

KB

Koninklijke Bibliotheek

Submission form Databank of Digital Daily newspapers RFI

<i>Request for Information DDD-project</i>	<i>Version 1.0</i>
<i>Koninklijke Bibliotheek, Research & Development</i>	<i>Date: 11 May 2007</i>
<i>Status: public document</i>	

The KB invites suppliers to provide a report which addresses the questions below. If your company specializes in one or more parts in particular, reports which address only these questions are appreciated as well.

The KB welcomes responses that address the questions as precisely as possible, but the report may be fairly brief.

Company name	
Address	
Postal code	
City	
Country	
Contact person	
Function	
Telephone	
E-mail	
URL	

1. Company profile

- a. Can you – in max. 3 lines – characterize your company?
- b. In what areas does your company specialize? Please provide a short description.
- c. Please provide details on the number of people your company employs for digitisation, OCR and related services. What is the annual turnover (in Euros) of your company and/or your digitisation department?
- d. Does your company have in-house procedures for quality control? If so, please provide a brief description.
- d. How do you keep up-to-date with the latest technological developments? Do you have a Research & Development branch?
- e. How does your company cope with technological requirements that change throughout the course of a project, e.g. as a result of new technological advancements?

2. Digitisation

- a. What technological means and current practices does your company have at its disposal for large-scale digitisation of historical newspapers? Please provide a brief description.

- b. Which quality standards does your company maintain when digitizing historical newspapers?
- c. Does your company have experience in scanning of historical newspapers from the original? If so, please supply a brief description.
- d. Does your company have experience in scanning of historical newspapers from microfilm? How does your company cope with high-contrast and low-contrast microfilm? Please supply a brief description.
- e. Does your company have any experience with techniques to enhance the quality of the images in order to enhance the quality of the OCR? If so, please supply a brief description.
- f. Do you have any further suggestions concerning the digitisation of the DDD-newspapers?

3. OCR and metadata

- a. Does your company have experience with OCR of historical texts? If so, please supply a brief description (e.g. historic period of the texts, kinds of texts and percentage of correct OCR).
- b. Does your company have specialized techniques to enhance the quality of the OCR of historical texts? If so, please supply a brief description. How does your OCR software handle historic text fonts?
- c. Does your company have techniques to recognise separate articles on newspaper pages during OCR ('zoning')? If so, please supply a brief description. To what extent is this an automated process? How would your software deal with articles that are split across different newspaper pages (e.g. an article starts on the front page and continues on page 5)?
- d. Does your company have software that can recognise basic metadata, like newspaper title, date, article headers and page numbers? If so, please supply a brief description. To what extent is this an automated process?
- e. Does your company have software that supplies the position of text elements on the page (e.g. the position of articles, paragraphs and words on the page)? If so, please supply a brief description. Does your software support the ALTO format?
- f. Can your company supply structural metadata which describes the hierarchical relationship between items (newspaper, issue, page, article)? If so, please supply a brief description. In which (XML-)format can structural metadata be supplied? Does your software support MPEG21-DIDL?
- g. Do you have any further suggestions concerning the OCR of the DDD-newspapers?

4. Searchability

- a. Do you have any suggestions how to deal with historic spelling variants in the texts? Does your company have any techniques concerning this issue? If so, please supply a brief description.
- b. Do you have any further suggestions to enhance the searchability of the DDD-newspapers?

5. Presentation

- a. Do you have any suggestions with regard to the navigation of the newspapers on a website? Does your company have software which supports navigation? If so, please supply a brief description.
- b. Do you have any suggestions concerning zooming and zoning in the presentation on a website? Does your company have software to support these issues? If so, please supply a brief description.
- c. Supposing the scans will be TIFF-images, which derived image-format would you advise for presentation on the website? JPG, PNG or anything else? Please state the arguments briefly.
- d. Do you have any further suggestions for the presentation of the DDD-newspapers on a website?
- e. Please provide URLs to relevant projects your company has been involved in or is currently working on.

6. General

- a. To what extent is your OCR- and presentation software scalable? Can it easily cope with 8 million pages of newspaper texts? Is it scalable to potential new mass digitisation projects that will possibly be even larger in volume?
- b. In case of search-and-retrieval or presentation software, to what extent is your software open? Can it cooperate with software from third parties, e.g. search engines? Is it modular and is it possible to substitute components with new ones, e.g. components which may be developed by the KB in the future?

Please send this document **before June 1 2007** to ddd@kb.nl