Integrating Scholarly Repository Services into Consortial Organizations and Statewide University Systems

Tyler Walters
Presented by: Sara Fuchs & Keith Gilbertson

Georgia Institute of Technology
Atlanta, Georgia, USA

INSTITUTE of **Museum** and **Library** SERVICES

The Problem

❑ Much of the University System of Georgia's intellectual product never appears in a permanent printed form

❑ Exists as disorganized pockets of digitally born objects & media scattered among individual hard drives, departmental servers, and removable storage media across an institution

❑ While USG's digital scholarly output reflects substantial investment of resources, assets and effort, it:

- Lacks curatorial stewardship;
- May be inaccessible;
- Exists on unsustainable hardware, software, or individual support;
- Needs future-proofing migration strategies.

INSTITUTE of MuSEUM and Library SERVICES

Open Repositories 2010

Types of Content

- Electronic theses and dissertations
- Research and technical reports
- Pre-print research
- Post-print research
- Working papers
- Conference papers
- Open access journals
- Service publications
- Research proposals

- Research data
- Organizational documents
- College and university archives
- Organizational reports
- Institutional planning and evaluation documents
- Learning objects
- Other resources that reflect the quality and diversity of the USG

INSTITUTE of **Museum** and **Library** SERVICES

Open Repositories 2010

GALILEO Knowledge Repository

❑ GALILEO: **G**eorgi**A LI**brary **LE**arning **O**nline

❑ Proposing a system-wide approach to institutional repositories as well as a collaborative strategy for promoting open access to scholarly information

❑ Concept developed by Regents Advisory Committee on Libraries (RACL) in August 2004

INSTITUTE of
Museum and **Library**
SERVICES

GKR: The Project Basics

❑ 3-yr project to finalize several components toward establishing a comprehensive statewide repository

❑ $1,664,650 project total: $857,005 (IMLS) & $833,946 (institution share)

❑ Leads: Georgia Institute of Technology & University of Georgia

❑ Participants: Georgia State Univ., Georgia Southern Univ., Kennesaw State Univ., Medical College of Georgia, Valdosta State Univ., Albany State Univ., College of Coastal Georgia

❑ Scholarly Communication Survey: North Georgia College and State University

INSTITUTE of MuseumandLibrary SERVICES

GKR Project Components

❑ Build a repository of standardized metadata harvested from IRs within the USG (GT and GKR Technical Committee)

❑ Hosting service for IRs at three USG institutions (GT)

❑ Establish IR-related services: copyright research (GT/UGA); digitization (UGA); content submission (GT/VSU); and preservation (GT)

❑ Partner with NGCSU to conduct an assessment survey of the USG faculty's usage and perceptions of IRs

❑ Document and make available the GKR organization model to others interested in establishing statewide IRs

❑ Develop and implement a statewide and consortial repositories symposium and workshop

Timeline

2009: Hire GKR Manager and Developer

2010: Create GKR metadata repository; Establish IR-Related Services

2011: Provide IR-Related Services

2012: Dissemination (guide and workshop)

