

100000

# 100 000 Bildminnen

## Slutrapport



**WIKIMEDIA**  
SVERIGE

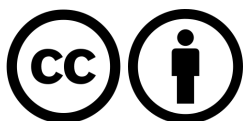


**100 000 Bildminnen**  
**Slutrapport 6 februari 2024**

**Författare:** André Costa

**Medförfattare:** Alicia Fagerving, Josefine Hellroth Larsson

**Omslagsbild:** "[Ultuna - Nordiska museet - NMAx.0000452](#)" av K.W. Gullers, 1944.  
Digitaliserad av Nordiska museet. CC BY 4.0.



**Licens:** Om inget annat anges är detta dokument och de bilder som inkluderas i det licensierade under en Creative Commons Erkännande 4.0 licens. Du har tillstånd att distribuera, dela och bygga vidare på detta verk under förutsättning att du ger erkännande till författaren av detta dokument. Du hittar hela licenstexten på <https://creativecommons.org/licenses/by/4.0/legalcode.sv>

**Attribution:** 100 000 Bildminnen - Slutrapport av André Costa et al. (2024)

**Finansierades av:** [Nordiska museet](#)

**Projekt team (Wikimedia Sverige):**

[André Costa](#), Projektledare och utvecklare

[Alicia Fagerving](#), Utvecklare

[Axel Pettersson](#), Eventledare

[Josefine Hellroth Larsson](#), Eventledare och kommunikation

**Samarbetspartner:**

[Nordiska museet](#)

[KulturIT](#)

**Kod:**

De verktyg och den kod som har utvecklats för projektet av Wikimedia Sverige har publicerats som öppen källkod under en MIT licens på följande platser.

[Wikimedia-Sverige/commons-diff](#) på GitHub

[Wikimedia-Sverige/nordiska-2022](#) på GitHub

# Innehåll

<b>Innehåll</b>	<b>2</b>
<b>Bakgrund</b>	<b>3</b>
<b>Att förstå konserverings-, digitaliserings- och uppladdningsprocessen</b>	<b>3</b>
Insyn i konserverings- och digitaliseringsprocessen	4
Inledande pilotuppladdning	5
Utbildning av museipersonal	5
<b>Leverans, urval och uppladdning</b>	<b>6</b>
En inledande not kring urval	6
Uppladdningen och förberedelser på Wikimedia Commons	7
Det ursprungligen planerade flödet	8
Första batcherna (januari–mars 2023)	8
Nästkommande batcher (maj–oktober 2023)	10
Sista batcherna (november–december 2023)	12
<b>Berikning</b>	<b>13</b>
Inledande matchning av metadata	13
Metadataskrivstuga	14
Veckans tävling	15
Organiskt av gemenskapen	17
<b>Återförsel av information från Crowdsourcing</b>	<b>19</b>
Svårigheter kopplade till att ta ut och förstå ändringar	19
Extraheringsverktyg	20
Återbruk	21
<b>Statistiska mått</b>	<b>22</b>
Bild- och metadatauppladdning	22
Filanvändning	23
Filvisningar	24
Interaktioner, förbättringar och engagemang	26
Övrig statistik	28
Segmentering	28
<b>Synlighet och kommunikation</b>	<b>29</b>
<b>Sammanfattning, slutsatser och framtida arbete</b>	<b>30</b>
<b>Appendix</b>	<b>32</b>
Appendix A: Uppladdningsbatcher	32
Appendix B: Statistik	35
Bild- och metadatauppladdning	35
Filanvändning och filvisning:	35
Interaktioner, förbättringar och engagemang	36
Veckans tävling	38
Appendix C: Mediaomnämmande och övrig kommunikation	39

# Bakgrund

Nordiska museet och Wikimedia Sverige bedriver tillsammans projektet 100 000 Bildminnen. Inom ramen för detta projektet kommer Nordiska museet att nydigitalisera och publicera en större samling av fotografier. Syftet med Wikimedia Sveriges deltagande i projektet är att öka effekten av digitaliseringsprocessen samt göra Nordiska museets samlingar mer användbara och tillgängliga för allmänheten. Projektet är experimentellt och arbetet anpassas kontinuerligt.

Det material som digitaliseras är en delmängd av de negativ på plastbas som i januari 2021 drabbades av en vattenskada och som i samband med konserveringsarbetet även kommer att digitaliseras. Samlingarna inkluderar bilder från några av Sveriges mest namnkunniga fotografer.

Wikimedia Sverige arbetar tillsammans med Nordiska museet för att ladda upp de digitaliserade bilderna på Wikimedia Commons. När materialet är uppladdat kan Wikimediagemenskapen, de volontärer som är aktiva på plattformarna, fritt använda materialet för att t.ex. illustrera artiklar på Wikipedia. Genom sin interaktion med materialet bidrar de även med ytterligare information om, och kontext kring, bilderna.

Det som för Wikimedia Sveriges del gör denna insats unik är att vi får hjälpa till i ett så tidigt skede av digitaliseringsprocessen. Det ger oss en bättre förståelse av både konserveringsprocessen och gör att vi tillsammans kan identifiera tidiga insatser som underlättar för tillgängliggörandet och berikar digitaliseringsinsatsen. Det gör det även möjligt för oss att redan från början lägg grunden för att kunna hämta tillbaka tillagd information från Wikimediaprojekten och återintegrera dem med materialet i Nordiska museets samlingar.

