



University of Hertfordshire  
College Lane  
Hatfield  
Hertfordshire  
AL10 9AB

October 2001

## DEF Digitisation Project Proposals - a Review

Simon Tanner

Brian Robinson  
HEDS Manager  
University of Hertfordshire  
College Lane  
Hatfield  
Hertfordshire  
AL10 9AB

Tel: 01707 284166 [+44 1707 284166]  
Fax: 01707 286079 [+44 1707 286079]  
Email: [b.p.robinson@herts.ac.uk](mailto:b.p.robinson@herts.ac.uk)



**Contents**

<b>1</b>	<b>Introduction to the proposed projects.....</b>	<b>2</b>
<b>2</b>	<b>Weilbachs Kunstnerlesikon.....</b>	<b>3</b>
<b>3</b>	<b>Danish posters.....</b>	<b>4</b>
<b>4</b>	<b>Archive for Danish Literature (ADL).....</b>	<b>6</b>
<b>5</b>	<b>Illustreret Tidende .....</b>	<b>8</b>
<b>6</b>	<b>Danish journals.....</b>	<b>10</b>
<b>7</b>	<b>Aerial photographs of Denmark .....</b>	<b>11</b>
<b>8</b>	<b>Conclusion.....</b>	<b>14</b>

## 1 Introduction to the proposed projects

The DEF have asked the HEDS to investigate the feasibility, costs and planning implications of the following projects. HEDS has worked closely with the project proposer in each case to develop the information contained in this report. Many thanks are due to the efforts of the project proposers in providing vital information to enable this review.

However, the opinions and recommendations in this report reflect HEDS view of the projects and thus the proposers are not responsible for any error of omission, fact or difference of opinion the contents of this report may represent.

The purpose of this report is to enable the DEF Steering Committee to make decisions regarding the projects proposed at the last meeting. This report will detail for each of the projects listed below:

- Project purpose
- Overview of project plan
- Overview of project costs
- Project risks and mitigating factors

Project costs are estimates at this time and should be considered as figures which the projects would not reasonably be expected to exceed. Also all figures are excluding taxes and so assume in house working or non-taxable effort. Should outsourcing to commercial agencies be considered then taxes may be incurred in addition to the estimated costs.

### 1.1 The proposed projects

- **Weilbachs Kunstnerlesikon** - proposed by The State Museum of Art (Statens Museum for Kunst)
- **Danish posters** - proposed by the Library of Art and design at the Museum of Decorative Art (Kunstindustrimuseets Bibliotek)
- **Archive for Danish Literature (ADL)** - proposed by the Royal Library (Det Kongelige Bibliotek)
- **Illustreret Tidende** - proposed by the Royal Library (Det Kongelige Bibliotek)
- **Danish journals** - proposed by the Royal Library (Det Kongelige Bibliotek)
- **Aerial photographs of Denmark** - proposed by the Royal Library (Det Kongelige Bibliotek)

## 2 Weilbachs Kunstnerlesikon

### 2.1 *Proposing institution*

The State Museum of Art (Statens Museum for Kunst).

### 2.2 *Project purpose*

To make freely available to all the artist information contained in the Weilbachs Kunstnerlesikon via the Danish Art Index (Kunst Indeks Danmark <http://www.kid.dk/>)

### 2.3 *Overview project plan*

The dictionary of Danish artists and architects is known as the Weilbach's Kunstnerleksikon. The latest edition (the 4th) comprises nine volumes and documents about 9,000 artists and architects active in Denmark and known by name, from the Romanesque period until the present day. These 9,000 records were delivered to the State Museum for Art in electronic format (WordPerfect 5.1). The project proposes to convert these to a format suitable to be integrated into the Danish Art Index to enrich the artist records held in this database.

The project process in overview is:

1. Extract each artists data from the WordPerfect file.
2. Reorder the data into the format required for the Art Index.
3. Tag the data in the suitable XML encoding for the Art Index.
4. Upload the data into the Art Index database.