GALILEO Knowledge Repository
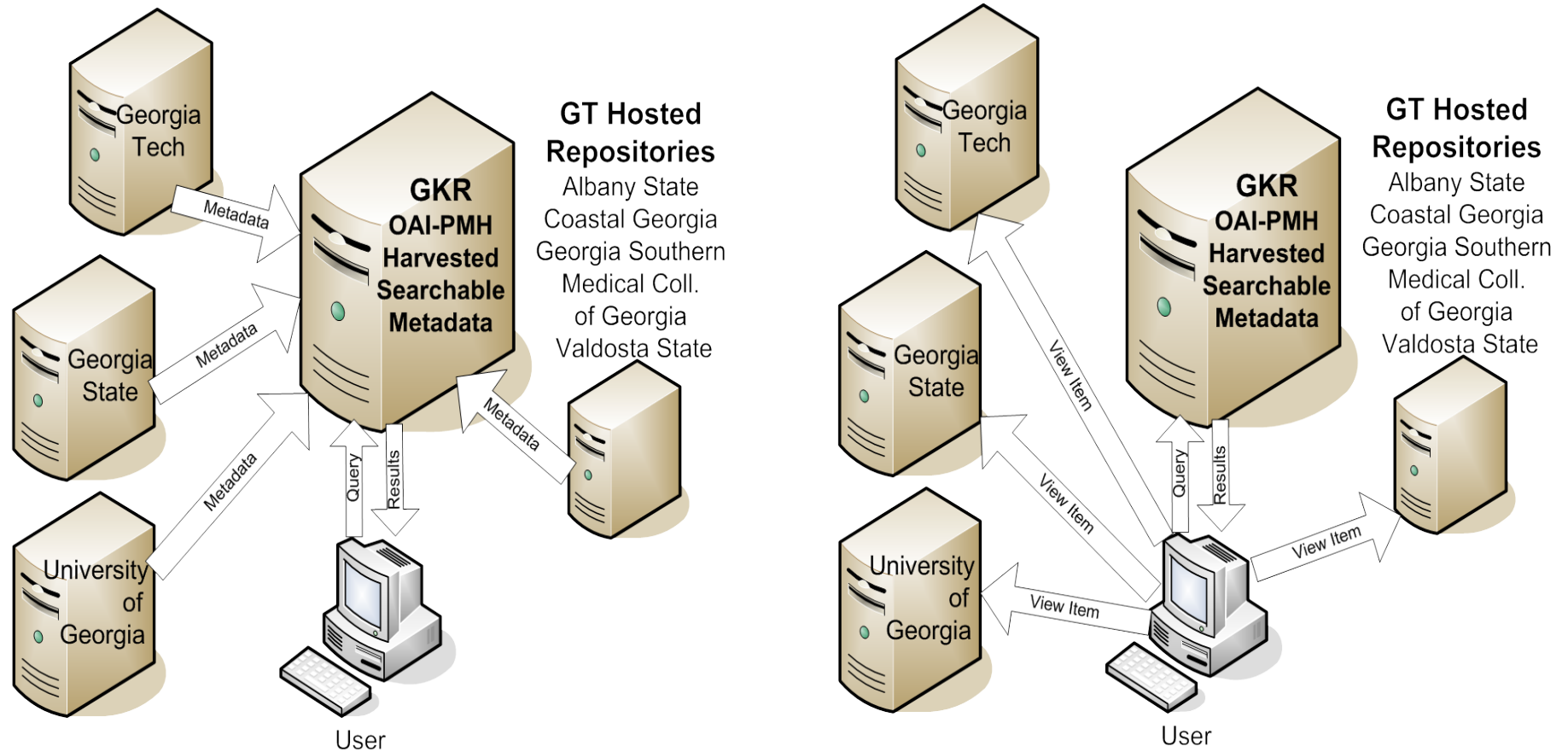
Technology Goals - Reduce, reuse, recycle

❑ Certain technological problems come up when creating a consortial repository system

❑ These problems have already been solved by others

❑ Let others do the work!

Open Repositories 2010
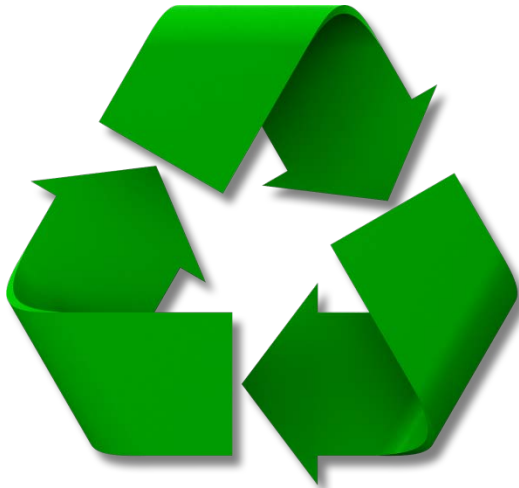
Work the Technical Lead hasn't done…

