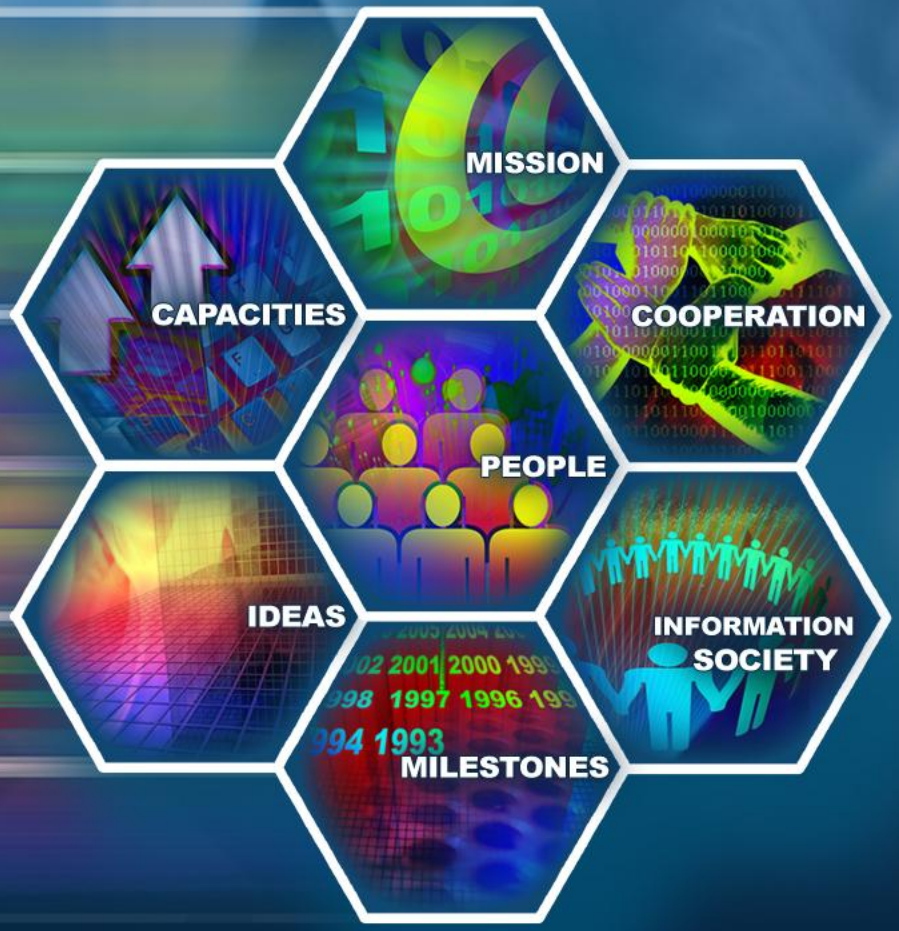
POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER



POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

ACCESS IT Plus – Training workshop in Veria (March ’12)
Session 1: Building digital collections

Speaker: Marcin Werla (mwerla@man.poznan.pl)

Education and Culture DG
Culture Programme

How to create good collections?

- Base material for this session is „A framework of guidance for building good digital collections” prepared by NISO
  - http://framework.niso.org/
  - http://www.niso.org/publications/rp/framework3.pdf

  It is cited in several places on these slides.
- The framework is organised around four core types of entities:
  - collections - organized groups of objects
  - objects - digital materials
  - metadata - information about objects and collections
  - Initiatives - programs or projects to create and manage collections

How to create good collections?

- Creating a good, useful collection is directly associated with choosing objects for digitisation
- Consistent selection of digital objects facilitates their discovery and use
  - A good digital collection should be useful and relevant in the context of the needs of users
  - The meaning of a good collection changes over time

How to create good collections?

- Other factors considered among critical features of a good collection include
  - cost/value issues
  - interoperability
  - sustainability
  - trust
- Collections and their objects should no longer be seen just in the context of a project, but also as building blocks that others can reuse

How to create good collections?

- Goodness now demands:
  - interoperability,
  - reusability,
  - persistence,
  - verification,
  - documentation,
  - and support for intellectual property rights.

How to create good collections?

