

1 **Slide 1** **Policy issues**

2

3 This is where we start to get into the fun stuff. There is a lot of overlap between policy
4 issues and how you market and promote an IR. Some of these concepts will be covered
5 from a slightly different angle in both of the next two sessions.

6

7

8

9 **Slide 2** **Document and publicize policies**

10

11 Whatever policies you have, document and make them publicly available. This will
12 save you a lot of time and is an essential part of becoming a TDR.

13

14

15

16 **Slide 3** **Division of responsibilities ...**

17

18 One of the policy decisions you'll need to make is deciding who will be doing what. I
19 would recommend flexibility in this, as you'll probably find that many things change as
20 you go along.

21

1 **Slide 4** **Target group and users**

- 2
- 3 In the last session I identified some of the roles within the hosting institution.
- 4 Your target group and users can help with:
- 5 • Identifying and submitting content
 - 6 • Verifying copyright
 - 7 • Serving as advisers
 - 8 • Letting you know of service problems

9

10

11

12 **Slide 5** **Steering committee**

13

14 You might find it useful to have a steering committee. This group will provide a forum

15 for making policy decisions. Our original task force morphed into a steering committee

16 and has gained some new members

17

18

19

20

21

1 **Slide 6** **Structure, definition, name**

2

3 Another key policy issue is how you plan to structure the archive, organize the materials
4 that you collect.

5

6

7

8 **Slide 7** **Structure of the archive**

9

10 The structure of the archive depends on a lot of things. The structure will depend on the
11 software and your technical support

12

13 • Communities, collections, titles, files – what kind of structure does the software
14 support easily?

15 • Does it facilitate both author and mediated submissions?

16 • Do you need to make any local modifications to the software?

17 It also depends on your vision of the IR and the resources you have available

18 • Target group(s)

19 • Type of content

20 • Staffing

21

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21

Slide 8 Software implications

This is the home page of UO's IR, Scholars' Bank. The software we use, DSpace, has determined the basic structure of the archive. DSpace allows several levels of organization: communities, sub-communities, and collections. Individual titles within a collection can have single or multiple files. Without modifying the code, you can customize the look and feel to some degree, particularly at the top and the right-hand side. But the basic structure is the same for all DSpace installations, unless the institution has invested the time and money needed to modify the source code.

One policy decision if you use open-source software is if you are willing and able to modify the source code to give yourself a more customized approach? If you have the technical expertise, when will you do it? We have in-house technical expertise for modifying the DSpace source code but have only done it twice. Once was to fix an indexing error that we submitted to SourceForge for enhancing the next software release and the other was to give us greater flexibility for something that we considered a high priority for our marketing efforts.

1 **Slide 9** **Software implications**

2

3 This is the University of Utah's IR home page. They use CONTENTdm for
4 their archive and that has affected the structure of their archive, just as
5 whatever software choice you make will affect yours.

6

7

8 **Slide 10** **software implications**

9

10 Just to show you an example of how the structure might work. This is the hierarchical
11 structure of the DSpace software. Individual items (titles) can also be mapped to
12 multiple collections.

13

14 The use of the out-of-the-box software will require you to make any number of policy
15 decisions – will you use all of the software's features? Will you accept its organization
16 and world view? Because a software package very definitely does present a world view.
17 DSpace reflected the particular assumptions that MIT had about scholarly
18 communication, an assumption of author self-submission, and the nature of academic
19 communities. Although we shared those assumptions initially, we found them being
20 challenged and stretched almost immediately and thus found the DSpace world view
21 sometimes limiting and annoying. It is getting better.

1 **Slide 11 software implications**

2

3 This shows how a community page looks in our IR – it presents all the sub-
4 communities and collections that are part of that community. DSpace
5 software allows you to customize the content of these pages and add logos or
6 images to give a collection a specific brand or look. It also supports different
7 **policies** of submission and use for each collection. I think you will want to
8 have at least this level of flexibility, because one size does not fit all. No two
9 groups we have worked with has wanted the same thing done the same way.
10 In fact, they want a dizzying variety of things, many of which you will find
11 yourself unable to provide because of the limitations of the software. (And
12 sometimes you will find yourself offering up a silent prayer of thanks that
13 you can blame the software for not being able to implement some of their
14 visions.

15

16

17

18

19

1 Slide 12

2

3 This shows you a record for an individual item or title from the Ural State
4 University in Russia. They use DSpace software. In that software package,
5 each title (or what they call item) can have one or multiple files. Even if you
6 can't read Russian, this will seem very reminiscent of a library catalog
7 display. Although to get the labels and search boxes to appear with Cyrillic
8 script required modifying the underlying source code.

9

10 Different software packages will present you with a different structure and
11 challenges out of the box. Be alert to this because the software will affect the
12 conversations you have with your users and determine the options you can
13 present them with. It helps to have made some key policy decisions before
14 you start collecting content. You can modify them as you go along, but at
15 least be aware of some of the issues and have thought through your ideal
16 answer.