Thanks, Georgia Tech and USG Faculty and Staff!
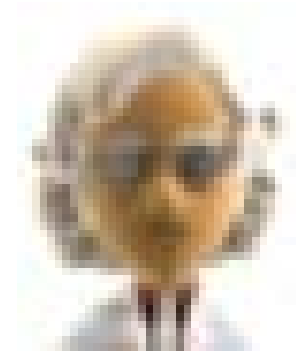
Current Infrastructure

Two servers at Georgia Institute of Technology

❑ 5+ terabytes of storage

❑ 24 Gigabytes of RAM/system

❑ Powerful CPU

❑ DSpace Repositories hosted at other Institutions

❑ Digital Commons Sites hosted on other servers

Thank you repository software developers!

INSTITUTE of
Museum and Library
SERVICES

Harvesting Process (Original Prototype)

Metadata from member repositories were pulled into the GKR using:

❑ PKP Harvester component

❑ MySQL data dump of harvested metadata

❑ Shell and awk scripts to format dumped metadata

❑ Command Line DSpace Import (check for errors and repeat if necessary)

Current Process (DSpace 1.6 Harvester)

Click the button!

Thank you Texas A&M!

Edit Collection: University of Georgia - New Version

Edit Metadata Assign Roles **Content Source**

Content source:
- ○ This is a standard DSpace collection
- ◉ This collection harvests its content from an external source

Harvested Collection Location

OAI Provider:	http://tolson.libs.uga.edu.:8080/oai/request
OAI Set id:	all
Metadata format:	DSpace Intermediate Metadata
Content being harvested:	Harvest metadata and references to bitstreams.
Last Harvest Result:	Harvest from http://tolson.libs.uga.edu.:8080/oai/request sucessful on 2010-06-11 11:06:54.122
Current harvest status:	Collection ready for harvest

[Change Settings] [Import Now] [Reset and Reimport Collection]

[Save] [Return]

Mapping Tool

Mapping Tool

Quickly Getting Content into Medical College of Georgia

MCG has Open Access articles in other databases

❑ [SWORD submission test from BioMed Central]

❑ Worked beautifully!

❑ MIT has SWORD publisher workflow project

❑ SWORD workshop 2:15 – 4:30 PM Friday July 9

Thank you SWORD, BioMed Central, and MIT pioneers!

