

NOA: Ein Forschungsprojekt trifft Wikimedia

Lucia Sohmen, TIB Hannover

Zusammenfassung

Im NOA-Projekt werden Abbildungen aus wissenschaftlichen Open-Access-Artikeln gesammelt und zu Wikimedia Commons, einer Plattform für frei nachnutzbare Medien, hochgeladen. Gleichzeitig können die Bilder in einer projekteigenen Suchmaschine gefunden werden. In diesem Artikel wird der Uploadprozess erläutert und über Erfahrungen mit den technischen Schnittstellen sowie mit der Wikimedia-Community berichtet. Zum Schluss wird der Impact der hochgeladenen Medien auf Wikimedia Commons evaluiert.

Summary

The NOA project collects images from open access articles and uploads them to Wikimedia Commons, a platform for freely reusable media. The images can also be found by means of a newly developed search engine. This article explains the upload process and gives an account of the experiences with the technical interfaces and the Wikimedia community. The last part evaluates the impact of the uploaded images on Wikimedia Commons.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/2019H4S204-215>

Autorenidentifikation: Sohmen, Lucia: ORCID: 0000-0002-2593-8754

Schlagwörter: Open Access; Bilder; Wikimedia

1. Einleitung

An der Technischen Informationsbibliothek Hannover besteht seit 2016 das Projekt NOA, bei dem Bilder aus Open-Access-Artikeln gesammelt und zu Wikimedia Commons hochgeladen werden. In diesem Artikel werden die Erfahrungen vorgestellt, die in der Interaktion mit Wikimedia Commons gesammelt wurden, sowohl auf technischer Ebene als auch auf der Community-Ebene. Dies soll eine Hilfe für mögliche andere Projekte sein, die etwas Ähnliches vorhaben.

In den folgenden Abschnitten wird die Projektarbeit mit Wikimedia erläutert. In Kapitel 2 werden das Projekt und seine Ziele kurz vorgestellt. In Kapitel 3 werden Wikimedia Commons und Wikidata sowie ihre Relevanz für das Projekt erläutert. Außerdem wird beschrieben, was bei der Arbeit mit diesen Plattformen zu beachten ist. In Kapitel 4 wird der Prozess erklärt, mit dem die Bilder hochgeladen werden. Dabei wird auch darauf eingegangen, welche Lösungsmöglichkeiten es bereits gibt und welche Lösung schließlich im Projekt entwickelt wurde. Kapitel 5 bespricht den Aspekt der Communityarbeit, deren Wichtigkeit und welche Anstrengungen im Projekt unternommen wurden. Abschließend wird in Kapitel 6 der Impact der bereits hochgeladenen Bilder evaluiert. Kapitel 7 bietet die Zusammenfassung und einen Ausblick auf die zukünftige Arbeit.

2. Vorstellung des Projekts

NOA ist ein Projekt¹, das von 2016 bis 2019 von der DFG gefördert wird. Der Name steht für „Nachnutzung von Open-Access-Abbildungen“. Das Prinzip des Projekts ist, Abbildungen aus wissenschaftlichen Artikeln mit freier Lizenz zu extrahieren, um diese besser durchsuchbar zu machen und damit die Nachnutzung zu fördern. Sie werden der Öffentlichkeit in einer projekteigenen Suchmaschine² zugänglich gemacht. Dort werden Daten aus den Artikeln sowie automatisch erzeugte inhaltsbeschreibende Metadaten angezeigt. Langfristig sollen die Bilder auch in die Wikimediaprojekte integriert werden, um nachhaltig einem größeren Publikum zur Verfügung zu stehen. Das endgültige Ziel ist es, wissenschaftliche Inhalte untereinander zu verknüpfen und offener zu machen.

Es wurden bisher ca. eine Million Artikel und mehrere Millionen Bilder erschlossen. Die Quellen dafür sind Verlage, die große Mengen an Artikeln mit Lizenzen veröffentlichen, die eine uneingeschränkte Nachnutzung erlauben. Aus den Artikeln werden neben den Bildern die bibliografischen Daten (Autor, Titel, Zeitschrift, etc.) extrahiert. Hinzu kommen die Textstellen, die sich auf die Bilder beziehen und die Bildunterschrift. Anschließend werden mit automatischen Methoden die Bildtypen (zum Beispiel Diagramm oder Foto) ermittelt. Inhaltlich werden die Bilder anhand der Bildunterschrift und der umgebenden Textstellen auf Wikipedia bezogen erschlossen. Jedes Bild wird dabei bis zu fünf Wikipediakategorien zugeordnet.

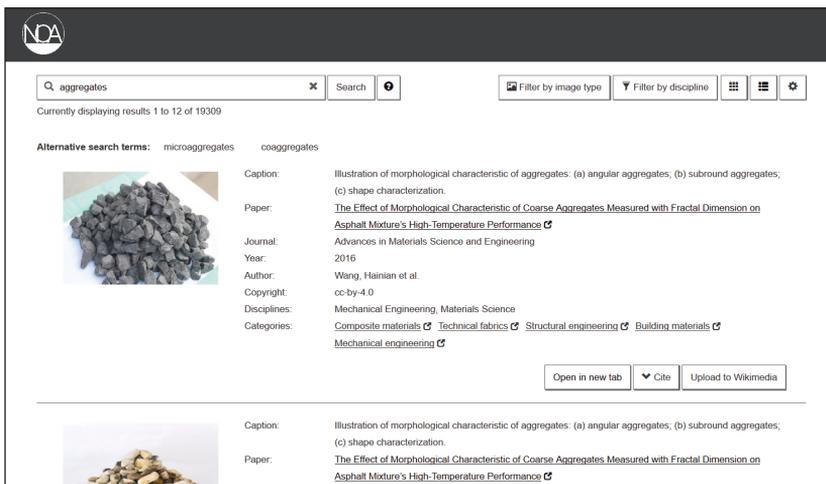


Abb. 1: Darstellung der Bilder in der NOA-Suchmaschine. Bilderquelle: Wang, Hainian et al: *The Effect of Morphological Characteristic of Coarse Aggregates Measured with Fractal Dimension on Asphalt Mixture's High-Temperature Performance*, in: *Advances in Materials Science and Engineering*, 2016, <<http://dx.doi.org/10.1155/2016/6264317>>.