17

18

19

20

1 **Slide 13 software implications**

2

3 This is a record for an item with Utah's IR – different software, different
4 structure. They use CONTENTdm which provides more flexibility out of the
5 box in the choice and display of metadata than does DSpace. This has clear
6 policy implications. Sometimes you can have too much flexibility. One
7 policy implication is how much you are willing to customize the field labels,
8 order of fields, and type of information you collect for each collection.

9 Imagine what it would be like to set up different cataloging and OPAC
10 display standards for each academic department or discipline. That's the kind
11 of flexibility that CONTENTdm gives you. There are pros and cons for all of
12 the software packages available to you.

13

14 With DSpace such customization is much more difficult, requiring
15 modification of the source code and not just choosing options through the
16 administrative interface. This can be both a blessing and a curse, as I've said.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

Slide 14

This shows an actual file that has been opened from within our IR. How does the software handle the ingest, storage, and display of individual files? This will affect the policies you set up. Does the software support batchloading of files and metadata? Does the software perform checksums on files as they are loaded? Does each file have its own unique and persistent identifier? DSpace provides a unique identifier at all levels of its hierarchy – communities, collections, individual titles, and the separate files that make up a title all have their own unique handle that can be linked to and cited. Knowing the strengths and weakness of your software of choice will affect many of your policies.

1 **Slide 15 definition**

2

3 How you define your archive will affect all of your policies. This is how the
4 UO defines its institutional repository. Each institution will have a slightly
5 different definition. For us, the current definition is:

6

7 • The content is somehow affiliated with the UO

8 • Cumulative and perpetual

9 • Open access

10 • Interoperable

11

12 And ours has changed dramatically over time in relation to the final element
13 of the definition. I'll discuss this more in the marketing session.

14

15 • Academic content, or in support of the academic mission

16

17 Opening up our definition in this way has enabled us to collect a wide variety
18 of campus publications, such as newsletters and planning documents.

19

1 **Slide 16 The name**

2
3 If ever there was a political issue, the name you give your IR is one. How
4 you refer to your archive will also be a major factor in how you market it.
5 The name should reflect your archive's basic orientation or definition. But
6 make no mistake – this is a major policy issue, as well.

7
8 Scholars' Bank was intended to convey a safe place to deposit scholarly
9 content. Some faculty think that the term "scholars" automatically excludes
10 student work.

11
12 Interestingly, when I presented a variation of this pre-conference in Spain
13 last September, the Spaniards had a very negative connotation attached to the
14 word "bank"

15
16 With our name now, the most common challenge I face is getting other
17 people to spell it right. It's now a toss-up whether I spend more time telling
18 people the correct spelling of my last name or the correct spelling of
19 Scholars' Bank. That apostrophe following the "s" is a real killer.

1 **Slide 17 definition of communities, collections, users**

2

3 How you define your target groups – your communities, the collections, the users are
4 key policy issues. We require a group to have some kind of official affiliation with the
5 University before we will set them up in our IR. One locally-based group which actually
6 has a lot of good academic content wanted to archive their work in our IR. Even though
7 the group was headed by a member of the faculty, it did not itself have any affiliation
8 with the university. That and the fact that there were numerous copyright concerns led
9 us to a decision not to allow them to deposit their materials in our archive.

10

11

12 **Slide 18 General guidelines**

13

14 We have general guidelines that we let everyone see. Not that anyone but us looks at
15 them but you need to think through these issues. Because your vision will be stretched
16 almost every day once you start to have some success. People will want to use the
17 repository in creative ways. I don't know if you can read these guidelines, but one key
18 one is that we expect the material deposited in the archive to be free of copyright
19 restrictions. But how we understand that and how faculty understand it is not the same.
20 You can come to our site and see all of these guidelines and other supporting
21 documentation.

1 **Slide 19 Definition of communities, collections, users**

2

3 You'll need to think about:

4

- 5 • Definition of community – what IS a community? Do communities match the
- 6 departmental structure of your campus or city? Do you have sub-communities?
- 7 Who makes the decision about what a community is?
- 8 • Limits on the number and type of collections?
- 9 • Different rules for different groups of users – students, faculty, local, distant
- 10 • Commercial use permitted or not? Any cost recovery models you want to apply?

11

12

13

14 **Slide 20 Establishment of communities**

15

16 You'll need to have policies in place for how a group can be established in your archive.
17 Regardless of the software you use, you'll need to decide how to organize the content
18 and how different groupings can be set up. There will be dozens of small and large
19 decisions that will come up as you start to identify potential authors or content within
20 your institution. The watchword is, be prepared to change your definition.

21

1 **Slide 21 Academic department**

2

3 This is a sample community in our IR that represents an academic department. This was
4 our very first non-library community. We used to have long, intensive discussions
5 before setting up such communities and required the department head to sign off on it.
6 Now, I set them up routinely whenever any faculty member wants to contribute to the
7 archive. The long back and forth discussions always ended up with the same result – the
8 departments would accept, without modification, the draft sites that I set up for them.
9 The only difference is that it used to take far longer. Now, I tell them after the fact and
10 ask if they would like anything modified.

