DEFINING ESSENCE OF USER-GENERATED CONTENT FOR PRESERVATION

Ayoung Yoon
School of Information and Library Science

Why Preserving UGC?

UGCs now become part of personal digital collections on the Web since they often contain personal memories, activities, thoughts, and even profiles. UGC warrants much attention due to its impact on society as well as the need for a “bottom-up” approach to collection development in archives. Since more individual stories are now reflected on the Web and diffused through the Internet, archiving and preserving them seems a way to preserve human traces of this era.

What is UGC?

- Is publicly available on the Internet
- Reflects a certain amount of creative effort
- Is created outside of professional routines and practices (by OECD)

What to Preserve?

Understanding Creators’ Perspectives

- For archivists, understanding the original meaning and context of content is always important in terms of provenance and collection management.
- For tool developers, before developing preservation tools for repositories and fully adopting these tools in workflow, it is better to understand the creators’ perspectives and reflect or integrate them into the tools.
- For creators, knowing about preservation can be helpful to develop personal digital archives.

Understanding Significant Properties

- Defining the “essential characteristics” or “significant properties” of digital objects is one mechanism for identifying the characteristics that must be preserved to remain meaningful.
- E.g. Content, context, appearance, behavior, and structure of digital objects (InSPECT project)

Research Questions

- What values bloggers give to the self-created content?
- How they define what is important for preserving their blogs?

What is blog?

Blog is a type of personal web page, managed and maintained by individuals. As one of the most popular forms of UGC, majority of blogs are personal journals or diaries that often include personal and private content, while they exist in the public realm.

FINDINGS

Purpose of Preservation: Value of blogs and Expected Audience

<table>
<thead>
<tr>
<th>Property</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Express self, personal memories, family records</td>
<td>4.30</td>
</tr>
<tr>
<td>Useful information to others</td>
<td>3.77</td>
</tr>
<tr>
<td>Culturally and historically significant</td>
<td>3.23</td>
</tr>
</tbody>
</table>

- Other values include Entertainment value and sharing value.
- Family (68.8%), Friends (81.6%), and General Public (67%) are the most largest expected audience group.
- 78.8% of respondents are interested in “making their blog posts available for future use” since “they are historical records for future generations... and it captures sort of the views and ideas of the populations and related to a particular time.” (U06)

Composition for Preservation: What is important to preserve?

<table>
<thead>
<tr>
<th>Preserve</th>
<th>Preserve SOME Content &amp; Context</th>
<th>Preserve SOME Structure</th>
<th>NOT Preserve Appearance</th>
<th>NOT Preserve Behavior (hyperlink, updating information)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Photo that I owned and posted</td>
<td>4.54</td>
<td>3.72</td>
<td>2.84</td>
<td></td>
</tr>
<tr>
<td>2. Text that I wrote</td>
<td>4.53</td>
<td>3.60</td>
<td>2.84</td>
<td></td>
</tr>
<tr>
<td>3. Video that I owned and posted</td>
<td>3.92</td>
<td>3.32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>