# Overall goal of the Wikidata for DigiPres project

- Describe software

- Describe emulated environments (== less useful for AV?)

- Describe file formats

# Possible AV Hack Day outcomes:

- Define necessary properties for audiovisual files

- Points to documentation on file formats

- Find out about existing registries for AV formats not covered yet

- (ie. not PRONOM, UDFR, LoC, NARA)

- Must have identifier

- Try to group all audiovisual formats (Katherine can SPAQRL Query / find out what way of linking is best)

- Create a bot to scrape existing software info from the Wikipedias

- Compare / scrape info already on dbpedia

- Connect AV Artifact Atlas if useful

- adding file formats to this list? <https://query.wikidata.org/#SELECT%20%3Ffileformat%20%3FfileformatLabel%20%0A%0AWHERE%20%7B%0A%20%20%3Ffileformat%20wdt%3AP31%20wd%3AQ235557%20.%0A%20%20%0A%20FILTER%20NOT%20EXISTS%20%7B%20%3Ffileformat%20p%3AP1195%20%3Ffileextension%20%7D%0A%20%20%0A%20%20SERVICE%20wikibase%3Alabel%20%7B%0A%09%09bd%3AserviceParam%20wikibase%3Alanguage%20%22en%22%20.%0A%20%20%20%20%0A%7D%0A%20%20%7D>

# General info

1. Google Group: <https://groups.google.com/forum/#!forum/wikidata-for-digital-preservation>
2. Blog Post about Yale University Library’s WikiData Project: <http://openpreservation.org/blog/2016/09/30/wikidata-as-a-digital-preservation-knowledgebase/>
3. Here is a link to the WikiProject about file formats: <https://www.wikidata.org/wiki/Wikidata:WikiProject_Informatics/File_formats>
4. pronom to wikidata property mapping: <https://www.wikidata.org/wiki/Wikidata:WikiProject_Informatics/File_formats/PRONOM>
5. <http://fileformats.archiveteam.org/wiki/Video>
6. this query: <http://tinyurl.com/z7tmmy8> gives us 906 file types and we could just filter out those with audio\* video\* as part of their mime type
7. PRONOM Information Model Doc: <http://www.nationalarchives.gov.uk/aboutapps/fileformat/pdf/pronom_4_info_model.pdf>
8. <https://tools.wmflabs.org/hay/propbrowse/>
9. <https://tools.wmflabs.org/wikidata-game/distributed/>
10. <https://tools.wmflabs.org/mix-n-match/>
11. LoCFDD descriptions as XML: <http://www.digitalpreservation.gov/formats/schemas/fdd/v1/fdd-v1-0.xsd>
12. All the MIME types <http://www.digipres.org/formats/mime-types/>
13. LocFDD: <https://www.wikidata.org/wiki/Wikidata:WikiProject_Informatics/File_formats/LocFDD>
14. PRONOM: <https://www.wikidata.org/wiki/Wikidata:WikiProject_Informatics/File_formats/PRONOM>
15. File format families on wikidata: <https://www.wikidata.org/wiki/Wikidata:WikiProject_Informatics/File_formats/Lists/File_format_families>

Archivematica Format Policy Registry- this is tricky to access unfortunately without an Archivematica installation :/ You can see it on the sandbox though:

<http://sandbox.archivematica.org> login: [demo@example.com](mailto:demo@example.com), password: demodemo

Click on Preservation Planning- this is a user interface to the FPR.

# Properties

Per archive team wiki: *kinds* of video files

FIDO / SIEGFRIED > both use PRONOM

# Comparison of Properties

How much information from <other sources> is useful to copy over / import as statements into a WikiData description?

→ this is hard to answer since it will be deliberated on by the whole group

→ the process is: you propose the property proposals one-by-one to the committee and then they have a week to approve the proposal or not

→ there is no room for prose; the shorter(and less freeform) the property, the better

→ the vision for the LoC Format Sustainability project & WikiData is to write a bot that will scrape the info in XML (KateMurray @ LoC has XML already written) and then the bot could populate WikiData with this info

<gonna do this in sheets - wait for the link: <https://docs.google.com/spreadsheets/d/12cJL58t3-w7gWsUgJ1_KJ-LDUaoxA62s3WbhkE-3e-4/edit#gid=0>

Thank you very much, participants! It was great to meet you and hack with you today! -Katherine

**WikiData User Names**

* [speltzman](https://www.wikidata.org/wiki/User:Speltzman) - Shira Peltzman
* sromkey - Sarah Romkey
* [verwinv](https://www.wikidata.org/wiki/User:Verwinv) - Erwin Verbruggen
* kmbarbera – Kate Barbera
* Jgrazley - Jana Grazley

**Other**

http://www.digitalpreservation.gov/formats/intro/resources.shtml

Schema.org

**Item - Property - Value**