



Coração de Suíno
Técnica: Diafanização

GLAM & Structured Data

Jonathan Morgan,
Wikimedia Research

January 5, 2017

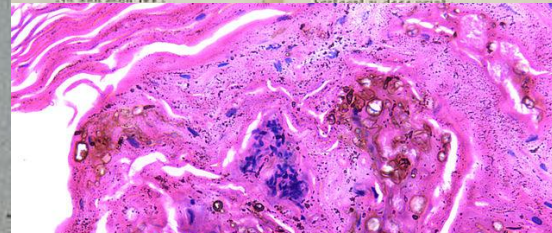
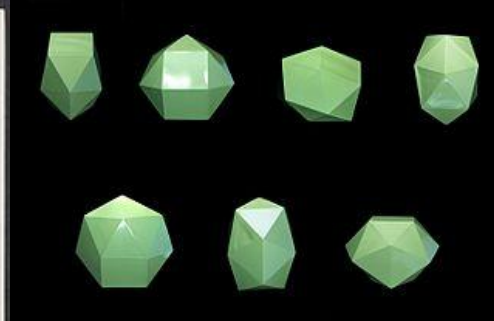
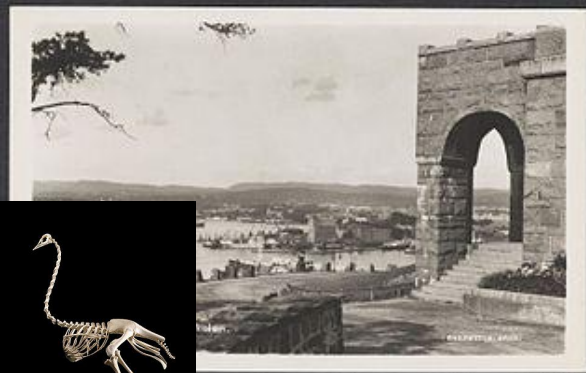


Background: Why GLAM projects?

https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions

GLAM projects are (even) more diverse
than the acronym suggests

Content diversity



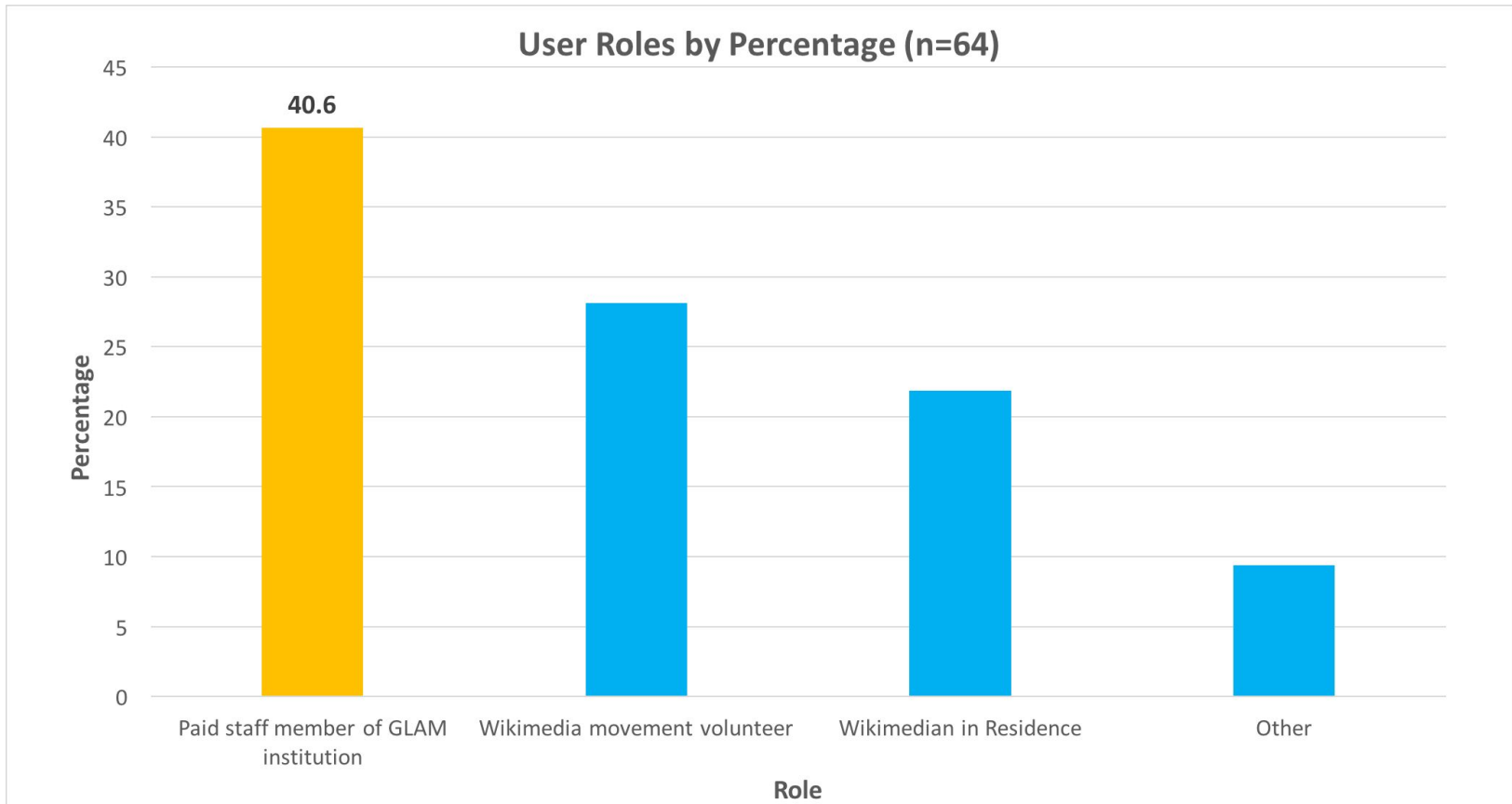
- Process diversity
- Resource diversity
- Metadata diversity

Methods: Mapping diversity

Interviews

- Archive (4), Museum (3), Library (2), Hospital (1)
- Wikimedian in Residence (5), GLAM technical staff (3), Wikimedian (3)
- Europe (6), Latin America (2), Africa (1), Oceania (1), South Asia (1)

Survey



Research themes

Theme #1

Preserving important metadata

Issue #1: Categories

“Categorizing on Commons requires insider knowledge.”

“I don’t have a complete knowledge of how categories in Commons work. I have this big generic collection of media from the museum. I came up with the categories by thinking 'if I were looking for this media, which steps would I take?’”



Category:Glass_plates_by_Baldomer_Gili_i_Roig_at_Museu_d'Art_Jaume Morera



Category:Glass_plates

+

Category:Baldomer_Gili_i_Roig

+

Category:Collections_of_the_Museu
_d'Art_Jaume Morera

||

Category:Glass_plates_by_Baldomer_Gili_i_Roig_at_Museu_d'Art_Jaume Morera

Category:Glass plates

From Wikimedia Commons, the free media repository

This category is NOT for photographs made on glass plates; use [Category:Glass plate positives](#) or [Category:Glass plate negatives](#) as appropriate!

