Slide 1 Chicken or an Egg

The title of my section: "A chicken or an egg" goes back to the quandary about which came first, the chicken or the egg.
I use this phrase in relation to planning a digital project because so often there is no planning – it's a case of doing first and planning mid-stream or thinking about planning for the next project. Digital projects tend to take on a life of their own and once you have a basic infrastructure in place, they can multiply like the proverbial rabbits.

I'm going to cover some of the things you should think about as you get ready to tackle a digital project. My experience with digital collections is from my previous position at Oregon where I was head of Metadata and Digital Library Services. Many of the examples I show you to illustrate my points will be taken from Oregon's collections and experiences.

Slide 2   Why, what, who, how, when

These are some of the who, what, why, how, when

Slide 3   Why Build digital collections

The key step in planning a digital project is being clear about why you're doing it. Should you do it just because everyone else is? No! You may feel pressured because of that factor but digital collections should meet certain needs.

Slide 4   Why – to provide access…

One of the strongest reasons for building a digital collection is To provide access to and awareness of under-utilized materials (special collections)

Slide 5   provide access

Materials in Special Collections, such as 400,000 photographs, Medieval manuscripts, unique manuscripts such as Oregon Trail diaries that we wanted to make our users and the general public aware of.

Slide 6   provide access - fafard

At Regina we've digitized slides of Joe Fafard's artwork and these will form the basis for U of R's first digital collection. This will increase awareness of our collections and of the artist's work outside the province

Slide 7   Why – physically fragile

To broaden access to physically fragile materials – things like glass plate negatives, manuscripts

Slide 8   **physically fragile - mss**

This is an image from a collection of medieval manuscripts that Oregon was building with the collaboration of several key faculty. The goal was to allow more people to have access to these kinds of rare or unique materials, including scholars from around the world who might otherwise not be able to study them at all.

Slide 9   **why – preserve at-risk**

To preserve some at risk materials (glass plate negatives, fragile archival material)

Slide 10   **at-risk**

This is an image that was scanned from a glass-plate negative. In addition to creating a digital image of this negative, Oregon created contact prints and production and master negatives for the selected images so that the original glass plates do not need to be handled again and so that the photographer's work will not be lost if the original is damaged or destroyed.

Slide 11   why - partnerships

To strengthen and build partnerships and collaborations with other cultural heritage institutions.

There are many opportunities for such collaborations within the province and granting agencies are fond of collaborative projects. Collaborations that start up around a digital project can often lead to broader collaboration in other areas.

Slide 12   partnerships – ed chapman

This image is from Oregon's Picturing the Cayuse collection which represents a unique partnership with the Confederated Tribes of the Umatilla Indian Reservation and the University of Oregon Libraries.

Slide 13   partnerships - basket

This image represents a collaborative project with the University of Oregon's Museum of Natural and Cultural History where the library is helping them to build a digital collection of some of their artifacts.

Library staff are teaching them how to use the digital content management software system, teaching them about scanning and metadata standards, and also hosting the collection for them.

Slide 14   why - instruction

To expand support for instructional programs.
Oregon's University Librarian, Deborah Carver, is adamant about developing collections that are not just collections of pretty pictures but that, instead, tie in directly to the University's instructional mission. All collections there are built with an eye to working with academic departments and even specific courses.

If you are serving an educational institution, this should probably be one of your reasons for planning a digital project

Slide 15   instruction – black panthers

This is an image from a collection that Oregon built as part of a specific class for their Honors College. Students selected primary source materials, primarily documents, from the UO presidential papers in the University Archives. The library scanned them and made them available within 24 hours with full-text searching capability. The students wrote research papers based on the primary source materials and the library used their research to assist in the selection of key documents from the period.

Slide 16   why – new delivery mechanisms

To explore new delivery mechanisms for content (full text searching of text)

Slide 17   new delivery – full text

Full-text searching of digitized text allows you to provide deeper access to some traditional materials. Many software packages allow you to store the full-text of scanned documents so that it can be searched. Be aware that the process of creating full-text is very labor-intensive.