11

12

13

14 **Slide 22 Academic program**

15

16 This community represents an academic program and not a department.
17 Setting up communities like this requires great flexibility. Trying to nail
18 them down to an official name or to identify their place in the institution's
19 organizational structure is a challenge. So, really, my background as a
20 cataloger was ideal preparation for this outreach to our campus.

21

1 **Slide 23 administrative dept.**

2

3 This is a campus administrative unit, not an academic unit at all. One policy
4 decision that you can see reflected on this page is a decision to link out to the
5 group's own web site. Typically, in setting them up I visit their web sites, lift
6 a few paragraphs that seem to explain succinctly what they're about, and
7 then link to the full site for more information.

8

9

10

11

12 **Slide 24 Student community**

13

14 This is a student community, or one run on behalf of students. This is the
15 only community devoted entirely to student work, although there are many
16 student collections connected to other communities.

17

18

19

1 **Slide 25 Faculty community**

2

3 This is a broad-based community that we have defined that doesn't match
4 any particular organizational community but that allows us a single place to
5 create collections for individual faculty. I'm showing you some examples of
6 some of our communities because they each necessitated one or more policy
7 decision on our part. Often, I was making policy decisions without realizing
8 that's what I was doing.

9

10 If I set up a collection for an individual, I also always map his or her work to
11 the academic department of which they are a member as well. This was a
12 policy decision that I made because it seemed logical and it also gave me an
13 opportunity to expand the communities of academic departments without
14 having to have long discussions first.

15

16

17

18

19

1 **Slide 26 sub-community**

2

3 Will you have sub-communities? Dspace software allows this, as I
4 mentioned. Just because it's allowed, do you want to do it? If you do it, how
5 will you define the sub-community? Most materials can be organized
6 multiple ways. If any of you have a cataloging background, you are aware of
7 this.

8

9 A policy decision must be made about whether you want to have a consistent
10 format and organization for similar types of content, or will you permit each
11 group to make its own decision about the organization of their community
12 and collections?

13

14 We basically permit each group to decide how it wants to present itself and
15 organize itself within the IR. But this can be a challenge when you get into
16 interdisciplinary programs. On several occasions, it almost came down to
17 tossing a coin to decide how and where to affiliate a collection to some
18 appropriate community. This is highly political.

19

1 **Slide 27 administrative documents**

2

3 This is a sub-community within Special Collections and University Archives.
4 Of course, Special Collections is organizationally a sub-community within
5 the libraries but we decided to set them up as their own community in the
6 archive because of software issues and also because of the level of
7 prominence we wanted to give them. So, in setting up your groups, think
8 about how closely you plan to follow the existing organizational structure of
9 your institution. Are you willing to present your institution's resources in
10 new ways?

11

12 Once you set up a structure it can sometimes be hard to modify it. We have
13 occasionally gone into the underlying database and remapped collections to
14 entirely different communities because the people involved changed their
15 minds.

16

17

18

19

1 **Slide 28 Limits on the number or type**

2

- 3 • Will you place limits on the number or types of collections.
- 4 • If so, who makes the decision
- 5 • What are the determining factors?

6

7 We usually do not place limits but sometimes faculty are SO creative that it
8 can make your head hurt. One community has managed to reuse the same
9 content 2 or 3 times on their own web site to make it look as if they had
10 accomplished a lot more than they had. And their mappings to their own
11 different collections were not logical or consistent. I drew some lines
12 working with them simply because I could see that it would continue to be a
13 nightmare to keep up with their frequent redefinition and presentation of
14 themselves.

15

16

17

18

1 **Slide 29 Types of collections**

2

3 These are some of the possible collection types that we have incorporated
4 into our IR. Each one of these represented a policy decision when we decided
5 to accept or solicit this kind of content. I'll show you some examples to
6 illustrate what I mean but I won't show you examples from all types.

7

8

9

10

11 **Slide 30 Collection guidelines**

12

13 This page from the Australian National University's IR gives clear guidance
14 on the guidelines and standards for collections. I highly recommend this kind
15 of clarity and transparency about your policies. To be truthful, we find
16 ourselves often making ad-hoc decisions and then going back and
17 legitimizing them with official policy statements.

18

19

1 **Slide 31 Informational**

2

3 This is an example of what I call an informational collection – a campus
4 newsletter. We have numerous newsletters now. Acquiring this publication
5 was a turning point for us – it became easy and natural to acquire these
6 publications. And they are extremely useful and heavily consulted through
7 our IR.

8

9

10

11 **Slide 32 administrative**

12

13 This is an administrative collection – campus planning documents (which we
14 initially resisted putting in the IR but which we've found gets very heavily
15 used.) This particular plan has been consulted 170 times since we first started
16 tracking the statistics a year ago.

17

18

19

1 **Slide 33 Primary resources**

2

3 Will you collect raw data or other primary resources? We acquired this
4 collection when the agency was defunded. If you accept this kind of content,
5 will you attempt to provide any user assistance for using the data? We accept
6 the files and try to preserve them, but we draw the line at helping people use
7 them. I have received emails from around the country about this specific
8 collection and have tried to point users to places where they can get
9 assistance reading such files.

