

## **Data Mining *The American Archivist* via JSTOR**

To-do #18 from our February 2012 meeting was to explore with MetaPress and JSTOR the possibility of downloading the whole AA corpus so that SAA could enable new studies of the journal's content and get professionals and scholars with the right data analysis skills interested in the back run of the journal.

JSTOR can facilitate such text mining. JSTOR has a website up right now called the Data for Research (Dfr) service: <http://dfr.jstor.org/>. The service is in a beta right now, but it was built as a way to allow scholars and researchers to perform analysis across the content that is on JSTOR. The site allows for full text and fielded searching and it does allow the researchers to download limited data sets. Below is the paperwork that researchers would have to sign when they get access to large data sets using the Data for Research site.



### **DATA FOR RESEARCH QUESTIONNAIRE**

To facilitate creating a Data for Research agreement between JSTOR and an institution or individual, JSTOR's Legal Unit has created the following questionnaire. This questionnaire is designed to help answer the most common questions we have about the specifics of JSTOR's data contribution.

Please type in the answers to the following questions in a much detail as possible and add additional information as needed, including any pertinent proposals or descriptive materials that would assist in further describing the nature of the data contribution.

Please note that the information you provide herein will become part of the agreement between ITHAKA and the recipient of the data, so please be as complete and accurate as possible.

**INSTITUTION and PRIMARY RESEARCHER:**

Please specify: \_\_\_\_\_

## SCHEDULE A

### GENERAL

1. Project Description:

---

---

---

Academic Institution

Non -academic Institution

2. Is the Project subject to a grant?

Yes

No

Source of the grant funding and details about the grant:

---

---

3. Nature of the data to be provided (check all that apply):

a. Citation Reference

b. OCR

c. Metadata

d. Usage Statistics

e. Other:

Please explain: \_\_\_\_\_

4. Scope and nature of the data to be provided (check all that apply):

a. The data will be limited by date   
Please describe \_\_\_\_\_

b. The data will be limited by journal titles   
Please describe \_\_\_\_\_

c. The data will be limited by discipline   
Please describe \_\_\_\_\_

c. The data will be limited by regional location   
Please describe \_\_\_\_\_

d. The data is sensitive   
Please describe \_\_\_\_\_

Other: \_\_\_\_\_

**ACCESS TO DATA**

5. "Key researchers" and their affiliation:

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

6. Others, if any, who are expected to have access to the data (for example, graduate students or undergraduates)?

\_\_\_\_\_  
\_\_\_\_\_

7. Delivery Format of the Data:

a. Delivery in tangible medium

b. On-line delivery mechanism

c. Data to be delivered in a tangible medium by:

JSTOR

Licensee

d. Return of delivery device?

Yes

No

8. Will other parties require access to this data?

Yes

No

Please explain: \_\_\_\_\_  
\_\_\_\_\_

9. Return or destroy data after a particular period of time?

Returned

Destroyed

Please explain: \_\_\_\_\_  
\_\_\_\_\_

10. Use of data:

Check all that apply.

Research

Classroom

Other

Please explain: \_\_\_\_\_  
\_\_\_\_\_

**RESULTS OF THE RESEARCH**

11. Results of the research using the data will be used as follows:

Publications

Conferences

Other

Please explain: \_\_\_\_\_  
\_\_\_\_\_

12. If publication is intended, are the researchers likely to want to specify the names of publications/publishers or institutions that are the source of the data?

Yes

No

13. Invention or other product, device, or process developed from the use of the data?

Yes

No

Please explain: \_\_\_\_\_  
\_\_\_\_\_

JSTOR interested in method of research or the results?

Yes

No

Please explain: \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

15. Reports or meetings regarding progress of the Project:

Yes

No

How often: \_\_\_\_\_

Upon request by JSTOR?

Final report?

Yes

No

16. Licensee Notice Address

---

---

---

---

---



## DATA FOR RESEARCH AGREEMENT

THIS DATA FOR RESEARCH AGREEMENT (this “Agreement”) is entered into by and between JSTOR, a part of Ithaca Harbors, Inc., a not-for-profit corporation formed under the laws of the State of New York (“JSTOR”), and \_\_\_\_\_ (“Licensee”) (hereinafter referred to as the “Parties”) as of the Contract Date specified below (the “Effective Date”).

