Good Practices in Cost Reduction for Digitisation

Resources for Minerva and Minerva Plus WG on Good Practices

Dov Winer

dovw@jazo.org.il

1. JEDS Resources

Costing a Digitisation Project

http://heds.herts.ac.uk/resources/costing.html

It is possible to define the factors that will determine the price of conversion. There are three threads: (1) What you want from the information content of the originals (2) Balance between costs, technology and benefit goals (3) Most of all – depends on the nature of the original material itself.

Digitisation: How Much Does it Really Cost (Simon Tanner and Joanne Lomax Smith, HEDS) paper for the Digital Resources in the Humanities Conference 1999

http://heds.herts.ac.uk/resources/papers/drh99.pdf

The issues tackled are: (1) What are the major cost factors? (2) How the choice of original and technical specification will affect costs? (3) Whether to outsource or use in-house resources (4) Examples of comparative costs between the two approaches.

HEDS Costing Matrix

http://heds.herts.ac.uk/resources/matrix.html

Key: (1) Typical specification (2) Preparation time (3) Handling (4) Automated processing (5) Skill/Experience rating (6) Optimisation costs (7) Resource costs (8) Quality assurance (90) Filesizes.

2. The costs of Digital Imaging Projects

(Steven Puglia, National Archives and Records Administration) from RLG DigiNews, October 15, 1999, Volume 3, Number 5.

---

http://www.rlg.org/preserv/diginews/diginews3-5.html#feature

This document discuss in detail the components for planning and budgeting a digital imaging project: (1) Selection (2) Preparation (3) Metadata creation (4) Preservation/conservation of the physical object (5) Production of intermediates (6) Digitization (7) Quality control of images and metadata (8) Technical infrastructure (9) On-going maintenance of images and metadata.

3. Economic Factors of Digital Libraries

An special issue of the Journal of Digital Information (http://jodi.ecs.soton.ac.uk), edited by Simon Tanner, is dedicated to Economic Factors of Digital Libraries.

S. Tanner (June 2003)
Editorial: Economic Factors of Managing Digital Content and Establishing Digital Libraries

With its economics theme, this issue is about the choices made by individuals, institutions and communities with regard to alternative uses of scarce resources to satisfy the desire for digital content and digital libraries. The effective utilization of resources is among the most important of management activities and in the context of digital libraries has several components:

- immediate start-up costs of either creating or purchasing digital content;
- further implementation costs for establishing a digital library or even just basic access to bought resources;
- costs implicit in preserving, managing and maintaining a digital resource in the longer term.

S. Chapman (May 2003)
Counting the Costs of Digital Preservation: Is Repository Storage Affordable?

Formats are significant, but not sole factors in determining preservation costs in these models. Owners' definitions of content integrity and tolerance for risk, which can change over time, are also important variables in the complex equation of preservation costs and

affordability.

4. National Initiative for a Networked Cultural Heritage

http://www.nyu.edu/its/humanities/ninchguide/index.html

The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials

By adopting community shared good practice, project designers can ensure the broadest use of their materials, today and in the future, by audiences they may not even have imagined and by future applications that will dynamically recombine 'digital objects' into new resources. Such projects can be produced economically and can be maintained and managed into the future with maximum benefit for all.

Project Planning

http://www.nyu.edu/its/humanities/ninchguide/II/

Some guidance on cost estimation is offered later in this section, and also in the sections on specific digitization areas (Sections V, VI, and VII). You should make sure in researching costs to take into account all of the startup and infrastructural costs the project will incur-costs for initial planning, choosing data specifications, building or choosing tracking and documentation systems, training staff, and so forth-as well as the incremental cost of digitizing the materials themselves.

5. Budgeting a Digital Project

Conceiving and Planning Digital Projects (Howard Batchelor, UCLA)

http://digital.library.ucla.edu/about/planning/planning_menu.html

http://digital.library.ucla.edu/about/planning/budget.html

This section offers a general model for projecting the direct costs of a digization project by dividing tasks into five major categories:

(1) Selecting & Preparing Materials for Digitization

(2) Imaging Requirements

(3) Digitization Costs

(4) Metadata Requirements

(5) Post-Digitization Processing and other Local Costs

Examples are given within each category of the kind of work that may be involved, depending on the nature of the source material and the context of the project. Because no two projects are alike, there can be no prescriptions that meet all circumstances.

6. Library Preservation at Harvard: Planning Digitization Projects Bibliography

http://preserve.harvard.edu/bibliographies/digitalplanning.html

http://preserve.harvard.edu/bibliographies/textconversion.html

http://preserve.harvard.edu/resources/imagingsystems.html

A Decision Making Matrix for selection for digitizing is provided at:

http://preserve.harvard.edu/bibliographies/matrix.pdf

7. Cost Comparisons for Digitization Projects

http://clrc.org/lstadigital/CostComparDigitizProjRev.pdf

The CLRC (Central New York Library Resources Council) funded digitization projects from 1997-2002. This document summarizes what has been learned to date concerning costs; most of the projects were pilots or demonstrations and only a few are "production" projects.