## Att förstå konserverings-, digitaliserings- och uppladdningsprocessen

Tidigt i projektet säkerställdes att både Wikimedia Sverige och Nordiska museet fick god insikt i de processer som den andra arbetar enligt, samt de begränsningar som processen, materialet eller plattformen medför. Detta kombinerades med en mindre pilotuppladdning av en handfull bilder till Wikimedia Commons.

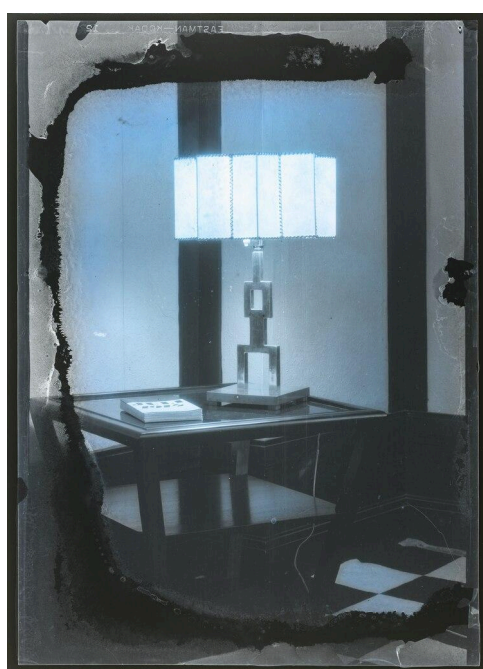
Denna transparens och ömsesidiga förståelse medförde:

- Rimliga förväntningar kring metadata.
- En bättre informerad diskussion om formatering av metadata för uttag av Wikimedia Sverige.
- Bättre förståelse av de skador materialet kan ha samt vilka typer av skadat material som ändå är intressant på Wikimedias plattformar.
- Ändrad lösning för hur länkarna till Wikimedia Commons kopplas till bilderna.

## Insyn i konserverings- och digitaliseringsprocessen

Wikimedia Sverige bjöds in för en guidad tur av det rum, samt en demonstrering av de processer, som satts upp för att konservera och digitalisera negativerna efter att de tagits ut ur frysen (de frystes ner för att förhindra ytterligare vattenskador).

Detta gav Wikimedia Sverige möjlighet att se hur både konserveringen och digitalisering fungerar i praktiken. Det fysiska ursprunget för bildernas metadata visades vilket gav en bättre förståelse för de inneboende begränsningar det innebär. I detta fall består samlingen av kartonger vilka i sin tur innehåller flera kuvert där vart kuvert i sin tur innehåller flera bilder. Det finns en liten mängd data på var kartong, samt en liten mängd information på vart kuvert, men ingen information per bild. Var kartong, brev och bild får en unik referens och denna data, tillsammans med metadata om själva digitaliseringen, sammanställs i ett kalkylark för att sedan sparas ned i Exif-data<sup>1</sup> för den digitaliserade bilden.



En skadad bild kan beskäras och restaureras och därigenom vara till nytta på Wikimediaplattformarna trots skadan.

[Foto: Erik Holmén / Nordiska museet, CC BY-SA 4.0.](#)



Skadade bilder kan även användas för att illustrera själva skadorna och därigenom ge en värdefull inblick i konserveringsarbetet.

[Foto: Erik Holmén / Nordiska museet, CC BY-SA 4.0.](#)

På grund av en vattenskada som föranledde digitaliseringen, men även på grund av materialets ålder, har delar av materialet i olika utsträckning blivit skadat. Genom direkta observationer, samt en genomgång av den skadekarta Nordiska museet tagit fram, fick Wikimedia Sverige en god inblick i de olika typer av skador som materialet kan ha drabbats av. Detta är en unik inblick då skadat material vanligen har sållats bort innan Wikimedia Sverige kommer in i bilden. Detta föranledde en diskussion där Nordiska museet fick en bättre insyn i hur även skadat material kan vara av intresse för Wikimediagemenskapen som

<sup>1</sup> [https://sv.wikipedia.org/wiki/Exchangeable\\_image\\_file\\_format](https://sv.wikipedia.org/wiki/Exchangeable_image_file_format)

t.ex. kan vilja restaurera små skador eller välja att beskära mer skadade bilder där intressanta detaljer ändå finns kvar.

Utöver huvudsyftet, att ge båda organisationerna en bättre förståelse av varandras arbete, så ledde det fysiska besöket och de resulterande diskussionerna till två direkta förändringar i projektet. Metadata som levereras till Wikimedia Sverige strukturerades om och utgick istället från det format som skapats för kalkylarket, istället för exif-formatet. Istället för bara de oskadade bilderna så skulle alla bilder levereras, tillsammans med en skadegradering, till Wikimedia Sverige som sedan skulle återkoppla om det fanns en skadenivå där materialet är så pass skadat att det ej längre är av nytta för Wikimediagemenskapen.<sup>2</sup>

## Inledande pilotuppladdning

För att få ett praktiskt exempel på vilket ytterligare diskussioner kring uppladdnings-, och återföringsprocessen kunde baseras genomfördes en pilotuppladdning av 20 bilder<sup>3</sup> med en normal-mängd metadata.

Pilotuppladdningen möjliggjorde även manuella manuella insatser för att berika metadata (se [Berikning](#)). Detta innebar i sin tur att dessa bilder kunde användas som referensobjekt under framtagandet av ett verktyg för att hämta ut berikade metadata (se [Extraheringsverktyg](#)).

