Audiovisual Metadata Platform (AMP) tool for large-scale metadata generation

Amy Rudersdorf, AVPreserve / @avpreserve
Juliet L. Hardesty, Indiana University / @jlhardes
DLF Forum
October 25, 2017
What is AMP?
Where did the idea come from?
Core AMP Team

- Adeel Ahmad - AVPreserve
- Tanya Clement - UT-Austin
- Jon Dunn - IU
- Julie Hardesty - IU
- Chris Lacinak - AVPreserve
- Amy Rudersdorf - AVPreserve
Mellon-funded workshop

3 Days
16 People
1 Tech Platform Plan
Workshop Participants

- Kristian Allen, UCLA Library
- Jon Cameron, IU Libraries
- Maria Esteva, Texas Advanced Computing Center, UT at Austin
- Mike Giarlo, Stanford University Libraries
- Brian McFee, Music & Audio Research Laboratory, NYU
- Scott Rife, Packard Campus for Audio Visual Conservation, Library of Congress
- Sadie Roosa, WGBH Media Library & Archives
- Felix Saurbier, TIB/German National Library of Science & Technology
- Brian Wheeler, IU Libraries
- Maria Whitaker, IU Libraries
Business Requirements

- Automate analysis of av content & human-generated metadata
- Provide an intuitive interface
- Offer a variety of metadata outputs
- Build a community of developers
- Offer easy-to-use APIs
- Process multi-TB batches of content at a time
- Leverage multiple, integrated, best-of-breed software tools in a single workflow
### Technical Requirements

- Data Model (42)
- APIs (42)
- Open source (41)
- Scalable (41)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queueing</td>
<td>38</td>
</tr>
<tr>
<td>Data output</td>
<td>38</td>
</tr>
<tr>
<td>Access to metadata</td>
<td>36</td>
</tr>
<tr>
<td>Authentication</td>
<td>36</td>
</tr>
<tr>
<td>Versioning</td>
<td>36</td>
</tr>
<tr>
<td>Modularity</td>
<td>35</td>
</tr>
<tr>
<td>SEO</td>
<td>33</td>
</tr>
<tr>
<td>GUI/User Interface</td>
<td>32</td>
</tr>
<tr>
<td>Ease of configuration</td>
<td>31</td>
</tr>
<tr>
<td>Multi-tenancy</td>
<td>30</td>
</tr>
</tbody>
</table>
Non-technical Requirements

- Clear service/cost model
- Wide adoption
- Robust documentation
- Active dev community

Also:

- Proven stability of the product
- Open licensing
- Established governance model
- Robust outreach & marketing program
- Training available
- Semantic versioning
- Support for internationalization
Metadata Generation Mechanisms (MGMs)

Identified systems that could perform NLP, content matching, facial recognition, OCR from video, music/speech detection, and more
Metadata generation scenarios

Given actual media and asked to map out how metadata would be generated from media and related documentation ingest / creation to export.
Next Steps

- White paper about workshop outcomes coming soon
- Grant application to build AMP
Thanks!

Amy Rudersdorf
amy@avpreserve.com
@avpreserve / @rudeamy

Julie Hardesty
jlhardes@iu.edu
@jlhardes