WHEREAS, JSTOR is dedicated to helping the scholarly community discover, use, and build upon a wide range of intellectual content in a trusted digital archive. JSTOR’s overarching aims are to preserve a record of scholarship for posterity and to advance research and teaching in cost-effective ways;

WHEREAS, JSTOR operates a research platform that deploys information technology and tools to increase productivity and facilitate new forms of scholarship. JSTOR collaborates with organizations that can help it achieve its objectives and maximize the benefits for the scholarly community;

WHEREAS, in connection with operating its research platform, JSTOR has generated a corpus of data that is of value to the scholarly community;

WHEREAS, Licensee is a research organization that aims to conduct research to improve and advance scholarship; and

WHEREAS, in furtherance of its mission, JSTOR shall provide certain data to Licensee;

NOW, THEREFORE, in consideration of the premises and the mutual promises contained herein and other good and valuable consideration, the receipt and sufficiency of which are hereby acknowledged, the parties agree as follows:

### 1. DEFINITIONS

For purposes of this Agreement, the following definitions shall apply:

1.1 “Data” shall mean the JSTOR Data as set forth in Schedule A, attached hereto and incorporated herein.

1.2 If applicable, the “Key Researchers” shall be those individuals as set forth in Schedule A.

1.3 The “Project” shall be as described in Schedule A.

### 2. USE OF DATA

2.1 Use of the Data shall be limited to Key Researchers solely for internal evaluation and research purposes within the scope of the Project. Data shall be accessed by Key Researchers solely on an as-needed basis. Data shall not be accessed by or reproduced, distributed, communicated, or displayed by Licensee to any third party. In the event that Licensee and JSTOR jointly determine that it would be mutually beneficial for the Data to be provided to other parties as designated in Schedule A in connection with the Project (or for any other mutually agreed upon purpose), those such parties shall sign and submit to JSTOR the JSTOR Confidentiality and Non-Disclosure Agreement, a copy of which is attached hereto as Exhibit A.

2.2 The parties further agree that any article or reports about the Project developed for publication shall be provided to JSTOR for approval prior to submission to a conference, journal, or other publication. Any communications, reports, articles made, distributed, or published in connection with the Project shall uphold the non-disclosure terms of Section 5, and the trademark provisions of Section 6, and shall be modified as reasonably requested by JSTOR to ensure that non-disclosure is maintained with respect to the Data and trademarks.

2.3 **If applicable**, the parties further agree that any article or reports about the Project developed for publication shall be provided to JSTOR for approval prior to submission to a conference, journal, or other publication. Any reports or articles distributed or published by the parties designated in Schedule A in connection with the Project shall uphold the non-disclosure terms of Section 5 and the trademark protection provision of Section 6, and shall be modified as reasonably requested by JSTOR to ensure that non-disclosure is maintained with respect to the Data and trademarks.]

2.4 As needed pursuant to Schedule A, Licensee shall communicate the terms and conditions of this Agreement to Key Researchers, including, but not limited to, the terms set forth in Section 2 and Section 5.1

### 3. DELIVERY OF DATA

JSTOR shall deliver the Data or cause the Data to be delivered to Licensee to the attention of the contact designated herein within thirty (30) days of execution of this Agreement or as otherwise set forth in Schedule A. The Data shall be delivered in a medium to be mutually agreed upon by the Parties as set forth in Schedule A.

### 4. PROJECT COMMUNICATIONS

4.1 During the term of this Agreement, representatives of Licensee and JSTOR shall cooperate in discussing the progress and findings of the Project from time to time. During the period of this Agreement, Licensee shall provide JSTOR with a written progress report concerning the findings of the Project as set forth in Section 16 of Schedule A. If required pursuant to Schedule A, a final written report including a complete summary of the Project and setting forth the results achieved under the Project shall be submitted by Licensee to JSTOR within forty-five (45) days of termination of the Project. The Parties agree that an article about the Project developed for publication shall suffice as a final written report and that such article shall be provided to JSTOR, with a reasonable period for review, prior to submission to a conference, journal, or other publication

4.2 The Parties agree to enter into the Project and discussions about the future use of the Data in good faith; however, the Parties also agree that neither the holding of discussions nor the exchange of material or information shall be construed as an obligation of either party to perform any work, enter into any license, business engagement or other agreement with the other party in the future. The Parties also reserve the right, in each of their sole discretion, to reject any and all proposals made by the other party or its representatives with regard to a transaction or relationship between the parties and to terminate discussions and negotiations at any time.