- 1 Nachnutzung von Open-Access-Abbildungen (NOA) – Technische Informationsbibliothek (TIB), <<https://www.tib.eu/de/forschung-entwicklung/projektuebersicht/projektsteckbrief/nachnutzung-von-open-access-abbildungen-noa/>>, Stand: 25.06.2019.
- 2 NOA – Scientific Image Search, <<http://noa.wp.hs-hannover.de/>>, Stand: 25.06.2019.

3. Wikimedia Commons, Wikidata und die Ausgangslage

Abgesehen von der allseits bekannten Wikipedia werden von der Wikimedia Foundation einige andere Projekte betrieben, die verschiedene Inhalte bereitstellen und teilweise untereinander verknüpft sind. Zwei dieser Projekte sind im Projekt besonders relevant: Wikimedia Commons, eine Plattform für Medien, und Wikidata, eine Plattform für Daten. Beide sollen hier kurz vorgestellt werden. Wikimedia Commons ermöglicht Nutzer/innen, frei lizenzierte Medien hochzuladen und zu veröffentlichen. Das sind größtenteils Bilder. Hier sollen auch die Bilder aus NOA hochgeladen werden. Nutzer/innen werden dazu angehalten, die Bilder mit verschiedenen Metadaten anzureichern. Dabei ist es ihnen freigestellt, wie sie die Felder benennen, in welchem Format sie sie befüllen und wie viele Felder sie verwenden. Für die Dateneingabe gibt es je nach Medientyp verschiedene Vorlagen. Das kann zu einer sehr unterschiedlichen Qualität der Metadaten führen. Zurzeit gibt es etwa 54 Millionen Mediendateien.³

Das zweite Projekt, in dem Daten aus NOA hochgeladen werden sollen, ist Wikidata. Hier entsteht eine Wissensbasis, die ähnlich wie Wikipedia das Wissen über die Welt abbilden soll, dieses aber in Form von Linked Data Items speichert. Das Ziel ist also, einen Wissensspeicher mit maschinenlesbaren, untereinander verknüpften Daten zu erstellen. Bisher gibt es etwa 57 Millionen Items.⁴ Zur Beschreibung der Daten gibt es sogenannte Properties. Das ist eine begrenzte Menge von benannten Metadatenfeldern, die genutzt werden können, um die Eigenschaften der einzelnen Items wiederzugeben. Wie bei Wikimedia Commons auch ist aber nicht festgelegt, welche Items wie beschrieben werden sollen. Innerhalb von Wikidata hat sich die Community Wikicite⁵ gebildet, die sich vorgenommen hat, sämtliche bibliografischen Daten der wissenschaftlichen Literatur in Wikidata abzubilden. Jeder Artikel bekommt ein eigenes Item, das mit bibliografischen Daten, aber auch mit Bildern ergänzt wird.

3.1. Auswahl der Bilder und Metadaten

Bevor der Uploadprozess begonnen werden kann, muss zunächst ausgewählt werden, welche Bilder und Daten hochgeladen werden. Im Fall von Wikidata muss keine weitere Auswahl getroffen werden, da hier Metadaten zu allen wissenschaftlichen Artikeln erwünscht sind. Es kann also die Gesamtmenge aus dem Projekt dort integriert werden. Bilder können allerdings nur zu Artikeln hinzugefügt werden, wenn sie bei Wikimedia Commons gehostet werden.

3.2. Automatischer Upload vs. Intellektuelle Auswahl

Der oben beschriebene Zustand führt zu der Frage, ob überhaupt alle Bilder dort hochgeladen werden können. Anders als bei Wikidata gibt es bei Wikimedia Commons keinen Gesamtheitsanspruch. Die Anforderung an die erwünschten Bilder wird allgemein so beschrieben, dass sie „realistically useful for an educational purpose“⁶ sein sollen, also einen realistischen bildenden Anwendungszweck haben sollen. An der gleichen Stelle wird außerdem betont, dass die Qualität des Bildes eine Rolle

3 Wikimedia Commons, <https://commons.wikimedia.org/wiki/Main_Page>, Stand: 24.06.2019.

4 Wikidata, <https://www.wikidata.org/wiki/Wikidata:Main_Page>, Stand: 24.06.2019.

5 WikiCite - Meta, <<https://meta.wikimedia.org/wiki/WikiCite>>, Stand: 28.06.2019.

6 Commons:Project scope - Wikimedia Commons, <https://commons.wikimedia.org/wiki/Commons:Project_scope>, Stand: 24.06.2019.