10

11

12

13 **Slide 34 Images**

14

15 This is from an image collection in the Australian National University's IR.
16 Will you accept this type of collection in your IR? This is a policy decision
17 you will need to make. We have so far resisted using our IR to archive single
18 images that are not connected with a broader resource. But we are rethinking
19 that because of a community request.

1 **Slide 35 Society publication**

2
3 This is another type of collection – the electronic journal produced by a society of
4 scholars. I find it endlessly fascinating to discover the variety of content that our
5 institution is creating.

6

7

8 **Slide 36 Faculty collection**

9

10 This is a collection of faculty works – also archived in their disciplinary archive, called
11 RePEc. This was our first collection and right from the start we faced 3 different policy
12 decisions:

- 13 1. whether our archive had to be the exclusive repository for content
- 14 2. whether we would harvest metadata and files and submit materials on behalf of
15 authors
- 16 3. whether we would accept works where the authors were not themselves affiliated
17 with the University

18

19 Another major policy decision was to be very proactive in going after content. We
20 periodically check for new papers and harvest and load them into our IR.

21

1 **Slide 37 Faculty collection**

2

3 This is an individual faculty member's collection? Will you allow or support
4 this type of collection? Remember the peacock. (Poor John, he's becoming
5 famous, unbeknownst to him.)

6

7 **Slide 38 Student collection**

8

9 This will be a major policy decision – will you have collections of student
10 work? If so, who determines what goes in? We require individual faculty or
11 departmental sponsorship and authorization for student collections. I have
12 had 2 students attempt to upload materials that were not authorized and the
13 review process we have in place stopped them.

14

15 Some people find it problematic to mix student and faculty work in the same
16 archive.

17

18

19

1 **Slide 39 Theses**

2

3 While some institutions balk at most student content, they can accept the idea
4 of collecting theses and dissertations. If you do make a policy decision to
5 accept them, I wish you luck in actually capturing them. Graduate Schools
6 can be some of the most change-resistant groups you will ever deal with –
7 this is not just the UO experience.

8

9 We have worked long and hard to get the Graduate School to assist us with
10 collecting electronic theses and dissertations. Every time we think we've
11 reached a solution, a new problem is found. We have created special pages,
12 special guides, special forms, special processes for them. Each time they set
13 us a new task, we meet it – almost immediately.

14

15 Story of “moral obligation”

16 Story of digitizing and mixture with self-submissions

17

18

19

1 **Slide 40 Different rules**

2

3 Will you have different rules and standards for different groups?

4 For instance:

- 5 • One set of rules and services for students
6 • Another set of rules and services for faculty
7 • Internal users versus external

8

9 We do have some different rules for different categories of authors, but almost
10 everything in the archive is fully open to anyone who wants to look at and download it,
11 whether they're a member of the UO community or not. Faculty get a higher level of
12 service and have essentially a lifetime commitment to their careers.

13

14

15 **Slide 41 Services for different users**

16

17 If you do have different categories of users with different rules and services, document
18 the different policy decisions and consider providing entry pages or resource pages for
19 the different groups. We target our services to the different groups.

20

1 **Slide 42 Commercial use**

2

3 As you get into this, you will be surprised at the different groups that come
4 out of the woodwork that want to put their material in your archive and
5 relieve them of long-term responsibility. We never anticipated having to
6 decide on whether commercial use would be made of our archive, but we
7 were approached by several groups who wanted us to host their content and
8 provide an e-commerce interface for them. We made a policy decision not to
9 accept commercial content, but we made the decision more because of
10 staffing and resource limitations than because of philosophical concerns. The
11 type of content otherwise met our IR's definition and vision.

12

13 As you can see here, MIT has decided to sell access to some of the content in
14 their IR to people who are not from MIT. We are now reconsidering this.

15

16

17

18

19

1 **Slide 43 Control of content**

2

3 Another major policy decision is who controls the content in an IR In terms
4 of reviewing it for quality, revising it, withdrawing it.

5

6

7

8 **Slide 44 Control of content**

9

10 These are some of the issues you will need to consider regarding the content:

11

- 12 • Who sets the standards for the content? – *Vice Provost's comment –*
13 *garbage in, garbage out – tell the full story*
- 14 • What type of material is acceptable?
- 15 • Who owns the content once deposited?
- 16 • Will you restrict access to any content?

17

18

19

1

2 Slide 45 Type of content accepted

3

4 Whatever you decide, formalize it somewhere on your site and document it.

5

6

7

8

9 Slide 46 Type of content accepted

10

11 Will you digitize content? Will it be your policy to accept only born digital
12 content or will you accept digitized content? We started out accepting only
13 born-digital material but we now routinely digitize content. This was a policy
14 decision I made single-handedly since my staff have expertise and access to
15 equipment for scanning materials and creating high-quality digital
16 surrogates.