Subcategories

This category has only the following subcategory.

G

- ▶ [Glass plates by Baldomer Gili i Roig at Museu d'Art Jaume Morera \(13 C, 182 F\)](#)

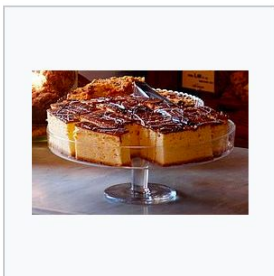
Media in category "Glass plates"

The following 200 files are in this category, out of 369 total.

(previous page) ([next page](#))



"McKinley & Roosevelt"



01 Typische



04520 Christmas Eve



04526 Christmas Eve



1985 HR-typeGlassSara



Buttered toast.jpg

- (cur | prev) ○ 07:24, 9 November 2012 Jmarchn (talk | contribs) m . . (99 bytes) (-2) . . (removed Category:Buildings in Lleida; added Category:Museums in Lleida using HotCat) (undo)
- (cur | prev) ○ 13:49, 24 May 2012 Jordiferrer (talk | contribs) m . . (101 bytes) (+31) . . (added Category:Culture of Lleida using HotCat) (undo)
- (cur | prev) ○ 13:49, 24 May 2012 Jordiferrer (talk | contribs) m . . (70 bytes) (+13) . . (removed Category:Lleida; added Category:Buildings in Lleida using HotCat) (undo)
- (cur | prev) ○ 18:20, 16 May 2012 Kippelboy (talk | contribs) . . (57 bytes) (+57) . . (←Created page with 'Category:Lleida Category:Art museums in Catalonia') (undo)

Issue #2: Templates

“I feel like I'm just guessing at what the best format is.”



Size of this JPG preview of this OGG file: 800 × 450 pixels. Other resolutions: 320 × 180 pixels | 640 × 360 pixels | 1,024 × 576 pixels | 1,280 × 720 pixels | 1,920 × 1,080 pixels.

[Original file](#) (Ogg multiplexed audio/video file, Theora/Vorbis, length 24 min 24 s, 1,920 × 1,080 pixels, 3.53 Mbps overall)

Summary [\[edit \]](#)

Author	VPRO
Title	Guy Consolmagno: Extra footage 1
Date	1 January 1960
Medium	Moving Image
Dimensions	PT24M23S
Current location	VPRO; Nederlands Instituut voor Beeld en Geluid
Accession number	oai:openimages.eu:1026848
Place of creation	Vatican City
Credit line	Guy Consolmagno
Notes	geology; meteorites; cosmology; physics; astronomy; Space exploration; Space; Science; philosophy; Religion; Christianity; Guy Consolmagno explains his fascination for meteorites and talks about combining science and religion; closeups of the Vatican meteorite collection; examples of special kinds of meteorite; a 3D scan of a meteorite

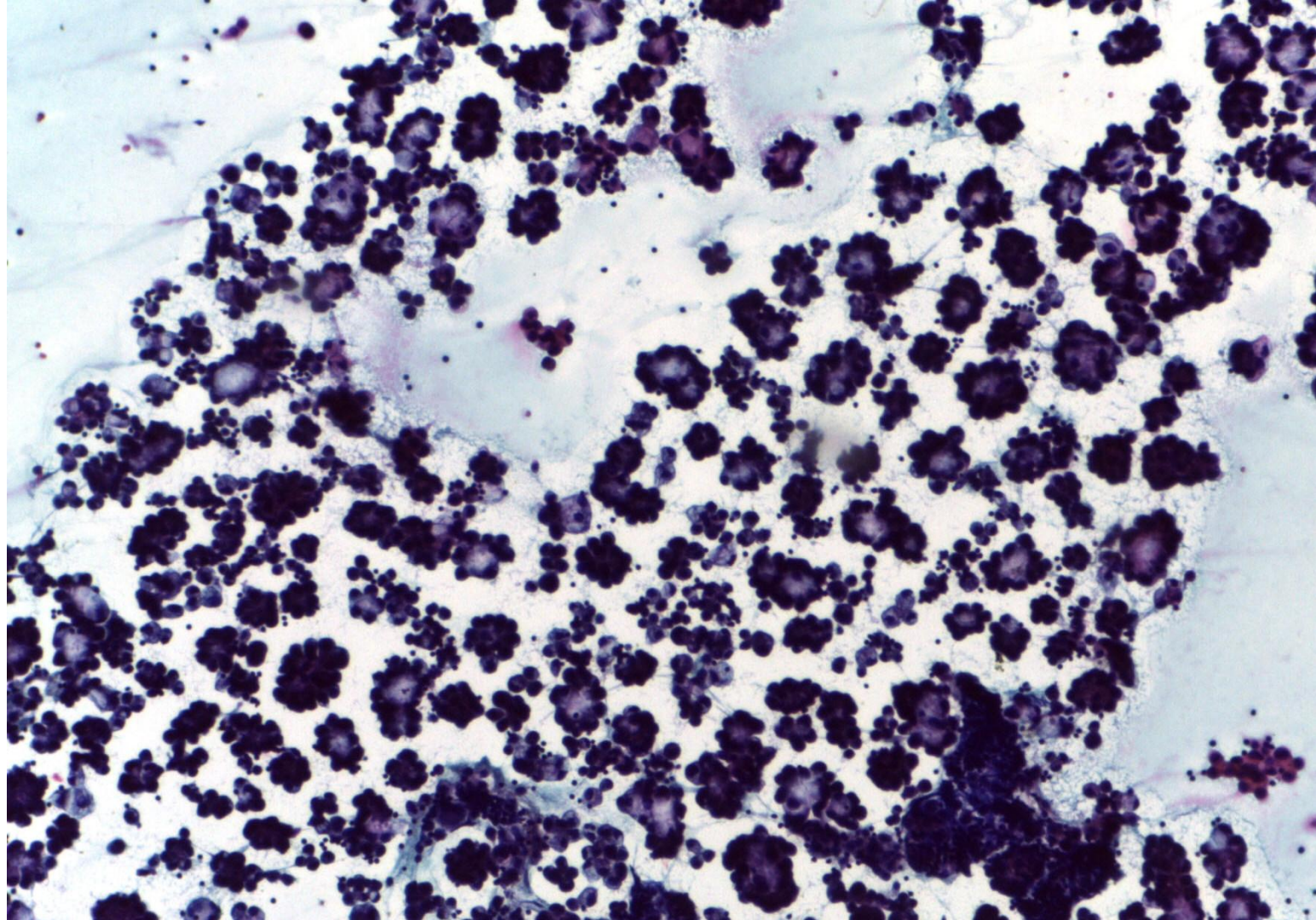


Size of this JPG preview of this OGG file: 800 × 450 pixels. Other resolutions: 320 × 180 pixels | 640 × 360 pixels | 1,024 × 576 pixels | 1,280 × 720 pixels | 1,920 × 1,080 pixels.
[Original file](#) (Ogg multiplexed audio/video file, Theora/Vorbis, length 24 min 24 s, 1,920 × 1,080 pixels, 3.53 Mbps overall)

“I could have created categories for each tag, but would have been entirely manual and taken a huge amount of time. But we wanted to capture it somehow, so that we could theoretically go back later.”