INSTITUTE *of* **Museum** *and* **Library** SERVICES

PubMed Central Open Access Import

Import full-text articles from PubMed Central

❑Innovation by Scott Lapinski at Harvard Medical School Library

❑ Find and download articles

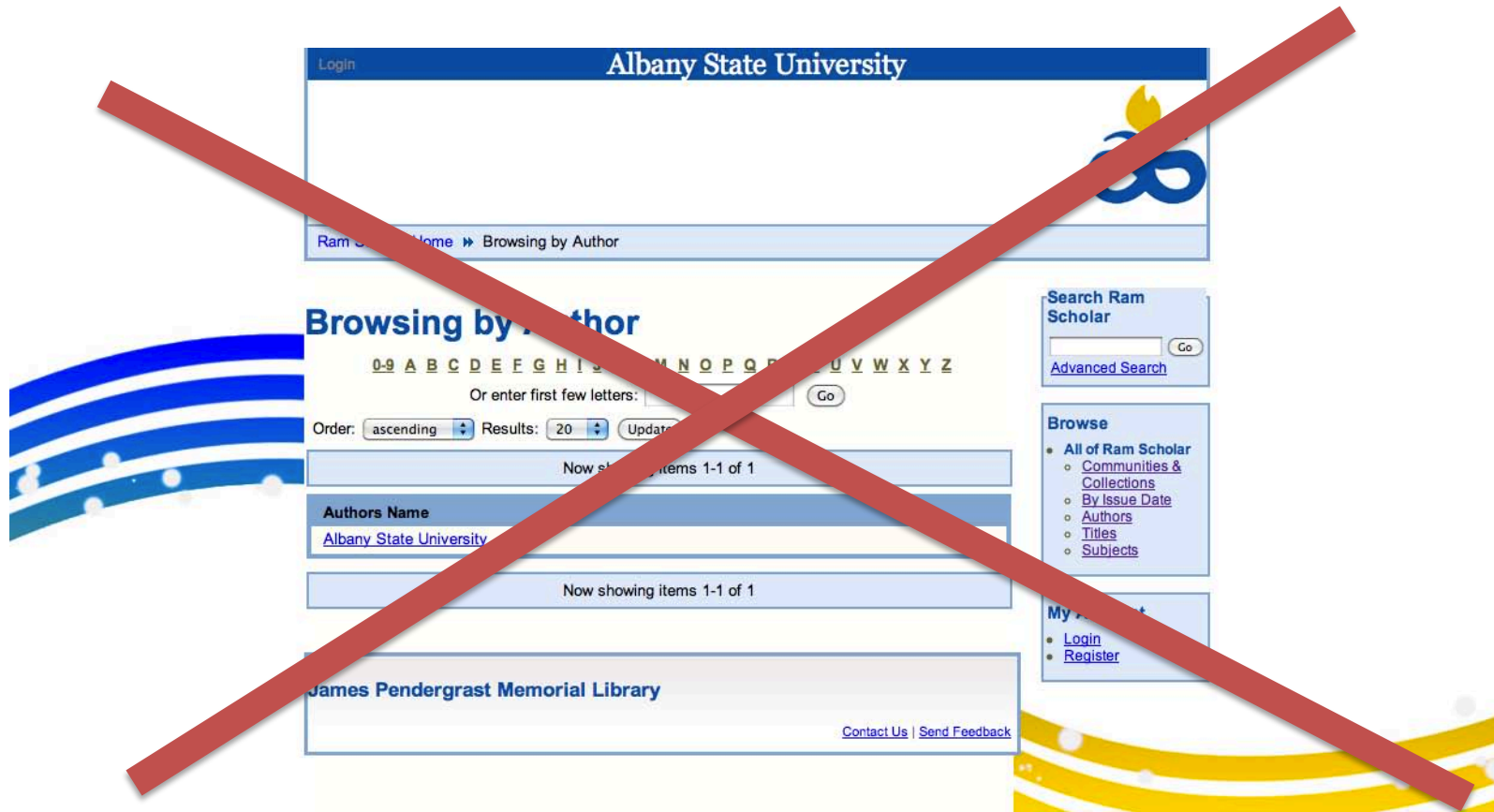❑ Limit to open access records only

❑ Convert metadata into DSpace format and import items

Thank you Scott Lapinski! (Harvard Medical School)
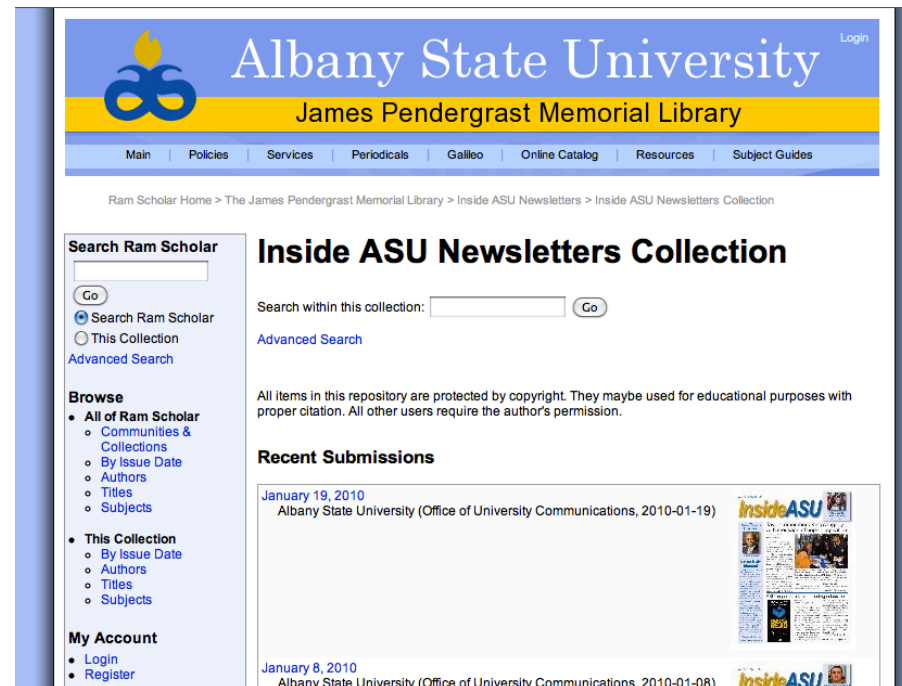
Theming with the DSpace xmlui ("zoom-louie")…