### 5. NON-DISCLOSURE

#### 5.1 GENERAL

In connection with the use of Data for purposes of conducting the Project, Licensee shall not, during the term of this Agreement or any time thereafter, disclose, or permit any of its staff, students, employees, agents, or assignees to disclose, to any other person or entity any Confidential Information (as defined herein) of or provided by JSTOR. "Confidential Information" shall mean all of the Data, and other data, reports, instructions, directions, interpretations, specifications, and records containing or otherwise reflecting information concerning JSTOR, its affiliates, and subsidiaries that is not available to the general public and is provided by JSTOR to Licensee in the course of their dealings, together with all analyses, compilations, studies, or other documents or materials, whether prepared by JSTOR or by others, that contain or otherwise reflect such

information, along with any other information that is identified as proprietary and exclusive by JSTOR or otherwise understood by its nature to be proprietary and exclusive when disclosed. “Confidential Information” does not include information which (a) was or becomes generally available to the public other than as a result of a disclosure by Licensee or its representatives or (b) was or becomes available to Licensee on a non-exclusive basis from a source other than JSTOR or its advisers, provided that such source was not known by Licensee to be bound by any agreement to keep such information restricted, or otherwise prohibited from transmitting the information to Licensee by a contractual, legal or fiduciary obligation.

## 5.2 PERSONAL IDENTIFYING INFORMATION

JSTOR is committed to protecting user privacy and has made best efforts to anonymize the Data. Nonetheless, during the course of the Project, Licensee may discover it has the capability to discern from the Data certain personal identifying information about JSTOR users, participants, or content contributors. Under no circumstances shall Licensee discern or attempt to discern such information or disclose or communicate information from which a user, participant, or content contributor can be identified, whether directly or indirectly.

## 5.3 CORPORATE CONFIDENTIALITY

JSTOR may request from time to time under certain circumstances that its contributions or participation in certain research efforts be acknowledged. Licensee shall in good faith consent to and further support any such reasonable acknowledgment requested by JSTOR, and such consent shall not be unreasonably withheld. Such acknowledgment as requested by JSTOR may be in the public or private forum, and may be in the form of a publication, media announcement, online posting, or any other method that JSTOR deems appropriate.

## 6. INTELLECTUAL PROPERTY

6.1 Subject to the intellectual property rights of third parties, JSTOR shall retain all rights to the Data. Except as expressly granted herein, under no circumstances shall anything in this Agreement be construed as granting, by implication, estoppel, or otherwise, to Licensee a right or license to JSTOR’s name, logo, design marks, trade names, or service marks (collectively, the “Trademarks”), issued patents and patent applications, copyrights and copyrights registration and applications, rights in ideas, designs, works of authorship, derivative works, or any other rights or license relating to the other party’s intellectual property or services. Licensee shall not at any time during or following termination of this Agreement use Trademarks of JSTOR, participants, or content in any third-party accessible advertising, reports, or other documentation or presentations, whether written or oral, without the prior written consent of JSTOR.

6.2 The parties agree that should there be any resulting software or other product, device, process or method arising out of the use, whether directly or indirectly, of the Data (the “Invention”), such Invention shall be subject to the right of JSTOR to a worldwide, royalty-free, perpetual license and any subsequent modifications thereof created by or on behalf of Licensee.