### 2.4 *Overview of project costs*

The project costs are basically:

- Development of the Danish Art Index to allow for the addition of the Weilbach information.  
Estimated at 130,000 dkk
- Cost of data extraction, conversion to XML and formatting for the Danish Art Index.  
Estimated at 100,000 dkk
- Copyright: The authors of the items contained in the Weilbach have to be compensated for the electronic publication. Their agreement with Weilbach ensures they cannot restrict electronic publication, but they automatically receive 25% of the royalty initially agreed for the paper publication.  
Estimated at 940,000 dkk of which DEF would be asked to contribute ~ 200,000 dkk.

**Total overview cost is estimated at: 440,000 dkk**

### 2.5 *Project risks and mitigating factors*

#### **Low risk**

There are no new technologies or techniques in this project and the copyright issues are already resolved.

### 3 Danish posters

#### 3.1 Proposing institution

The Library of Art and Design at the Museum of Decorative Art (Kunstindustrimuseets Bibliotek).

#### 3.2 Project purpose

To make freely available to all a unique record of historic and artistic content expressing Danish culture through posters.

#### 3.3 Overview project plan

The collection, which is a separate unit in the Library of the Danish Museum of Decorative Art, contains about 20,000 Danish and international posters from the past 150 years. These cover a wide subject area, ranging from posters signed by the artist, to graphic and photographic posters and on to commercial, political or culture-historical ones. The proposed project would focus on Danish material of which the library has about 8,000 Danish posters. Selection would be concentrated on the group of major, Danish poster artists who covered all aspects of poster design.

The project proposes to make available online an image of the poster with associated deep metadata providing detailed information about the poster, the artist and the history of the original. Due to the size and condition of the original posters the project will photograph the posters required and then digitise from the photograph.

1. Select the posters required to be digitised.
2. Photograph those posters and deliver 4" x 5" transparencies.
3. Scan the photographic transparencies as 24-bit RGB TIFF images (resolution =1000 dpi).
4. Downsample the TIFF images into JPEGs for online delivery.
5. Convert the existing catalogue data into an XML format suitable for online delivery.
6. Upload the images and metadata together into the delivery system.

HEDS suggests that the delivery mechanism already in place through the State Museum of Art's virtual gallery would be the most suitable means of delivering the content<sup>1</sup>. HEDS recommends this solution as the system is already in existence with well-established methods of data update and image delivery.

#### 3.4 Overview of project costs

The project costs are not fully available, as there is an issue of copyright in the photography that would have to be resolved (currently the photographer owns the copyright). Also the delivery mechanism is not fully resolved although the State Museum of Art resource seems extremely viable.

- Costs of photography should not exceed 600 dkk per poster<sup>2</sup>.

---

<sup>1</sup> <http://www.detvirtuellekunstmuseum.dk/>

<sup>2</sup> 600 dkk is the current maximum price of photography at the Library of the Danish Museum of Decorative Art. The State Museum of Art showed HEDS metrics that delivered a photograph and digital images with no copyright difficulties for a cost of 580 dkk.

- Dedicated project management is required and should not exceed an estimated cost per poster of 200 dkk.
- Image scanning and metadata extraction should not exceed 200 dkk per poster.
- Delivery mechanism preparation, data uploading etc should not exceed 100 dkk per poster.

**Overview cost is estimated at: 1,000 dkk per poster**

HEDS suggest that at least 500 posters be considered to maximise the cost efficiency of this project, thus totalling 500,000 dkk.

### **3.5 Project risks and mitigating factors**

#### **Medium risk**

- The copyright in the posters means that poster selection will have to be carefully considered. HEDS suggests selecting from pre-1930's materials to avoid this issue.
- Copyright in the photography will have to be resolved before the project could proceed. It is unacceptable for the photographer to retain copyright in the images. The State Museum of Art has successfully resolved this issue and may be able to advise and assist this project.
- Written and formalised agreements must be in place to guarantee any copyright agreements before any photography for digitisation can proceed.
- There is a lack of experience at the Library of the Danish Museum of Decorative Art in managing this type of project or technology. The appointment of an experienced project management resource possibly calling on the experience of either the State Museum or the Royal Library would help mitigate this risk.
- There is no delivery mechanism in place yet. The use of the State Museum of Arts delivery mechanisms would mitigate this risk and deliver DEF objectives of wide access. This solution would not restrict the Library from other implementations of the data within their institution.

