



Invitation to Tender: Purchase of micro-level publication data for Danish scientists during the period from 1970 to 2017

1. Introduction

The aim of the Danish Ministry of Higher Education and Science's (UFM) present tender is to receive offers for publication data for all Danish scientists (in both the public and the private sector) during the period from 1970 to 2017. The purpose is to acquire a dataset that may be used together with other data sources (e.g. register data, patent data, etc.) to conduct studies and research projects that will help elucidate the socio-economic effects of research.

Above all, there must be three levels to the dataset (described in detail in Section 6): 1) Microdata on all authors of scientific publications who have been affiliated with a Danish institution or company, 2) microdata on all publications by these authors, and 3) microdata on these publications' references and citations.

The tender process will be conducted in collaboration with Professor H. C. Kongsted, Copenhagen Business School (CBS). Professor H. C. Kongsted, Copenhagen Business School (CBS) will have the same rights as the Danish Ministry of Higher Education and Science to utilize the data.

Tenders for this project are invited pursuant to Section 191 and 192 of the Danish Procurement Act, cf. Act no. 1564 of 15 December 2015.

2. Appendices

Appendix 1: Draft contract

3. Background

The background for this Invitation to Tender is multi-faceted:

- In 2016, the Danish Ministry of Higher Education and Science (UFM) initiated a major interministerial analysis programme with the purpose of strengthening the academic basis for evaluating existing public research. Among other things, the analysis programme will examine differences in effect across a variety of types of research, thus helping improve prospective prioritisation and organisation of public research. As part of this analysis programme, a broad spectrum of data will be collected for use in the sub-analyses that will support the overall analysis programme. In this context, bibliometric data on scientific publications that can be attributed to Denmark due to the author are of crucial importance.
- UFM is responsible for ensuring continual development of the research and innovation sector. Basic factual knowledge of the sector is a condition for being able to conduct a well-informed dialogue with the sector's stakeholders and players. Purchase of the dataset at author level and with author ID enables identification of individuals/scientists. These data may be enriched with, for instance, register data. This will enable analysis of a range of hypotheses in Danish research, which to date are untested. For example, it will be possible to present new, clarifying analyses of the mobility of researchers and the link to



scientific productivity and quality. Thus, it will be possible to trace the correlation between publication behaviour and horizontal and vertical career mobility.

- The dataset will be included in current and future research projects at CBS. The current project attempts to gain a better understanding in general of the way in which universities, companies and research-funding institutions interact. The project focuses on the individual scientist and the incentives and barriers facing the scientist when conducting research-related activities such as publication, patenting and start-up of new companies.

-

4. The contracting authority

The contracting authority is:

The Danish Ministry of Higher Education and Science
Børsgade 4
1215 København K

Postboks 2135
1015 København K

Web: www.ufm.dk

Danish Agency for Science and
Higher Education

5. Tender deadlines and other important conditions concerning the [tender/announcement]

Tenders for implementation of the project must be received by UFM no later than 12.00 midday on 9 November 2017.

Tenders must be sent to UFM as a PDF document by email to Lasse Nielsen (lnl@ufm.dk). Please enter 'Tender for author-based publication data' in the subject field.

All enquiries, including questions, must be sent by email to UFM's contact, Lasse Nielsen (lnl@ufm.dk).

To ensure that we are able to respond to any questions from interested parties, please send your preliminary expression of interest no later than 30 October 2017, indicating the email address to which UFM may send questions and answers.

Questions will be answered in writing, and both questions and answers will be sent in anonymised format to tenderers who have submitted a preliminary expression of interest. It is expected that questions will be answered on a continual basis. Questions submitted no later than 1 November 2017 will be answered no later than 3 November. Only a brief response may be expected to questions submitted after this date.

The financial framework for this project is DKK 500,000.

Finally, the dataset must be delivered, cf. the tender specifications, to UFM before the end of November.



UFM evaluates tenders received on time in two stages. Firstly, the suitability of all tenderers will be assessed, cf. Section 7. Tenders received from tenderers who have been found suitable and who have submitted tenders in compliance with contract conditions, cf. Sections 8 and 9, will then be evaluated.

After 9 November, UFM may choose to enter into negotiations with the tenderers who have submitted tenders in compliance with contract conditions. If UFM chooses to initiate negotiations, all relevant tenderers will be invited individually. After a round of negotiations, each tender will be reassessed based on the negotiations.

UFM will then decide which tenderer has submitted to UFM the economically most advantageous tender, cf. Section 9. All tenderers will be notified of the result in writing, as quickly as possible and simultaneously.

Any reservations regarding parts of the tender documents must be clearly stated. Reservations may not be made for basic elements or factors that cannot be priced.

Please note that no legally binding agreement or promise thereof exists before the final contract has been signed. The legal framework for the project is based on UFM's standard contract, which is attached as Appendix 1.

6. Description of the project

The customised data should cover authors and publications in Denmark, in all fields, during the period from 1970 to 2017. The main specifications for the dataset are listed below.