17

18

19

1 **Slide 47 Type of content accepted**

2

3 A major policy issue will be what types of files you accept and then what
4 type of support you expect to provide for those files. Do you have a plan for
5 preserving different file types? Here you see multi-media files.

6

7 More and more people are coming to us wanting to archive not just the files,
8 but also their particular interactive interfaces to those files. This is really
9 stretching our expertise and our resources enormously.

10

11 Think about how you would handle such situations. Define your limits and
12 stick to your guns. It's better to provide good follow-through on a limited set
13 of services than it is to raise expectations that you can't in good faith meet.

14

15

16

17

18

1 **Slide 48 What happens when someone leaves?**

2

3 Since academics don't always spend their entire lives at a single university –
4 especially as universities employ more and more adjunct faculty – you need
5 to think about how you'll handle the content of people who move on and go
6 elsewhere. Our decision was once a faculty member, always a faculty
7 member. The content will stay in the archive and, if the individual wishes, he
8 will be able to continue to add new content from wherever he is in the future.

9

10 We've made a policy decision to support the faculty member throughout his
11 career. This policy decision has brought us some faculty members' content
12 just as they were leaving the University – and in one case this content is
13 receiving the highest use of any items in the repository.

14

15 We don't allow alumni to submit materials – unless they produced those
16 materials when they were students and the work is sanctioned by some
17 existing department or faculty member. So we don't have the same lifetime
18 commitment to students as we do to faculty.

19

1 **Slide 49 Submission and withdrawal**

2
3 Major policy decisions will be needed related to the submission process.
4 Who gets to submit materials to the archive, what is the workflow for
5 submission, will you allow any revision of content or withdrawal of files?
6
7
8

9 **Slide 50 submission and withdrawal**

- 10
11 • Some of the issues that you'll need to make policy decisions about are
12 whether or not your institution will provide any assistance with the
13 submissions. And if you will mediate the submissions, how far will you
14 go? Will you convert files to a preferred format, will you digitize from
15 hard copy, will you help authors investigate copyright?
16 • Do you expect that authors or content owners will handle all
17 submissions themselves? Will you be able to provide the service
18 support to help them resolve problems when they screw up during the
19 submission process? **(Continued ...)**

- 1 • Whoever does the submission, will you have any review mechanisms in
2 place? Someone who reviews the content to make sure it's appropriate,
3 someone who reviews files to make sure they're in accepted formats,
4 someone to review the metadata to make sure it's accurate and follows
5 whatever standards you've set up?
6
- 7 • Will you allow any replacement of files? The idea of revision is one
8 that comes up often when we discuss the archive with authors. What
9 will your policy be? The features and limitations of the software may
10 play as much of a role in your policy decisions as your philosophy of
11 what constitutes an IR
12
- 13 • Will you ever allow withdrawal of files? And if you remove all files,
14 will you remove the metadata as well?
15
16
17
18
19

1 **Slide 51 Policies for submission and withdrawal**

2

3 At UO, we have developed our policies for submission and withdrawal and have
4 documented those decisions as part of the contextual information we provide for the IR.
5 I highly recommend doing this.

6

7

8

9 **Slide 52 Policies: UO example**

10

11 These are some of our specific policy decisions that we have made regarding
12 the submission process.

- 13 • Every community determines the specific submission policies for its
14 collections
- 15 • Someone, whether from the library or the community itself, will revise
16 new submissions to make sure the content is appropriate for the
17 collection where they were submitted (**continued...**)
- 18 • Authors may submit their own work or they may ask the library for
19 assistance

1 **Slide 53 Policies: UO example**

2

3 These are more of the policy decisions we've made regarding submission
4 and withdrawal.

- 5 • At the start of every new collection, someone from the library reviews
6 the first few submissions to give feedback and advice on metadata and
7 file formats
- 8 • The library reserves the right to remove content – for example, in cases
9 of copyright violation
- 10 • The library reserves the right to withdraw content and return it to the
11 author or his representative in the event that the repository cannot be
12 maintained (**continued...**)
- 13 • The library reserves the right to change its guidelines without telling
14 users

15

16 Your institution's decisions will be different but the policy issues you'll face
17 will be similar. You'll need to think about these same issues.

18

19

1 **Slide 54 Challenges with mediated submissions**

2

3 If you decide as a policy that you will mediate submissions to your archive – to submit
4 files for the content owners – you will find yourself faced with a slew of policy
5 decisions.

6

7

8

9 **Slide 55 Capturing hyper links**

10

11 If you mediate submissions and harvest electronic web-based publications, you'll need
12 to decide how far you'll follow the links in the publication. This is an example of an
13 issue of our campus newsletter which we harvest from the web and convert into a stable
14 pdf (with the links working internally in the document). This particular issue covered
15 blogs and wikis and the size of the file kept growing as we followed links. Of course,
16 following links to external sites brings up issues not just of file size but also of
17 ownership of the content.

18

19 Our policy decision is that we ordinarily follow links one level down but
20 deactivate deeper links.

1 **Slide 56 Broken or erroneous links**

2

3 If you're harvesting legacy html publications, as we do, you will encounter
4 broken links or links that take you to content that has been revised
5 substantially since the original publication was produced. This was a link we
6 followed from a 1999 web publication that led to a page that had been
7 updated in 2005.