## 4 Archive for Danish Literature (ADL)

### 4.1 Proposing institution

The Royal Library (Det Kongelige Bibliotek). An associated partner in the project is the Danish Society for Language and Literature.

### 4.2 Project purpose

To provide classic Danish literature from the best critical editions online for access, research and education. The digital version will also become the authoritative, stable reference resource.

### 4.3 Overview project plan

The Royal Library has already achieved part of the total vision of the ADL project through Phase One where 16 authors were completed<sup>3</sup>. In total there are 54 further authors under consideration for this proposal with a potential of 120,000 pages to be converted.

The ADL proposal delivers for each author covered a "portrait" of the author work, it's importance and an introduction to the author content. There are usually about 6 critical editions per author that are digitised in full and images of every page is made available online. The most important editions are further processed for free text retrieval and access. This resource provides a unique, stable and informative information resource relating to classic Danish literature with supportive historical research/learning aids.

The project process in overview is:

1. Select an author.
2. Order an author portrait by relevant specialist. The Royal Library will convert this to XML.
3. Select the best critical editions.
4. Scan the editions as 600 dpi, 1-bit, Group 4 compressed TIFF images.
5. Select items for free text inclusion.
6. All other items are given basic "empty" XML mark-up to aid retrieval.
7. Free text items are given OCR and full enhanced XML mark-up.
8. Select additional material from the Royal Library collections to enhance the author portrait. Clear copyright for additional materials.
9. Digitise additional materials.
10. Load and deliver all the content to the database and ADL delivery mechanism.

This is a well established project process with no new techniques or technologies.

### 4.4 Overview of project costs

The proposed project costs can be broken down into three segments. The initial costs of hardware and software plus management are present whatever the total volume of authors are converted. For every author there is an overall cost to the portraits which is required however

---

<sup>3</sup> <http://www.adl.dk/>

many books are included. Finally, there are the costs of converting the variable number of critical editions available for each author selected.

The project costs are basically:

- Project management and management of the author portrait process = 360,000 dkk
- Hardware and software upgrades to cope with the additional materials = 340,000dkk

The above costs are not linear and would be incurred for the project whatever the likely volume of authors works are converted.

- Cost per author of author portrait, additional materials, processing costs, copyright clearances and other costs = 20,805 dkk per author.
- Cost per author for conversion of the authors works assuming an average of 6 works per author = 23,190 dkk per author.

**Cost Summary:**

Headline costs = 700,000 dkk

Author costs = 43,995 dkk per author added.

**Some Totals:**

To complete the 54 remaining authors would cost 3,075,730 dkk.

To complete 30 authors would cost 2,019,850 dkk.

To complete 7 authors would cost 1,007,965 dkk.

**4.5 Project risks and mitigating factors**

**Low risk**

This is an extremely well established project process with the delivery technology in place. No new technologies or techniques are required but software and hardware improvements would be necessary to deliver the additional data load.

The start date for this project is critical to the success of subsequent projects. HEDS recommends this project be given permission to proceed as quickly as possible to enable early completion and thus release staff time and equipment for subsequent projects.

## 5 Illustreret Tidende

### 5.1 Proposing institution

The Royal Library (Det Kongelige Bibliotek).

### 5.2 Project purpose

To provide access to a digital facsimile of a unique source of Danish history between 1859 and 1924. The original is both fragile and rare, thus this project will extend access to a wide potential audience including scholars and students.

### 5.3 Overview project plan

The Illustreret Tidende is a unique cultural, news and pictorial information resource about Denmark. It has approximately 58,000 pages across 65 bound volumes plus 9 index volumes (5 general indexes and 4 topographic indexes). The pages are roughly A3 sized and the content changes across the years from all black and white to include more continuous tone content (for instance photographs) as the year's progress.

