Security Through Obscurity No Longer

Tim Pyatt
Eberly Family Special Collections Library
Penn State University
MARAC Oct. 26, 2012
Security through obscurity

Sensitive information could be hidden anywhere in the files, but no fast way to find it. Must sort through the files one by one.
Sensitive data concerns reported in a 2012 survey of ARL special collections

Below are ethical/appraisal issues that may be encountered while managing born digital materials. Please indicate which of these issues are addressed by your library’s ingest policies or procedures. Check all that apply.

- Whether to retain (under restriction) or destroy personally identifiable information (PII)
- Whether to make files available for research use without having screened them for PII on the file level
- Whether to retain or destroy file fragments and deleted content in the absence of explicit guidance in the donor agreement
- Whether to preserve log files, preferences, browser caches and other types of ambient data in the absence of explicit guidance in the donor agreement
- Whether to preserve e-books, software, digital music, and other copyrighted content
- Other issue, please describe
What are the privacy concerns with a hard drive or other e-media received from a donor?

- Personally identifiable information (pii) such as SSNs and credit card numbers – liability for both the donor and the repository
- Hidden and deleted files
E-records language for deeds of gift

Deed of Gift Addenda for collections with electronic records

The Donor acknowledges that the Library acquires the materials with the intent of making them available for an ongoing or indefinite period of time. In order to accomplish this, the Library may need to transfer some or all of these materials from the original media as supplied by the donor to new forms of media to ensure their ongoing availability and preservation. The donor grants the library rights to make preservation and access copies of materials in the collection and to make those copies available for use.

The Library may contract with university staff or outside contractors to store, evaluate, manage and or analyze materials in the collection. Any such arrangements must abide by the terms of this agreement.

Does the Library, or contractor as outlined above, have permission to crack passwords or encryption systems, if any, to gain access to electronic data received as part of the materials? 
____ Yes  ____ No
If no, such materials will not be retained by the Library.

Does the Library have permission to recover deleted files or file fragments, if any, and provide access to them to researchers?
____ Yes
____ Yes, under the following conditions
____ No

Does the Library have permission to preserve and provide access to log files, system files, and other similar data that document your use of computers or systems, if any are received with the materials?
____ Yes
____ Yes, under the following conditions
____ No

Privacy
The Library will review the materials in the collection in an attempt to identify items that contain private information. Please indicate below your awareness of materials that may sensitive electronic information.
Digital Forensics Workflow

Accession & Store
- Survey materials (number of discs, types, known or estimated volume)
- Create disk images
- Virus scan
- Search for PII

Arrange/Describe
- Consider: Series, depth of description (group, media, top level folders).
- Extent: volume (MB), file count, folder count
- Describe materials (survey contents of files, sampling)
- Access restrictions: special software, local use only, sensitive content

Access
- Researcher use agreement
- Copy requested material to a reading room computer
  - Provide disk image
  - OR Export data
  - OR Create Virtual Machine
FTK (Forensics Tool Kit)
Example of pii scan

Identity Finder Status

Searching: Search Completed in 0 days, 0 hours, 3 minutes, and 48 seconds
Progress: 100% Overall: 01/01
Locations Containing Identity Matches/Total Locations Searched: 128/5329 Total Identities Found: 766

Files: 5325 Messages: 0 Browser Data: 0
Compressed: 4 Attachments: 0 Registry: 0

Social Security: 4 Bank Account: 0 Phone: 31 Passport: 0
Credit Card: 0 Driver License: 0 E-Mail Addr: 731 Maiden Name: 0
Password: 0 Date of Birth: 0 Address: 0 Custom Typer: 0
So what are options for access?

• Locked down computer in reading room
• Authenticated repository where researchers has to agree to terms of use
• Future options – authenticated access to truly virtual reading room with copy protected access to all collections (not there yet)
Rushdie papers at Emory: one approach to managing sensitive e-content

- Locked down computers with emulated environment
- Pre-screened content
- No access outside of reading room
Virtual Reading Room for e-records

Application for the Virtual Reading Room in UCIspace @ the Libraries

Privacy Notification

The State of California Information Practices Act of 1977 (effective July 1, 1977) requires the University to provide the following information to individuals who are asked to supply information:

The principal purpose for requesting the information on this form is for administration of the use of Special Collections and University Archives. University policy authorizes maintenance of this information.

Furnishing all information requested on this form is mandatory. Failure to provide such information will delay or may even prevent use of Special Collections and University Archives. Information furnished on this form may be used by various University departments for identification purposes and will be transmitted to the State and Federal governments if required by law.

The person responsible for maintaining the information contained on this form is the Head, Department of Special Collections and University Archives.

* Indicates required field

<table>
<thead>
<tr>
<th>Requesting access to:</th>
<th>Select a collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>* First name:</td>
<td>Middle initials:</td>
</tr>
<tr>
<td>* Last name:</td>
<td></td>
</tr>
<tr>
<td>* Email address:</td>
<td>(e.g. <a href="mailto:jsmith@duke.edu">jsmith@duke.edu</a>)</td>
</tr>
<tr>
<td>* Phone number:</td>
<td></td>
</tr>
<tr>
<td>* Street address:</td>
<td></td>
</tr>
</tbody>
</table>
Future Directions

Hypatia — a Hydra application — supports accessioning, arrangement, description, discovery, delivery, and long term preservation of born digital archival collections. This instance of Hypatia is a prototype of the open source software, and is meant to demonstrate the potential of managing digital archives holistically, in a repository-powered platform. It includes parts of 14 digital collections developed as part of the AIMS Project ("Born-Digital Collections: An Inter-Institutional Model for Stewardship"), funded by the Andrew W. Mellon Foundation. The full contents of each of these collections are hosted at their respective home institutions.

Featured Collections

Stephen Jay Gould papers

Stephen Jay Gould was a renowned evolutionary biologist, paleontologist, historian of science, educator, popular science author, polymath, and an enthusiastic collector. His papers include, but are not limited to: subject files; coursework and research from his graduate studies at Columbia University; his teaching and administrative files from Harvard University and New York University;
Helpful URLs


Rushdie Archives at Emory: http://marbl.library.emory.edu/innovations/salman-rushdie

UC Irvine Virtual Reading Room: http://special.lib.uci.edu/using/virtual-reading-room-application.html

Hypatia demo: http://hypatia-demo.stanford.edu/
Presentation URL:

https://scholarsphere.psu.edu/files/cn69m429f

Penn State Gift Agreement addenda:

https://scholarsphere.psu.edu/files/0k225b067