spielt, sowie die Frage, ob der gleiche Inhalt bereits durch andere Medien abgedeckt wird. Andere Beschränkungen beziehen sich auf Format und Lizenz der Datei. Daraus ergibt sich, dass eine intellektuelle Sichtung der hochzuladenden Daten angebracht ist. Allerdings gibt es im Projekt keine Kapazitäten, um mehrere Millionen Bilder zu sichten. Gleichzeitig könnte diese große Bildermenge auch für Unmut in der Wikimedia Community sorgen, weil die Bilder, die immerhin 10 Prozent der vorhandenen Daten ausmachen würden, das Projekt überwältigen würden. Dabei ist zu bedenken, dass die Qualität der Bilder sehr unterschiedlich ist. Viele haben geringe Auflösungen, sind schlecht beschriftet oder ohne Lektüre des Artikels überhaupt nicht zu verstehen. Solche Bilder werden mit hoher Wahrscheinlichkeit nie benutzt, nehmen aber trotzdem die Aufmerksamkeit der Mitglieder in Anspruch, wenn es darum geht, neu hochgeladene Bilder zu kontrollieren, um zum Beispiel falsche Lizenzen zu finden oder Kategorien zu vergeben. Für diese Problematik muss eine Lösung gefunden werden, die sowohl technisch machbar ist, als auch die Bedürfnisse der Wikimedia Community und die Verfügbarkeit der Kapazitäten im NOA-Projekt berücksichtigt.

4. Der Uploadprozess

Unter Berücksichtigung der oben beschriebenen Ausgangslage wurde im Projekt ein Workflow entwickelt, um Bilder und Metadaten bei Wikimedia Commons einzustellen. Dort gibt es bereits einige Projekte, die Massenuploads durchführen, darunter auch Medien von einigen Bibliotheken und Archiven, wie zum Beispiel die niederländische Nationalbibliothek oder das Kieler Stadtarchiv.⁷ Doch normalerweise werden Medien zu Wikimedia Commons hochgeladen, weil Nutzer/innen sich entscheiden, dass bestimmte Bilder oder Videos dorthin gehören und sie einfach selbst hochladen. Im NOA-Projekt haben wir beschlossen, diese Praxis zu nutzen und die Userinnen und User in unseren Uploadprozess miteinzubinden. Dazu werden auf einer Unterseite der Suchmaschine zufällig ausgewählte Bilder aus dem Projekt angezeigt, und es kann mit einem Klick bestimmt werden, ob das angezeigte Bild hochgeladen werden soll oder nicht. Zusätzlich sollte die Möglichkeit geschaffen werden, direkt aus der NOA-Suchmaschine heraus Bilder hochzuladen.

Um diese Uploadfunktionen zu entwickeln, müssen zunächst alle Bilder mit den relevanten Metadaten angereichert sein. Das wichtigste ist die Lizenz, die in einer standardisierten, maschinenlesbaren Form vorliegen muss. Als nächstes muss die Quelle des Bildes angegeben werden, da dies von den meisten Lizenzen verlangt wird. Als Bildbeschreibung verwenden wir die Bildunterschrift aus dem Artikel. Auf Wikimedia Commons erfolgt die Inhaltserschließung zusätzlich über Kategorien, die hierarchisch aufgebaut sind. Da diese zu einem großen Teil den Wikipedia-Kategorien entsprechen, die im Projekt bereits für jedes Bild ermittelt wurden, können sie leicht aus den bisherigen Daten generiert werden. Auf Abbildung 2 ist zu sehen, wie ein fertig hochgeladenes Bild auf Wikimedia Commons aussieht.

⁷ Für weitere Informationen über Massenuploads und Beispiele für bereits durchgeführte Projekte siehe Commons: Batch uploading – Wikimedia Commons, <https://commons.wikimedia.org/wiki/Commons:Batch_uploading#2018>, Stand: 28.06.2019.

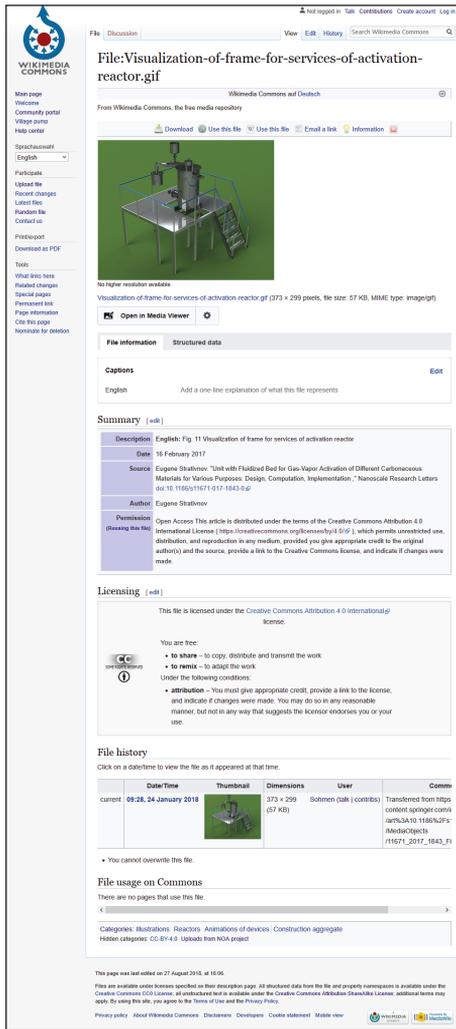


Abb. 2: Screenshot eines Bildes aus NOA auf Wikimedia Commons. <<https://commons.wikimedia.org/wiki/File:Visualization-of-frame-for-services-of-activation-reactor.gif>>. Bild aus: Stratvinov, Eugene: Unit with Fluidized Bed for Gas-Vapor Activation of Different Carbonaceous Materials for Various Purposes: Design, Computation, Implementation, in: Nanoscale Research Letters 12, Article 122, 2017 <<https://doi.org/10.1186/s11671-017-1843-0>> .