The content of the Illustreret Tidende will be scanned as 1-bit bitonal where the content is text and purely black and white pictures. Where there is continuous tone content such as photographs etc then the scanning will be in 8-bit greyscale. The indexes will be rekeyed into a database format to be used as the finding aid. There will be no OCR of text to enable searching as the accuracy is likely to be low and the Danish language is not supported by fuzzy search engines to improve search results.

The project proposal has to consider whether the Illustreret Tidende can be disbound or not. If it cannot be disbound then microfilming would be suitable to enable scanning in greyscale. If it can be disbound then the scanning be done directly from the original. HEDS recommends that wherever possible a version that can be disbound be used.

Overview of project plan:

1. Preparation to enable the Illustreret Tidende to be scanned. This might be microfilming if no disbinding is possible or may be the procurement of a copy for disbinding.
2. Scanning the page content. The overall content is in the following proportions:
  - ~2/7 black and white - scanned as 1-bit, 600 dpi only.
  - ~2/7 is black and white with some greyscale content - scanned in 1-bit, 600 dpi and 8-bit, 400 dpi.
  - ~3/7 is mainly greyscale with some black and white - scanning in 8-bit, 400 dpi and 1-bit, 600 dpi.
3. Post processing of the images for online delivery including basic XML structuring.
4. Rekeying of the indexes into a suitable XML format for database feed.
5. Development of the ADL/Centaurus system architecture to enable delivery of Illustreret Tidende.
6. Load and deliver all the content to the database and ADL delivery mechanism.

#### 5.4 Overview of project costs

The project costs will be dependent upon whether a disbindable version of Illustreret Tidende can be found. This report suggests the maximum costs in the project assuming scanning using bound volumes and offers options should an unbound version be found.

The project costs are basically:

- Rekeying the indexes and establishing the delivery mechanism: specification, rekeying, validation and database load.

Estimated at 89,000 dkk.

- Preparation to enable the Illustreret Tidende to be scanned. This could either be the cost of microfilming or the cost for the Royal Library to procure a disbindable version of Illustreret Tidende.

Estimated at 58,000 dkk.

- Cost of scanning:

1. Bitonal only total cost estimated: 28,169 dkk

2. Greyscale scanning total cost estimated: 186,412 dkk<sup>4</sup>

- Post processing of images and data load to delivery mechanism.

Estimated at 232,000 dkk.

**Total overview cost is estimated at: 593,381 dkk**

#### 5.5 Project risks and mitigating factors

##### Low/medium risk

The main risk is to do with the availability of an unbound original. Should there be an unbound version available then this project becomes low risk.

The risk in a bound version is that there are additional stages (microfilm the content and then scan the microfilm) and that microfilm scanning involves technology and techniques which the Royal Library are less experienced with.

These risks can most easily be mitigated by seeking and procuring a complete (or even substantial portions of) the Illustreret Tidende.

All other processes technologies and techniques are well-established and thus low risk. The use of the existing ADL delivery mechanism ensures that current solutions are used to reduce risk and increase cost efficiency.

---

<sup>4</sup> Greyscale cost based on scanning from microfilm. Should a disbound version be available then this cost would cover the scanning and the purchase of a suitable scanner for the greyscale content.

## 6 Danish journals

### 6.1 *Proposing institution*

The Royal Library (Det Kongelige Bibliotek).

### 6.2 *Project purpose*

To provide online access to complete Danish journals of special historical, scholarly or research importance to Denmark.

### 6.3 *Overview project plan*

As there was inadequate information at the time of HEDS review regarding which journals made up the top 10 Danish journals only general planning was possible. The assumptions for this project proposal are that the journals may be disbound - any other option is considered uneconomic and too time consuming for the project period available. The project plan and processes are based upon the ADL model, but with the exclusion of the author portraits and delivery of the pages as images only.

