

Agreement Elements for Outsourcing Transfer of Born Digital Content

#bdreport

Ricky Erway
OCLC Research

Ben Goldman
Pennsylvania State University

Matthew McKinley
University of California, Irvine



© 2014 OCLC Online Computer Library Center, Inc.

This work is licensed under a Creative Commons Attribution 3.0 Unported License.

<http://creativecommons.org/licenses/by/3.0/>



August 2014

OCLC Research

Dublin, Ohio 43017 USA

www.oclc.org

ISBN: 1-55653-482-5 (978-1-55653-482-9)

OCLC Control Number: 885368946

Please direct correspondence to:

Ricky Erway

Senior Program Officer

erwayr@oclc.org

Suggested citation:

Erway, Ricky, Ben Goldman and Matthew McKinley. 2014. *Agreement Elements for Outsourcing Transfer of Born Digital Content*. Dublin, Ohio: OCLC Research.

<http://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-born-digital-content-transfer-2014-a4.pdf>.

Acknowledgments

The authors acknowledge the many contributions of their collaborators and advisors:

- Jim Austin, Jim Austin Computer Collection
- Fran Baker, University of Manchester
- Mary Elings, University of California, Berkeley
- Matthew Kirschenbaum, University of Maryland
- Dave Thompson, Wellcome Library
- Stephen Torrence, Museum of Computer Culture

They also appreciate the guidance and editing advice of Jackie Dooley, OCLC Research.

Introduction

In 2012, OCLC Research published *Swatting the Long Tail of Digital Media: A Call for Collaboration*.¹ It made the case that few institutions can be expected to have the hardware, software, and expertise to be able to read all digital media types. In order to make real progress in preserving and providing access to born-digital content, libraries and archives need to leverage specialized resources and expertise across the community.

Working with OCLC Research, a group of archival practitioners embarked on a pilot project to test outsourcing of the transfer of content from physical media they couldn't read in-house. Some considered working with another cultural heritage institution, while others investigated outsourcing to commercial entities. Early in the pilot they realized the need for agreements between repositories and service providers to spell out the terms of such collaboration. The group began compiling a list of elements that should be considered when creating these agreements. That list has evolved and been fleshed out to the point where we feel it worth sharing with the community.

This document suggests elements that should be considered when constructing an agreement (or memorandum of understanding) for outsourcing the transfer of born-digital content from a physical medium, while encouraging adherence to both archival principles and technical requirements. It has five sections:

1. General Provisions
2. Information Supplied by Service Provider
3. Information Supplied by Client
4. Statement of Work
5. Cost and Liability

The document assumes that two parties are involved: the Client and the Service Provider. The parties should collaborate in compiling the agreement and agree on which elements are appropriate for the particular relationship and the project at hand. The specifics of each element are flexible, allowing for development of an agreement appropriate to the

1. Erway, Ricky. 2012. *Swatting the Long Tail of Digital Media: A Call for Collaboration*. Dublin, Ohio: OCLC Research. <http://www.oclc.org/research/publications/library/2012/2012-08.pdf>.

type of work and risk level of the material involved. The Client often won't know the nature of the content on the media or what hardware and software was used to create it, so some decisions (such as whether to create a disk image or to transfer files individually, or whether to convert the files to a contemporary format) may best be made after a test batch is processed. A test batch can also be undertaken to test the effectiveness of the planned workflow and documentation specifications.

This document does is not intended to provide legal guidance. The Client should consider these suggestions in the context of local procedures and with appropriate legal advice.

1. General Provisions

A. Identification of parties and statement of roles

For each party (Client and Service Provider), provide

- name and postal address
- name and contact information of primary contact
- names and roles of other individuals involved

B. General description of the work covered by agreement

Include a summary description of the material that will be sent to the Service Provider and the service to be provided (e.g., *The Client will provide the Service Provider with 5½-inch disks. The Service Provider will create disk images of the disks and provide them to the Client via FTP.*)

C. Desired outcome

Describe the desired outcome and how the Client will determine whether the service was provided successfully (e.g., *The goal is for the Client to obtain a copy of the contents of the disks for preservation purposes, and from which access-copies can be derived. The project will be deemed a success if the disk images can be read on x hardware using y software and the files extracted using z software.*)

D. Communication

Indicate the means of communication between the two parties at various stages (e.g., acknowledging receipt of material; confirmation of time the job is likely to take following initial examination; notification of errors or issues).