4.1. Vorhandene Lösungen

Sowohl Wikimedia Commons als auch Wikidata bieten verschiedene Möglichkeiten an, größere Datenmengen automatisiert hochzuladen. Für Unerfahrene kann das sowohl eine Hilfe als auch ein Problem sein, da ihnen eine Vielzahl an Optionen offensteht, die alle evaluiert und verstanden werden müssen. Im folgenden Abschnitt soll eine kurze, nicht vollständige Übersicht über die zur Verfügung stehenden Schnittstellen und Tools gegeben werden.⁸

Beide Projekte verfügen über eine API (application programming interface)⁹, auf der die hier vorgestellten Tools beruhen. Da der Zugriff darauf kostenlos ist und nur wenigen Einschränkungen unterliegt, ist die Programmierung von eigenen Tools für alle Nutzer/innen möglich.

Für Wikidata gibt es das Tool Quickstatements¹⁰, mit dem viele Items zeitgleich bearbeitet und erstellt werden können. Dieses hat eine eigene Syntax, mit der Befehle formuliert werden. Vor der Benutzung ist also eine Einarbeitung notwendig. Einen leichteren Zugang bietet SourceMD¹¹, das allerdings nur bibliografische Daten zu Artikeln und Autor/inn/en hinzufügt und bearbeitet. Dafür muss hier keine besondere Sprache gelernt werden. Alles, was das Tool benötigt, ist eine Liste mit Identifiern (zum Beispiel DOI oder ORCID) und es fügt automatisch alle Informationen hinzu, die über öffentliche Schnittstellen zugänglich sind. Dieses Tool wurde im Projekt benutzt, um Metadaten zu allen verwendeten Artikeln zu Wikidata hinzuzufügen. Dabei war die Benutzung problemlos. Allerdings gab es Beschwerden über einzelne Artikel von der Wikidata Community, da es wohl hin und wieder Probleme

- 8 Eine Übersicht der Tools ist hier zu finden: Commons:Upload tools - Wikimedia Commons, <https://commons.wikimedia.org/wiki/Commons:Upload_tools>, Stand: 28.06.2019.
- 9 Wikimedia Commons API, <<https://commons.wikimedia.org/w/api.php>>, Stand: 24.06.2019; Wikidata API, <<https://www.wikidata.org/w/api.php>>, Stand: 24.06.2019.
- 10 QuickStatements, <<https://tools.wmflabs.org/quickstatements/#/>>, Stand: 24.06.2019.
- 11 SourceMD, <<https://tools.wmflabs.org/sourcemd/>>, Stand: 25.06.2019.

mit der Kodierung von Sonderzeichen und der Darstellung von Formeln im Titel des Artikels gibt. Wer diese Probleme umgehen möchte, muss doch auf das Tool Quickstatements zurückgreifen, um eine bessere Kontrolle über die eingegebenen Daten zu haben.

Die zwei Wikimedia-Tools für den Massenupload von Bildern, die am häufigsten empfohlen werden, sind Pattypan¹² und das Glamwiki Toolset¹³. Letzteres wurde von Europeana entwickelt und richtet sich an Angestellte von sogenannten GLAMs („galleries, libraries, archives and museums“). Die Benutzung erfordert wenig technisches Fachwissen, ist aber trotzdem relativ aufwendig. Wer das Toolset benutzen möchte, muss zuerst sowohl auf Wikimedia Commons als auch auf Commons Beta (eine Installation von Wikimedia Commons, auf der Uploads und Änderungen getestet werden können, ohne die eigentliche Seite zu beeinflussen) ein Profil anlegen und Nutzungsrechte für das Tool beantragen, was auf vier verschiedenen Wegen erfolgen kann. Die Nutzungsrechte für Wikimedia Commons werden erst nach einem erfolgreichen Test auf Commons Beta gewährt. Solch ein Test kann wiederum erst durchgeführt werden, wenn die URL, unter der die Bilder gefunden werden können, zum Upload freigegeben wird, was an einer weiteren Stelle beantragt werden muss. Pattypan ist ein Desktoptool, das auf der Basis von Metadaten in Tabellen funktioniert. Um das Programm zu benutzen, müssen Metadaten und Bilder in eine vorgegebene Form gebracht werden und können dann mit wenigen Schritten zu Wikimedia Commons hochgeladen werden. Pattypan wurde im Projekt getestet und funktioniert für den Massenupload von Bilddateien. In der Vergangenheit haben einige Projekte auch Bots für den Upload benutzt.¹⁴

4.2. Lösung im Projekt

Im NOA-Projekt wurde ein Tool benötigt, das einzelne Bilder in Echtzeit hochlädt und dabei den Nutzer/innen so wenig Klicks und Entscheidungen wie möglich zumutet. Weiterhin soll es ihnen möglich sein, vorgeschlagene Kategorien aus der Suchmaschine zur Beschreibung auszuwählen und am Ende des Vorgangs eine Erfolgsmeldung und ihr hochgeladenes Bild auf Wikimedia Commons zu sehen. Diese Bedürfnisse können durch existierende Tools nicht abgedeckt werden, weshalb eine eigene Lösung entwickelt wurde.

Die bereits vorhandenen Tools sind schwer anpassbar. Aus diesem Grund wurde eine eigene Funktion entwickelt. Sie benutzt das OAuth-Protokoll¹⁵, um Nutzeraccounts zu authentifizieren und greift zum Großteil auf die dort vorgegebenen Befehle zurück. Dadurch können Nutzer/innen mit ihrem eigenen Account Bilder von der Suchmaschine aus hochladen. Der Uploadprozess ist in den Abbildungen 3 bis 5 dargestellt.