1. Select the journals to be scanned.
2. Clear copyright where required.
3. Scan the journals as 600 dpi, 1-bit, Group 4 compressed TIFF images. Greyscale content scanned at 400 dpi, 8-bit, TIFF images.
4. All Journals are given basic "empty" XML mark-up to aid retrieval.
5. Create information regarding journal title, contents and articles to aid retrieval.
6. Load and deliver all the content to the ADL type delivery mechanism.

### 6.4 *Overview of project costs*

The project costs are based upon the ADL cost models. The costs estimated for the scanning and delivery per page of journals are:

- To scan and deliver black and white pages is estimated at 6 dkk per page.
- To scan and deliver greyscale pages is estimated at 10 dkk per page.
- The cost of delivery is estimated at 500,000 dkk or 50,000 dkk per journal title.

It is not possible to cost a complete project until the exact journals are known, metrics such as number of pages and type of content are discovered and copyright is cleared.

The cost of copyright clearance is not included in these costs.

### 6.5 *Project risks and mitigating factors*

#### **Low/medium risk**

The only significant risk is the large number of potential pages to be processed in the timeframe available. Any delay in starting the digitisation element of the project past January or February 2002 would seriously risk project completion for the end of 2002. Possible delays might include elapsed time in committing funding or the time required gaining copyright agreements. The Royal Library's resources may be stretched beyond their capacity by this project and further planning, as more information becomes available about the 10 journals is recommended.

## 7 Aerial photographs of Denmark

### 7.1 Proposing institution

The Royal Library (Det Kongelige Bibliotek). Possible partnerships with municipal communities and the KMS (Kort and Metrikelstyrelsen) are also proposed.

### 7.2 Project purpose

To create a unique cultural resource showing the changes over time and space in the Danish landscape, industry, towns and country. To provide free access to aerial photographic materials to as wide an audience as possible through enhanced finding tools and potentially interactive services.

### 7.3 Overview project plan

There are approximately 2.5 million aerial photographs held by the Royal Library, who also own the copyright in them. The aerial photographic collection is characterised by the wide range of photographs in print form but also negatives, glass plates and 35mm slides. The quality of prints is good but very fine detail is not present in prints or in the negatives – therefore very high resolutions should not be necessary. It is estimated that about 10% may be in colour.

In many cases there is very good supporting information such as place, owner and date - but all supporting information is either hand written or printed - there is no data in electronic format. Most pictures have a film number and picture number on them and also the name of property owner at the time.

The vision of the Royal Library is to create a resource where the user can make visual selections via a clickable map with metadata to support selection and also includes a timeline information resource. Further development of this vision would include interactive services where the user is able to select images and via a clickable map interface submit suggested co-ordinates for the exact location of the property photographed. Users would also be able to provide descriptions and annotations to the properties in the photographs.

During discussions it became apparent to both the Royal Library and HEDS that 2 solutions would have to be offered to the DEF due to the costs and development cycles involved in the proposed project. These are characterised as the "Showcase" project and the "Pilot" project.

The Showcase delivers for a limited set of images (covering some 2,500 villages from the island of Fyn) a system that enables the finding of photographs, use of timelines, clickable maps as finding aids but no interaction or user input. It is intended to show what could be achieved and to demonstrate to a wide audience for the first time the wealth of aerial photographic material available. The Pilot moves forward from the Showcase to develop true interactive services for the user with a wider range of images (up to 80,000) from the island of Fyn. This Pilot would show the range of images available and also be a very exciting cultural development linking the electronic resources with the user and their community.

Therefore there are 2 project plans offered.

#### **Showcase:**

1. Procure hardware and software Geographic Information Systems (GIS).
2. Procure digital map data from KMS.
3. Selection of images from the aerial photograph archives.
4. Design and development of the underlying GIS management system and delivery mechanism.

5. Implementation of the basic application.
6. Preparation and handling of photographs in archive prior to scanning.
7. Scanning of ~2,500 photographs into suitable formats for the GIS system.
8. Post processing of images to enable zoom facilities, etc.
9. Basic metadata keyed to enable finding aids.
10. Basic co-ordinates assigned to locate photograph within the interface.
11. Load and deliver all the content to the database and GIS delivery mechanism.