8

9 What will you do in those cases? Will you attempt to find the correct content,
10 either by tracking down broken links and capturing the correct content? Will
11 you go to the Internet archive and try to capture the content as it existed at
12 the time of the original publication?

13

14 These are all policy decisions that you will find yourself making, either
15 explicitly or implicitly. We didn't know what we were getting into, which
16 was probably just as well because we might have been paralyzed before we
17 started. As it was, we were in the middle of the process and committed to the
18 collection when these issues came up – which meant we had to find a
19 solution and make a policy decision quickly.

1 **Slide 57 Revision or replacement**

2

3 Will you permit the replacement of files? If so, in what circumstances?

4 Within Dspace, replacing bitstreams is limited to people with fairly high

5 authorizations – it's not something that the individual author would have

6 permission to do, unless we were prepared to open up the entire archive. So,

7 all bitstream replacements are done by library staff. If you permit it and are

8 prepared to facilitate it, under what circumstances will you permit it – and

9 how far will you go to facilitate it?

10

11 In our case, we will allow files to be replaced in the event of errors, but the

12 authors have to give us the corrected files. We will not go into an existing

13 file and make corrections to the file itself.

14

15 We will also allow additional files to be added to an item. Again, this is

16 something that we have to do on behalf of authors because of the nature of

17 the software. But it's a policy decision about whether you will allow this.

18

19

1 **Slide 58 Revision of submissions**

2

3 As always, we try to document our policy decisions and make that
4 information available to our public. That way, if there are questions, we can
5 refer people to the policy. It becomes far less personal that way.

6

7

8

9 **Slide 59 New version**

10

11 We don't allow replacement of files just because the author has come up
12 with a newer version. This was actually one of the earliest discussions we
13 had with authors, regarding working papers. They initially wanted to replace
14 earlier versions with the later revised versions. We talked them out of that
15 approach, noting that libraries have long collected paper versions of working
16 papers and do not routinely remove earlier versions or editions of works
17 when a newer version is published. Here you can see side by side two
18 different versions of the same working paper. Both are in the archive and
19 they have their own unique handle – URL.

1 **Slide 60 Withdrawal**

2

3 Will you allow any content to be withdrawn? In what cases? Will you remove just the
4 files or will you also remove the metadata record describing the file content?

5

6 We will remove files – at the request of the author or on our own initiative – if there are
7 serious errors or inaccuracies in the content. We have also once or twice removed files
8 when the author changed her mind and wanted the content removed. We no longer
9 remove the metadata, however, because the URLs and the metadata are harvested by
10 multiple services very quickly and we don't want people to be drawn into broken links.

11

12

13 **Slide 61 Example of a withdrawal**

14

15 This is an example of an item where the content was removed at the author's
16 insistence. **(CLICK)** Notice that we modified the metadata to alert users to
17 this fact. We have also since cataloged the paper version and added a note to
18 this record indicating it is available for circulation and ILL. This item has
19 been looked at 719 times by people wanting to read the paper. And it was a
20 great paper!

1 **Slide 62**

2
3 That last slide provides a nice segue into the next major policy area,
4 metadata.

5

6

7

8 **Slide 63 Metadata standards**

9

10 I come out of a cataloging background. But I have not taken a cataloging
11 approach to the IR's metadata. You'll see what I mean in a minute when I
12 show you some examples.

13

- 14 • In setting up your IR, I think it's important that you follow some
15 metadata standards. Many IRs support Dublin Core metadata and are
16 starting to support some other metadata standards. But even having that
17 as a given, you still have a lot of decisions to make. If your IR supports
18 Dublin Core, you'll struggle with how to describe many standard
19 bibliographic items, such as articles and books, using the Dublin Core

1 metadata elements set. There isn't a one-to-one match with MARC. So,
2 you'll need to make some policy decisions about where to put certain
3 kinds of information.

- 4
- 5 • Some of your decisions on metadata will be determined by what your
6 software can and cannot do – and you'll see some examples of what I
7 mean in a minute.
 - 8
 - 9 • If you allow or encourage author self-submission (which was the
10 original dream of most people going into IRs), how will you handle the
11 metadata? Will you review self-submissions and “clean up” the
12 metadata?
 - 13
 - 14 • Different disciplines have different terminology. How do you handle
15 the terminology from different disciplines within the same archive? Do
16 you worry about it and try to follow some universal controlled
17 vocabulary?
 - 18

- 1 • Will you be willing to do – as we are – whatever is necessary with the
2 metadata in order to find, organize and display the files appropriately?
3 We have the advantage with our IR in that I come out of a background
4 of almost 30 years involvement with cataloging, having been the head
5 of several cataloging departments. I was involved with our effort from
6 the beginning – originally as the co-chair of the group charged to
7 investigate the feasibility and now as the coordinator of the entire
8 archive.
9
- 10 • So, when I say, it's not cataloging, there is no one to argue with me.
11 And it isn't cataloging. I think it would be a mistake to try to turn it
12 into cataloging. We do follow some general cataloging principles, but
13 we don't sweat it. And I'll explain some of that as I show you a few
14 examples.
15
16

1 **Slide 64 Authors**

2

3 We try to follow basic cataloging standards – AACR/RDA – for personal and
4 corporate names. That means, more than anything for us, that we want to
5 have only one form of name for an author in the archive.