E. Method for transport of media

- Explicitly state the means of transporting the media between the parties. If a third party will be used, identify which party is responsible for arranging, insuring and paying for transport at each stage of the process.
- Address international shipping issues such as customs documents, fees and possible detrimental effects of screening equipment.

F. Unforeseen complications

Due to the uncertain nature of data recovery, unforeseen complications can arise (e.g., discovery of pre-existing damage to the media or data, or inability to provide the needed hardware or software solution).

- Acknowledge that in dealing with old and/or obsolete media, risk of failure is significant, and the Service Provider will not be penalized for unavoidable loss of data.
- State what action should be taken (e.g., *If complications make completion of the project impossible, after communicating the problem, Service Provider should document what was tried and learned, return the documentation and media to the Client, and bill for time spent (x% of project cost or x dollar amount)*).
- Specify that if the Service Provider needs to consult a third party for assistance they may share general information about the project but cannot share the media or content without permission of the Client.
- Indicate whether subcontracting is allowed (with written permission and amendments to address ramifications related to costs, shipping, and insurance, as well as how security, privacy, and other contractual responsibilities will extend to the subcontractor).
- Specify the method for achieving resolution in the event that either party fails to meet its obligations as set out in the agreement.

G. Liability

Indicate liability for damages, (e.g., *Service Provider accepts liability for the protection of the media and its contents during storing, handling and transfer of the data*). Specify responsibilities for insurance.

H. Glossary

Include a glossary of terms used that may not be known by both parties (e.g., disk image).

2. Information Supplied by the Service Provider

The Client may ask the Service Provider to provide the following information:

A. General information

If there has been a call for proposals or other bidding process, this information will likely have been exchanged in advance of the agreement phase.

- documentation describing both the services they can provide and services they cannot provide
- schedule of charges
- indicative turnaround times for the various services that they offer
- list of previous clients and evidence of past successes/failures in similar projects
- summary of appropriate approaches for dealing with archival/historical material, including details of processes that are undertaken
- statement about compliance with data protection legislation

B. Confirmation that the project data will be handled in accordance with the Client's requirements and/or legal requirements.

The Service Provider should explicitly state:

- who will have access to the data while in custody of the Service Provider
- how physical media will be securely stored and handled during the course of processing
- how any intermediate files or data copied or created during data recovery (e.g., migration formats) will be securely deleted from hardware and software, describing the means, what software will be used, and how destruction will be confirmed
- how any intermediate media used in data recovery will be securely destroyed, describing the means (e.g., shredding or burning)
- similar assurance of secure destruction of source media, if requested by Client
- policy for handling any recovered data that may be illegal or obscene (Service Provider should clearly state what they will do in such cases and how it will be communicated to the Client)

C. Statement of business terms and conditions

3. Information Supplied by the Client

A. Data protection and security

- Indicate whether personal, sensitive or confidential data is likely to be present. Define it, state the required procedures and refer to relevant documentation (e.g., the UK Data Protection Act²).
- Provide confirmation that the Client has the right to process such data and use an external third party to do so.
- Include a statement of intellectual property rights confirming the Client owns the rights or has permission from the rights-holder(s) to copy/extract the data for preservation purposes, or provide other rights status information.

B. Overview of content

- Describe the data expected to be found on the media or provide a statement that the content is unknown. Indicate whether the Client has “read” or attempted to read the media. (e.g., *The Client has not attempted to read the disks. The content is believed to have been created by a single author using WordPerfect. Some image files may be present.*).

C. Summary of media

Provide the total quantity and types of media (e.g., *132 “3½-inch” disks, 57 “5¼-inch” disks, and 4 CD-ROMs*).

