

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



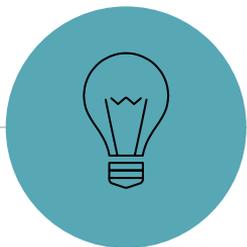
THE
ANDREW W.
MELLON
FOUNDATION



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)

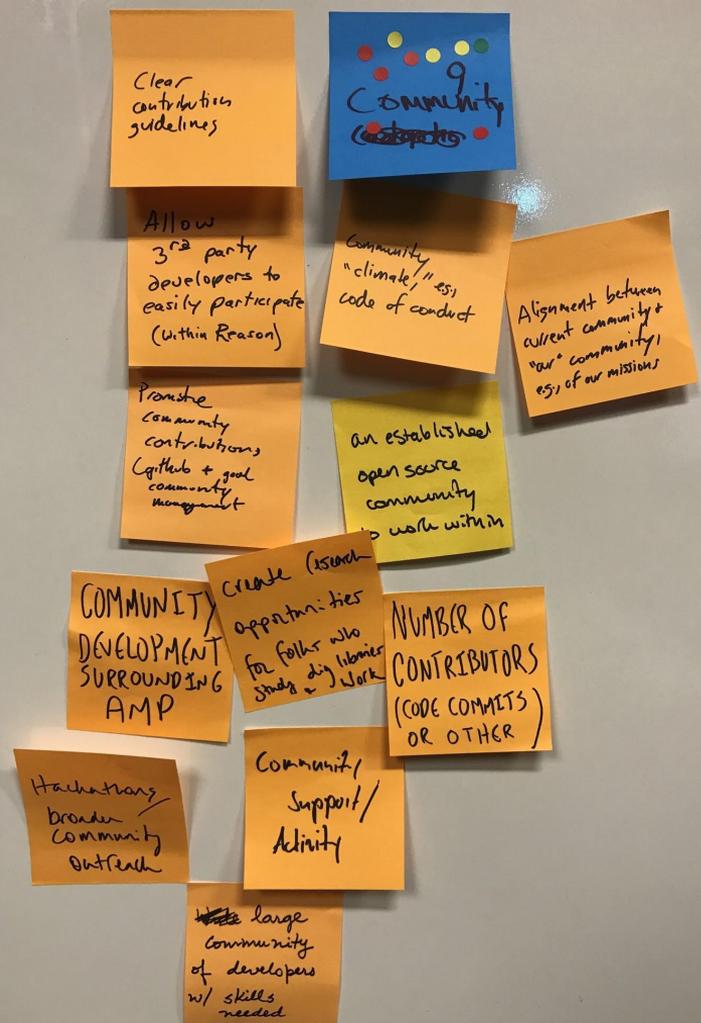
Queueing	38
Data output	38
Access to metadata	36
Authentication	36
Versioning	36
Modularity	35
SEO	33
GUI/User Interface	32
Ease of configuration	31
Multi-tenancy	30

Non-technical Requirements

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

Also:

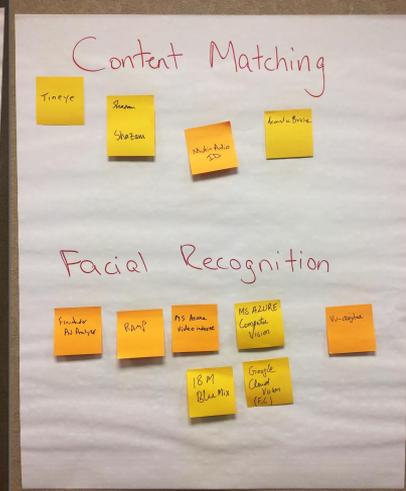
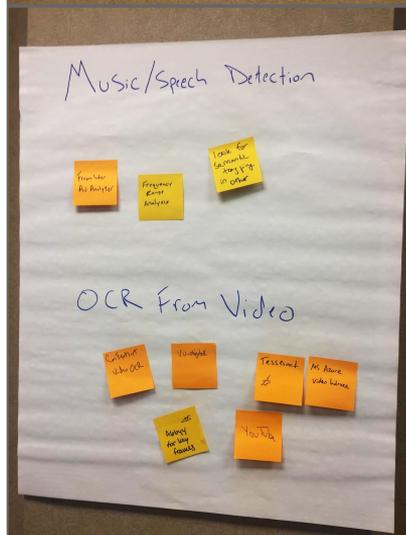
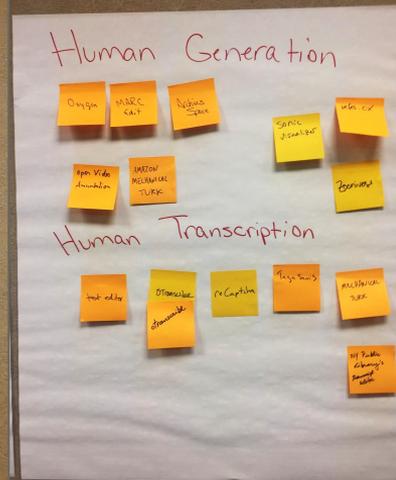
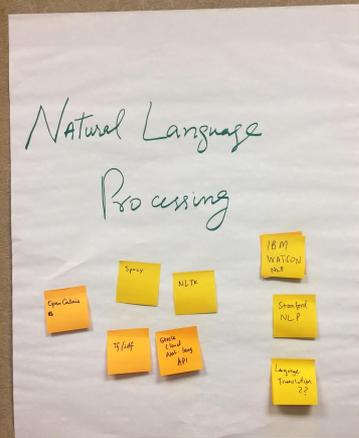
- Proven stability of the product
- Open licensing
- Established governance model
- Robust outreach & marketing program
- Training available
- Semantic versioning
- Support for internationalization





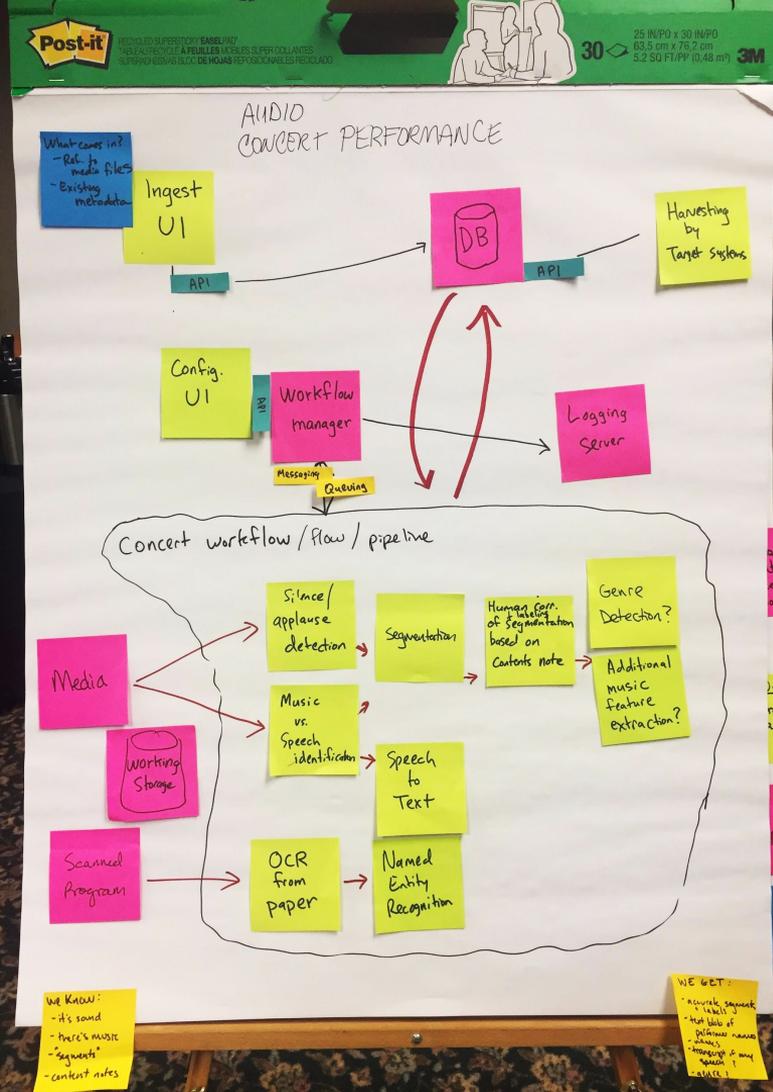
Metadata Generation Mechanisms (MGMs)