8. A question of cost: choices on the road to digitization – Simon Tanner (KDCS,KCL)

http://www.kcl.ac.uk/humanities/ cch/kdcs/pubs/conf2003/tanner.pdf

---

Considerations of creating a surrogate that need resolution are; transport, conservation preparation, indexing, handling, potential damage, insurance and risk management. Cost scenarios of possible routes, such as microfilm versus photography, must be explored and compared before any aspect of preservation can begin. A clear picture of the preservation intention and the costs that will be incurred by either method of surrogate creation is essential

9. **The Cost of Digital Image Distribution: The Social and Economic Implications of the Production, Distribution and Usage of Image Data – Chapter 2: The Cost of Creating Digital Images and Metadata by Museums**

http://sunsite.berkeley.edu/Imaging/Databases/1998mellon

Although this report is dated from July, 1998 the detailing of the methodology adopted in the research that undertook to access the costs of the different phases in the process of digitization and distribution.

10. **US Art Museums: charging models & policy for digital resources: A KDCS Project for the Andrew W. Mellon Foundation**

http://www.kcl.ac.uk/humanities/cch/kdcs/content/USart_vision.htm

The project is a continuation study to explore the cost and policy models adopted by cultural institutions in the USA in arriving at pricing structures for delivering surrogates of unique or rare items as digital objects. This study aims to examine the new market realities and opportunities cultural institutions face due to the transition to digitised collections. Further, it aims to discover the key factors that affect the willingness to collaborate and enable digital content to be shared.

11. **Digitization: Is it worth it? By Stuart D. Lee at Computers and Libraries**

http://www.infotoday.com/cilmag/may01/lee.htm

"The simplest example would be to take an item, consider the costs and benefits of digitizing it, and then compare this with not digitizing it. The easiest quantifiable method to use would be how this affects access. Let's take an article (for which copyright has been

cleared) that is 10 pages long, printed in black and white with no graphics. Our table suggests that the unit cost of digitizing this would be $1.80 (10 pages at 18 cents each), but our *real* costs would be more in the region of three times that amount, or $5.40. Even so, for this amount we would now have a digital facsimile of the article that we could place on the Web and make available to our readers. Now let's look at what would happen if we did not digitize the item."…

12. **Digital Work Flow Management: The Lester S. Levy Digitized Collection of Sheet Music – G. Sayeed Choudury et al.**

http://www.firstmonday.dk/issues/issue5_6/choudhury/

This document describes the development of a set of workflow management tools to reduce the manual input necessary to manage a large-scale digitization project. It also support the path from physical object and/or digitized material into a digital library repository by providing effective tools for perusing multimedia elements.

13. **Digitization and Preservation Online Resource Center**

Include resources that provide good models, guidelines, best practices, tools, tutorials, and standards.

http://digitalcooperative.oclc.org/digitize/goodpractice.html

OCLC: Digitization and Preservation – Products, Services and eContent

http://www.oclc.org/services/preservation/default.htm

Cost components for newspaper projects (from Position Paper on the Digitization of Historic Newspapers by Marilyn Deegan)
http://digitalcooperative.oclc.org/digitize/digitalnewspaper.html

The costing of digitization projects is something that many libraries are still struggling with, and there are problems inherent in the apportionment of fixed, variable, and 'sunk' costs. (See Deegan and Tanner, chapter 4 for some guidance on cost modeling.) Everything that an individual or institution does in the course of a digitization project has a

cost. Some of the costs can be made explicit: digitization, purchase of software, hiring of new staff, etc. Other costs are often 'sunk': time of existing staff, premises, overheads, etc. Still other costs are unknown and to some extent unknowable (though reasonable assumptions can be made): costs of on-going support and long-term preservation.

Marilyn Deegan and Simon Tanner, **Digital Futures: Strategies for the Information Age**, Library Association Publishing ISBN 1856044114

14. OTA – The Oxford Text Archive

Creating and Documenting Electronic Texts: A Guide to Good Practice by Alan Morrison, Michael Popham and Karen Wikander

http://ota.ahds.ac.uk/documents/creating/

Scoping the Future of the University of Oxford Digital Library Collections

Funded by the Andrew W. Mellon Foundation – Final Report by Stuart D. Lee

Appendix E – details the main issues in the digitization chain. Although dated at 1999 it is very comprehensive and illuminating

http://www.bodley.ox.ac.uk/scoping/report.html

15. Exploring Charging Models for Digital Library Cultural Heritage (Simon Tanner and Marilyn Deegan)

http://www.ariadne.ac.uk/issue34/tanner/

This article describes the results of a study to investigate some of the underlying financial and policy assumptions being made in the move from previously analog photographic services into the realm of digital capture and delivery in cultural heritage institutions. 14-Jan-2003

16. David Dawson provided a collection of documents from a planned project.

The following link with the documents is not public (it is not connected to any public page): http://www.ejewish.info/reka/minerva/docs/documents_from_d_dawson.zip

a. The HEDS Matrix of Potential Cost Factors
http://heds.herts.ac.uk/resources/matrix.html

b. Contribution – letter BS 2_pages_summary_17_2.doc

c. Digitisation Chains chains_dd.doc

d. Contribution to FP6 common project Contribution to FP6 common project.doc

e. Meeting report 2003_03_03FP6.doc

f. Minutes of meeting me030131.doc

g. Participation of French laboratories labs and researchers_12_2.doc