Slide 18   why – digital landscape

Another reason you might want to get involved in digital projects is to help shape the digital landscape - things such as participate in standards testing and development.

Slide 19   landscape - WWDL

Through some of their collaborative projects, such as the Western Waters Digital Library which UO developed as a member of the Greater Western Library Alliance, Oregon worked with other institutions to refine digitization and metadata standards.

The aggregated collection uses CONTENTdm's multi-site server. In a 2005 presentation at CNI, I and a colleague from Utah talked about some of the challenges regarding metadata and others aspects that a multi-institutional project faces.

Slide 20   why – user demand

Our users are asking us to provide more and more digital content. And they want us to provide access to some of our own resources, not just to licensed resources that are available commercially. This could be a very powerful reason you have for starting a digital project.

Slide 21   user demand – presidential papers

This collection was built after the users started building their own collection using our materials. An Honors College Class was using papers from our archives and had scanned documents and photographs from the archives and had put them up on a web site. The University Archivist and I met with the professors and planned for the next class and developed a sustainable digital collection around the class work.

Slide 22   what to select

Once you are clear about why you want to create a digital collection, you then have to decide what to digitize.
Chances are, you will have more suitable materials than you could ever digitize. The idea that everything will one day be available in digital form is somewhat naïve.

Library and Archives Canada lists its selection criteria that reflect their particular mission. Your selection will be based on your own goals and resources.

Slide 23   select – also from LAC

Also from LAC

Slide 24   select – collection development criteria …

There are as many different guides on selection criteria as there are collections. Perhaps more. Some factors that you might consider are:

The digital collection should fit within your institution's overall collection parameters. Materials might be selected for their visual impact, historical importance, representation of a particular theme or practice, or because they meet the need of a special or high-profile user group.

Slide 25   select - copyright

In a Council on Library and Information Resources 1998 pub **Selecting Research Collections for Digitization** by Dan Hazen and others highlight these factors in helping to select what to digitize or preserve in a digital collection. The publication provides a decision-making matrix.

Slide 26   DPC coalition

The Digital Preservation Coalition provides an interactive assessment tool to help you select what to digitize.

Clearly defined selection policies will enable cost savings in terms of time taken to establish whether or not to select and also potential costs further down the track.

This Decision Tree may be used as a tool to construct or test such a policy for your organization. The decision process represented in the tree should be addressed by your policy for selection of digital materials for the long-term.

Slide 27   challenges – web design

As you plan your digital collection, you'll need to be aware what challenges you face.

One of the challenges is Web design – many institutions undertaking digital collections don't have specialized staff for this. At Oregon, we learned as we went along.

Slide 28   browsable subject lists

This is one of the specialized interfaces Oregon built on top of the software. You can build any kind of a web interface on top of the server if you have selected the right software.

Slide 29   drop-down navigation

Another example of different search interfaces Oregon developed.

While they tried to impose similar design principles for the interface, each collection has a different target audience and calls out for different search interfaces.

Slide 30   challenges - context

Getting the proper context for the collection – building the contextual wrapper – is a major challenge. It is what makes your digital project into a collection, as opposed to a set of pretty pictures or unrelated texts.

Slide 31   context - moorhouse

This is just one page of the contextual information Oregon provided about one collection. Contextual information for this collection was written by members of the UO Libraries or by the partner group with whom they were building this collection.

Slide 32   challenges – multiple collections

Keeping track of where you are on a specific collection is an ongoing challenge, if you find yourself in the position of starting work on another collection before you have finished an earlier one.

Slide 33   uo dig collections home

Project management is a challenge when different collections involve different partners. Oregon was always working on at least three collections simultaneously and each one involved collaborations with different external groups. As I mentioned at the beginning, digital projects have a way of multiplying once you've got the first one started.

Slide 34   challenges - technical

There are many technical challenges, and others will go over some of them in detail in subsequent sessions.

Slide 35   technical issues - list

These are some of the technical issues that you will need to be aware of and plan for how to deal with them.