Summary [\[edit \]](#)

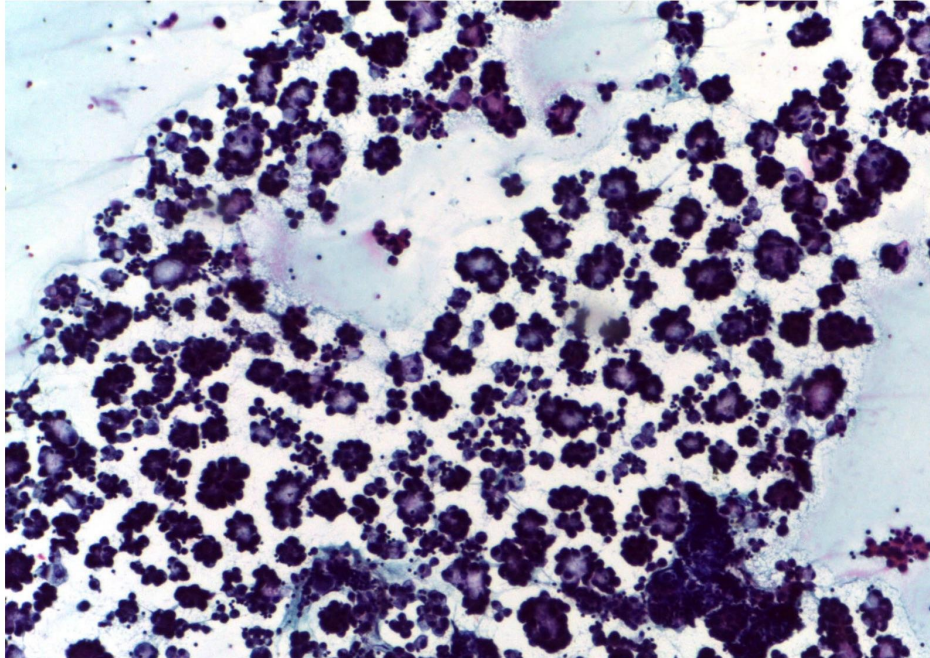
Author	VPRO
Title	Guy Consolmagno: Extra footage 1
Date	1 January 1960
Medium	Moving Image
Dimensions	PT24M23S
Current location	VPRO; Nederlands Instituut voor Beeld en Geluid
Accession number	oai:openimages.eu:1026848
Place of creation	Vatican City
Credit line	Guy Consolmagno
Notes	geology; meteorites; cosmology; physics; astronomy; Space exploration; Space; Science; philosophy; Religion; Christianity; Guy Consolmagno explains his fascination for meteorites and talks about combining science and religion; closeups of the Vatican meteorite collection; examples of special kinds of meteorite; a 3D scan of a meteorite



File:Adenocarcinoma cytocentrifuge preparation 100x.jpg

From Wikimedia Commons, the free media repository

[File](#) [File history](#) [File usage on Commons](#) [Metadata](#)



Size of this preview: 800 × 562 pixels. Other resolutions: 320 × 225 pixels | 640 × 450 pixels | 1,024 × 720 pixels | 1,280 × 900 pixels | 1,765 × 1,241 pixels.
Original file (1,765 × 1,241 pixels, file size: 675 KB, MIME type: image/jpeg); ZoomViewer: [flash/no flash](#)

 [Open in Media Viewer](#)



Download

all sizes



Use this file

on the web



Use this file

on a wiki



Email a link

to this file



Information

about reusing



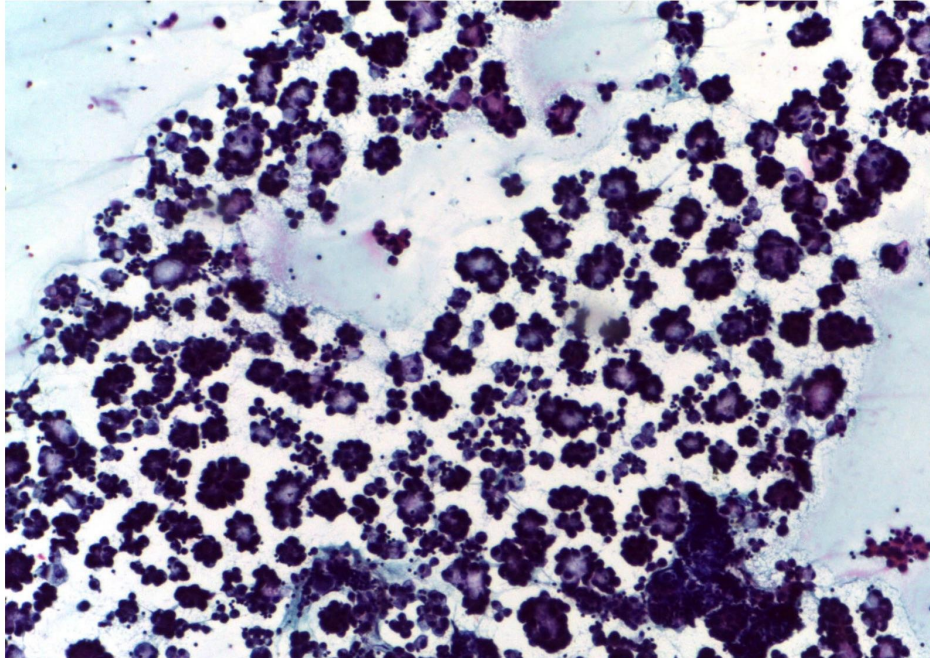
Summary [\[edit \]](#)

Description **English:** Low power view showing malignant cells arranged in glandular pattern. Note the high cellular yield compared to normal cytology fluid preparation (Papanicolaou, 100X)

File:Adenocarcinoma cytocentrifuge preparation 100x.jpg

From Wikimedia Commons, the free media repository

[File](#) [File history](#) [File usage on Commons](#) [Metadata](#)



Size of this preview: 800 × 562 pixels. Other resolutions: 320 × 225 pixels | 640 × 450 pixels | 1,024 × 720 pixels | 1,280 × 900 pixels | 1,765 × 1,241 pixels. Original file (1,765 × 1,241 pixels, file size: 675 KB, MIME type: image/jpeg); ZoomViewer: *flash/no flash*

 [Open in Media Viewer](#) 