12 Pattypan, <<https://commons.wikimedia.org/wiki/Commons:Pattypan>>, Stand: 25.06.2019.

13 GLAMwiki Toolset, <https://commons.wikimedia.org/wiki/Commons_talk:GLAMwiki_Toolset>, Stand: 25.06.2019.

14 Commons:Guide to batch uploading – Wikimedia Commons, <https://commons.wikimedia.org/wiki/Commons:GLAMwiki_Toolset>, Stand: 28.06.2019.

15 OAuth, <<https://tools.wmflabs.org/oauth-hello-world/>>, Stand: 25.06.2019.

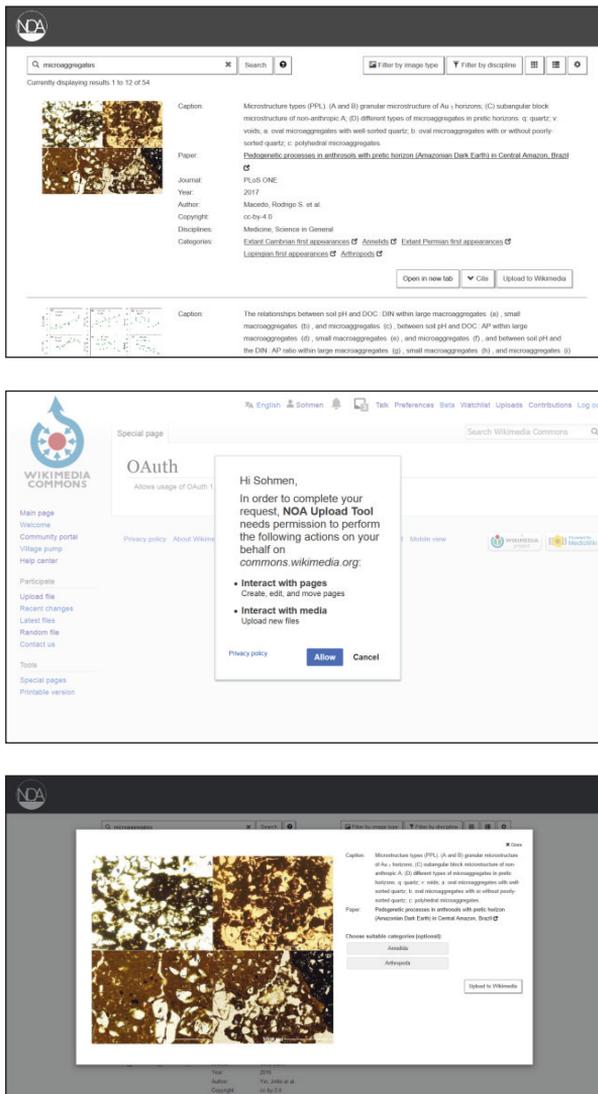


Abb. 3-5: Funktion des Uploadtools. Eigene Screenshots.

5. Communityarbeit

Die Einbindung der Nutzercommunity ist in diesem Projekt aus zwei Gründen besonders wichtig. Zum einen wird die aktive Mitarbeit der Community benötigt, um sicherzustellen, dass möglichst viele qualitativ hochwertige Bilder ausgewählt, mit Kategorien beschrieben und hochgeladen werden. Zum anderen sollen die Projektergebnisse bekannt gemacht werden, damit viele die Möglichkeit zur

Nachnutzung haben. Gleichzeitig ist es wichtig, die Akzeptanz aller Nutzer/innen zu gewinnen. Da Wikimedia Commons auf einer Moderation durch die Community selbst baut, ist es ratsam, diese Community im Projektverlauf miteinzubeziehen, um mögliche Probleme zu identifizieren, die mit den Bildern selbst, den Metadaten oder dem Uploadprozess vorliegen könnten. Es gibt also zwei Ziele: Erhöhung des Bekanntheitsgrades bei gleichzeitiger Qualitätssicherung unter Einbindung der Community.

5.1. Die Wikimedia Community

„Die Community“ von Wikimedia Commons ist allerdings keine feste Größe. Es gibt eine kleine Zahl von Leuten, die sehr aktiv sind und ihre Meinung bei jeder Frage einbringen, während es andere gibt, die ausschließlich Medien hochladen oder nutzen, sich jedoch um Entwicklungsfragen keine Gedanken machen. Erstere sind vor allem wichtig, wenn es um Feedback und generelle Akzeptanz des Projekts geht. Die Gruppe der aktiven Nutzer/innen besteht aus etwa 40.000 Personen.¹⁶ Der Großteil dieser Gruppe lädt wahrscheinlich hin und wieder ein Bild hoch oder ergänzt ein paar Metadaten, hält sich aber aus aktiven Diskussionen zurück. Hier gibt es Potenziale, einige von ihnen für das Hochladen aus NOA zu gewinnen. Die Zahl der Seitenaufrufe ist ungleich höher und liegt für die meisten Monate im Bereich von etwas über 100 Millionen Aufrufen.¹⁷ Dazu gehören allerdings mehrmalige Aufrufe derselben Personen, sowie höchstwahrscheinlich viele, die Wikimedia Commons als Projekt nicht wahrnehmen, sondern nur über Bildersuchen oder Links dorthin gelangt sind, um ein bestimmtes Bild zu sehen. Es ist nicht bekannt, wie groß die Menge der Leute ist, die Wikimedia Commons kennen und gezielt ansteuern, wenn es um Bildersuchen geht. Diese Leute können aber potenziell vom NOA-Projekt erreicht werden und die Plattform verstärkt für die Suche nach wissenschaftlichen Abbildungen nutzen. Andere werden wahrscheinlich nur zufällig, zum Beispiel über eine Bildersuche auf Google, auf Bilder aus NOA stoßen, können also nicht über verstärkte Communityarbeit erreicht werden.