6

7 **CLICK HERE**

8

9 With Dspace, the software treats any variation as a different author, so
10 entries without a period display separately from those with a period. I
11 periodically go in and do cleanup on authors to avoid split files. Will you do
12 that? What standards will you have and how strictly will you follow them?

13

14 Other software packages make it possible to do more sophisticated authority
15 control. In those cases, you'll still need to decide which controlled list of
16 terms and names you'll be using within your archive, if you decide to use
17 them at all.

18

19

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

Slide 65 Description

Our IR uses Dublin Core metadata which is not as finely delineated as MARC, as I said. So, if you want to provide many of the same descriptive elements as are found in MARC records, you'll have to make decisions on which fields to put them in.

CLICK HERE

For those of you who are catalogers, you'll see that we combined the elements of a 300 and a 502 field into the description field. These are policy decisions, whether you make them implicitly or explicitly. I would recommend that you formalize and document your off-the-cuff decisions as you go along because it ultimately saves everyone time.

1 **Slide 66** **Appropriate level of metadata**

2
3 You'll need to make decisions about what is the appropriate level of
4 metadata. As I said, it's not cataloging.

5
6 **CLICK HERE**

7
8 If you enter individual issues of a serial publication, how deeply will you
9 index it through the metadata you provide? Will you provide indexing down
10 to the article level? Will you always be consistent? For us, we let our
11 communities dictate many of these decisions. Which means, that we **are not**
12 **consistent** throughout the archive, but try to maintain consistency within a
13 single collection – for every issue of the same journal.

14
15 We also do far less indexing of journal issues now that we have implemented
16 full-text searching capability for our text-based files.

17
18 These are policy decisions in part determined by the functionality of the
19 software.

1 **Slide 67 Logical and useful presentation**

2

3 In our collection of metadata, we are guided by the desire for a logical and
4 useful presentation. So, for instance, that means that we replicate the journal
5 title in the record for every issue of a journal. We also provide volume and
6 issue numbering as part of the title field.

7

8 This is definitely not cataloging but it allows for a clear, logical display of
9 the issues within our IR.

10

11

12

13 **Slide 68 Numbering**

14

15 This is another example of where we have been guided by the desire for a
16 clear display and have had to make decisions because of the software, as well
17 as the nature of the original publication.

18 **(CONTINUED ...)**

1 We were faced with two different problems here. One that the software cannot count
2 beyond 0-9, so to get numbers ten and higher to display in correct numerical order,
3 means that all numbers in a collection have to have the same number of digits – so we
4 now supply leading zeros for the earlier numbers.

5

6 The second problem we faced here –and I know that this will come as a great shock to
7 all your serialists – the publication itself was not consistent in its description of itself.
8 Sometimes it provided numbers, sometimes it just used seasonal designations.

9

10

11

12 **Slide 69 Numbering**

13

14 In coming up with an approach for the metadata, I consulted with my head of serials
15 cataloging and we came up with this approach – all of which was necessary to get things
16 to display in a logical order. So our policy decision for serial publications is to supply
17 whatever metadata is necessary to arrive at a logical, orderly display. This will mean
18 different things depending on the nature of the publication – as well the type of
19 organization that the owning community wants for the collection.

20

1 **Slide 70 Dissociation**

2

3 With this journal title, the editor wanted us to digitize each article separately
4 and to replicate the structure of the journal. So we created a separate
5 collection for each issue.

6

7 **Slide 71 Dissociation**

8

9 Then within the issue we were faced with the challenge of how to get the articles to
10 display in the order that they appeared in the journal – another requirement of the
11 content owner. **ARROW** You can see what we did here – all titles begin with the
12 journal title, followed by the volume and issue number we then included page numbers
13 – and had to make up page numbers for the cover, and tables of contents. Then we
14 added a subtitle with the title of the individual article. We also added in an alternate title
15 field with the article title (which you can't see on this display)

16

17 So, again, this reflects our policy of being guided by what the content owners or
18 community wants and then doing whatever is necessary to create a logical and useful
19 display. You'll have to make similar policy decisions, even if your decision
20 is different from ours.

1 **Slide 72 Institutional commitment**

2

3 A major policy decision that needs to be made early on (others of these can
4 be made and changed over time) is the level of institutional commitment. Are
5 you in this for the long haul?

6

7 I think it's realistic to expect this to take a couple of years to take hold on a
8 campus and it will take a lot of marketing. If you're not in it for the long
9 haul, I think you probably shouldn't even start.

10

11 And, certainly one of the primary marketing issues for us is that we are
12 saying that we will archive and preserve these materials – or return them to
13 the content owner if we find we cannot do so.

