Access IT Training

Part 1.
What is all this digitisation about and which objects should be digitised?

Digitisation

- … is the conversion of analogue materials into a digital format
  - It is a **complex** process
  - It is an **expensive** process
  - Involves **risk** in many areas
  - So… why to digitise?

Digitisation

- So… why to digitise?
  - To provide better access to unknown or little used collections
  - To offer better search and retrieval facilities for an image collection
  - To provide a better understanding of original works through improved indexing or some form of digital image enhancement
  - To create resources that are suitable for use in learning and teaching
  - To enhancing the public knowledge, recognition and understanding of the collection
  - To preserve collections for the future in digital format
  - …

Digitisation as a project

- Digitisation of a collection is a project
  - Each project is UNIQUE!
    - More or less, but always
  - The (uniqueness of) the project is defined by
    - The scope
    - The resources
    - The time (which is a specific resource by the way)

Digitisation as a project

- Planning is crucial for each project
- Initially the following questions should be answered
  - **What** work needs to be done?
  - **Who** should do it?
  - **Where** should it take place?
  - **When** will it take place?
  - … and finally…

Digitisation as a project

- **How** will it be done?
  - Decisions made at the time of digitsation have fundamental impact on the
    - Manageability
    - Accessibility
    - Viability

    of the created digital resources

Basic steps of the digitisation project

1. The selection of materials for digitisation
2. The physical preparation of materials for digitisation
3. The digitisation process

Basic steps of the digitisation project

1. **The selection of materials for digitisation**
2. The physical preparation of materials for digitisation
3. The digitisation process

The selection of materials for digitisation

- „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"

NISO collections principle 1

http://framework.niso.org/node/9

The selection of materials for digitisation (high level)

- Material may be chosen in order to:
  - Meet the criteria of an external funding body
    - E.g. CIP ICT PSP (http://ec.europa.eu/information_society/activities/ict_psp)
      - Theme 2: Digital Libraries
      - Objective 2.3: Digitising content for Europeana

      „The aim is to support targeted digitisation actions in synergy with and complementing collections already accessible through Europeana, in particular digitisation of masterpieces of Europe's cultural heritage. The material to be digitised should be selected through a thematic approach. The themes must be of interest to a broad public and not just to a limited set of specialists. The resulting complementary content should be made available for citizens through Europeana."

The selection of materials for digitisation (high level)

- Material may be chosen in order to:
  - Enable cooperation with another institution
    - E.g. Manuscriptorium project (http://www.manuscriptorium.com)

    The Manuscriptorium project creates a virtual research environment providing access to all existing digital documents in the sphere of historic book resources (manuscripts, incunabula, early printed books, maps, charters and other types of documents). These historical resources, otherwise scattered in various digital libraries around the world, are now available under a single digital library interface. The service provides seamless access to more than 5 millions of digital images.

    The content providers can use various services and tools developed for easy aggregation and integration of heterogeneous digital content.

The selection of materials for digitisation (high level)

- Material may be chosen in order to:
  - Coincide with particular anniversary
    - E.g. DARWIN200 (http://www.darwin200.org/)

    Darwin200 is a national programme of events during 2009 celebrating Charles Darwin, one of the world's most creative and influential thinkers.

    Many events took place on and around 12 February 2009, which was 200 years since Darwin was born. We are now approaching another important anniversary on 24 November – 150 years since the publication of his famous book, On the Origin of Species.

    http://www.darwin-online.org.uk/

    http://ebuw.uw.edu.pl/dlibra/news?news=full#4fc95c20-b58a-401b-bb18-9935cbe25bd9

The selection of materials for digitisation (practical level)

- Material may be chosen in order to:
  - Increase the accessibility of this material, which would be otherwise unavailable or of limited availability
    - E.g. Material to valuable to be available publicly
    - E.g. Lack of exhibition space for material
    - E.g. Material very popular, requiring wider access

The selection of materials for digitisation (practical level)

- Material may be chosen in order to:
  - Ensure the preservation of this material
    - E.g. Preservation of delicate originals by making digital versions available as alternative
    - E.g. Condition and amount of originals requires further preservation in digital form

The selection of materials for digitisation (practical level)

- Material may be NOT chosen because of:
  - Copyright and IPR
  - Availability of existing digital versions
  - Cost of digitisation…