-  [Download](#)
all sizes
-  [Use this file](#)
on the web
-  [Use this file](#)
on a wiki
-  [Email a link](#)
to this file
-  [Information](#)
about reusing

Summary [\[edit \]](#)

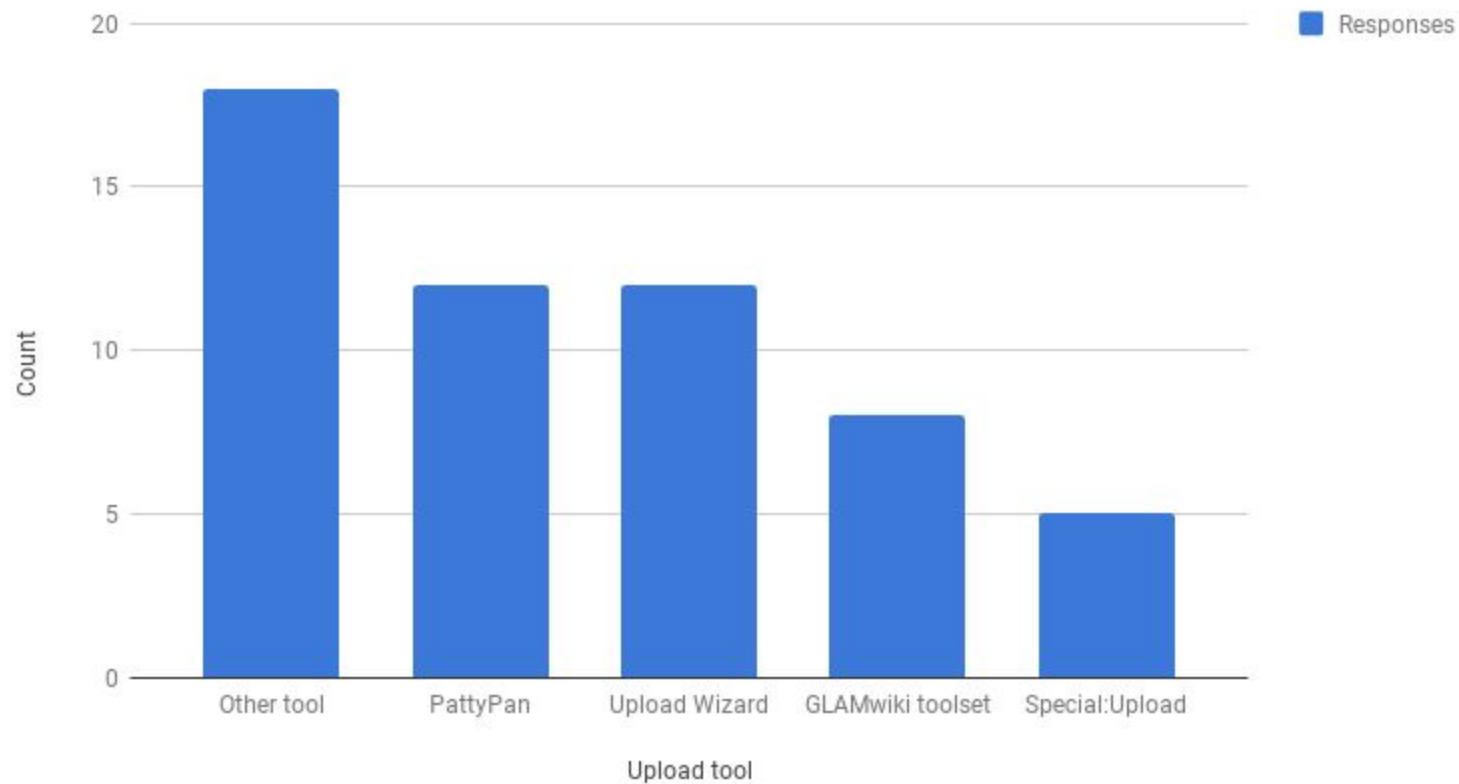
Description **English:** Low power view showing malignant cells arranged in glandular pattern. Note the high cellular yield compared to normal cytology fluid preparation (Papanicolaou, 100X)

Summary

- Categories are challenging for almost everyone. They aren't specific enough, or they are too specific. And they tend to move around.
- Specialized information templates fit metadata better than categories, if you are aware of their existence, and they have the right fields, and you are comfortable editing them.
- GLAMs don't like to leave important metadata on the table; given enough time and expertise, they will figure out a way to add it.

Theme #2:
Upload tools: Functionality and usability

Tool used for upload (n=55)



Issue #1: Complexity & Usability

“We don't have the skills within our organisation to use any of the tools to systematise uploads.”

Issue #2: Metadata mapping

“You can only use the default [Information] template, not artwork, or books, etc.”

Issue #3: Validation and error handling

“[Manually] correcting errors with the media files which hadn't been flagged up in the preview/test viewer. For something like 500+ images that was a little intense!”