**Pilot:**

All the steps from the Showcase with the addition of the following steps spread throughout the process.

1. Procure and develop relational database software, including Oracle upgrade.
2. Develop the interactive element of the GIS system.
3. Prototype testing with user group to validate the interactive elements.
4. Implement the enhanced system with interactive elements available to users.
5. Scanning of ~80,000 photographs into suitable formats for the GIS system.
6. Implement visible watermarking scheme.
7. Load and deliver all the content to the database and GIS delivery mechanism.

**7.4 Overview of project costs**

- Basic underlying system costs for both plans.  
Estimated at 250,000 dkk
- Additional costs of delivering the Showcase - includes 0.5 person year for development, scanning, metadata, co-ordinates etc for 2,500 images and related digital data from KMS.  
Estimated at 318,000 dkk

**Total cost of the Showcase = 568,000 dkk**

- Additional costs of delivering the Pilot - includes 1 person year for development of basic system plus interactive elements, relational database upgrade, scanning, metadata, co-ordinates etc for 80,000 images, watermarking and related digital data from KMS.

Estimated at 2,080,000 dkk

**Total cost of the Pilot = 2,330,000 dkk**

All costs based upon ARCIMS and ARCInfo software from ESRI.

## 7.5 Project risks and mitigating factors

### Medium/High risk

There is no GIS in place at the Royal Library – therefore the whole project is at risk from the heavy development required to deliver the vision envisaged in either plan. It is conceivable that the development may not deliver the desired functionality or will overrun severely.

All the technology and techniques involved in this proposal are new to the Royal Library. They do not have the required experience with GIS development to guarantee delivery to a set timescale. The Royal Library does have staff they believe are qualified and trained who would mitigate this risk to some extent.

Timescales are critical to this project. The Royal Library intends to employ different staff from those carrying out the other proposed projects so there should be no risks of overlap or delay in waiting for equipment. However the level of development required in either plan is such that any timescale is at risk of overrun.

The Royal Library feels that if the project is able to start by Christmas 2001 then the Showcase application could be available by Autumn 2002. HEDS feel that this may be optimistic but is a realistic goal to aim for and further system planning is required now to mitigate the risk of project overrun.

HEDS feels that it is not feasible to consider the Pilot as achievable by the end of 2002. HEDS estimate that the Pilot may be achievable in an 18 month cycle (or by addition of 8 months to the Showcase). HEDS recommends that the aerial photographs project be considered as a 3 year project with distinct deliverables at each stage which are a development or expansion upon the deliverable of the previous stage.

The selection of photographs to scan is a risk area identified by HEDS. There are obviously orders of magnitude more photographs available than could possibly be digitised in the time period and funds available. HEDS feels that selection and handling of the photographs could be a significant risk factor, which risks extending timescales for delivery. The Royal Library disagree with HEDS about the extent of this risk factor.

The Royal Library has a high confidence in the potential for success in this proposal. Good planning and a strong vision of the project goals back up their confidence. HEDS consider the risk factors higher than the Royal Library because of the high public and political profile of this type of project, the very limited timescale available and the amount of previous experience available to the Royal Library.

## 8 Conclusion

The DEF have selected projects which are both innovative and generally achievable within the funds and timescale available. The integration of proposed resources, such as the posters with the Danish Art Index and the Weilbachs Kunstnerlesikon, provides support for projects with lesser technical infrastructures in place. The Royal Library has a series of well established technologies which may be used to deliver exciting and worthy new content such as Danish literature, journals or the Illustreret Tidende. The aerial photographs project is very innovative and would add new technologies and capabilities to the Royal Library's means of delivering information resources to a wide Danish audience.

The feasibility and risk factors in these proposed projects has been discovered and HEDS recommends that levels of funding and timescales for delivery of exact outputs are agree as early as possible. Further detailed planning is recommended. This will be to fill in gaps in the technology development cycle, to resolve any outstanding copyright risks or to develop more accurate metrics regarding the selection of originals and their contents.