research
All 10 partners have contributed in some way, by serving on committees, participating in the symposium and workshop, sending projects through the digitization and content submission services, etc.

These are the GKR partners which brought particular expertise to the project.

Georgia Tech
In 2004, when the GKR concept first came to be, Georgia Tech had just started its SMARTech repository, which was (and still is) run on the DSpace platform.
Georgia Tech was also a founding member of the MetaArchive Cooperative, a digital preservation service.
And Georgia Tech staff in scholarly communication and digital curation services also bring substantial experience dealing with copyright issues.

UGA
Brad Baxter, a programmer for GALILEO based at UGA, developed the open source collection mapping tool, which is an integral part of the GKR central metadata repository.
UGA is also home to the Digital Library of Georgia, which has both the equipment and the expertise to run the grant-funded digitization service for the GKR.
They also have extensive experience dealing with copyright issues.

VSU
Valdosta State University has operated Georgia’s only ALA-accredited library school since 2007. Part of the GKR content submission service is an MLIS student internship, where a small group of VSU students work with GKR partners to submit content and metadata into their repositories, as part of a summer internship. Last year seven VSU students submitted 450 records and completed several related projects as well.
Not all of the GKR partners have been USG institutions or contributed to the shared metadata repository.

The USG faculty-wide survey was designed by Jennifer Campbell-Meier when she was on the faculty at North Georgia College and State University. Jennifer is currently an SLIS professor at University of Alabama, and the results of the survey will further her research there.

Two statewide library organizations have played a major role in supporting the GKR:
1. The Regents Academic Committee on Libraries (RACL), which is an organization of academic libraries from the 35 colleges and universities in the university system.
2. GALILEO (which stands for Georgia Library Learning Online), the state’s virtual library.

Our partner for digital preservation is the MetaArchive Cooperative, a Private LOCKSS Network.

Chain Bridge Group, a consulting firm, helped us draft part of our sustainability plan.
The idea for the GKR precedes my involvement with the project – by quite a lot. The GKR concept has been around since the first repositories in Georgia were being designed. The initial minds behind the GKR were from Georgia Tech, UGA and the Medical College of Georgia (now GHSU).

In 2004, the GKR idea received the blessing of RACL, and a steering committee was formed.

2004-2005, a metadata committee was formed and began working on what are now the GKR metadata guidelines.

In 2005 the Technical Committee was formed, and began creating the GKR’s technical specifications.

In 2007, a conceptual model of the GKR was produced, and in November of that year, a GKR stakeholders meeting was called. The IR-related services (digitization, etc.) were first outlined based on feedback from this meeting.

In 2008, a test version of the open source repository collection mapping tool developed. Also, the first IMLS grant application was rejected. A second attempt was made the following year.

September 2009: IMLS grant awarded
In addition to the Steering, Metadata and Technical committees already mentioned

Outreach and Evaluation committee oversees the faculty survey and focus groups, and the usability testing of the repositories.

Symposium and Workshop committee oversees today’s and tomorrow’s events.
Project staff
Kathy Tomajko: Principal Investigator (PI)
Jeff Carrico and Toby Graham: Co-PIs

Grant funded staff positions
GKR Manager
Technical Writer
DSpace Technical Lead

Other staff are providing in-kind support, including information technology and development, supervision of student assistants and interns, administration and management, and so on.

Next we’ll look at some of the services we’ve established.
The GKR metadata repository is an instance of DSpace 1.8.2, hosted at Georgia Tech. The content is metadata records, harvested by OAI-PMH and OAI-ORE protocols, from 10 GKR partners. Some partners use DSpace – among those there are different versions of DSpace in use – and two partners use Digital Commons software from Berkeley Electronic Press.

The GKR is arranged by discipline. The disciplines are actually DSpace communities and sub-communities. We harvest collections from the partner IRs, which are mapped to the disciplines using the mapping tool. We configured the GKR not to show collection names, so the user only experiences communities and items. We also broke the architecture of DSpace in order to map collections to multiple communities, and it turns out this works fine!

Search and browse results as well as item records are branded with institutional logos, so users can see where the content comes from as they search. Item records contain links to the original items. When a user wants to see the item bitstream, the link goes to the original item record, so the home institution gets a tick in its IR statistics.
This is a shot of the mapping tool. Users log in and see a list of their IR’s collections. When they click on a collection name, they see a list of the GKR disciplines. They map the collection to one or more disciplines using check boxes. Then the mapping tool publishes code that we use in harvesting to link the original collections to the GKR disciplines. We don’t display the collection names in the GKR because it would create confusion for users; the collection names don’t make sense outside of the context of the home IR.
We host four independent instances of DSpace on behalf of Albany, GHSU, CCGA and GGC. Each institution chose its own theme, and we have done a little customization in some cases. Each institution also received training on DSpace – how to create communities and collections, how to submit items, and how to interpret the GKR metadata guidelines.

Some of the content has been submitted by library staff at these sites, and some has been added by student interns from Valdosta State as part of the content submission service. Some of the content was also created through the digitization service, by scanners at the Digital Library of Georgia.
The digitization service is overseen by the content and metadata committee and by Andy Carter at the DLG, who supervises the student workers who do the scanning.

Anyone participating in the GKR can submit a project for digitization, using an online form. The C&M committee selects projects about once a semester – generally 3-4 – based on a weighted evaluation system as well as the capabilities of the DLG.

The collections that have been digitized so far include a historic journal from GHSU, a literary magazine from CCGA, yearbooks from Albany State, and many others. Most collections have been historic documents and photographs.
The content submission service, as envisioned in the grant narrative, has two components: first, metadata specialists from GKR partner libraries were available for hire – funded by the grant – to submit items to hosted repositories where the home institution lacked staff resources. However, there was not enough demand for this service, and it was never implemented.