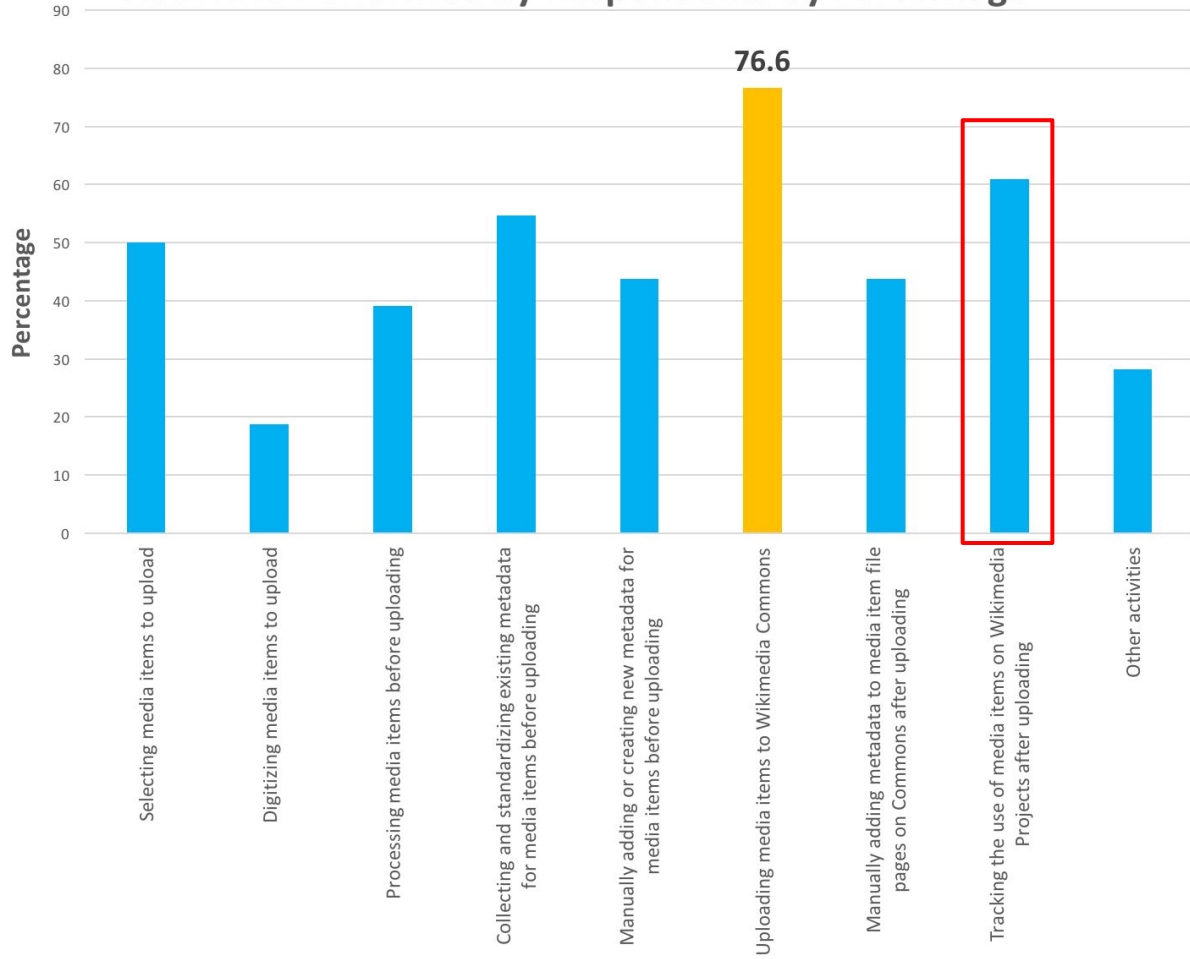
Summary

- There is a rich ecology of batch upload tools out there that meet a variety of needs, but you have to know where to look for them and how to use them.
- There is often a substantial trade-off between usability and sophisticated functionality, there's no 'one size fits all' solution.
- Correcting errors after batch upload is a manual (and painful) process.

Theme #3

Monitoring activity and impact

Activities Performed by Respondents by Percentage



Issue #1: Available metrics

“For videos, we would love to see not just pageviews, but also logging whether people clicked the play button.”

“For museums, getting a share of the potential media views generated by WP is an important reason to share their collections.”

Issue #2: Integration

“We tried to find statistics about views of the uploaded files. It was difficult to see if they actually were used/viewed.”

“I've tried but the existing tools never seemed to work and/or I just couldn't figure out how to make them work! I would LOVE training on this if it's available.”

Category:Picture books from Koninklijke Bibliotheek



From Wikimedia Commons, the free media repository

[Image uptake stats](#)

Subcategories

This category has the following 191 subcategories, out of 191 total.

A

- ▶ [Een aardig prentenboek met leerzame vertellingen](#) (22)
- ▶ [Johns Diorama](#) (11 F)
- ▶ [Het jonge volkje - beelden uit het kinderleven](#) (22 F)
- ▶ [Jongensrampen - schetsen met bijschriften](#) (18 F)
- ▶ [Nieuw prenten boek voor brave kinderen](#) (10 F)
- ▶ [Nieuw-jaars geschenkje voor kinderen](#) (36 F)
- ▶ [De Nieuwe Rijschool](#) (18 F)

Issue #3: Reporting outcomes

“We don't have human resource to track the results.”

“We want to make sure our content being used, so we use GLAMorous. It doesn’t track your data, you type in your category and it gives you a number at that point in time. I used to every month go back and type in the numbers for 30 organizations in the Netherlands that have collections on Commons.”

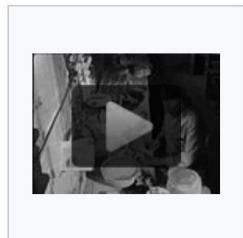
From Wikimedia Commons, the free media repository

Media in category "SVupload122015"

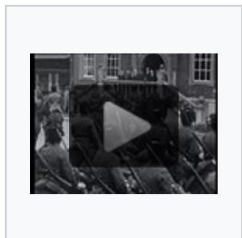


The following 200 files are in this category, out of 392 total.

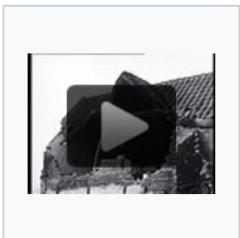
(previous page) ([next page](#))



"Wat Friese gleiers bakken"-516759.ogv
2 min 27 s, 768 × 576;
28.49 MB



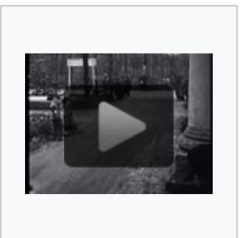
115 jarig bestaan van het regiment huzaren-517769.ogv
1 min 23 s, 768 × 576;
19.4 MB



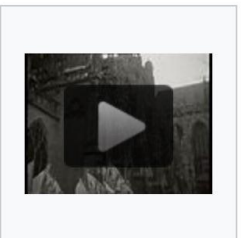
12 huizen afgebrand-522774.ogv
34 s, 768 × 576; 8.8 MB



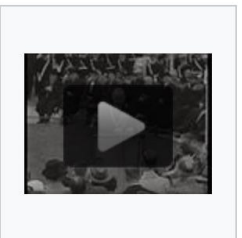
150 reddingsjaren KNZHRM-150 years of rescue by the KNZHRM-525031.ogv
3 min 36 s, 320 × 240;
12.98 MB



24 uur prestatierit van de vrijwillige burgerwacht-524791.ogv
56 s, 768 × 576; 9.87 MB



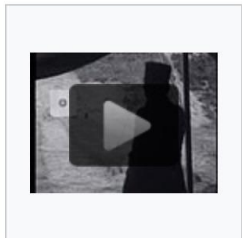
2500 pelgrims maken een boetetocht voor de vrede-518443.ogv
58 s, 768 × 576; 11.34 MB



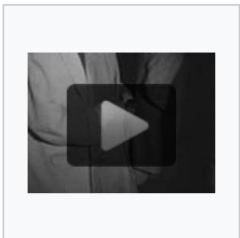
60 jaar Philips-Philips exists 60 years-510328.ogv
2 min 0 s, 768 × 576;
17.88 MB



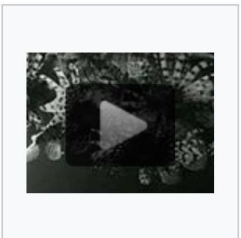
700 jaar stad-700 years a city-512149.ogv



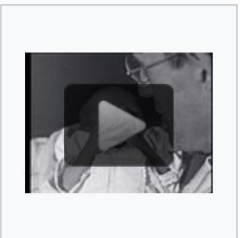
75 jaar bestaan van schietschool-524783.ogv



75 jaar stille omgang in Amsterdam-75 years of



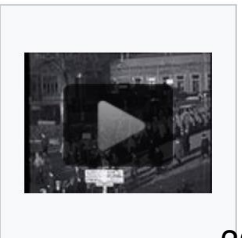
75-jarig bestaan van Artis aquarium-The 75



750.000ste amsterdammer geboren-



Aanbieding van het petitionnement voor een



Aankomst deelnemers aan de

Summary

- Monitoring collection status and measuring impact is very important to GLAMs.
- There are gaps in what current tools can measure.
- Tools are not well-integrated; hard to access metrics about a File or a Collection in context.
- Reporting change over time (what GLAMs want) is not well-supported.