Slide 36   technical expertise

Hopefully you are all at institutions where you have a Systems Dept or other in-house expertise, can work with a computing center, or can hire technicians to bring up and maintain the system. These are some of the broad technical skills you need to make sure you have in place or can get ready access to. The specifics will vary depending on many factors. This is by no means a comprehensive list but is intended to help you begin thinking about some of these issues for your digital project planning.

- Someone will need a knowledge of operating systems and servers

- You will need to have someone who understands the construction and operation of databases. Most likely, whatever software you choose will have certain reporting and administrative functions built in. Chances are also good that the standard functions will not be enough to meet all of your needs. You will want to extract different kinds of information,

- you will need to work around bugs and make things work smoothly so someone needs to be able to troubleshoot – and do it quickly. With digital collections, you want to be very responsive to any problems and get them resolved quickly.

- You'll want to make sure that your technical people have a good understanding of existing and emerging standards. The strength of digital collections is the discoverability of the metadata and the contents and that means making sure that the objects and the metadata meet basic standards.

Slide 37   hardware and software

There are a number of options available to you, things you will need to consider relating to the hardware and software.

• Do you want to go open source (meaning the software is freely available for download and installation). If it's open source, is there a strong user community that you can turn to for technical advice? Is there a community that is committed to improving the underlying code and sharing their updates and fixes, formally or informally?

• If you purchase or license software, what is the technical support like? Again, is there a user community you'll be able to tap into and get help from.

• Are you going to mount the collections on your own servers or will you contract with a vendor to do it for you? Either way, you need to be sure of having quick turnaround on resolving problems

• Whether you do this yourselves or contract it out, you'll want to know that the equipment is reliable and up to the task and that you won't suddenly run out of space or that the server won't be able to handle the load. Oregon was plagued at times by both scenarios.

• And since digitization is an expensive process if you do it right, you'll want to be sure that the backup mechanisms are robust.

Slide 38   software requirements

As you start your plan, think about the software requirements.

- Support any type of file – not just text-based files – the software should be able to handle a wide variety; You'll need to make policy decisions about whether you should accept all types of files, even if the software permits it
- Does the software allow you to batch load objects and metadata or does each object have to be added and described individually
- Customizable user interface –you want to be able to put your institution's or group's brand on the collection and present it to your users the way that you think is best
- Modular – so that if you do any local customization of the software retaining those customized elements will not require too much effort when you do upgrades
- Flexible system administration – maintain the application quickly and easily. Ability to manage authorizations in groupings, make global changes to a group of users;
- Granular authorizations – different users are allowed to do different things. Ideally this should be flexible enough so that you can determine the combination of features that different users can do.

Slide 39   software requirements

- You'll want to consider if the software can carry out searches based on standard metadata – is the metadata mapped to a standard set of data elements such as Dublin Core?
- Does the software provide any kind of support for controlled terms, either locally created or using external sources? Does it provide any kind of support for authority control? Does it provide for flexible metadata capture, edit, and display – things like the ability to edit metadata easily; ability to set up defaults on a collection-by-collection basis and override those defaults; ability to modify the elements that are searched or displayed on a collection-by-collection basis; ability to change labels for different metadata elements
- Is it possible to do any sort of global changes, either across collections or across the entire system?
- Does it provide useful statistics and reports?

Slide 40   digital preservation

Preservation is the single greatest challenge we face. If you aren't prepared to think seriously about digital preservation and how you will plan for it, you should not be creating digital collections. I'm going to condense what I learned in three different week-long digital preservation workshops into five slides. Your handout lists many more resources for you to inform yourself. The single best thing is the Cornell digital preservation tutorial that is on the Web.

This quotation is from the Digital Preservation Coalition and addresses one of the fundamental differences between digital preservation and preservation of analog materials.

A 2003 OCLC publication: *The Incentives to preserve digital materials* points out that Preservation of analog materials has traditionally been achieved through something akin to a "***crisis management***" strategy.

But an ad hoc approach to digital preservation is almost certain to prove inadequate. You can't wait for digital materials to start to deteriorate. You have to think about preservation from the start by following standards and having a clear, well-documented process.

Slide 41   digital preservation - bitstreams

Digital preservation entails many different processes. In the most simplistic sense it's about ensuring the long-term stability of a bitstream

Slide 42   dig pres – viability, etc.

