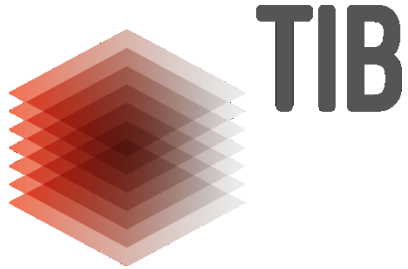
LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK



TIB

What do users expect from image repositories? – Focus group impressions

Lucia Sohmen, Ina Blümel, Lambert Heller
Open Repositories 2019
12.6.2019

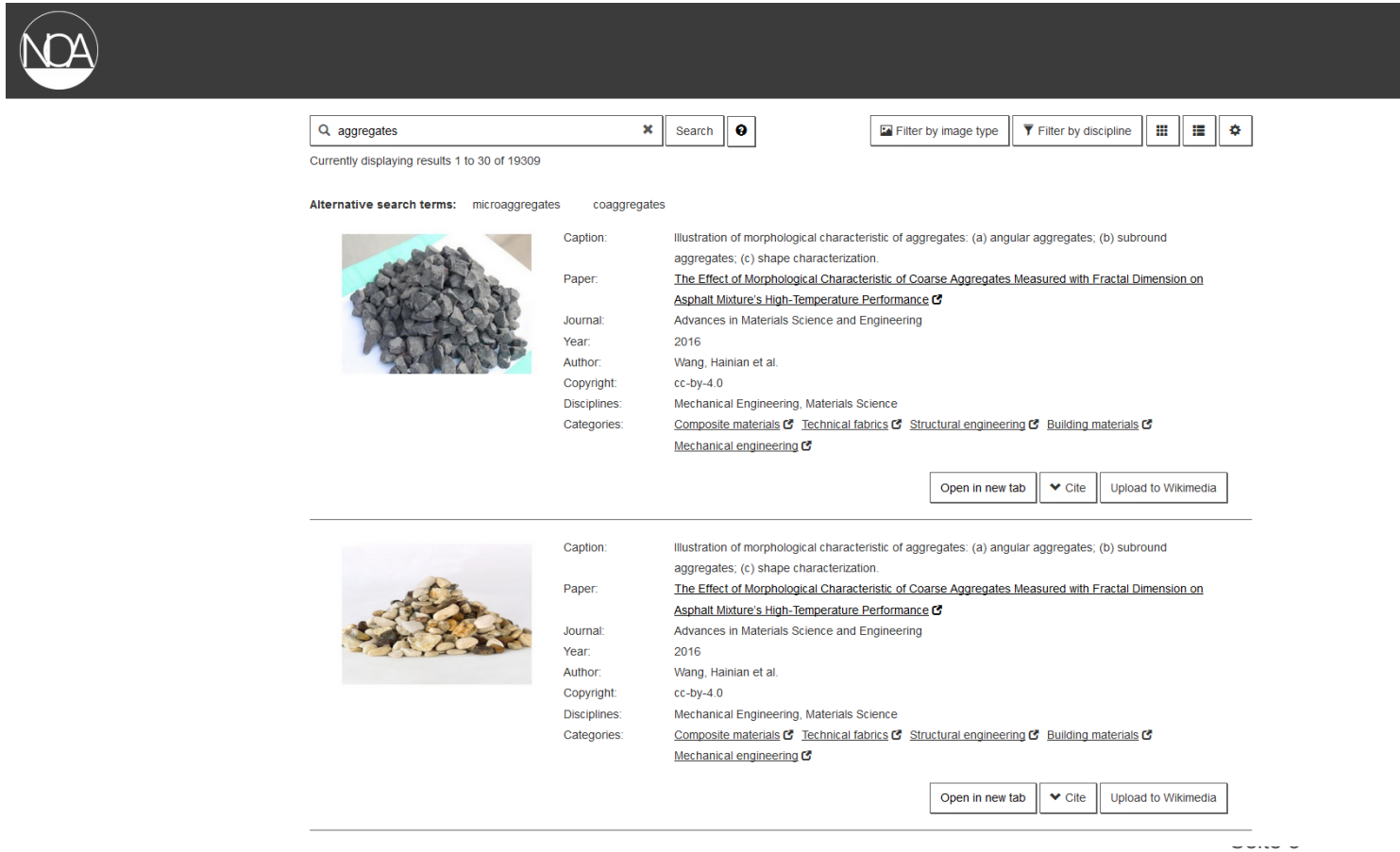
The NOA project

- This project aims to:
  - Extract **images** from **open access articles**
  - Extract and enrich **metadata**
  - Make them **findable** and **reusable** with a search engine
  - Upload to **Wikimedia Commons**