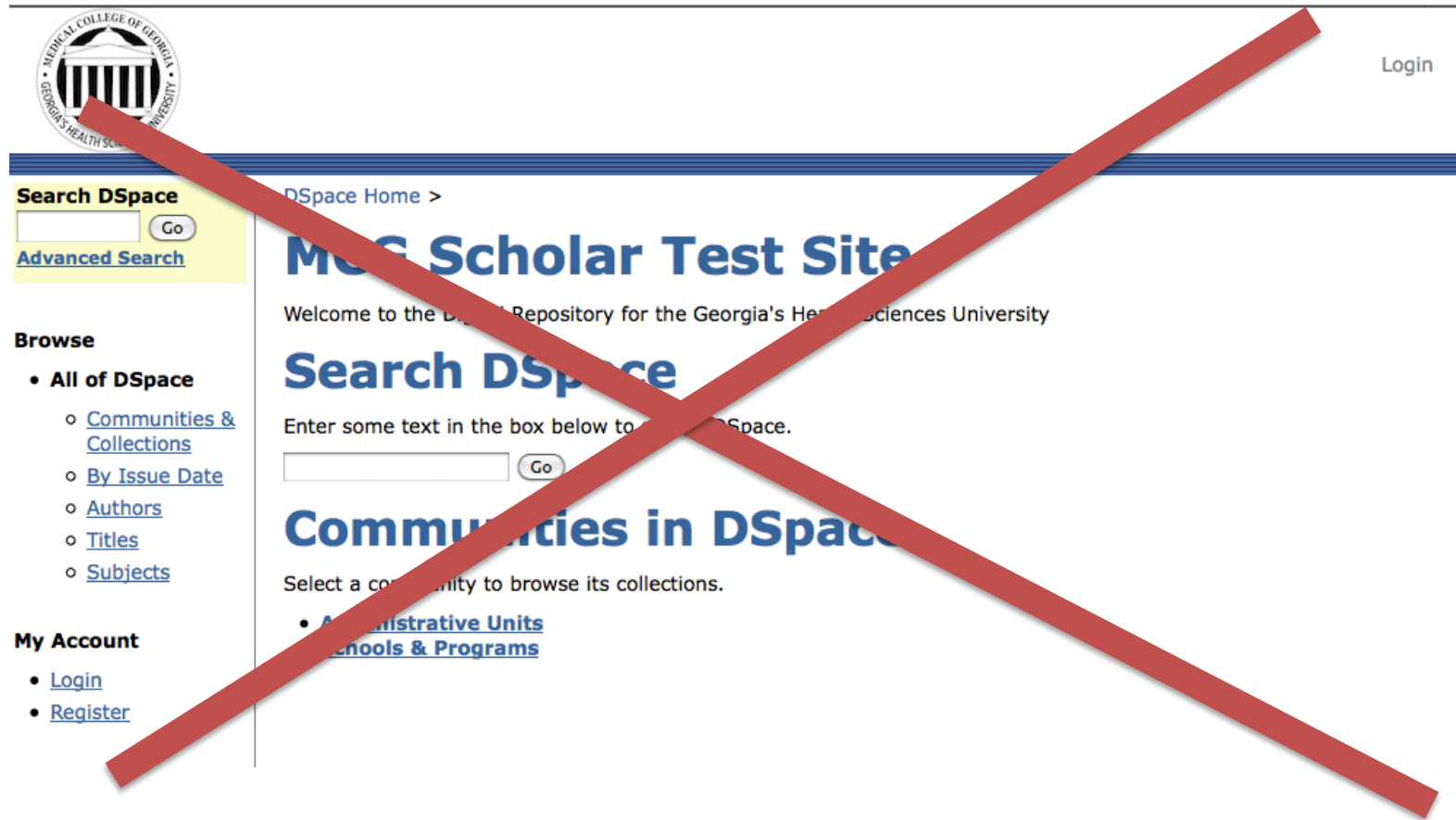# The Big, Bubbly Ribbon

"New" Albany State Theme

Thank you Jeff Craft (OhioLINK), Kent State,
Bill Anderson (Georgia Tech)!

Medical College of Georgia, Strike 1

Medical College of Georgia, Strike 2

MCG Theme

Medical College of Georgia wants the DSpace repository to be viewed as <u>the</u> *Institutional* Repository:

❑ Theme will *not* look like the library website; it will look more like the main institution website

Overheard at OR2010

❑ Institutional Repository software could benefit greatly from the work of design and interface professionals

Work that the Technical Lead is still waiting for others to do…

Open Repositories 2010

Backup and Recovery Plan

Currently in development:

❑ Necessary for credibility as a trusted service

❑ Rescue from hardware problems, developer mistakes, natural disasters, unnatural disasters

❑ Must be tested!

Early thank you to DuraCloud developers and anyone who sends me a sample backup plan – please?

INSTITUTE of
Museum and Library
SERVICES

Inline Media Playing
on Remote Repositories

❑ Some IRs now allow the user to play embedded multimedia content from the browser (example: @mire A/V streaming module for DSpace)

❑ We wish to allow playback of audio and video files stored in remote repositories (Example: Play a University of Georgia file from the GKR metadata site)

INSTITUTE *of* **Museum** *and* **Library**
SERVICES

Full-Text Searching of Remote Repositories

DSpace Harvester includes OAI-ORE support

❑ Trick 1: Import Thumbnails for Harvested Records

❑ Trick 2: Allow full-text search of articles for which GKR would have had only the metadata records