5.2. Outreach

Es wurde früh damit begonnen, das NOA-Projekt auf Konferenzen der Wikimedia Community vorzustellen. Ein erster Diskussionsanstoß fand auf der Wikicite 2016¹⁸ statt, bei der Lambert Heller erstmalig die Projektidee vorstellte und mit anderen Teilnehmenden weiterentwickelte. Zwei Jahre später wurde auf der gleichen Konferenz das Uploadtool zusammen mit anderen ausprobiert. Durch diese Vorstellungen können besonders interessierte und engagierte Mitglieder der Community erreicht werden, die als Multiplikatoren und Unterstützer infrage kommen. Ein breiteres Publikum wurde mit diversen Blogbeiträgen angesprochen, unter anderem auf dem Blog von Wikimedia Deutschland.¹⁹ Im September 2019 ist ein Beitrag auf dem Kickoff-Workshop im Fellow-Programm Freies Wissen²⁰

16 Zu diesen Leuten gehören alle, die in den letzten 30 Tagen eine Aktion ausgeführt haben, siehe Statistics – Wikimedia Commons, <<https://commons.wikimedia.org/wiki/Special:Statistics>>, Stand: 25.06.2019.

17 Page Views for Other Projects, Raw data, <<https://stats.wikimedia.org/wikispecial/EN/TablesPageViewsMonthlyOriginal.htm>>, Stand: 25.06.2019.

18 WikiCite 2016, <https://meta.wikimedia.org/wiki/WikiCite_2016>, Stand: 25.06.2019.

19 Schwarzkopf, Christopher: Wie macht man 5 Millionen wissenschaftliche Open-Access-Abbildungen frei nutzbar?, Wikimedia Deutschland Blog, 26.03.2019, <<https://blog.wikimedia.de/2019/03/26/wie-macht-man-5-millionen-wissenschaftliche-open-access-abbildungen-frei-nutzbar/>>, Stand: 25.06.2019.

20 Fellow-Programm Freies Wissen, Stifterverband, <<https://www.stifterverband.org/freies-wissen>>, Stand: 25.06.2019.

geplant, bei dem der Kontakt zu Wissenschaftlerinnen und Wissenschaftlern gesucht wird, die sich für Wikimedia und Open Science interessieren. Zusätzlich wurde gezielt der persönliche Austausch mit einzelnen Freiwilligen und Mitarbeitenden von Wikimedia gesucht.

Außerdem wurde die Präsenz von NOA auf Wikimedia Commons und anderen Wikimediaprojekten selbst ausgebaut. Dazu gehört zum Beispiel die Dokumentation der Ideen, die 2016 vorgestellt wurden.²¹ Es wurde auch eine eigene Kategorie für die Uploads aus dem Projekt angelegt²², in der alle hochgeladenen Bilder gesammelt werden. Dort sind auch kurze Informationen über das Projekt sowie Links zu weiteren Details zu finden. Es gibt in dieser Kategorie mehrere hundert Uploads, von denen die ersten manuell hochgeladen wurden, während spätere Bilder über das Uploadtool hochgeladen wurden. Durch die Reaktionen auf diese Uploads und eventuelle Änderungen, die von anderen Nutzer/innen durchgeführt wurden, konnten Erfahrungen für die weitere Projektarbeit gesammelt werden. Dies wird im nächsten Kapitel ausführlicher dargestellt.

6. Evaluation

Um den Impact von NOA auf die Wikimediaprojekte zu evaluieren, müssen zunächst geeignete Methoden gefunden werden. Da in unserem Projekt der Schwerpunkt darauf liegt, Bilder für die Nachnutzung zu erschließen, liegt es nahe, die Anzahl der Nachnutzungen zu messen. Diese Zahlen müssen aber immer unter der Voraussetzung betrachtet werden, dass die Bilder für eine lange Zeit auf Wikimedia Commons liegen und der Nachnutzung zur Verfügung stehen sollen, aber während der Projektlaufzeit natürlich nur ein begrenzter Zeitraum evaluiert werden kann. Um diese Zahlen zu ermitteln, nutzen wir das Glamorgan-Tool²³, das die Verwendung von Bildern in den Wikimedia-Projekten dokumentiert. Gleichzeitig wollen wir auch wissen, wie die Wikimedia Community unser Projekt aufgenommen hat. Das kann allerdings nicht in Zahlen gemessen werden und relevante Interaktionen mit den Bildern werden oft nur zufällig entdeckt.

6.1. In Zahlen

Laut dem Glamorgan-Tool wurden bisher 23 von 739 hochgeladenen Bildern auf Seiten aus beispielsweise der englischen Wikipedia, Wikidata oder dem Specieswiki verwendet. Diese Seiten wurden seit der Einbindung der Bilder insgesamt ca. 230.000 Mal angesehen. Diese Zahlen sind bereits ein guter Anfang. Dennoch gibt es Potenzial für eine Erhöhung der Nachnutzung, was sowohl die Zahl der nachgenutzten Bilder als auch die Zahl der insgesamt hochgeladenen Bilder betrifft. Besonders positiv ist, dass die Bilder nicht nur auf englischen oder deutschen Seiten verwendet werden, sondern eine große Bandbreite an Sprachen wie Japanisch, Arabisch oder Französisch abdecken.

21 Heller, Lambert: WikiCite 2016/Proposals/WikiSource and Wikidata as a hub for collaborative annotation and reuse of Open Access literature – Meta, 2016, <https://meta.wikimedia.org/wiki/WikiCite_2016/Proposals/WikiSource_and_Wikidata_as_a_hub_for_collaborative_annotation_and_reuse_of_Open_Access_literature>, Stand: 16.12.2016.