The selection of materials for digitisation

- Mass digitization projects may involve little selection below the collection level
  - E.g. Digitisation of institution collection of 19th century newspapers
    - This is also good example for selection based on criteria related to digital preservation and condition of originals

The selection of materials for digitisation

- Sometimes selection process may be moved to end users
  - E.g. Digitisation on demand
  - E.g. Community projects like "Great War (Poetry) Archive" http://www.oucs.ox.ac.uk/ww1lit/gwa

    The Great War Archive contains over 6,500 items contributed by the general public between March and June 2008. Every item originates from, or relates to, someone's experience of the First World War, either abroad or at home. Contributions were received via a special website and also through a series of open days at libraries and museums throughout the country.

    The deadline for contributions has now passed. However, you can still share any images that you have by posting them to The Great War Archive Flickr Group. To post to the group you will need to set up a free Flickr account.

The selection of materials for digitisation

- So – however selected – a good digital collection consists of digital objects that have been developed according to a collection development policy
  - Step 1: Establish selection criteria
  - Step 2: Evaluate each candidate against established criteria
    - … But if some important or critical object will be rejected, this may be the reason to review selection criteria
      - Far more important than the project plan document, is the project planning process itself

Basic steps

1. The selection of materials for digitisation
2. **The physical preparation of materials for digitisation**
3. The digitisation process

Cataloguing originals

- Every physical object should be catalogued before being digitised
  - If it was not done earlier, it should be done during the digitisation project

Cataloguing originals

- Cataloguing is important for:
  - Knowledge about and interpretation of the object to be digitised (also for preservation purposes)
  - Contextualisation of the object
    - The catalogue links the object with the collection or family of objects it belongs to
  - Finding and understanding of the original object and of the digital resource representing it

Movement and handling of original materials

- The original materials may need to be cleaned or conserved before digitisation takes place
  - The time and cost of any such works should be taken into consideration in the project plan

Movement and handling of original materials

- Different aspects should be considered when establishing digital capture workflow
  - The originals may require different approaches to digitisation (e.g. flat vs 3D)
  - The originals may require special handling (like very fragile, very large or very precious objects)
  - This is a subject for internal digitisation guidelines (discussed later)

Basic steps

1. The selection of materials for digitisation
2. The physical preparation of materials for digitisation
3. **The digitisation process**

The digitisation process

- The digitisation may be carried out in-house by the institution or it may be outsourced
- Some of factors involved in this choice
  - The volume of similar original material to be digitised
    - Small volume – DIY
    - Large volume – outsource

The digitisation process

- The digitisation may be carried out in-house by the institution or it may be outsourced
- Some of factors involved in this choice
  - The fragility of the material and the risk of moving it outside the institution
    - Moving it far (using cheap labour for mass digitisation)
    - Moving it not so far (specialized digitisation labs)
    - Not moving it at all (but still outsourcing)
    - Not moving it at all (and not outsourcing)

The digitisation process

- The digitisation may be carried out in-house by the institution or it may be outsourced
- Some of factors involved in this choice
  - The complexity of the digitisation process
  - The availability of trained staff
  - The availability of hardware and software
  - … And last but not least **the continuity of funding**
- More details about the digitisation process will be presented in next parts of the tutorial

The digitisation process

- The outcomes of the digitisation process are stored in digital repositories and described with metadata for preservation and further usage
- The repositories may be created separately for each digitisation project or may be prepared during the first project and used in next projects
- More details about digital repositories and metadata will be discussed later on

Digital-born material

- Repositories with digitised objects may also contain cultural heritage which is born-digital
  - Newly written books or articles
    - Includes scientific materials uploaded by authors ("self-archiving")
  - Various present events recorded by digital means
    - E.g. Video recordings from theatre performances
    - E.g. Recordings of oral history – may combine video and audio recordings with photographs, notes and digitized personal documents

Digital-born material

- The fact, that the material is born-digital does not mean that it does not require processing prior to depositing in a digital repository
- High quality born-digital material may consume large amounts of storage space

Intellectual Property Rights (IPR)

- All digitisation projects must respect IPR held in the materials they work with
  - The rights of the owners of the source materials that are digitised
  - The rights of the owners of the digital resources
  - The rights or permissions granted to a service provider to make digital resources available
  - The rights or permissions granted to the users of the digital resources