- Our goals are:
  - More **visibility** for scientific figures
  - New ways to **discover knowledge**
  - **Open Science!**

The user interface

- Basic search engine (Solr)
- Can be filtered by subject, image type
- Search by field, boolean operators
- Search for Wikipedia Categories



Evaluation goals

How much value does our service provide to users?

- Target group: Researchers, people who teach classes / produce scientific output
- Is our content **useful** and **relevant**?
- Are the images **easily findable**?
- Do we provide enough **information** about them?
- Are any **important functions** missing?

Focus group setup

- 1 day workshop, **focus group discussions** mixed with presentations about the project

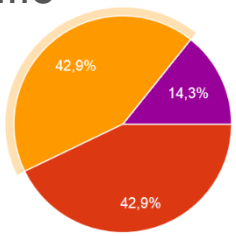
- Participants: **7 researchers**
  - 3 engineers (teaching focus)
  - 2 natural scientists
  - 2 technical writers

- Part one: Survey about **image seeking behavior** (written and discussion afterwards)
- Part two: **Testing** the live systems (written and discussion afterwards)
- Part three: Presentations and discussions

Part one – image seeking behavior of researchers

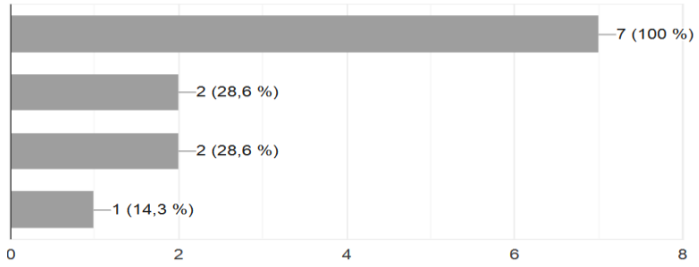
How often do you look for scientific images?

- Never (0)
- Several times a year (3)
- Several times a month (3)
- Other: „More often in winter, because that is when I give a relevant class“ (1)



Where do you reuse the images you find?

Presentations, seminars
Peer-reviewed articles
Non peer-reviewed articles
Other (terminologies)



Why do you look for images?

To reuse them (All participants)

To gain knowledge from the content

>> “Find explanations for unknown terms“

>> “Images sometimes say more than long texts“

>> “I want to see how some things are labelled“

„For terminologies“

„A fast way to find textual information“

„To compare my images to others“

Part one – image seeking behavior of researchers

Name some recent image searches you did. What was your experience?

- „Photo and video technology“
- „None“
- „I don‘t know“
- „Schematics of computer clusters. I couldn‘t find any so I had to make my own for a seminar“
- „I looked for images from my own publications and found them immediately on Google Images“
- „oven lamps, UML-diagrams – I could not find good images of oven lamps without brand labels“
- „saturation absorption spectroscopy - successfull, Band-pass filter - successfull, line pulling effect – not successfull (too specific, but I did find something on optical clocks)“

Part one – image seeking behavior of researchers

What search engines do you use for images?

- Google (all participants)
- DuckDuckGo because of privacy (2)
- Wikimedia Commons (1)
- Search function in Safari (1)

Do you use Wikimedia Commons or Wikipedia to find images?

- No (1)
- Wikimedia Commons (3)
- Wikipedia (6)

Do you know any other image search engines?

- No (3)
- „Europeana, but I never used it“
- Springer Images
- Image boards, material design libraries
- Pixabay

Many participants do not search for images directly but find them in the literature that they are already using

Part one – image seeking behavior of researchers

Which aspects are important to you when choosing an image?

- Content (all participants)
- Image properties like size, color, format (4)
- Title / description (1)
- Context: Where / by whom was it published (2)
- License (4)

Do you prefer searching in English or your native language?