IR-Related Services

To reduce barriers to recruiting scholarly content:

❑ Copyright Research Services

❑ Digitization Services

❑ Content Submission Services

❑ Digital Preservation Services

Copyright Research Services

❑Provide guidance and assistance in resolving intellectual property concerns regarding IR submissions

❑Emails sent to the GKR Manager; Georgia Tech and University of Georgia will assist with requests

INSTITUTE of
MUSEUMandLibrary
SERVICES

Digitization Request Form

Significance

Current use: *

	Daily	Several times a week	Once or twice a month	Several times a year	Once a year
How often is the material accessed in analog form?	○	○	○	○	○

Potential use: *

	1	2	3	4	5
One a scale of 1-5, 1 highest and 5 lowest, to what extent would your institution benefit from access to the material in digital format?	○	○	○	○	○

Barriers: *

	1	2	3	4	5
One a scale of 1-5, 1 highest and 5 lowest, to what extent does the analog form limit use of the material?	○	○	○	○	○

Context: *

	Yes	No
Is the collection part of a larger digital collection?	○	○
Would scanning of these materials complete an existing digital collection?	○	○
Are there related digital collections online?	○	○

INSTITUTE of
Museum and Library
SERVICES

Open Repositories 2010

Content Submission Services

Interpreting/applying GKR Metadata Guidelines (modified Dublin Core):

❑ Resolving content submission issues in DSpace
❑ Submitting content and metadata when needed

Valdosta MLIS: 2-4 student interns as content submitters:
❑ Students earn 3 credit hours, gain valuable work experience
❑ GKR / VSU MLIS relationship unique in statewide repositories