Intellectual Property Rights (IPR)

- There may be also some rights arising from the particular terms and conditions of a digitisation programme within which a particular project operate
  - Projects funded in some of the EC programmes are required to give free access (open access) to all of its outcomes

Intellectual Property Rights (IPR)

- Copyright generally lasts for 70 years from the death of the creator of the work
- Before digitisation a written permission from the copyright owner must be obtained
- Works having no copyright become part of "public domain"
  - Works which can be used without IPR limitations
  - It is not a official legal term, it is more like an idea assuming that the existence of such public domain is necessary for unconstrained development of culture and science
  - Not in every legal system the public domain term is offcialy defined (e.g. in Poland it is not)

Intellectual Property Rights (IPR)

- If the owners of the copyright are unknown or untraceable the **"orphan works"** come to existence
  - In such case a standardised, well documented and recorded process of finding the owners should be established
  - If owners will not be found during this process, the institution may take a risk and digitise the orphan works
  - Such approach offers no protection under the law, but may be really important mitigating circumstance in case of legal problems
  - This may be the only solution in case of mass digitisation of such works

Intellectual Property Rights (IPR)

- In case of objects being results of works undertaken by an institution's staff as a part of their normal duties, the right remains the property of that institution
  - Sometimes assigned to external publishers (in scientific institutions)
  - Unpaid volunteers retain the copyright unless they sign away their rights

Intellectual Property Rights (IPR)

- All rights existing in the materials should be identified and recorded in order to be managed
  - 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

Intellectual Property Rights (IPR)

- If someone is wondering on which license his/her works should be published on-line, the Creative Commons initiative can be useful
  - It defines a set of re-usable licenses in order to enable rights-owners to allow re-use of material under conditions chosen by the rights-owner
  - The CC licenses typically allow re-use in educational or non-profit contexts
  - Each CC license type is clearly explained and for each license a long legal document is prepared expressing the idea behind the license
  - The legal documents are translated and adjusted to suit many legal systems world-wide
  - As the licenses are developed, new versions are published

Intellectual Property Rights (IPR)

- If someone is wondering on which license his/her works should be published on-line, the Creative Commons initiative can be useful
  - Allow commercial use?
    - Yes
    - No
  - Allow modifications?
    - Yes
    - Yes, but share alike
    - No
  - http://creativecommons.org/choose/

Intellectual Property Rights (IPR)

- Identifying rights and negotiating licenses is not enough
- Very often projects must (try to) ensure that their rights and the rights of other parties are protected
  - It can be done by taking steps to block the unauthorised use of materials
  - Technical aspects are often connected with DRM (Digital Rights Management) technologies
    - Watermarking (visible/invisible)
    - Fingerprinting
    - Limitations in access (IP addresses, low resolution)
    - Limitations in use (e.g. Blocked printing)
    - Limitations in copying/distribution
    - …
  - A balance between objects' usability, protection of rights and technical possibilities is often very problematic

Intellectual Property Rights (IPR)

ALL LEGAL ISSUES (INCLUDING IPR)
SHOULD BE CAREFULLY ANALYSED
IN THE CONTEXT OF **LEGAL REGULATIONS OF A PARTICULAR COUNTRY**
IN WHICH
THE DIGITISTATION PROJECT TAKES PLACE!

Coordination of digitisation

- There may be a need to coordinate digitisation efforts between several institutions or projects
  - Especially in libraries – e.g. many copies of the same edition of a book
    - Of course in some specific circumstances it may be worth to digitise several copies of the same edition

Coordination of digitisation

- Possible solutions
  - Central database on a country or even on a continent level
    - Strictly connected with funding of digitisation
    - May be also connected with the storage of high quality digital objects
  - May be introduced by funding body like Ministry of Culture
  - If connected with significant funding, may be quickly adopted

Coordination of digitisation

- Possible solutions
  - Automated synchronization of data between distributed digital repositories
    - Works in Poland as an effect of lack of central solution
    - Not connected with funding
      - Based on good will of cooperating institutions
    - Improves cooperation between institutions
    - Requires good interoperability on digital library software level
    - Examples:
      - http://fbc.pionier.net.pl/owoc/planned-report
      - http://fbc.pionier.net.pl/owoc/duplicates-list-report

Q&A