D. Inventory

Itemize each physical media object and include each of the following characteristics that is relevant:

- unique identifier
- type of media (e.g., CD-ROM, 9-inch tape, 8-inch floppy disk)
- label information and identifying features of the item (e.g., *3M label with ‘My documents’ written on it*); include photograph, if helpful
- estimated size of data (or capacity of media if indicated or known)
- known problems with the media or data

2. <http://www.legislation.gov.uk/ukpga/1998/29/contents>.

- for each media object, provide (if known):
 - general description of content
 - original operating system (include version number and update level)
 - software (include version and modifications) used to create the data
 - model of hardware used to record the data
 - filename to be used for disk image
 - file-naming convention for individually transferred files (if different than source filenames)
 - for each file, provide (if known)
 - filenames of source files
 - formats of source files
 - date that data files were originally stored and date they were last accessed
 - filenames for individually transferred files (if different than source filenames)

4. Statement of Work

The parties should agree upon a clear set of requirements regarding the services that the Service Provider is to provide. These may include:

A. Processing

- virus scanning (before and after transfer)
- write blocking
- checksum verification, specifying the hashing algorithm
- file copying, retaining original file hierarchies and filenames
- disk imaging, specifying the disk image format and disk image naming

Note: Since there may be only one opportunity to capture the content from the original media, any subsequent manipulation of files should be from copies of the transferred disk images or files.

- extraction of files from the disk image(s)

- renaming files, and changes to file hierarchy arrangement
- migration to contemporary file format(s)
- any required content normalization (e.g., character set conversion)
- metadata generation (including retention of significant properties from source file)
- destination medium for transferred content (including type and size)

B. Exception handling

- what to do when viruses are detected
- handling of duplicate disks or files
- handling of files that had been intentionally deleted
- handling of hidden files
- what to do in the case of problematic formats (e.g., *In case of an unidentifiable format, provide the bitstream.*)
- handling of blank media
- handling of unreadable media
- special handling required due to value or sensitivity of the data
- handling of unreadable/unrecoverable content
- urgency of recovery and appropriate level of service (e.g., *No more than three hours should be invested in forensic investigation on a problematic disk.*)

C. Documentation

- Service Provider should use the inventory provided by the Client to record the sequence of activities that was performed throughout the course of the process and should include the completed document among the deliverables at the conclusion of the project. For each physical media object and for each file, this may include:
 - appraisal of condition of media upon receipt
 - method of content transfer
 - date each process was completed and by whom it was completed
 - hardware and software used to read and write disk images and/or files

- checksum type and value
- disk image file name
- hierarchy and filenames of individually transferred or extracted files with indication of file format
- whether metadata was successfully extracted and supplied to the Client
- additional notes about the process
- Client should detail their requirements for Service Provider's summary technical documentation, which is to be provided at conclusion of project

D. Delivery

- delivery schedule (e.g., *A test batch of x will be provided to the Client within two weeks of receipt of the media, and all remaining deliverables will be provided to the Client within 4 weeks of acceptance of the test batch.*)
- indication of whether original media will be returned at the end of the process or securely destroyed
- delivery method for transferred data (e.g., *Files and metadata are to be packaged with BagIt and made accessible on Service Provider's FTP site or Files and metadata are to be transferred to Client's hard drive and shipped via DHL Express.*)

E. Acceptance

- method of quality review (e.g., extent of review, turnaround time, hardware, access tools, and other software to be used)
- metrics for determination of acceptability (e.g., disk images created and viewable, migrated files legible in target software with no or minimal artifacts, extant metadata retained)
- action to be taken for unacceptable transfer (rework, turnaround time, financial ramifications,)

F. Completion

After Client has reviewed the deliverables and completed quality control, Service Provider will provide:

- return shipment of original media, if appropriate
- assurance of secure destruction and deletion of project media and data, as well as of any copies made during the project

5. Cost and Payment

- Include an agreed schedule of setup charges, unit charges, and/or hourly rates; consider a unit rate for everything that processes according to plan and an hourly rate for problem solving and any unanticipated steps. Establish a maximum hourly rate for problem-solving per disk or batch.
- Explicitly state which party is responsible for costs incurred (e.g., media, FTP hosting, transportation).
- Attach billing and payment instructions (e.g., *After the Client accepts the work, the Service Provider will invoice for the full amount, to be due within 30 days of invoice date.*)