Second, a virtual student internship, run through the Valdosta State University MLIS program, allowed for students in library school to get hands-on experience with DSpace, and allowed GKR institutions who lacked staff resources to increase the content in their IRs. This was a great project for experimenting with “virtual” student internships, as the students were allowed to work remotely, communicating with their advisor and site supervisors via e-mail or phone, and accessing files via Dropbox.

The faculty supervisor at Valdosta is Changwoo Yang. Dr. Yang and Dr. Oguz from UNC-Greensboro plan to publish a paper on virtual internships, and two of Dr. Yang’s students created a poster about their internships, which is on display at this event.
We offered assistance with copyright questions as part of the grant project. This was not a heavily used service. We created a clearinghouse of copyright, author’s rights and open access resources on our web site, and we set up a form to take in questions. However, most questions came via e-mail directly to me, or by phone, and often the topic was not copyright to start with. For example, someone might have a question about digitizing a faculty member’s DVD – a technical question about converting DVD to a digital file – not realizing that we’d have to investigate the copyright status of that DVD before getting to the nuts and bolts of digitization. I also introduced people to SHERPA/RoMEO for those cases where an interested faculty member had approached them about self-archiving. In 2 ½ years I have answered 8 questions.

We do not plan to continue this service after the grant, for liability reasons. However, we will keep our lists of resources online.
“The negotiated cost between the GKR and the MetaArchive Cooperative is $7,000 for three years ($2/GB x 3.5TBs = $7,000). After three years, the GKR institutions will pay the costs of the service. Georgia Tech will provide assistance in ingesting content into the MPN; it has experience with this process, having been a founding MetaArchive partner since 2004.”
Outreach

Surveys
Site visits
Focus groups
Usability test – spring 2012

We are using OBE to evaluate GKR services, which is an IMLS requirement.

Outcome 1: Expand the use of IRs to more colleges and universities in the U.S.

Outcome 2: Expand library professionals’ knowledge of consortial IR models, technologies, and practices in the U.S.

Outcome 3: Multiply the impact of USG scholarship.
GKR will meet its deliverables. We have also applied for and expect to receive a no-cost grant extension. We have been particularly frugal and will be under spent in September. Some of this is on purpose – for example, we have not made as many overnight trips as we originally planned, and we have not had a need to hire any graduate assistants as we originally budgeted for. Some of it was unforeseen – for example, savings due to staff resignations.

We were able to use some funds originally budgeted for a graduate assistant, to hire a consultant to do a cost analysis and business plan. We will use some travel funds to supplement the symposium and workshop budget. The rest will go toward a no-cost grant extension, which if approved by IMLS will start in October.

Our sustainability options look promising. I can’t say much, we’re still discussing our plans, but we think we may have found a low-cost solution through a vendor partner that may even allow other institutions not currently participating to join the GKR.
Tomorrow’s session will culminate in a panel discussion of lessons learned, but I can give you a few of my own ideas now.
We’ve had great partners.

First of all, RACL. RACL has been around since 1968. It is an advisory committee for the board of regents of the university system. The GKR concept was first introduced to this group, and several RACL members are on the GKR steering committee. RACL members have a long history of working together to exchange information, coordinate on library operations, and create statewide library services, including GALILEO. RACL supported the GKR because it was a consortial model that other states could replicate.

Next, GALILEO. GALILEO grew out of a RACL IT advisory committee, formed in 1989 and officially launched in 1995. Among its accomplishments are:
The GALILEO shared database service
The GALILEO Interconnected Libraries (GIL) ILS
The PINES online library catalog
The New Georgia Encyclopedia
The Digital Library of Georgia

Merryl Penson, GALILEO’s executive director, has said “The GKR continues this pattern of library cooperation, leveraging the working relationships that GALILEO has fostered among Georgia's college and university libraries to extend the benefits of Institutional Repositories across the university system.”

Finally, IMLS has been a terrific partner. IMLS program officers worked with Georgia Tech before and after the grant, offering feedback and suggestions, and supporting the GKR as we asked for additional funds or reallocation of funds to test ideas or make improvements.
The GKR has also had several champions, some of whom I’ve mentioned:

Tyler Walters
Julie Speer
Merryl Penson

And some others I have to mention as well:

Toby Graham, deputy university librarian at UGA and head of the DLG, is a co-PI on the GKR grant and one of our most helpful and dedicated supporters. Toby also has a lot of experience with grants.

Catherine Murray-Rust, VP for Learning Excellence and Dean of Libraries at Georgia Tech has been instrumental in creating our sustainability plan, with the assistance of associate deans Kathy Tomajko and Jeff Carrico – PI and Co-PI on the GKR.

The GKR steering committee has not only advised but also advocated for us. I have a few quotes to share from some of them.
Flexibility

One last thing I’ll mention: Flexibility

We could not have gotten to where we are without being able to adapt to changing conditions.

For example:

• One institution, originally set up in the grant as a GKR hosted site, decided to host their own IR. But we were able to find an additional partner, whose IR we are now hosting.
• 4 out of 5 committees have changed chairs
• The statewide Faculty survey – originally a pre-grant activity – did not actually launch until April 2012.
• The Metadata repository had more technological hurdles than expected due to DSpace itself, in addition to losing its dedicated developer.
• We’ve had to try out and reject several sustainability ideas.
• We’ve found ourselves with extra money to spend in the coming year and planning for activities not originally described in the grant narrative.
Thank you!

Marlee Givens
marlee.givens@library.gatech.edu

CC image via Flickr user Milanov: https://creativecommons.org