Theme #4

Preparing media for upload

“The work of mapping and data wrangling can only be made easier but not eliminated. The conceptual mappings are an intellectual effort that can’t be rationalised away when you map data from Schema A to B. Data is always messy. Always expect to wrangle. Most Digital Humanist scholars I know say that their projects average 80% time spent on datawrangling.”

Theme #5

Working with Wikimedians

“I wrote to the editor who had requested deletion, asked how we can do it right. He gave me a long list of instructions that I followed very strictly about how to upload properly. Now I’m an expert on OTRS, but I had to contact a lot of people to figure out I was doing it right. It was very frustrating - I was embarrassed in front of the museum director. OTRS were very polite, though.”

“Commons editors have been responsive when I had questions. They were helpful. I was also talking to them on IRC chat. They were very strict on copyright—which I appreciate. It’s kept us focused and sharp. It has sharpened our ideas on open publishing.”

Recommendations

Short term

- **Publish and advertise a single source** for good quality information for high-volume contributors.
- **Improve the default upload experience** to handle a more diverse set of structured (MediaInfo) and semi-structured (template-based?) metadata mappings; provide better mechanisms for previewing and validation.
- **Publish guidance for community tool developers** to help them improve the usability and functionality of their upload and monitoring tools.

Longer term

- Explore options for making impact and activity metrics accessible for non-Wikimedians, especially in a reporting context with multiple metrics.
- Consider ways that SDC and WMF can improve the OTRS process by improving communication and transparency, and increasing throughput.
- Consider metadata that don't fit well into current information templates when prioritizing new MediaInfo properties.

Questions & Discussion

How college students use



Coração de Suíno
Técnica: Diafanização

(like Wik

Thank you!

Special thanks: Niharika Ved and
Sandra Fauconnier



More information

Research report

[https://meta.wikimedia.org/wiki/Research:Supporting Commons contribution by GLAM institutions](https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions)

Image credits

1. https://commons.wikimedia.org/wiki/File:Domestic_pig_heart.jpg Wagner Souza e Silva, CC by SA 4.0
2. [https://commons.wikimedia.org/wiki/File:Mechelen, Belgium, Deventer, 1550.png](https://commons.wikimedia.org/wiki/File:Mechelen,_Belgium,_Deventer,_1550.png) public domain
3. https://commons.wikimedia.org/wiki/File:Chromoblastomycosis_40x.jpg Calicut Medical College, CC by SA 4.0
4. [https://commons.wikimedia.org/wiki/File:BASA-RS-59-1-296-Nikola_Obretenov, 1890.jpg](https://commons.wikimedia.org/wiki/File:BASA-RS-59-1-296-Nikola_Obretenov,_1890.jpg) public domain
5. [https://commons.wikimedia.org/wiki/File:Baldomer Gili Roig. Baldomer Gili Roig vestit de roma%CC%80 \(Roma\), 1900 - 1904.jpg](https://commons.wikimedia.org/wiki/File:Baldomer_Gili_Roig._Baldomer_Gili_Roig_vestit_de_roma%CC%80_(Roma),_1900_-_1904.jpg) public domain
6. https://commons.wikimedia.org/wiki/File:Adenocarcinoma_cytocentrifuge_preparation_100x.jpg Calicut Medical College, CC by SA 4.0

Unused slides

Process diversity

- ✓ How many files do you have?
- ✓ What type are they?
- ✓ What size are they?
- ✓ Can you demonstrate copyright status?
- ✓ Have you decided what upload too to use?
- ✓ How do you plan to track usage of your media?
- ✓ ...

