Collecting Emerging Formats

Giulia Carla Rossi
Curator, Digital Publications
(She/her)
DPC:WAPWG – 16 January 2020

@giugimonogatari
Collecting Approach to Complex Digital Publications

Outline:

• What we collect
• Why we collect it
• How we collect it
1. WHAT
The Emerging Formats Project

“The term 'emerging formats' refers to types of publication that are in scope to collect under the UK’s Non-Print Legal Deposit Regulations, but whose content and structure are more challenging compared to those currently collected.”

(https://www.bl.uk/projects/emerging-formats)

Two year project (April 2017 – March 2019) to:

• Identify publications that are in scope to collect
• Identify and explore the collection management needs of these more-complex digital publications

1. WHAT
What do we mean by ‘emerging formats’?

- Born-digital with no print counterpart
- Often consist of more than one media type
- Device dependent
- Comprised of non-standard format and metadata types
- Not typically part of existing collections
- At risk of obsolescence

Focused on three types of content:

- Web-based interactive narratives
- Book as mobile apps
- Structured data
CHAPTER 2

Christ I'm careless.

http://editionsatplay.withgoogle.com
2. WHY
UK Non-Print Legal Deposit (NPLD)

- In April 2013 the Regulations come into force
- UK Legal Deposit Libraries have been collecting various types of born-digital publications since 2013:
  - eBooks
  - eJournals
  - archived websites
  - geospatial datasets
  - digital sheet music

Benefits for libraries:

- NPLD helps mitigate the possibility of a digital black hole
- Ensure we represent the changing nature and diversity of the UK digital culture
- Develop capability that can be applied to more ‘mainstream’ collections
- Share knowledge and experience with other cultural institutions to address related challenges
Benefits for researchers:

• Ensure long-term preservation of complex digital publications
• Provide a collection of original and innovative publications
• Support access to emerging formats
Benefits for publishers and creators:

- Ensure their works can be collected and made accessible over the longer term
- Provide evidence of past uses of technologies
- Provide examples of creative and technical innovation
3. HOW
How to collect complex digital objects in a way that ensures preservation and access over the long-term?

At the acquisition stage we have to consider:

- users expectations and needs (to determine what)
- technical complexities and dependencies (to determine how)

**GOAL:** maximise long-term reusability & provide a meaningful access experience
We collect:

- Selectively:
  - High cost
  - Low level of standardisation
- Under Non-Print Legal Deposit regulations
- Following prioritisation criteria agreed upon by all LDLs:
  - Risk of loss identified
  - Engagement from creators and publishers
  - Existing capability that we can build on
# NPLD collection methods

<table>
<thead>
<tr>
<th>Collection method</th>
<th>Offline publications</th>
<th>Browser-based publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>File transfer</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Download via access code or password</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Web harvesting tools</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Description as a collection method</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

3. **HOW**
Challenges…

*Identify:*
- Horizon scanning
- Geographical scope
- Authorship and copyright definition
- Digital publishing environment is in continual change

*Collect:*
- DRM
- Multiple versions
- Partial capture
- Lack of standards for how content is produced

*Preserve:*
- Technically and structurally complex publications
- Strong hardware and software dependencies
…and more challenges

Describe:
- Lack of standards for how metadata is produced
- Publications consist of multiple media types

Access:
- Access to be provided within Legal Deposit Regulations
- Strong hardware and software dependencies for access
- Hardware collection and maintenance
- Re-use

The challenges presented by these publications are not unique to Legal Deposit Libraries, but cultural organisations worldwide
What we collected

• Interactive Narratives collection on the UK Web Archive

• Inkle’s *80 Days* mobile app (Android app & PC-compatible version of the app)

• Contextual information


80 Days © inkle

Breathe © Kate Pullinger, Editions at Play. Photo by Giulia Carla Rossi
Next steps

- Create new collections on the UK Web Archive + keep adding to existing collections
- Further research into discovery and description and developing a standard for catalogue records
- Providing access to offline emerging formats
- Identify new formats to consider for collection and related challenges (social media, chatbots, XR, AI, fanfictions, etc.)
Thank you!

💻 giulia.rossi@bl.uk
@giugimonogatari

Emerging Formats webpage:
https://www.bl.uk/projects/emerging-formats

Image by Giulia Carla Rossi