Detta medförde även att Wikimedia Commons-länkarna för de uppladdade bilderna kunde levereras tillbaka till Nordiska museet. Detta föranledde en diskussion om hur dessa länkar borde lagras och med det en ändring i projektplanen där man valde att frångå idén om att dessa länkar skulle ligga i bildens exif-data. Istället kom dessa att lagras i samlingsdatabasen.

Två aspekter av processen som inte testades var storskalig leverans av bildfiler till Wikimedia Sverige samt uppdateringen av bilderna på Wikimedia Commons med länkar till DigitaltMuseum<sup>4</sup>. Det senare då bilderna inte tillgängliggjordes på DigitaltMuseum förrän långt senare under projektet.

## Utbildning av museipersonal

Tidigt under projektet höll Wikimedia Sverige även en utbildning för personal på Nordiska museet om vad Wikimediarelsen är, och hur den fungerar, samt grunderna i Wikipedia, Wikimedia Commons och verktyget OpenRefine (se [Uppladdningen och förberedelser på Wikimedia Commons](#)). Detta gjordes för att ge samtliga deltagande personal en bättre förståelse för vad som är, och inte är, möjligt att göra med materialet på Wikimediaplattformarna. Detta inkluderar även Wikimedia Sveriges roll där organisationen inte själv lägger in material på Wikipedia, men kan verka för att gemenskapen gör det samt att organisationen inte heller kan styra över den interaktion som sker med materialet efter att det tillgängliggjorts på plattformarna.

Efter att den inledande pilotuppladdningen skett genomfördes ytterligare en utbildning som fördjupade kunskapen om hur Wikimedia Commons fungerar – informationsmallar,

---

<sup>2</sup> OBS! Detta reviderades senare, se [Leverans, urval och uppladdning](#).

<sup>3</sup> [https://commons.wikimedia.org/wiki/Category:100\\_000\\_Bildminnen/Pilot\\_upload](https://commons.wikimedia.org/wiki/Category:100_000_Bildminnen/Pilot_upload)

<sup>4</sup> <https://digitaltmuseum.org/0210913270579/100-000-bildminnen>

kategorier samt strukturerad data på Commons (SDC)<sup>5</sup>. En översikt av pilotuppladdningen gavs för att exemplifiera hur museets metadata kan översättas till informationsmall på Commons. En demonstration av hur OpenRefine används för att interagera med materialet gavs även. Denna utbildning möjliggjorde konkreta diskussioner kring att optimera det format som användes för den metadata som produceras av museet för att i ett senare steg kunna underlätta bearbetning på Wikimedias plattformar.

## Leverans, urval och uppladdning

Kärnan i Wikimedia Sveriges insats i projektet består av att göra de bilder som digitaliserats av Nordiska museet tillgängliga på Wikimedia Commons. Detta innefattar dock ett flertal steg som tillsammans utgör ett flöde.

Inom ramen för detta projektet har vi upprepade gånger behövt revidera dessa flöden för att hitta en balans mellan den krävda arbetsbördan och resultatet. Ursprunget till denna komplexitet är dels samlingens särart (skadat material, en relativt liten mängd metadata och en stor bredd i vad materialet avbildar samt för vilka ändamål det har skapats). Det är huvudsakligen när det kommer till urval och leverans som dessa revideringar skett medan metoden för uppladdning och den uppladdade datastrukturen har varit relativt stabil.

Vi försöker här redogöra för hur processen såg ut med ett särskilt fokus på hur leverans och urval ursprungligen planerades och därefter utvecklades.

## En inledande not kring urval

Under projektets gång tog Nordiska museet ett beslut om att bara tillgängliggöra, på både sina egna och Wikimedias plattformar, en del av materialet (ca. 10%) parallellt med att digitaliseringen pågick. Resten av det material som digitaliserats görs tillgänglig vid ett officiellt *släpp* för att fira att målet med 100 000 digitaliserade bilder uppnåtts. I den utsträckning Wikimedia Sverige gjort ett urval av vilket material som ska laddas upp till Wikimedias plattformar var detta med förståelsen att det handlade om *när* det skulle laddas upp, och att allt material kommer att laddas upp i slutet av projektet.

Urvalet gjordes utifrån vad som i närtid bedömdes ha goda chanser att skapa engagemang samt leda till en berikning av materialet, något som var en central anledning till att vi var med i projektet från digitaliseringsprocessens start. Urvalet tittade även på synergieffekter med andra pågående projekt, då särskilt *Wikipedia för hela Sverige*<sup>6</sup>, genom vilka särskilda insatser kring lokalt engagemang kunde fokuseras.

Det enda underlag som fanns tillgängligt för att göra urvalet var det motivbeskrivning som finns angivet för var bild. Denna har en väldigt varierande grad av informativitet, från *Saab flygverkstäderna, Linköping* till *Dimma*. Fokus blev således på motiv som var så konkreta som möjligt där dels svenska orter (*Borås, Finspång*), dels svenska industrier (*Saab, Vårgårda Havregryn*) särskilt sticker ut. Utöver detta identifierade vi även en del motiv som kunde vara särskilt intressanta för samarbeten med andra aktörer såsom Sveriges lantbruksuniversitet eller Flygvapenmuseum.