## 7. TERM AND TERMINATION

7.1 JSTOR’s involvement with the Project shall terminate on any of the following conditions: i) upon termination or completion of the Project; ii) in the event of Licensee’s breach of the terms of this Agreement; or iii) at JSTOR’s sole discretion, with or without cause on ten (10) days notice. Upon termination, the Data shall be destroyed or returned immediately upon the request of JSTOR

7.2 Upon termination of this Agreement, all rights granted hereunder shall be terminated, except that the obligations set forth in Sections 5, 6, and 9 shall survive termination, as shall any other provisions which by their terms are intended to survive termination

## 8. NOTICES



All notices given pursuant to this Agreement shall be in writing and sent to the Notice Address specified in Section 16 of Schedule A for the Licensee, and as below for JSTOR. These addresses also are applicable for general contact between the parties. Notices may be delivered by hand, by overnight carrier, or shall be deemed to be received within five (5) business days after mailing if sent by registered or certified mail, postage prepaid. If any notice is sent by facsimile, then confirmation copies must be sent as specified above. Either party may from time to time change its Notice Address by written notice to the other party.

If to JSTOR:  
Nancy A. Kopans  
General Counsel  
JSTOR  
149 Fifth Avenue  
New York, NY 10010

9. REPRESENTATIONS AND WARRANTIES

9.1 Each party hereby represents and warrants that it is duly organized and validly subsisting and has full authority to enter into this Agreement and to bind the party to the terms and conditions herein. Each party further represents and warrants that it has caused this Agreement to be executed by a duly authorized representative.

9.2 **THE DATA IS PROVIDED ON AN “AS IS” BASIS, AND JSTOR DISCLAIMS ANY AND ALL OTHER WARRANTIES, CONDITIONS, OR REPRESENTATIONS (EXPRESS, IMPLIED, ORAL OR WRITTEN), RELATING TO THE DATA OR ANY PART THEREOF, INCLUDING, WITHOUT LIMITATION, ANY AND ALL IMPLIED WARRANTIES OF QUALITY, PERFORMANCE, COMPATIBILITY, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. JSTOR MAKES NO WARRANTIES RESPECTING ANY HARM THAT MAY BE CAUSED BY THE TRANSMISSION OF A COMPUTER VIRUS, WORM, TIME BOMB, LOGIC BOMB OR OTHER SUCH COMPUTER PROGRAM. JSTOR FURTHER EXPRESSLY DISCLAIMS ANY WARRANTY OR REPRESENTATION TO AUTHORIZED USERS, OR TO ANY THIRD PARTY.**

10. GOVERNING LAW

This Agreement shall be interpreted and construed according to, and governed by, the laws of the State of New York or United States Federal law, as applicable, excluding any such laws that might direct the application of the laws of another jurisdiction.

11. MISCELLANEOUS

11.1 This Agreement constitutes the entire agreement of the parties and supersedes all prior communications, understandings and agreements relating to the subject matter hereof, whether oral or written. No modification or claimed waiver of any provision of this Agreement shall be valid except by written amendment signed by authorized representatives of JSTOR and Licensee.

11.2 This Agreement and any amendments may be executed in one or more counterparts, each of which shall be deemed an original, but all of which together shall constitute one agreement.

11.3 Nothing contained herein shall be deemed to create an agency, joint venture, partnership or contractual relationship between the parties except as expressly set forth herein. Neither party shall use or attempt to exercise apparent authority concerning the other party in its dealings with third parties. Nothing in this agreement shall prohibit the Parties from providing any of their own information to third parties and entering into agreements with third parties.

11.4 Waiver of any provision herein shall not be deemed to be a waiver of any other provision herein, nor shall waiver of any breach of this Agreement be construed as a continuing waiver of other breaches of the same or other provisions of this Agreement.

11.5 If any provision or provisions of this Agreement shall be held to be invalid, illegal, unenforceable or in conflict with the law of any jurisdiction, the validity, legality and enforceability of the remaining provisions shall not in any way be affected or impaired thereby.

11.6 Neither party may assign, directly or indirectly, all or part of its rights or obligations under this Agreement without the prior written consent of the other party.

**LICENSEE**

**JSTOR, a part of ITHAKA**

**BY:** \_\_\_\_\_

**BY:** \_\_\_\_\_

**NAME:** \_\_\_\_\_

**NAME:** Nancy A. Kopans

**TITLE:** \_\_\_\_\_

**TITLE:** General Counsel

**DATE:** \_\_\_\_\_

**DATE:** \_\_\_\_\_