- Good digital collections are becoming an active collaboration between the information professionals and the user
- A digital collection can contain not only results of digitisation but also born digital materials
  - This is particularly important in the context of user-generated content related to the theme of the collection

Short exercise

- On the next slides we will discuss 9 collection principles
- Please rate each of these 9 principles in scale

  + 2 = It is very important and not so obvious, deserves to be “principle”

  +1 = I find it useful and it is worth to remember about it

  0 = Neutral

  - 1 = It is quite obvious for me

  - 2 = It is really obvious and does not deserve to be “principle”

  Ratings will be summarized after all principles will be presented – keep them in secret for now to avoid influencing each other

- The aim
  - is not to test how professional you are ☺
  - but to see if such sets of principles makes sense

Collections Principle 1

- A good digital collection is created according to an explicit collection development policy that has been agreed upon and documented before building the collection begins
  - The collection should support the organization's mission
  - Collection builders should identify the target audience but also think about unexpected uses and users
  - In some cases strict selection policy may not be required:
    - When?
      - E.g. digitisation on demand, mass digitisation
  - It is worth publishing the collection development policy in order to inform users what they can expect in the digital library

Collections Principle 2

- Collections should be described so that a user can discover characteristics of the collection
  - What can be included in the description of a collection?
    - its scope, format, restrictions on access, ownership, and any information significant for determining the collection's authenticity, integrity, and interpretation
- Publication of detailed information about the collection might be helpful also for others during their digitisation efforts
- A description can be created using a given metadata standard (e.g. Z39.91 Collection Description Standard, Dublin Core Collections Application Profile) or narrative text

Collections Principle 3

- A good collection is curated – its resources are actively managed during their entire lifecycle
- Curation may include:
  - creating, correcting and enhancing metadata,
  - correcting or enhancing the data itself,
  - adding annotations, linkages to other materials, or other enriching information.

Collections Principle 4

- A good collection is broadly available and avoids unnecessary impediments to use
- This principle is associated with three attributes: availability, usability and accessibility
  - Availability means that the collection should be accessible and usable for all authorized people. In particular, the collection should be available in the Web.
    - Availability does not imply that the use of all materials should be free and unrestricted.
  - Usability means that the collection should be easy to browse/search and use for most users.
  - Accessibility means that the collection interface should be available also for people with disabilities, i.e. visual impairments, loss of hearing.

Collections Principle 5

- A good collection respects intellectual property rights.

Collections Principle 6

- A good collection has mechanisms to supply usage data and other data that allows standardized measures of usefulness to be recorded.
- The digital collection should be evaluated periodically to monitor its usage, assess service effectiveness, demonstrate return on investment, inform collection development, inform strategic planning and support funding request.
- Evaluation can be performed using various methods including: surveys, interviews, observation, case study and log analyses.

Collections Principle 7

- A good collection is interoperable.
  - Collection developers should design their services to support interoperability, i.e. ability to share their metadata with external search engines like Google or OAI-PMH compatible services

Collections Principle 8

- A good collection integrates into the users own workflow.
  - Collection building (digital library building and development) should be included into existing staff workflows.
  - End users find information most useful when it integrates smoothly with their own patterns of work. A faculty member looking for research articles and a recreational genealogist building an electronic family tree will work in different places, at different times, and using different tools. However each will use a digital collection more comfortably if they can access it from the environment with which they are familiar.

Collections Principle 9

- A good collection is sustainable over time
- Sustainability of a digital collection is not only a problem of software and hardware. It is also important to secure organizational, financial and technical perspective.

Examples of collections

- Public Digital Archive of Agnieszka Osiecka
- Armenian Rare Books
- DART-Europe E-thesis Portal
- Darwin 200
- Cartoon Drawings: Herblock Collection
- Manuscriptorium

What do you think about it?

- What are the results of our exercise?
- Would you add some additional principles?

Summary of objects selection criteria

- The material may be chosen in order to:
  - Meet the criteria of an external funding body
  - Enable cooperation with another institution
  - Coincide with a particular anniversary
  - Increase accessibility of this material, which would be otherwise unavailable or of limited availability
  - Ensure the preservation of this material
- Beside of the above, material may NOT be chosen because of:
  - Availability of existing digital versions
  - Cost of digitisation
  - Copyright and IPR
  - …

Overview of IPR

- Disclaimer:
  - The following slides are not a legal advice ☺
  - Please consult it with someone well knowing the legal system of your country before doing some serious things

Overview of IPR

- Intellectual property rights (IPR) must be considered from several points of view
  - what rights the owners of the original source materials retain in their materials;
  - what rights or permissions the collection developers have to digitize content and/or make it available;
  - what rights or permissions the users of the digital collection are given, to make subsequent use of the materials.