14

15

16

17

18

19

1 **Slide 73 Institutional commitment**

2
3 How permanent is the archive? What can your content providers expect?

4
5 How do you plan to guarantee the integrity of the files and the stability of the
6 archive. I outlined some strategies when I talked about digital preservation.
7 This is a major challenge but we have no business doing this if we're not
8 prepared to take appropriate steps

9

10

11

12 **Slide 74 Copyright**

13

14 You will face many policy decisions on the issue of copyright, permission,
15 and access to files. Copyright is extremely complex and unsettled. There are
16 numerous resources for you to consult on your handout. If you get into an IR,
17 you will become well acquainted with copyright issues – and you will find
18 yourself always doing some risk assessment.

19

1 **Slide 75 Copyright**

2

3 There are multiple issues within this and you will need to have thought of how you plan
4 to address all of these issues and develop policies and guidelines relating to them.

5

6 Do authors have the right to post their work in an IR? Do they have copyright to articles
7 that have been formally published and can they legally put them into your IR. This is
8 one of the most challenging areas to discuss with faculty – either they are terrified of
9 violating their publishers’ agreements (often without reason) or they have so routinely –
10 and unknowingly – violated copyright for years by posting their work to their own web
11 sites that they can’t understand why you won’t let them put the same works into the IR.

12

13 Will putting the work in IR affect later publication? This is a big question that occurs
14 with students and with faculty who are accustomed to developing multiple versions of
15 an item over time, such as working papers or to turning theses into books or articles.

16

17 Are authors turning over copyright to IR? Be clear about what your license agreement
18 says. I recommend that you set up a non-exclusive distribution license –w hitch
19 basically allows authors the right to use their content anywhere else they want,
20 whenever they want. Our standard license agreement asks for permission to make the
21 work freely and publicly available, the right to convert the file to a format that is easier

1 to preserve. And we also ask them to assert that they have the copyright to what they are
2 depositing in the archive.

3

4 Who gets to use content in the IR? Will you allow some items to have restricted access?

5 We do. For instance, one of our authors has been granted the right to make her content
6 available to her students for classroom use but not to the world at large. Within our
7 environment, that means that the professor must tell us who is authorized, the
8 individuals must register in the archive, and then we must set them up to be able to
9 access those specific files. It's fairly time-consuming but so far we haven't had many
10 requests for such restrictions.

11

12 What guarantees are there against plagiarism? This is a growing question. The answer is
13 that there are no guarantees. We have a variety of answers we give to this question. The
14 IR provides a date and time stamp that can help to prove "ownership" of a work, we
15 have a general copyright statement that says rights are restricted and works must be
16 properly cited, unless otherwise stated. Some of our authors cite Creative Commons
17 licenses. And we are investigating software tools that can help detect plagiarism.

18

19 *Read the Educause article if there's time*

20

21

1 **Slide 76 Copyright and permission**

2

3

4 Within Dspace, a permission file gets attached to every submission, based on the email
5 address of the person doing the actual submission. This was based on the dream of
6 author self-submission. Since the vast majority of the content in our archive was
7 submitted by me or my staff on behalf of authors, we had to make a policy decision
8 about how we would document that we had permission to make things publicly
9 available.

10

11 We have done and continue to do a variety of things. We have web forms where authors
12 can grant us permission – those come to us as emails that we convert to text files, add
13 into the archive attached to the submission and code so it is handled like the system-
14 generated license files. Sometimes we get paper permissions which we digitize and add
15 into the archive just as we do for email permissions.

16

17 You will have to decide how you will handle this. What will your policy be? Will you
18 simply require self-submission? If not, how do you plan to track on the fact that you
19 have been granted permission to archive and make the content available?

20

21

22

1 **Slide 77 Going it alone**

2

3 Another policy decision – one which may change over time – is whether you plan to go
4 it alone or whether you want to coordinate and combine efforts with other institutions.

5 We decided to go it alone initially. We found it hard enough to reach consensus even
6 within a single institution. We have subsequently had some discussions with other
7 interested parties in one of our consortia, but it has proved too hard to decide on the
8 approach we wanted to take and nothing has yet come of those discussions. If we all
9 subscribe to the same OAI-PMH standards, there is always the possibility of sharing
10 metadata later and having a multi-site search interface. IN the meantime, we have
11 registered our archive with services such as OAIster and other registries so that more
12 people can discover our content.

13

14

15

16 **Slide 78 Recercat**

17

18 There are examples of multi-institutional repositories. Catalunya's research universities
19 have set up a multi-institutional IR that could serve as a model if you and some sister
20 institutions were interested in working together. Whatever challenges there are for a
21 single institution you can expect to be multiplied for a multi-institutional IR.

22

1 **Slide 79 CARL**

2

3 Canada's research libraries have also done this, using **Public Knowledge Project** open
4 source software that is a federally funded research initiative located at the University of
5 British Columbia and Simon Fraser University.

6

7 If you do this with other institutions, just multiply by some x factor the complexity of
8 the policy decisions you'll need to make.

9