22 Category:Uploads from NOA project – Wikimedia Commons, <https://commons.wikimedia.org/wiki/Category:Uploads_from_NOA_project>, Stand: 25.06.2019.

23 Glamorgan-Tool; Ergebnisse für das NOA-Projekt: GLAMorgan, <<https://tools.wmflabs.org/glamtools/glamorgan.html?&category=Uploads%20from%20NOA%20project&depth=12&year=2016&month=1>>, Stand: 25.06.2019.

Commons category: Uploads from NOA project Depth: 12

Year/month: 2016 1

Get file usage data (demo)

Overview

739 files in category tree
 15 files were viewed, out of 23 used.
 40 pages on 14 wikis use those files.
 229,816 file views in 2016-01.
 Data for 11 pages could not be loaded from the WMF pageview API (404 error).

Views per file

Image	Image title	Views	Used on pages
	Professor-Karl-Gustav-Lennander-in-1897-removing-a-pistol-bullet-from-the-occipital-lobe-of-the-brain-in-a-young-man-af1.jpg	97,945	enwikiX-ray 97,945
	Fig4-Billie-Jean-Kings-dress-by-Ted-Tinling-1973.jpg	61,450	enwikiBillie Jean King 35,597 enwikiBattle of the Sexes (tennis) 25,205 enwikiTed Tinling 688
	Schematic-of-a-tokamak-chamber-and-magnetic-profile.jpg	27,514	enwiki Tokamak 23,457 frwiki Tokamak 2,429 ptwiki Tokamak 1,209 enwiktionarytokamak 18 wikidatawiki Q188059 8 srwiki Tokamak 0
	Conceptual-map-illustrating-the-connections-among-nonhuman-nature-ecosystem-services-environmental-ethics-environmental.jpg	17,516	enwikiEnvironmental health 17,045 enwikiEcohealth 307 dewikiUmweltgesundhet 153 arwiki الصحة البيئية 11
	Cochlear-implant-external-part.jpg	11,232	enwikiDeaf culture 11,232

Abb. 6: Ergebnisse für das NOA-Projekt: Screenshot von GLAMorgan, <<https://tools.wmflabs.org/glamtools/glamorgan.html?&category=Uploads%20from%20NOA%20project&depth=12&year=2016&month=1>>, Stand: 25.06.2019.

6.2. In Anekdoten

Diese Zahlen zeigen die konkrete Nachnutzung einzelner Dateien, können aber nicht aussagen, inwiefern die Bilder von der Community wahrgenommen und angenommen werden. Nicht jede Interaktion kann gemessen und in Zahlen dargestellt werden. Anfangs fielen verschiedene Änderungen durch Nutzer/innen an den Bildern nur zufällig auf. Durch ein einfaches Skript können mittlerweile alle diese Änderungen zusammengefasst dargestellt werden, wobei Änderungen durch Bots ausgeschlossen werden, da diese zum großen Teil nur Kleinigkeiten wie die Verbesserung von „http“ zu „https“ in Links betreffen.

Die meisten Änderungen an den Bildern betreffen die Kategorien. Diese werden hinzugefügt, gelöscht oder verbessert. Dabei bearbeiten einige Leute nur ein Bild, andere gleich mehrere. Dabei wird die Zusammenarbeit der Wikimedia-Community sichtbar: Manche Bilder wurden von eine/r Nutzer/in der Kategorie „unidentified animals“ zugeordnet. Andere, die das Tier identifizieren konnten, verschoben die Bilder von dort aus in die richtige Kategorie. Manche verfeinerten auch bereits vorhandene Kategorien. Beispielsweise wurde die Kategorie „Human pelvis“ zu „Bones of the human pelvis“ geändert, da die Kategorien bei Wikimedia Commons immer so spezifisch wie möglich sein sollen. Solche Änderungen sowie das Hinzufügen von Kategorien sind gleichzeitig ein gutes Feedback für

den Algorithmus, der automatisch Kategorien zuweist. Die restlichen Änderungen sind zum großen Teil entweder sogenannte „License reviews“ oder kleine Formatierungen und Tags. License reviews bedeuten, dass ein Bild gekennzeichnet wurde, weil dessen Lizenz manuell überprüft werden soll. Dies geschah in den meisten Fällen problemlos. Zwei Bilder wurden allerdings gelöscht, wobei das eine nur durch ein behobenes Problem mit dem Uploadtool hochgeladen wurde, während das andere eine Collage war, bei der zu jedem einzelnen Bestandteil eine Quelle verlangt wurde. Dies konnte natürlich nicht erfüllt werden, weil nur die ursprünglichen Autoren diese kennen oder die Bestandteile vielleicht sogar selbst erstellt haben.

Zwei substantziellere Bearbeitungen, die die Bilder direkt betreffen, sollen hier vorgestellt werden. Eines der Bilder aus dem NOA-Projekt²⁴ wurde ursprünglich mit einem weißen Rand zu Wikimedia Commons hochgeladen, so wie es von den Autorinnen und Autoren erstellt wurde. Ein Nutzer fand offenbar, dass das Bild ohne diesen Rand besser aussehen würde, entfernte ihn selbstständig und lud eine neue Version des Bildes hoch (die alte Version kann immer noch in der Detaildarstellung des Bildes gefunden werden).