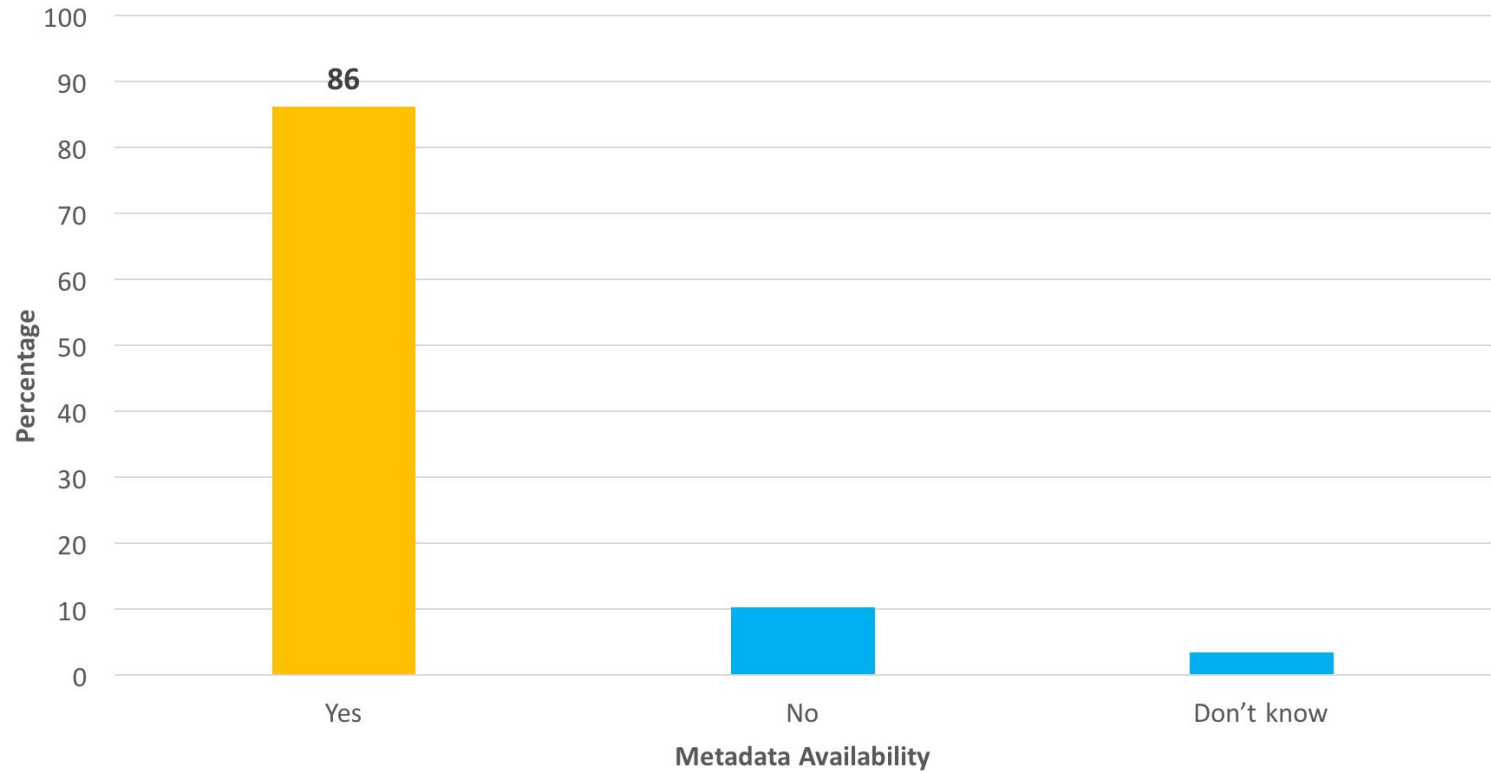
Resource diversity

- ✓ Is your project funded?
- ✓ How many people are involved?
- ✓ Does anyone involved understand copyleft licensing?
- ✓ Does anyone involved know how to code?
- ✓ Has anyone involved edited a wiki before?
- ✓ Is there a local Wikimedia chapter around?
- ✓ ...

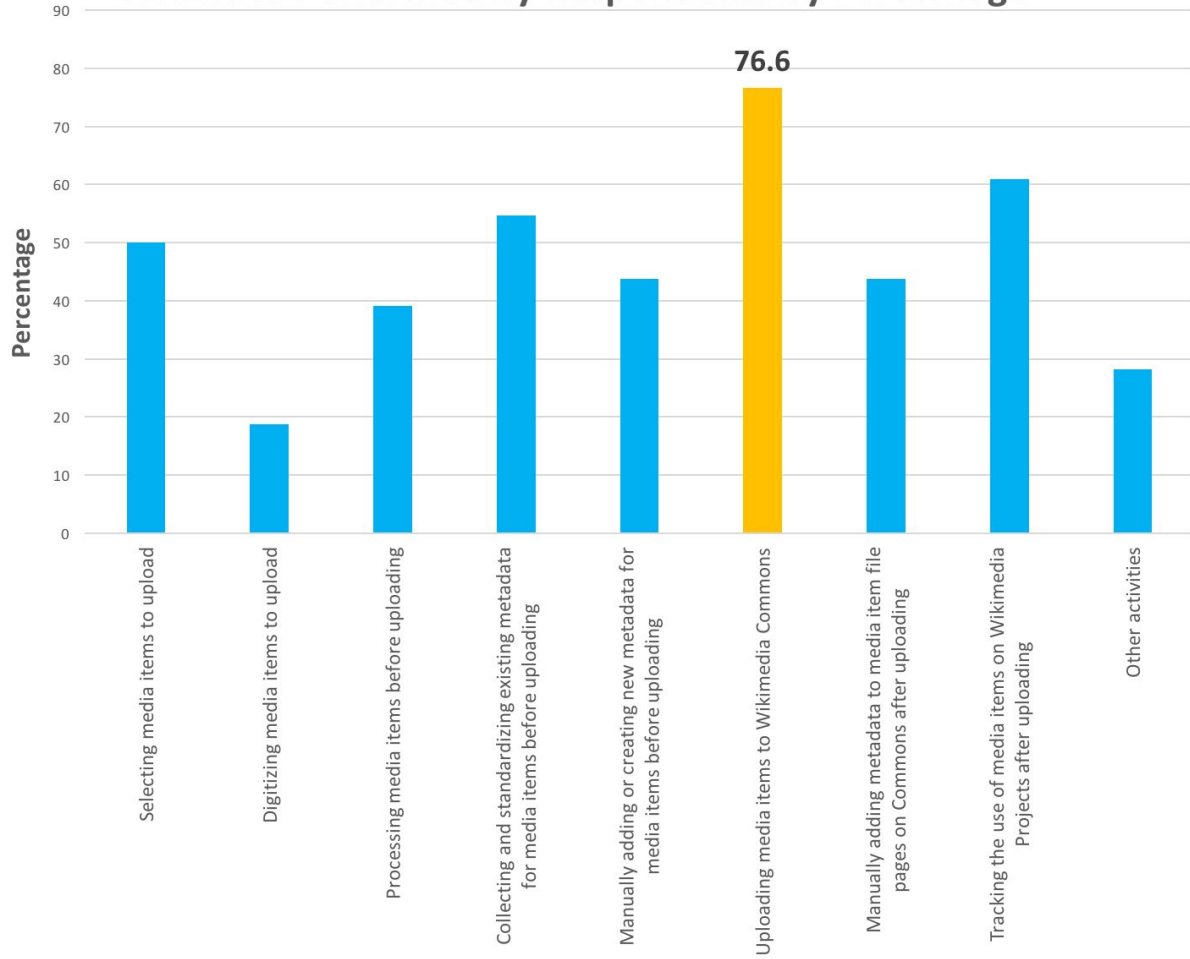
Metadata diversity

- ✓ What metadata is already available in the database?
- ✓ What metadata schema are you using?
- ✓ What metadata needs to be generated?
- ✓ What metadata absolutely must be preserved?
- ✓ How complete is your metadata?
- ✓ How correct is your metadata?
- ✓ ...