<sup>5</sup> [https://commons.wikimedia.org/wiki/Commons:Structured\\_data](https://commons.wikimedia.org/wiki/Commons:Structured_data)

<sup>6</sup> <https://wikimedia.se/2022/09/30/wikipedia-for-hela-sverige/>

## Uppladdningen och förberedelser på Wikimedia Commons

Innan bilderna kunde laddas upp till Wikimedia Commons krävdes det att en del strukturer kom på plats, såsom en projektkategori *100\_000\_Bildminnen*<sup>7</sup> som används för att samla alla bilder som laddas upp. Eftersom informationen om bilderna var enhetligt strukturerad önskade vi även att de skulle se likadana ut efter att de laddats upp. När man skapar fildidor organiserar man filens data i en informationsmall. Man kan välja en av de standardmallar som finns eller skapa en ny mall med en standardmall som grund, vilket var lösningen vi valde. Att använda en egen mall görs främst om man laddar upp en stor, sammanhållen samling och har då fördelen att mallens utseende kan kontrolleras samt att man kan sätta vissa fält till att alltid vara ifyllda med samma information. Den mallen som skapades<sup>8</sup> anger alltid att bilden kommer från Nordiska museet, med länk till en sida om Bildminnenprojektet, automatisk kategorisering i projektkategorin samt licensen CC BY-SA 4.0. Utöver detta innehåller den fält för den metadata som levererades med bilden; motiv, inventarienummer, arkivkod, år samt fotografens namn.

När det gäller *fotografens namn* behövde det först kontrolleras att fotografen redan är känd på Wikimedia Commons (med en egen kategori och Creatormall<sup>9</sup>), annars bör dessa skapas. Eftersom det endast förekom ett fåtal olika fotografier i de samlingar vi jobbat med kunde detta göras för samtliga fotografier.

*Motivfältet* användes för flera syften. Dels använder vi det för de dubbla beskrivningsfälten som finns i informationsmallen. Dubbla beskrivningsfält är nödvändiga för att både kunna ange museets ursprungsbeskrivning men även som ett startvärde för den beskrivning vi senare uppmanar gemenskapen till att förbättra. Mallen gör det tydligt att ursprungsbeskrivning ej bör uppdateras, även om det i praktiken inte finns något tekniskt hinder för dem att göra det.

Motivfältet ligger även till grund för att skapa lämpliga tematiska kategorier inom projektets kategoriträd, till exempel fick alla bilder med "Norrköping" i motivfältet kategorin *100\_000\_Bildminnen/Norrköping* även om det förekom olika motiv på dem, som *Norrköping Staden*, *Harvester. Norrköping*, *Skörde- och jordbruksmaskiner* och *Journalist P. Wange Norrköping för Se*. Dessa tematiska kategorier gör det enklare att senare kommunicera om materialet och komma med uppmaningar att berika det utifrån lokalkunskap.

Slutligen används motivfältet som en av komponenterna i det filnamn som behöver skapas när man laddar upp filer. Museets ursprungliga filnamn bestod enbart av arkivkoden och är därmed inte informativ för den som kommer utifrån. Wikimedia Sverige har en egen praxis i att namnge filer som gjorts tillgängliga av samarbetspartners enligt mönstret "Kortbeskrivning - Institutionens namn eller förkortning - Inventarienummer". Detta gör att det är lätt att avgöra var filen kommer ifrån, både från vilken institution och att den troligtvis laddats upp av oss. Vi följde detta mönster och således fick t.ex. en bild med motivet *Astra Göteborg* och inventarienumret *NMAx.0023707* namnet *Astra Göteborg - Nordiska museet - NMAx.0023707.tif*.

---

<sup>7</sup> [https://commons.wikimedia.org/wiki/Category:100\\_000\\_Bildminnen](https://commons.wikimedia.org/wiki/Category:100_000_Bildminnen)

<sup>8</sup> [https://commons.wikimedia.org/wiki/Template:Nordiska\\_museet\\_Bildminnen\\_image](https://commons.wikimedia.org/wiki/Template:Nordiska_museet_Bildminnen_image)

<sup>9</sup> t.ex. [https://commons.wikimedia.org/wiki/Creator:K.W.\\_Gullers](https://commons.wikimedia.org/wiki/Creator:K.W._Gullers)



Arkivkoden läggs dels till i informationsmallen, men användes också för att identifiera den bildfil som hör till var metadatapost. För att genomföra den faktiska bilduppladdningen används verktyget OpenRefine<sup>10</sup>. Verktygets huvudsakliga syfte är databearbetning men kan sedan 2022 även användas för att förbereda och genomföra bilduppladdningar till Wikimedia Commons. Eftersom verktyget även stödjer arbete med länkade data kunde relevanta delar av filens metadata även läggas till som SDC. Vi använde oss dessutom av egenskapen *på fokuslista för Wikimedia-projekt* för att ange att bilden hör ihop med Bildminnenprojektet.

## Det ursprungligen planerade flödet

I den första plan som togs fram var avsikten att Nordiska museet skulle skadegradera var bild, enligt en eller flera parametrar. Wikimedia Sverige skulle sedan sätta ett gränsvärde där alla bilder med lägre skadegradering framöver automatiskt skulle tillgängliggöras på Wikimedia Commons.

För dataleveranser var tanken att Nordiska museet löpande skulle kunna leverera metadata för det senast digitaliserade materialet och att Wikimedia Sverige därefter skulle meddela vilka bildfiler som var nästa på tur. Detta gjordes genom breda instruktioner i stil med "alla bilder med *Borås* i motivbeskrivningen".