But that bitstream is important because of the contents. And digital preservation is all about make sure that the contents remain viable (Information must be intact and readable from the storage media), that they can be rendered by a machine, and that they can be understood by a human being. It is one thing to preserve a bitstream, but quite another to preserve the content, form, style, appearance, and functionality

Slide 43   dig pres strategies

There are many strategies being discussed and employed for preserving digital content. This list is taken from **_some_** of the options outlined on the Cornell digital preservation tutorial.

- Bitstream copying and refreshing are the most basic strategies – good, robust backup procedures, refreshing the data periodically and checking for deterioration of the files, and making sure you have durable persistent storage media.
- Sometimes people resort to having analog backups.
- Digital archaeology is very expensive and few have the expertise for it.
- Migration is essentially tracking changes to file formats and migrating files forward to newer versions so that the files continue to be readable.
- Emulation - Combines software and hardware to reproduce in all essential characteristics the performance of another computer of a different design, allowing programs or media designed for particular environment to operate in a different, usually new environment. This is talked about as an ideal but would be very expensive

Slide 44   dig pres components

There are some components that are important for digital preservation that you should make sure that your plan and your digital content management system can accommodate, at a minimum. Time doesn't permit us to go into these but I'd be happy to answer questions afterwards or point you to some specific resources.

Slide 45   challenges - costs

There are a lot of costs in building a digital collection, not all of them obvious.

Slide 46   can you afford it

- How will you cover the personnel and other costs? Will you be able to (or do you want to) absorb the work within an existing unit
- or will you be able to hire new people?
- Will you use volunteers? Can you absorb the training time needed to make use of volunteers?
- Do you have the necessary technical infrastructure in place to support a digital collection or will you need to buy new hardware and software? Even open source solutions, and there are several, need specific hardware and software environments for them to work,
- Licenses and certificates
  - o You may need licenses and certificates for authentication and these carry costs, both for the purchase, as well as for the staff time to set them up
- Registration with other services
- Registering with harvesters and servers takes time and sufficient expertise to make sure your output is OAI-compliant; there may also be fees associated with some registries
- Attending meetings and conferences – stay abreast

Slide 47   challenges - copyright

Copyright and digital rights management is extremely complex in the digital world. An entire pre-conference would be needed to cover this. Copyright is extraordinarily complex because digital collections cross geographic boundaries but copyright laws are still bound by geography. If you get into digitizing materials, you will find that you will have to develop a certain amount of expertise on copyright issues.

Slide 48   challenges – training staff

Providing training to staff is a significant challenge.

At Oregon, we had several lead people. I was the lead on descriptive metadata, one of my staff was the lead on technical metadata, another was the lead on scanning. We had a lot of overlap and jointly worked on web design and training various staff members in one aspect of the work or another.

Slide 49   who will be involved

Who will be involved in your digital project?

Do you have all the skills you need within your organization to set up a digital collection?
• Are there groups with whom you want to partner for technical, political, or financial reasons?
• Do you want to set up a collection that is just for your institution or do you want to implement a collection that can take advantage of the skills and connections that exist in several institutions with similar missions?
• Are you planning to provide this as a completely free service or are you willing to consider some cost-recovery models?
• Are you willing to or interested in partnering with commercial vendors? ProQuest is one organization that is actively searching for good content to digitize and you may be approached by them or another company. There are a lot of opportunities and I recommend being open to them all, at least for discussion.

Slide 50   staffing roles

The staffing issues are going to vary depending on many things: the size of your project, the model you select for hosting the digital collection (in-house or off-site with a vendor), and the range of services you plan to offer with the collection. These are some of the roles you should think about who will fill:

• Coordinator with overall responsibility

• Web site designers – any web-based service needs to be designed, to have the look and feel that identifies it with your institution

