**DRAFT VERSION**

**Ford Foundation Digital Records Pilot Project**

The Rockefeller Archive Center (RAC) is requesting $xxxxxxx in support of a project that will transfer a select set of digitized records from the Ford Foundation to the Archive Center. The project has three main goals:

1. To transfer and preserve the specified materials and establish the RAC as the long-term custodian for these materials
2. To increase researcher access to the materials
3. To establish workflows and practices for future file transfers (i.e. digital accessions) between the RAC and the Ford Foundation

The RAC estimates that all six phases of the pilot project will require approximately 18 months to complete, primarily because of the complexities surrounding the Unpublished Reports metadata (phase 3).

**Materials to be Transferred and Preserved**

The digital materials that will be involved in this pilot project include:

* Annual Reports, 1950 – 2012: 64 PDFs, 1.4 GB
* Index Cards: 9036 PDFs, 7.05 GB
* Oral histories – digitized audio files: 275 wav/mp3, 290 GB
* Oral histories – scanned transcripts and agreements – 320 PDFs, 520 MB
* Unpublished reports – about 19,000 PDFs, about 100GB

**Process**

The overall process of this transfer entails six phases, each outlined below.

Phase 1: Donor Preparation

The Ford Foundation will, with guidance from the RAC, prepare the files mentioned above for transfer. Specific tasks will include naming of the files according to agreed upon conventions, grouping the files into manageable transfer sizes, and packaging the files using Bagit, a file transfer tool developed by the Library of Congress.[[1]](#footnote-1) Prior to transfer and packaging of the files, the Ford Foundation will also provide a directory listing (essentially a file-level inventory) of each of the transfers, noting any gaps or missing files. Any rights issues will be reviewed by Ford and communicated to the RAC prior to Phase 4. Transfer of the files to the RAC will be via a newly developed web deposit site. This site would automatically package the files using Bagit and ease overall submission of the files. This site would be developed externally and also include most of the functionality listed in Phase 2.

Development costs: $12,675

Phase 2: Acquisition

In Phase 2, the RAC will officially acquire the transfers. This includes: bag validations, virus checking, file verification, documentation of receipt, transfer documentation, recording of the accession, moving the transfers to a staging server and finally a second validation after the move.

Estimated time needed to process the acquisition of approximately 29,000 files totaling 400 GB: 41 hours

Phase 3: Data Management

After the acquisition phase is completed, the RAC will create descriptive information in the Archivists’ Toolkit or ArchivesSpace to aid in the management of the materials, and to produce online access tools (finding aids) for researchers. The Annual Reports, Oral Histories, and a portion of the Index Card materials will require the creation of original description. Another portion of the Index Cards will be described via a script generated from their filenames. Metadata previously provided by the Ford Foundation will be used to generate data for the Unpublished Reports. Although this will be automated as much as possible, massaging and cleaning up the Unpublished Report metadata will take a considerable amount of time. At this point, pertinent information about rights will be encoded into machine-actionable statements and each group of materials will be further grouped by their applicable rights status.

Estimated time for data management activities: 300 hours

Phase 4: Ingest to Archivematica

Phase four of the project will consist of ingest of the materials into Archivematica, the RAC’s digital preservation management system. At this time, the files will have the appropriate rights information attached to them, and will undergo a series of micro-services, which extract and create necessary metadata for archival packaging. Individual files will then be connected to descriptive information in ArchivesSpace, through an automated process developed with outsourced programming.

Estimated time for Archivematica ingest and ArchivesSpace linking: 26 hours

Quote from our Archivematica vendor for outsourced programming: $15,925

Phase 5: Ingest to MetaArchive

The RAC uses the MetaArchive service to distribute digital archival materials geographically to protect them from the preservation threats of natural disasters, technical errors, and malicious attacks. Archivematica packages will be uploaded to a MetaArchive-accessible server and then distributed across the MetaArchive system.

Estimated RAC handling time: 3 hours

Annual data storage costs at $1/GB/year: 400GB at $400/year

Phase 6: Upload to DIMES

The final phase of the project will consist of uploading new finding aids as well as the full content of unrestricted materials to the RAC online discovery system, DIMES.

Estimated time: 3 hours

**Staff Involved**

At the RAC, the project will be under the direction of Sibyl Schaefer, Assistant Director and Head of Digital Programs, who will act as the official project director for this grant (or contract). She will be assisted by Hillel Arnold, Digital Archivist, as well as Patrick Galligan and Laura Montgomery, Assistant Digital Archivists. Additional part-time staff may be hired as needed. Resumes for the project staff are included in an appendix.

**Project Budget**

60%: Asst. Archivist

30%: Archivist

10%: Asst. Director

1. http://sourceforge.net/projects/loc-xferutils/ [↑](#footnote-ref-1)