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

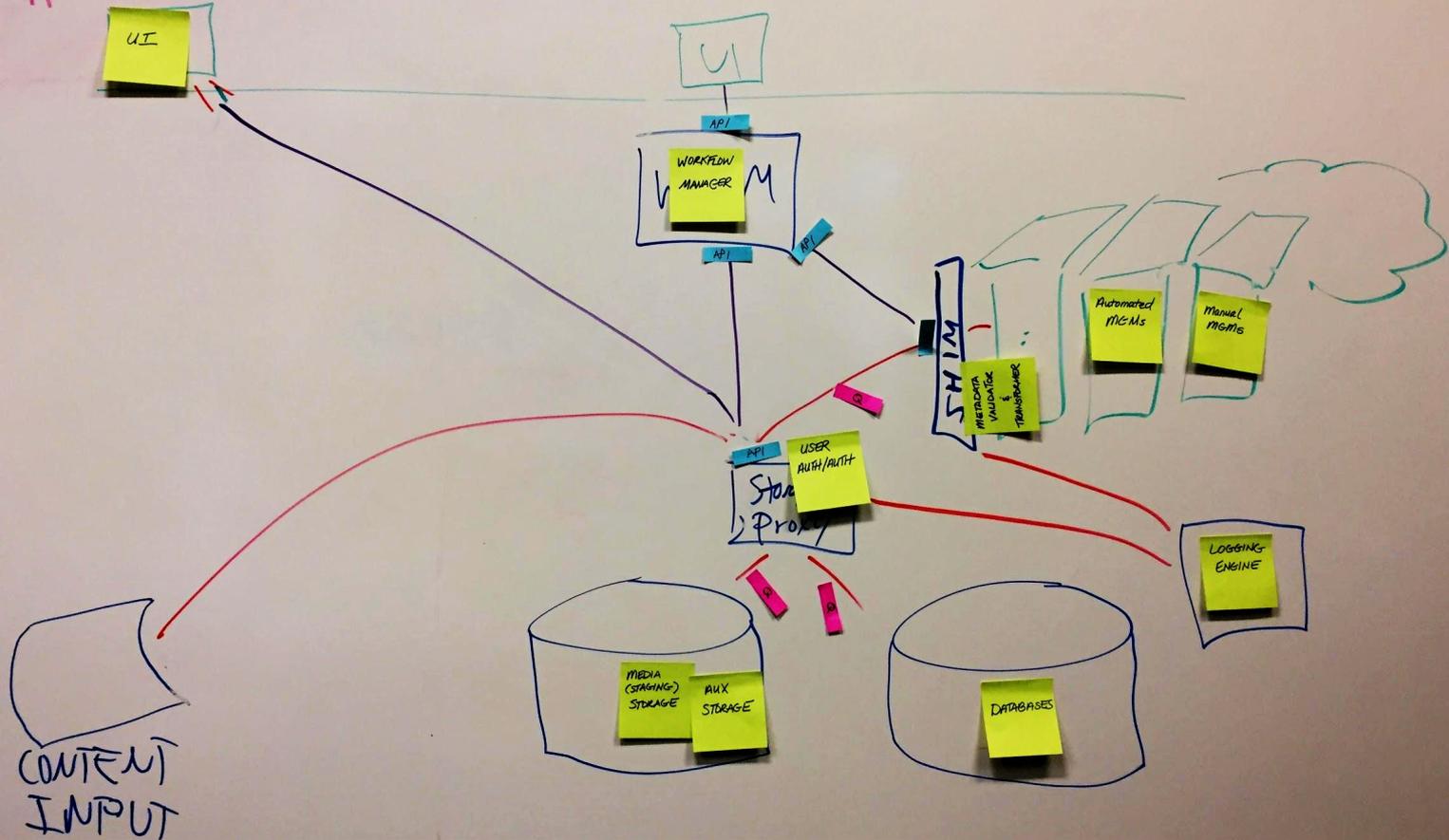


Metadata generation scenarios

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



COMPONENTS





Thanks!

Amy Rudersdorf

amy@avpreserve.com

@avpreserve / @rudeamy

Julie Hardesty

jlhardes@iu.edu

@jlhardes