Detta skulle vara startskottet för museets process för att på sin sida godkänna bilderna för publicering varefter de skulle göras tillgängliga på en yta som båda hade tillgång till. Därefter skulle Wikimedia Sverige processa metadata för bilderna, för att få fram data kompatibla med Wikimedia Commons, samt därefter ladda upp bilderna dit. Slutligen levererar Wikimedia Sverige en lista med länken till var uppladdad fil vilka Nordiska museet sedan kan importera till sin samlingsdatabas. När Nordiska museet sedan publicerar bilderna på den egna plattformen för publik delning, DigitaltMuseum<sup>11</sup>, skickar de i sin tur en lista med dessa länkar till Wikimedia Sverige som därefter lägger till dessa i bildinformationen på Wikimedia Commons.

För att enklare kunna kommunicera om de bilder som tillgängliggjorts, samt skapa engagemang kring dem, var avsikten att göra uppladdning i omgångar, så kallade *batcher*<sup>12</sup>, med ett tematiskt fokus allt eftersom materialet digitaliserades. Dvs. när det utifrån metadata gick att avgöra att det fanns en tillräcklig mängd av t.ex. Stockholmsbilder så skulle dessa kunna laddas upp som en samlad mängd. Tanken var att inledningsvis fokusera på mindre batcher för att därigenom kunna finjustera processen.

## Första batcherna (januari–mars 2023)

Redan innan den första riktiga uppladdningen blev det tydligt att en skadegradering per bild skulle innebära en orimlig arbetsbörda. I den första uppladdningen, med material från Nordiska Kompaniet<sup>13</sup>, skedde urvalet därför utan kunskap om materialets skadestatus och i detta material kan vi därför se exempel på tydligt skadat material. För senare batcher var det

---

<sup>10</sup> <https://commons.wikimedia.org/wiki/Commons:OpenRefine>

<sup>11</sup> <https://digitaltmuseum.org/s-nm/>

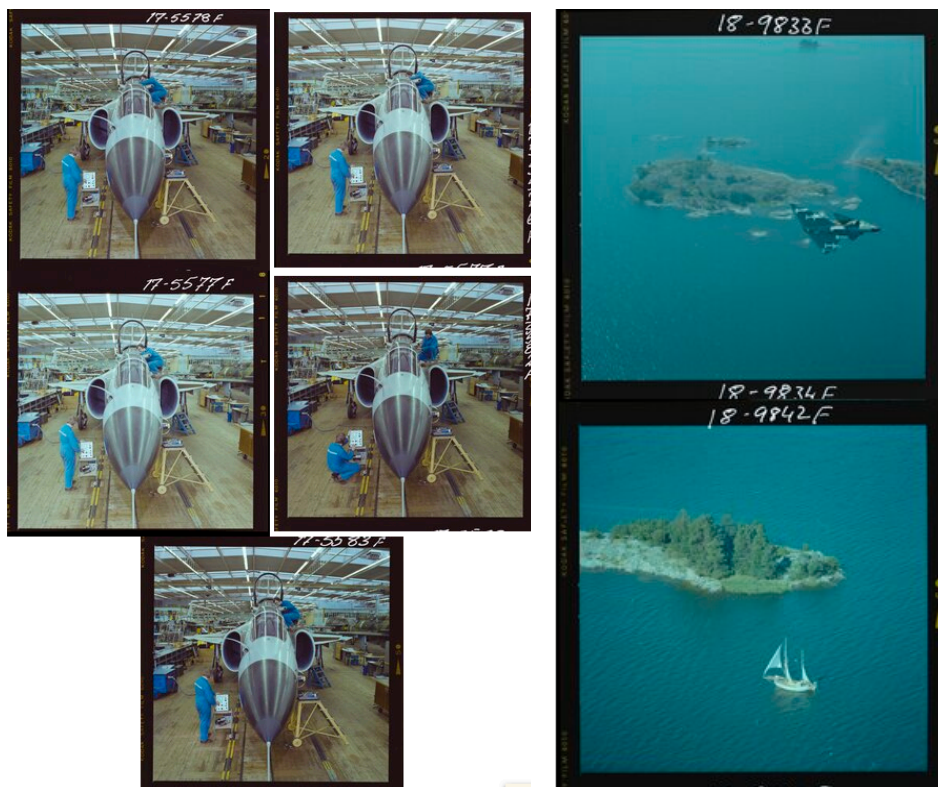
<sup>12</sup> För det faktiska innehållet i var batch se [Appendix A](#).

<sup>13</sup> [https://commons.wikimedia.org/wiki/Category:100\\_000\\_Bildminnen/Nordiska\\_kompaniet](https://commons.wikimedia.org/wiki/Category:100_000_Bildminnen/Nordiska_kompaniet)

mest skadade materialet exkluderat redan vid de leveranser av metadata som gick till Wikimedia Sverige.

Tanken kring att tematiskt samla material för att ladda upp det i batcher kunde inte helt realiserats då det i ett stort antal fall fanns otillräcklig metadata om vad bilden avbildar, eller den kontext inom vilken den skapades, för att det alls skulle gå att koppla dem till ett tema. Den tematiska grupperingen blev istället en indirekt effekt av att det fysiska materialet redan var tematiskt grupperat vilket medförde att denna gruppering återfanns efter att de digitaliserats.