Overview of IPR

- Infringement of copyright occurs when a person carries out an activity that is restricted by the copyright legislation without authorization from the copyright owner
- These restricted acts include:
  - Copying material
  - Issuing or making available copies to the public
  - Showing, playing or broadcasting or filming
  - Adapting or amending material

Overview of IPR

- In general, copying of protected material is prohibited but under certain circumstances it is possible to reproduce objects: such exceptions are called fair use.
- The main idea behind the fair use is that a given activity does not reduce economic benefits of the copyright holder.
- Examples:
  - Copying for non-commercial research or private study e.g. downloading images from the Internet and storing them on the hard disk of your computer - this is OK, as long as it is only personal use, and you must not put this image on your website
  - For criticism or review e.g. an image with a music album cover presented in a newspaper next to its review.
  - For the purpose of examination in an authorised educational establishment e.g. a lecturer can download a digital work and distribute it among class for the examination purpose.
- There is also an exception called "Library Privileges" which allows librarians and archivists to make a copy on behalf of their customers. This is fine as long as the copy is used for non-commercial research or private study

Overview of IPR

- Copyright protection is not endless, and its duration for any work can be calculated and depends from the law of country where given work of art was created.
  - For literary and artistic works if the creator is known, in most of the countries copyright lasts for 70 years from the death of the creator of the work.
  - In terms of published anonymous works, in most of the countries copyright lasts for 70 years since the publication date.

Overview of IPR

- Objects having no copyright become a part of “Public Domain“- can be used without IPR limitations.
  - The term "Public Domain" is present only in some legal systems (e.g. not in Poland)
  - The reason for Public Domain existence is that such an open domain is necessary for unconstrained development of culture and science
- If the owners of the copyright are unknown or untraceable, “orphan works” come to existence.
  - In such case the institution which is willing to digitize a given resource will take a risk of copyright violation

Overview of IPR

- Calculating the copyright risk (e.g. for putting orphan works on the Internet)

The financial risk = A x B x C x D

A – the chances that what has been done is infringement

B – the chances that the copyright owner becomes aware of such infringement

C – the chances that having become aware, the owner sues

- any ideas how to minimize it?

- “takedown policy” is the absolute minimum

D – the financial cost for such legal action

- This is just to provide new way of looking at copyright issues, although such approach is unthinkable in some countries

Source: http://www.slideshare.net/lisbk/innovation-and-the-social-web-learning-from-commercial-approaches

Overview of IPR

- Digitisation of public domain object is not creating new IPR
  - Any ideas for exceptions here?
    - As long as it is “non artistic” digitisation
- Of course it is not obligatory to put highest quality images on-line, but what is put on-line and comes from public domain can be freely used
  - Even if some security mechanisms are provided

Overview of IPR

- Useful web systems (for use at your own responsibility):
  - Choosing the right license for your works:
    - http://creativecommons.org/choose/
    - Jurisdiction is very important here!
  - Determining if object is still under copyright
    - http://outofcopyright.eu/calculator.html

Overview of IPR

- Your institution should maintain the registry of data about the rights that your institution hold and acquire.
- Such a registry should include at least the following information:
  - The identification of the resource itself
  - The name of the entity granting the rights
  - The precise rights that are being granted and any specific exclusions
  - The period of the time for which rights are granted
  - The groups of users permitted to use the resource
  - Any obligations of the users of the resource (including financial obligations)
  - All licensing agreements must be monitored and re-negotiated, if required
- All legal issues (including IPR) should be carefully analyzed **in the context of legal regulations of a particular country in which the digitisation project takes place**

POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER

End of session 1:
Building digital collections

Poznań Supercomputing and Networking Center
affiliated to the Institute of Bioorganic Chemistry of the Polish Academy of Sciences,
ul. Noskowskiego 12/14, 61-704 Poznań, POLAND,
Office: phone center: (+48 61) 858-20-00,
fax: (+48 61) 852-59-54,
e-mail: office@man.poznan.pl, http://www.man.poznan.pl