File history					
Click on a date/time to view the file as it appeared at that time.					
	Date/Time	Thumbnail	Dimensions	User	Comment
current	12:17, 16 January 2018		596 × 309 (266 KB)	Ariadacapo (talk contribs)	Cropped 10 % horizontally and 17 % vertically using CropTool with precise mode (removed white border)
revert	16:06, 11 January 2018		660 × 373 (206 KB)	Sohmen (talk contribs)	Transferred from https://www.hindawi.com/journals/SV/2017/6/103947.fig.004.jpg

• Upload a new version of this file

Abb. 7: Änderungshistorie von *File:Photo-of-the-steam-turbine-feed-pump.jpg* - Wikimedia Commons, <<https://commons.wikimedia.org/wiki/File:Photo-of-the-steam-turbine-feed-pump.jpg>>, Stand: 25.06.2019.

Ein anderes Bild erregte offenbar die Aufmerksamkeit seines Autors. Ein Jahr, nachdem es auf Wikimedia Commons hochgeladen wurde, lud ein Nutzer der Plattform eine neue Version mit wesentlich höherer Auflösung hoch. Gleichzeitig änderte er die Autoreninfo, die ursprünglich alle Autorinnen und Autoren des Artikels als Ersteller des Bildes angab, sodass nur noch einer der Autoren dort steht. Dazu kommentierte er „Author info updated (by author)“. Obwohl Wikimedia Commons es jedem erlaubt, Änderungen durchzuführen, und jeder sich als ursprünglicher Autor eines Bildes ausgeben könnte, wirken die Informationen glaubhaft, da der Nutzer über das Bild in einer höheren Auflösung verfügt und in der Bildunterschrift selbst auch nur der eine Autor als Ersteller angegeben wird.

Auch wenn sie nicht in Zahlen ausgedrückt werden können, sind solche Interaktionen sehr wertvoll für die Evaluation des Projekts, da wir durch sie sehen können, dass die Bilder im Wikimedia Universum angekommen sind auch wahrgenommen werden. Wenn die Bilder in Wikipediaartikeln verwendet werden, bringt dies das Projekt voran und trägt zur Erfüllung des hauptsächlichen Ziels bei: Die Nachnutzung von Bildern. Alle anderen Interaktionen sind Zwischenschritte auf diesem

²⁴ File:Photo-of-the-steam-turbine-feed-pump.jpg – Wikimedia Commons, <<https://commons.wikimedia.org/wiki/File:Photo-of-the-steam-turbine-feed-pump.jpg>>, Stand: 25.06.2019.

Weg, die ähnlich wichtig sind. Wenn jemand einem Bild eine Kategorie hinzufügt, macht er oder sie es einfacher für andere, dieses Bild zu finden. Direkte Änderungen an den Bildern machen es wahrscheinlicher, dass sie sich für eine Nachnutzung eignen. In einem Fokusgruppenworkshop zum Projekt gaben die Teilnehmenden an, dass Bildinhalt und Auflösung zwei der wichtigsten Faktoren seien, wenn sie Bilder zur Nachnutzung suchen. Auch die Überprüfungen der Lizenzen sind eine Hilfe. Entweder fallen sie positiv aus, wodurch bestätigt wird, dass die Bilder berechtigterweise auf der Plattform sind. Oder sie werden gelöscht, wodurch der Workflow angepasst werden kann, um solche Bilder in Zukunft nicht mehr hochzuladen.

7. Abschluss und Ausblick

In diesem Artikel wurde dargestellt, wie das NOA-Projekt wissenschaftliche Bilder für die Wikimediateprojekte verfügbar macht. Dabei wurden die technischen Lösungen sowie die Communityarbeit vorgestellt. Abschließend wurde evaluiert, wie der hochgeladene Inhalt auf Wikimedia Commons angenommen wurde.

In der jetzigen Phase des Projekts sind die Workflows für das Sammeln und Erschließen der Bilder, sowie das Hochladen auf Wikimedia Commons, weitestgehend abgeschlossen. Um den Projektabschluss zu sichern, liegt der Schwerpunkt der weiteren Arbeit auf der Information und Motivierung der Community und der Entwicklung von weiteren unterstützenden Funktionen für Nutzer/innen.

Weiteres Community Building ist wichtig, um neue Nutzergruppen zu erschließen. Dadurch sollen die Bilder aus dem Projekt eine größere Nachnutzung erfahren und mehr Leute von ihnen profitieren können. Gleichzeitig sollen mehr Nutzer/innen auch mehr Bilder aus dem Projekt hochladen und mit Kategorien und anderen Metadaten verbessern. Um das zu erreichen, sollen unter anderem Nutzer/innen über Workshops, Konferenzen und Blogposts erreicht werden, die wiederum als Multiplikatoren für andere agieren können.

Um die Nachnutzung der Projektergebnisse zu verbessern, sollen neue Funktionen entwickelt werden und bestehende Funktionen verbessert werden. Diese Funktionen werden das Auffinden und Nachnutzen der Bilder vereinfachen, vor allem für Autorinnen und Autoren von Wikipedia. Dazu gehört es auch, die Metadaten der Bilder an das neue Schema Structured Data on Commons²⁵ anzupassen, das gerade für Wikimedia Commons entwickelt wird und die Bilder mithilfe von Linked Data beschreiben soll.

Literaturverzeichnis

- Schwarzkopf, Christopher: Wie macht man 5 Millionen wissenschaftliche Open-Access-Abbildungen frei nutzbar?, Wikimedia Deutschland Blog, 26.03.2019, <<https://blog.wikimedia.de/2019/03/26/wie-macht-man-5-millionen-wissenschaftliche-open-access-abbildungen-frei-nutzbar/>>, Stand: 25.06.2019.

²⁵ Commons:Structured data – Wikimedia Commons, <https://commons.wikimedia.org/wiki/Commons:Structured_data>, Stand: 25.06.2019.