Utifrån levererad metadata identifierades ett par olika teman, urvalet av dessa avgjordes främst på om motivet var något väldigt konkret (som en industri) eller av om det kunde kopplas till de geografiska fokusområden som gällde för Wikipedia för hela Sverige, då detta projekt senare kunde marknadsföra materialet genom sina insatser. Särskilt värt att nämna är här den samling av bilder från SAAB som utgjorde fokusmaterialet för metadatataskrivstugan på Flygvapenmuseum.



Det fanns flera exempel med bildserier som var nästan helt identiska.

Foto [1, 2, 3, 4, 5]: Peter Gullers / Nordiska museet, [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/).

Två bilder ur bildserien *Viggen i action för Know How II*, varav en har tydlig avsaknad av Viggenplan.

Foto [1, 2]: Björn Enström / Nordiska museet, [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/).

Några av erfarenheten från dessa första batcher var:

- Att det fanns många exempel på bildserier med nästan identiska bilder. Detta är ett resultat av att delar av bildmaterialet utgörs av det råa bildmaterial som skapats och där enbart en av dessa bilderna senare använts. Detta är inte i sig ett problem för Wikimedia Commons men kan ha en negativ effekt när man önskar engagera en bredare allmänhet i att bidra till materialet.

- Den skrala metadata som fanns tillgänglig för de fysiska bilderna gjorde det ibland svårt att vara träffsäker i de tematiska urvalen. Som exempel kan ges en batch med bilder som valts ut för att de märkts upp med Helsingborg men där det senare visade sig att de samtliga avbildar en hemmiljö vilket inte gjorde dem lämpliga att användas för att engagera Helsingborgsbor att bidra med sin lokalkunskap.
- Att motivbeskrivningar fanns på kuvertnivå, snarare än bildnivå, resulterade i att motiven ibland inte alls var de man kunde förvänta sig från motivbeskrivningen. Som exempel kan ges en serie flygfoton på SAAB gripen som även inkluderar bilder av natur och båtar, troligen tagna i väntan på att flygplanet ska flyga förbi igen. Då motivbeskrivningen ligger till grund för både filnamn och initiala kategorier kan detta leda till förvirring.
- Processen med att Wikimedia Sverige först begärde bilder som Nordiska museet sedan manuellt gjorde tillgängliga på en FTP-server visade sig kräva mycket handpåläggning. Då processen till största delen var manuell ökade även risken för att fel filer delades.

För Wikimedia Sverige var dock det mest tidskrävande steget bearbetningen av metadata, inte för urvalet, utan för att bearbeta det så att det är kompatibelt med Commons (t.ex. identifiera grundläggande kategorier för materialet).

## Nästkommande batcher (maj–oktober 2023)



NMax.0014762  
1717/A Rom



NMax.0014763  
1717/A Rom



NMax.0014764  
1717/A Rom



NMax.0014765  
1717/A Rom



Ett exempel på en sida ur en bildrapport.

Foto (avbildade): K.W. Gullers / Nordiska museet, [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/).

Utifrån de tidigare identifierade problemen med bland annat dålig träffsäkerhet vid urval och många identiska bilder testade vi i denna omgång att Nordiska museet först levererade bildrapporter med 6-10 tumnagelbilder per sida jämte bildens id-nummer och motiv-beskrivning.

För att underlätta arbetet med att leverera filerna till Wikimedia Sverige gjordes samtliga filer, kopplade till den aktuella leveransen av metadata, tillgängliga på FTP-servern direkt. Wikimedia Sverige skickade sedan en lista med det önskade urvalet till Nordiska museet för att få dessa godkända för publicering varpå den metadata som fanns för filerna levererades till Wikimedia Sverige.

När de godkänts för publicering publicerades de även på DigitaltMuseum. Genom att tillgängliggöra dem där först kunde DigitaltMuseum-referensen läggas till redan under uppladdningen till Wikimedia Commons, istället för att som tidigare kräva en separat extra körning.

Några av erfarenheten från dessa batcher var:

- Bildrapporten gav en bra möjlighet att få en bättre bild av material som kunde vara visuellt intressant.
- Att utgå från bildrapporterna visade sig dock snabbt extremt tidskrävande, då det handlar om flera hundra sidor bildrapport att bläddra sig igenom, samtidigt som det också kan vara enklare att identifiera intressanta titlar i det mer kompakta kalkylarksformat som används för metadata. Det var även enklare att upptäcka eventuella konstigheter i metadata när det var denna man utgick från.
- Om man i bildrapporterna identifierat en större batch med intressanta bilder, som SAS med 900 bilder, blev detta orimligt att titta på varje sida. Istället var det, när batchens storlek blev tydlig, bara möjligt att göra ytterligare stickprov bland bilderna. Som ett resultat slank felaktigheter åter igenom här.<sup>14</sup>
- Att först göra urvalet för att sedan invänta godkännande om publicering medförde en paus och fördröjning i processen. I åtminstone ett fall visade det sig dessutom att några av bilderna i en motivkategori som valts ut av integritetsskäl inte borde publiceras.
- Efter att urvalet hade gjorts visade det sig vara komplicerat att gå mellan metadata och bildrapporten eftersom det krävde att man genom att bläddra i bildrapporten identifierade var batchen började och slutade för att sedan identifiera motsvarande bilder i kalkylarket. För större batcher kunde detta innebära upp mot 100 sidor som åter behövde bläddras igenom.
- Att ha tillgång till bilderna innebar även att vi vid åtminstone ett tillfälle identifierade en större samling bilder som såg väldigt intressanta ut men där motivbeskrivningen (enbart "Historieboken") var för svag för att vi skulle kunna använda materialet i en riktad insats eller ge det ens grundläggande kategorier på Wikimedia Commons.