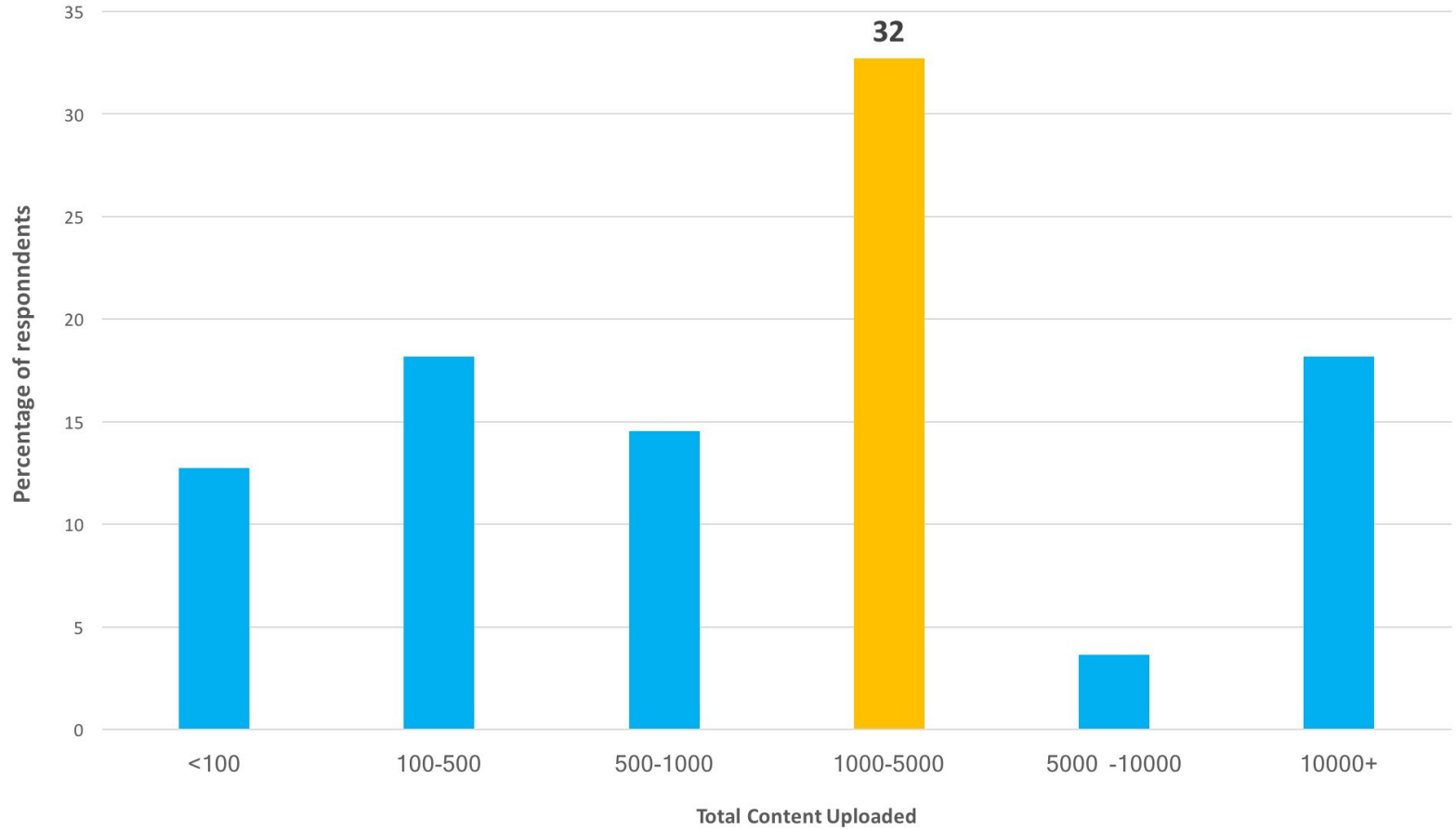
Metadata Available by Percentage (n=58)



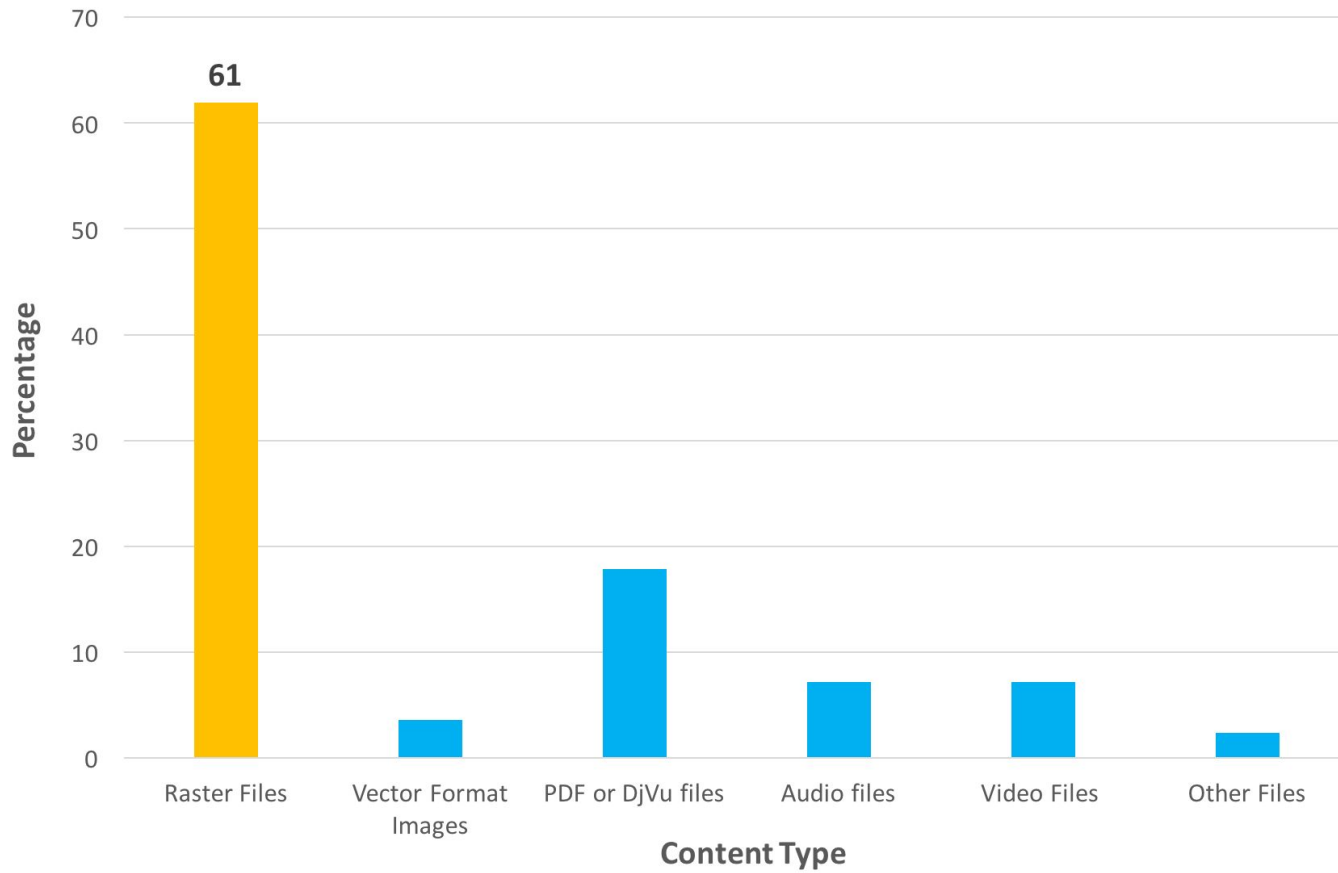
Activities Performed by Respondents by Percentage



Number of Media Items Uploaded by Percentage



Type of upload files - Percentage



“We’re a guest on their platform, so we obey their rules, even if the rules being asserted aren’t documented, or only shared by a few people, or only the priority of the loud/vocal person who’s talking to you. But we’ve found that being transparent is the best long-term solution.”

“We had an incident when a Commons admin flagged all our uploads for deletion, despite us using OTRS. I contacted the admin concerned and the items were not deleted, however if they had been it would have caused our project to be shut down. We worked very hard to build awareness of the value of contributing to Commons within our organisation and this would have killed it.”