- English (3)
- Native language (4)
- No preference (0)

Do you consider the license at all?

- Depends (2)
- Always (5)

- „Licensing is very inconvenient“
- „It seems like an artificial barrier“
- „Even if I am allowed to reuse something, it takes a long time to find out if and how I can use it“
- „I always cite a source, but I am probably doing it wrong“

Part one – image seeking behavior of researchers

If you could wish for anything regarding image searches, what would it be?

- „Easier way to find images suitable for reuse (Google Images has this functionality but filters out too many images that have the right license)“
- „Better findability in general“
- „Competition for Google Images“
- „Something like Google Scholar or BASE, but for images“
- „A snippet with the source information that I can copy and include in my work“

Part two – evaluation of the live system

The task

- Participants were given as much time as needed to perform searches in the live system
- They were asked to come up with searches relevant to their own work
- They were also free to choose from sample use cases
- 3 Use cases:
  - Search for relevant images until they are satisfied with the results or determine there are no good results
  - Document their experience in a survey
  - No further instructions or explanations of the search engine functionalities

Difficulties during the exercise:

- Some people could not think of any use cases or only a few
- Some started searching in their native language
- Some needed a lot more time than expected

Part two – evaluation of the live system

What did you search for?

[something from their own publications]

arousal valence

Geoid

„damp mashine“

Learning goal taxonomy

Oven lamp

User Stories Mapping

rotating calipers geometry

Saturated absorption spectroscopy

vignetting

Engineers‘ skills

UML diagram

Single Pole Double Throw

Convex hull

Diagrams about reproducibility crisis

Given use cases:

Cable-stayed bridge

Hornets attacking bees

Micrographs

Part two – evaluation of the live system

- Participants could **easily find images** related to the **given use cases** (as expected)
- However, the did **not feel confident judging the quality** of the results because the use cases were not in their expertise
- They were mostly **unsatisfied with the results from their own searches**
- Possible reasons suggested by the participants:
  - **Translation problems** / the concept does not really exist in English
  - Search term was **not explicitly scientific**
  - The **ranking** could hide the right results
- What they liked:
  - Possibility to **cite sources**
  - Being able to access **background information**
- Some noticed subject indexing with **Wikipedia categories** and liked it but wanted **better precision**
- Some noticed filtering by **image type**, liked the functionality but suggested **improvements**

Results

What is your overall impression?

- In general, participants liked the user interface and the functionalities
- **Wish**: More scope and scale (for example images from textbooks, more languages)
- **Wish**: Browser plugin for the search bar
- **Wish**: Being able to cite sources by copy-pasting an automatically generated snippet (implemented now)
- **Wish**: Direct link to images, find them in Google Images
- **Question**: Does the project see itself as competition to Wikimedia Commons, Google Images? If images are uploaded to Wikimedia Commons, why should they use our search engine?

Conclusion

- The **demand for finding and reusing scientific images** is there – not so much for peer-reviewed articles but for other content like presentations
- Users **want generic, abstract schematic images**, so a lot of images from scientific articles are not useful to them
- Users struggle with **licenses** and searching in **different languages**
- Users did not notice many advanced functionalities but liked them when they did -> How realistic is it for casual users to become expert users?

- **Where do we see ourselves going?**

The end

- Publications
  - Sohmen et al (2018). Figures in Scientific Open Access Publications. TPDL 2018, doi.org/10.1007/978-3-030-00066-0_19
  - Charbonnier et al (2018). NOA – A Search Engine for Reusable Scientific Images beyond the Life Sciences. ECIR 2018, doi.org/10.1007/978-3-319-76941-7_78
  - Heller et al (2016). Discovery and efficient reuse of technology pictures using Wikimedia infrastructures. A proposal. doi.org/10.5281/zenodo.51562

- Links:
  - Project description: https://www.tib.eu/de/forschung-entwicklung/projektuebersicht/projektsteckbrief/nachnutzung-von-open-access-abbildungen-noa/
  - Blog: https://blogs.tib.eu/wp/noa/
  - Search engine: https://noa.wp.hs-hannover.de

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

Contact
Lucia Sohmen
0511 762-14271
Lucia.Sohmen@tib.eu

Funded by Deutsche Forschungsgemeinschaft (DFG) – Grant number 315976924