The dataset to be delivered must include the information necessary to enable us to produce 3 linked datasets:

- 1) A list of **authors** who published (article, book, book chapter, conference paper, review) with a Danish affiliation of any type (university, company, public research organisation, etc.) at any point of time and in any field during the period from 1970 to 2017.

This list of authors will be also used to match individuals with other databases (Danish Register and Patent Data) and, therefore, must contain the following information: *First name, last name, unique author ID (to enable grouping of the author's activities across time and affiliations), affiliation name, affiliation address, affiliation city, affiliation country, affiliation ID (enabling affiliations to be grouped across years), year of publication with a Danish affiliation.* The affiliation data must be reported with reference to the time of publication. If the author has several different affiliations within a given year, all Danish affiliations must be reported.

Special attention will be given to response to the following:

- The consultant should describe the coverage of the chosen database, i.e. with regard to scientific field, coverage over time, coverage of local language and publication types.



- The consultant should describe how the unique author ID is constructed, including the quality of this variable over time as well as across institutions and countries.
 - The consultant should also describe the construction and quality of the affiliation ID that allows analysis of Danish affiliations, e.g. coverage at author level, quality of this variable over time, etc.
- 2) For the authors identified under item 1), the data must comprise **their life-time publication records** (article, book, book chapter, conference paper, review) with **any DK or non-DK affiliation** in any field. These data should include: content abstract, forward citations' ID, forward citation count, references, article title, document type, source title, ISBN, ISSN, volume, issue, page, publication date, DOI, PII, pubmed ID, ORCID, first author, affiliation name, affiliation city, affiliation country, affiliation ID, complete author list with affiliation name, affiliation city, affiliation country, affiliation ID, as well as other relevant and available data.

Special attention will be given to response to the following:

- The consultant should provide an estimation of the proportion of the relevant publications that have a unique publication ID, which can be linked to publications in leading international bibliometric databases.
 - The consultant should estimate the degree of coverage of the above-mentioned variables.
- 3) In addition, we request **citations received disaggregated for each year** for each article identified under item 2). This is the output we expect to receive or to be able to produce:

Publication ID	Citations with self-citation	Citations without self-citation	Year
X			1970
X			[...]
X			2017
Y			1970
Y			[...]
Y			2017

The customised data should be delivered in XML format. With regard to the delivery method, if size permits, we expect the data to be stored on an external hard drive. Alternatively, the data should be available through an FTP site.

7. Suitability requirements

In order to be considered for this project, the tenderer must meet the following suitability requirements:

- The company must have experience of similar projects



- The company must be familiar with and have access to international bibliometric databases
- Company resources, both financial and staffing: Does the company have sufficient resources to complete the project?

8. Procedural requirements for tender format

The tender is expected to cover no more than five (5) pages, without appendices.

The tenderer must include a test dataset in its tender, complying with the specifications described under items 1, 2 and 3. The test dataset must be based on all relevant scientists in a specific and defined scientific sub-field.

To facilitate comparison and to ensure equal treatment of tenders received, tenderers are requested to adhere to the framework below:

Danish Agency for Science and
Higher Education

Tender description

Contact details

- Company name
- Contact
- Address
- Telephone number and email address (to which correspondence will be sent).

Description of the solution proposal, including specific reference to the following items (see Section 6):

- The consultant should describe the coverage of the chosen database, i.e. with regard to scientific field, coverage over time, coverage of local language and publication types.
- The consultant should describe how the unique author ID is constructed, including the quality of this variable over time as well as across institutions and countries.
- The consultant should also describe the construction and quality of the affiliation ID that allows analysis of Danish affiliations, e.g. coverage at author level, quality of this variable over time, etc.
- The consultant should provide an estimation of the proportion of the relevant publications that have a unique publication ID, which can be linked to publications in leading international bibliometric databases.
- The consultant should estimate the degree of coverage of the above-mentioned variables.

Appendices

- Time schedule and activity plan
- Payment plan

9. Evaluation of tenders received

The main criterion for evaluation of tenders received is 'economically most advantageous tender', with the sub-criteria below:

- Database 60%



- Solution description 20%
- Price 20%

Each of the sub-criteria will be included in the overall evaluation. The tenderer that has submitted the bid that is economically most advantageous for UFM will then be selected to carry out the project.

The first two sub-criteria are briefly described below:

Database:

In the case of this criterion, it is important that the dataset is based on a bibliometric database with high coverage of Danish and international publications and high quality and a high degree of coverage for each individual variable. It is a minimum requirement that data have a viable author ID that can be linked to register data.

Danish Agency for Science and
Higher Education

Solution description:

In the case of this criterion, it is important that the consultant addresses the three levels described in Section 6 and the points requiring special attention, highlighted in Sections 6 and 8.

10. Tentative time schedule

Date	Activity
24 October 2017	Announcement of the Invitation to Tender at udbud.dk
30 October 2017	Deadline for expression of interest
1 November 2017	Deadline for questions relating to the Invitation to Tender
3 November 2017	Deadline for responses relating to the Invitation to Tender
9 November 2017	Deadline for receipt of tenders: 12.00 midday
10-15 November 2017	Evaluation of tenders received
16 November 2017	Selection of consultant