INSTITUTE of
Museum and Library
SERVICES

GALILEO Knowledge Repository Metadata Guidelines by Priority

Status abbreviations:M = Mandatory ; MA = Mandatory if applicable ; Rec = Recommended ; RA = Recommended if applicable ; Opt = Optional

MANDATORY

Element	Qualifier	Status	Scope Note	Instructions / example
date	accessioned	M	Date DSpace takes possession of item.	Provided by DSpace. Example: 2005-01-04T18:08:50Z
date	available	M	Date or date range item became available to the public.	Provided by DSpace. Example: 2005-01-04T18:08:50Z
date	issued	M	Date of publication or distribution.	Example: 2001-01
identifier	uri	M	Uniform Resource Identifier	Provided by DSpace. Example: http://hdl.handle.net /123456789/60
description	Provenance (submitted)	M	History of object.	Provided by DSpace.
description	Provenance (approved)	M	History of object.	Provided by DSpace.
description	Provenance (available)	M	History of object.	Provided by DSpace.
				Draft standard language regarding use

Digital Preservation Services

- Built on LOCKSS (supports "distributed digital replication" approach)

- Closed Archive (no direct public access)

- Automated format emulation tools

- Low Cost (planned minimal expense, low barriers to adoption for mid-size institutions)

- Flexible, adaptable multi-inst. model

- LC / NDIIPP partnership (1 of 8 initial)

http://metaarchive.org/

Education for Hosted Sites

❑ Site visits planned for each hosted site

- GKR staff meet with library staff and research faculty to introduce them to the GKR project, IRs and scholarly communication

❑ DSpace training via Wimba collaborative learning software

- Repositories, DSpace structure, creating communities and collections, item submission and searching

❑ Metadata training via Wimba

- Dublin Core best practices, applying GKR metadata guidelines in DSpace

INSTITUTE of
MUSEUM and Library
SERVICES

Open Repositories 2010

USG Faculty IR Interest Survey

❑ GKR sponsoring a USG-wide faculty survey to assess perceptions, experiences w/ IR use, author's rights issues, OA publishing activity

❑ Survey by Jennifer Campbell-Meier (NGCSU Library).
   ▪ Ph.D. work, Communication & Information Sci., Univ. of Hawaii

❑ Results analyzed with GKR Outreach and Evaluation Committee
   ▪ Used to improve the technologies, services, marketing

❑ Learn about perceptions held by academicians about IRs and Open Access models and build business strategies to address them

❑ No surveys of IR use at statewide level
   ▪ Will inform GKR's development / Serve as model survey

INSTITUTE of
MUSEUM and Library
SERVICES

The GKR Model

Last deliverable:

☐ Document the GKR organizational model

☐ Document the GKR technical model:

- Guidelines for searching, harvesting repository metadata
- Technical specifications for GKR technologies

☐ Communicate, disseminate to statewide and consortial organizations via:

- GKR web site
- Presentations at meetings and professional conferences
- Articles and announcements through library publications

IMPACT: GKR

❑ Increase USG visibility to its digital scholarship and research

❑ Promote information sharing and discovery of research from a single web site

❑ Improve access to learning for the citizens of Georgia at large

❑ Create outlet for new forms of instructional media and scholarship

❑ Provide stewardship for the least permanent (i.e. non-published) elements of the USG's intellectual works

❑ Demonstrate effectiveness of IRs to USG faculty

❑ Advance scholarly communication

INSTITUTE *of* **Museum** and **Library** SERVICES

Links

GKR Project Site:

http://www.library.gatech.edu/gkr/

Training and Staging Site:

http://gkr-dev.library.gatech.edu/

Thank You!

Tyler Walters

tyler.walters@library.gatech.edu

Sara Fuchs

sara.fuchs@library.gatech.edu

Keith Gilbertson

keith.gilbertson@library.gatech.edu

GKR Manager: marlee.givens@library.gatech.edu

INSTITUTE of
MuseumandLibrary
SERVICES

Open Repositories 2010