# Practical exercises/discussion/closing remarks

**ACCESS IT – Training, Veria, Greece, 16-18.02.2010**
Marcin Werla, Adam Dudczak

1. **What is all this digitisation about and which objects should be digitised?**
    a. Think about what will be digitised
        i. Check the IPR issues related with the digitised material
    b. Think about who, when and where will do it
    c. Think about the costs of
        i. Creation of basic infrastructure (like DL software and hardware, scanners, computers for digital librarians etc.)
        ii. Digitisation
            1. To outsource or not to outsource
        iii. Long term activity of your digital library
            1. Costs related with the server being on-line
            2. Technical support
            3. Software and hardware updates
    d. Think about how to do it (see point 2)
2. **How to digitize objects?**
    a. Remember about things mentioned in point 8
    b. Check if national/institutional digitization strategy exists
        i. If such a document is available, consider how it can reused in your project (requirements and guidelines).
    c. Check available equipment and its capabilities
        i. IT infrastructure – storage space, intranet capacity
        ii. Scanners/cameras
    d. Think what is important for future users?
        i. Full-text search? (in case of text resources)
            1. Master copy
                a. PDF/A or TIFF (LZW compression) for digital master copy
                b. Results of OCR can be stored as XML next to TIFF files

2. Web delivery
    a. Remember that:
        i. DJVU is not suitable for storing master copy
        ii. DJVU is not indexed by Google
    b. PDF should have reasonable size
        i. i.e. consider dividing a book into chapters
3. Choosing OCR engine
    a. Take into account availability of qualified IT stuff
        i. Open Source tools like Tesseract/OCRpus may require some IT knowledge to be used
    b. Usually commercial OCR engines provides a trial/demo version, it is reasonable to check how given OCR performs with digitized resource before buying a license.

ii. No text – only images
1. Choose the simplest solution:
    a. TIFF – (LZW compression) for a master copy
    b. JPEG – for web delivery

e. Store master copies on DVDs (and on a file server)

f. Write down a set of necessary guidelines, requirements and procedures. Discuss this with the rest of project staff, make them available for everyone who is involved.

i. Quality requirements (resolution, colour depth, file format etc.)
1. digital master copies
2. web delivery version of object
3. other derivatives

ii. Most important procedures
1. How to deal with scanned objects.
2. Which equipment should be used for scanning particular items.

g. When project staff is distributed consider a few additional solutions

i. Consider creation of a file server
    1. Everyone can connect to server
    2. This would simplify file sharing
ii. Consider coordination using shared internet document
    **1.** e.g. use http://docs.google.com
**iii.** Evaluate established procedure and guidelines after some time in terms of digitisation performance and end-user experience
    **1.** Try to find things which might be improved,
    **2.** Refine your project guidelines.

**3. How to describe digital objects?**

a. Choose your metadata schema
    i. Start with DC Simple (15 elements)
    ii. Check DC Terms
    iii. Check if there is anything else you want to add
        1. If yes, prepare mapping of these new elements to DC elements
    iv. Prepare the manual for the creation of metadata
        1. Try to use existing guidelines
        2. Try to be compatible with standards
        3. Try to find some useful dictionaries to use
        4. If possible (and makes sense) think about importing the existing metadata
b. Prepare the structure of on-line collections
c. Think how your digital objects will be used
    i. How do you want to publish it in the Internet
        1. File format, user interface possibilities
    ii. Prepare the license information for end-users

**4. How to create a digital repository?**

a. Choose software
b. Install software, play with it, do some tests on your resources and metadata
    i. You can use Live CDs, pendrive editions etc.
c. Choose hardware for DL server
d. Install the software on the DL server

e. Configure it and adapt for your needs and assumptions
    i. Remember about the usability and accessibility (see point 9)
f. Start publishing your digital objects

**5. How objects and metadata from digital repositories can be reused?**

a. Check OAI-PMH interface correctness using OAI-PMH repository explorer
    i. http://re.cs.uct.ac.za/
b. Register repository to OAI archives website
    i. http://www.openarchives.org/Register/ValidateSite
c. Make content of your repository available for suitable aggregator
    i. i.e. OAIster, ScientificCommons, DART-Europe etc.
d. Consider creation of digital library blog
    i. Described selected items in more details - show users what can be found in your repository

**6. How to (and why) prepare a repository for Europeana?**

a. Find a suitable national or vertical aggregator who may help you to join Europeana
b. Check aggregator's requirements in terms of metadata
    i. You may have to prepare mapping rules from your metadata standard to ESE or other format used by aggregator (i.e. Athena – museums will use LIDO)
        1. Remember about Europeana Metadata Mapping and Normalization Guide
    ii. In some cases aggregator may prepare metadata for Europeana without your help.
c. Join EuropeanaLocal technical support forum
    i. http://europeanalocal.avinet.no/

**7. How to set up a metadata aggregator? – introduction + practical exercises**

a. Before you start check existing aggregators
    i. If there is no national aggregator for you, consider joining vertical/thematic aggregator like Athena, Europeana Film Gateway, APEnet, more at http://group.europeana.eu/

b. Setting up metadata aggregator may require some technical skills
c. Choose tools which fits into your infrastructure (e.g. consider underlying technology)
d. Join EuropeanaLocal technical support forum
    i. http://europeanalocal.avinet.no/
e. Define rules for repositories which may want to join your aggregator
    i. OAI-PMH compliance
    ii. Only mature repositories will be aggregated
        1. existing domain address, no ports like 8080

**8. Should I know something more about digital preservation?**

a. Use simple, well established open standards
b. Track findings of projects like Digital Preservation Europe
c. Check out DRAMBORA repository audit framework
    i. http://www.repositoryaudit.eu/

**9. Why to evaluate accessibility and usability of digital libraries?**

a. Compare interfaces of other digital libraries
b. Make yours the best one
c. Test it (with hallway testing for example)
d. Monitor its use (Google Analytics)