För Wikimedia Sveriges del blev det mest tidskrävande steget nu istället urvalet (scrollandet i bildrapporterna) samt, efter urvalet, matchningen mellan bildrapport och metadatafil. Detta då de två filerna strukturerade sitt innehåll utifrån två olika identifierare.

---

<sup>14</sup> Se exempelvis bilden [SAS DC 7 i luften färg \(NMAx.0010983\)](#) som i själva verket avbildar Tower Bridge i London.

Ett bättre upplägg här skulle vara att leverera bildrapporterna tillsammans med filernas metadata. Då skulle bildrapporten kunna användas för att få en bättre bild av det material som via metadata redan identifierats som intressant.

## Sista batcherna (november–december 2023)

Flödena för de tidigare batcherna var samtliga arbetsintensiva och lämpade sig bättre för de mindre batcher vi fokuserat på där avsikten var att ta fram material som var särskilt lämpat för engagemang och berikningsinsatser. För de sista batcherna ville vi istället ladda upp en större mängd material och vi ändrade därför i flödet genom att eliminera urvalsprocessen.

När ett urval inte längre var aktuellt behövde Nordiska museet redan i förväg godkänna samtliga bilder för publicering innan de levererades till Wikimedia Sverige. Därigenom publicerades bilderna även på DigitaltMuseum vilket medförde att länkarna dit fanns tillgängliga i den metadata som levererad. Länkarna till DigitaltMuseum innebar att denna plattform kunde nyttjas på ett liknande sätt som de tidigare bildrapporterna om det fanns ett särskilt behov av en snabb visuell kontroll av en bild.

Genom att inte behöva fokusera på vilka motiv som kan vara särskilt intressanta, eller tolka vad de betyder (t.ex. om det är namnet på en ort eller industri) minskades arbetsinsatsen väsentligt. Ännu mer så då den visuella bedömningen av bildrapporterna också kunde elimineras.

Genom att prova tre variationer på samma grundläggande flöde kunde vi eliminera några upptäckta komplikationer, som att det t.ex. var lättare att göra alla filerna tillgängliga på FTP-ytan direkt samt publicera dem på DigitaltMuseum innan de laddas upp till Wikimedia Commons. Vi identifierade även flöden som fungerade bättre i vissa situationer än andra som bildrapporter vilka var nyttigare för mindre batchar samt att urvalsprocessen lämpar sig väl när man vill ta fram ett material särskilt anpassat för engagemang mot vissa grupper. Utvecklingsmöjligheter för framtiden skulle här vara om man kunde generera bildrapporter direkt utifrån ett urval i kalkylarket för metadata.

Vad som genomgående karakteriserat uppladdningarna var dock att skral information om bildernas motiv gör insatserna kring att matcha metadata mot strukturer på Wikimedia Commons betydligt mer arbetsintensiva än liknande insatser där det funnits tillgång till mer information. Unikt för denna samling har även varit de situationer där bilderna utgjort ett underliggande råmaterial för fotoreportage vilket medfört många snarlika bilder eller "andra bilder på samma rulle"-situationer där motivbeskrivningen kan vara missvisande för en delmängd av bilderna.

De bilder som laddats upp inom ramen för Wikimedias insats utgör bara en delmängd av de bilder som har digitaliserats inom ramen för 100 000 Bildminnen-projektet. I samband med släppet av dessa (planerad till våren 2024) planeras även en sista större uppladdning av bilder. Denna uppladdning ligger utanför ramen för denna rapport men vi ser redan nu att den stora mängden bilder kommer att medföra att flödet för uppladdning åter måste justeras något.

# Berikning

Att tillgängliggöra bildmaterialet på Wikimediaplattformarna fyller ett syfte genom att materialet blir tillgängligare för fler och därigenom även ser ytterligare vidareanvändning. Men det är också en plattform för crowdsourcing där materialet kan berikas med ny information. Detta kan ske både genom bidrag i fritext, strukturerad information eller med länkade data. Utöver detta kan berikningen ske genom redigeringar av själva bildfilen.

Fritextbidragen tar främst formen *bildbeskrivningar* på Wikimedia Commons samt *bildtexter* där bilderna används, till exempel på Wikipedia. Dessa bidrag kan ske på olika språk men kräver att man känner kontexten för att avgöra vilket språk det är.

Bidrag genom strukturerad information tar huvudsakligen form genom *kategorisering* av bilderna på Wikimedia Commons samt *captions*<sup>15</sup> på samma plattform. De kategorier som läggs till är på engelska men kan i vissa fall kopplas till Wikidata-objekt. De beskriver vad bilden avbildar eller den kontext under vilken den skapades. Captions är generaliserade bildtexter som även innehåller semantisk information om vilket språk de är på.

Den länkade data som läggs till bilderna kan ta flera former från att beskriva fotograf och när bilden togs till att beskriva var bilden togs och vad den avbildar. De koncept som används härstammar från Wikidata. Den vanligaste typen av denna berikning är *depicts*-påståenden, dvs. påståenden om vad bilden avbildar.

Inom ramen för detta projektet har vi huvudsakligen arbetat med fyra olika spår.