• Group to make policy decisions

• Staff to handle or review submissions

    o This can also include cleaning up or supplying metadata

• Staff to set up and maintain the system and resolve technical issues

    o As a high-profile new service, this needs to get top priority

• Group or individuals to make contacts and market the collections

• Staff for subsidiary services

Slide 51   how to plan

- Planning for a digital project will never be perfect. Even after you've done it a few times, there will be surprises along the way. But don't let that stop you.
- As we worked on our digital collections, we often talked about developing a business plan. We took a stab at it several times but it was always vague.
- In general, I would say that this is much easier said than done. As a profession, we aren't yet very comfortable with business models. We tend to be service organizations and we do what needs to be done to provide a service. We have some experience with buying services or products but not much experience with starting up new services from scratch. Business plans go against the grain for us in many ways.
- A project plan is a little easier.
- Whatever you decide, document it and try not to make decisions that you can't reverse or recover from because you will probably not do everything perfectly from the beginning.

Slide 52　business plan elements

In a publication for the Council on Library and Information resources in 2004, Liz Bishoff and Nancy Allen outline the elements of a business plan for cultural heritage institutions to develop a sustainable digital asset management program. It's a short publication of 55 pages and might be worth reviewing to help you think about the issues we've covered so far. Some of the elements of a business plan mentioned:

- Mission, Vision, Values, Goals
- Executive summary
- Product or service description - Pretend you have to sell this idea – what is it that a digital collection would provide?
- I'm in the midst of preparing this for the U of R campus now – the next step is for the proposal to be reviewed by a campuswide Information Technology Steering Committee
- Needs assessment or market research
- Environment and competition
  - What is the local environment like? Who or what are your competitors? Markets and services
- Identify your target audience and the services you will provide for them
- Organizational structure has a large impact on long-term sustainability

Slide 53   business plan elements – financial, etc.

And finally these last two elements

- Financial plans
    - Bishoff and Allen recommend having a separate financial plan, and not just incorporating these kinds of initiatives into their existing operating budget
    - They say that a financial plan should cover 3 to 5 years and should have expense and revenue components.
- Product evaluation and usability assessment
    - Evaluation should be done on a regular basis

Slide 54   project plan issues

These are key issues that I believe your project plan should address. We've covered some of these and others will cover other aspects.

Slide 55   project implementation

The University of Oregon had done some planning before it started. It had a group working for almost two years talking about the issues, evaluating software options, planning for the storage of digital objects, discussing roles that different people would play. Oregon's implementation involved a Metadata Implementation Group which began work in November 2002. Originally, this group was charged to look at metadata that would be needed for a particular collection of photographs and we included people from the Slide library, special collections, documents, and cataloguing.

We were building the henhouse carefully and systematically.

Slide 56   project implementation - cayuse

Then the first egg landed in our lap – before we'd even finished the henhouse.

The Metadata Implementation Group became aware of a grant-funded project in February 2003, two months before the grant was due to be completed – and before any work had been done. The group and the former Catalog dept (now Metadata and Digital Library Services) swung into action to rescue the grant. This became the first digital collection we built – and it was far more complex and challenging than the one we had been planning for.

This is what I mean about the chicken and the egg. The collection came into being because it had to, not because there was any systematic plan. We tried to stay one step ahead of the staff we were training to scan and describe the collection.

Slide 57   project management

As we went along, we discovered that managing the project and all the different stages of it is complex. Some tasks take less time than others and it's challenging to avoid bottlenecks in your workflow and to keep from having to touch the same digital object over and over again. At Oregon, we tried a number of different techniques to help us manage the projects, all with moderate success at different times.

- We used web pages to track our decisions and provide work process documentation.
- We used archived discussion lists to raise and resolve issues and document decisions.
- We used project management software.
- We used electronic folders and spreadsheets
- We had overlapping responsibilities to try to keep things moving

Slide 58   project evaluation

As we went along, we tried to develop ways to evaluate our projects and improve them.

Use is one measure

Usability of the site is another way of assessing its utility

Customer satisfaction surveys are another way

We tried all of these approaches

Slide 59   evaluative form

This is an example of an ongoing attempt at collection evaluation – building in comment forms that fed into an archived list that different project managers monitored.

Slide 60   when will you be done

Slide 61 - 68

I'm going to end by showing you some of the images from the Picturing the Cayuse collection. These are like eye candy.

Slide 69   contact information