## Inledande matchning av metadata

I samband med att bilderna först laddas upp till Wikimedia Commons (se [Uppladdningen och förberedelser på Wikimedia Commons](#)) sker vanligtvis ett inledande försök att automatiskt matcha den metadata (i fritext) som levererats med bilderna mot koncept på Wikidata eller kategorier på Wikimedia Commons. Detta görs dels för att underlätta för gemenskapen att upptäcka för dem relevanta bilder, dels för att bättre kunna dra nytta av deras bidrag genom att använda den för insatser som inte kan göras automatiskt.

Eftersom den motivbeskrivning som fanns kopplad till var bild var samma för samtliga bilder i ett kuvert, visade det sig dock snabbt att det för denna samling inte gick att tillförlitligt användas dessa för att kategorisera bilderna i "riktiga" kategorier. Dessutom var det inte ovanligt att motivbeskrivningen beskrev den kontext eller uppdrag inom vilken bilderna skapats, snarare än vad deras motiv vilket ytterligare gjorde dem olämpliga för automatiserad metadata-matchning.

Det visade sig istället lämpligare att, som tidigare beskrivits, samla bilder med liknande motivbeskrivningar i kategorier inom projektets kategoriträd. Detta gav gemenskapen en ingång för att arbeta med bilderna och i vissa fall kunde hela denna kategori sedan läggas i en riktig kategori.

---

<sup>15</sup> *Captions* kallas *Bildtexter* på svenska, men för att undvika förväxling med de bildtexter som används på exempelvis Wikipedia använder vi här den engelska termen.

Vi ser exempelvis de bilder som alla har motivbeskrivningen *Flygvapnet Västerås*<sup>16</sup>. En logisk automatisk kategorisering av dessa, utifrån motivbeskrivningen, hade varit *Västerås*, men det är inte bilder på Västerås. En användare ur gemenskapen kunde istället manuellt koppla alla dessa bilder till Västmanlands flygflottilj (F 1 Hässlö), vilket inte hade varit möjligt att göra automatiskt.

Försöken att automatiskt matcha metadata gjorde istället behovet av crowdsourcing extra tydliga och detta blev fokus för våra fortsatta berikningsinsatser.

## Metadataskrivstuga

Inom ramen för detta projektet, och Wikipedia för hela Sverige, tog Wikimedia Sverige fram ett helt nytt koncept, Metadataskrivstugan<sup>17</sup>. Det är ett tillfälle där deltagarna får dela med sig av sin kunskap och berika bilder på Wikimedia Commons med strukturerad metadata och beskrivande bildtexter.



Ett av fotona som fick uppdaterad metadata och beskrivning under arrangemanget. Bilden föreställer sammanbyggnadsarbete av Saab 35 Draken i hall 2, i Saab-fabriken i Linköping. Flygplanen har en väv på sig med röd plastfärg på, som skydd under arbetet när arbetarna går på dem. Siffran på fenan beskriver individnumret för varje flygplan. Nummer 824 och 825 slutgodkändes 2 maj 1963, så fotot är taget tidigare samma år.

[Foto](#): K.W. Gullers / Nordiska museet, [CC BY-SA 4.0](#).

Under arbetet med uppladdningarna hade ett stort antal bilder av Saab i Linköping identifierats.<sup>18</sup> Därför ville vi samla människor med stark koppling till Saab, som jobbat eller jobbar där själva, som kunde berätta vad bilderna föreställde. På så sätt skulle vi direkt kunna lägga till depicts-påståenden, captions och beskrivningar till bilderna.

Hos Flygvapenmuseum i Linköping samlades därför ett drygt tiotal personer med koppling till Saabs flygplanstillverkning på orten. Bland annat var föreningen för Saab-veteraner och

<sup>16</sup> [https://commons.wikimedia.org/wiki/Category:100\\_000\\_Bildminnen/Flygvapnet\\_Västerås](https://commons.wikimedia.org/wiki/Category:100_000_Bildminnen/Flygvapnet_Västerås)

<sup>17</sup> <https://sv.wikipedia.org/wiki/Wikipedia:Skrivstuga/Metadataskrivstuga>

<sup>18</sup> [https://commons.wikimedia.org/wiki/Category:100\\_000\\_Bildminnen/SAAB](https://commons.wikimedia.org/wiki/Category:100_000_Bildminnen/SAAB)

Östergötlands Flyghistoriska Sällskap inbjudna till arrangemanget, som anordnades av Wikimedia Sverige i samarbete med Nordiska museet och Flygvapenmuseum.

Under eventet användes verktyget ISA Tool<sup>19</sup> för att både sätta depicts-påståenden till bilderna och att lägga till captions. Verktyget funkade bra för detta men sämre när flera bilder liknar varandra och i princip skulle kunna ha fått samma taggar och beskrivningar, eftersom informationen inte går att kopiera på något smidigt sätt.



En av deltagarna tittar på [ett foto av sig själv](#) från 1976. På det fotot genomför han undersökningar på en nosvinge från ett Saab 37 Viggen-flygplan, troligen i samband med haveriundersökningar som gjordes efter att flera flygplan av den modellen fått haverier. Han hade fått tips om arrangemanget efter att en bekant hade sett honom i tidningen efter Wikimedia Sveriges pressmeddelande.

Foto: Sven Rentzhog, [CC BY 4.0](#).

Höjdpunkten på eventet var att en man som syntes på ett av fotografierna själv var på plats under kvällen! Han kunde berätta många detaljer om vad han gjorde när fotot togs, och varför. Övriga deltagare fyllde på med information om både den och andra bilder. De berättade bland annat om lokalerna, flygplansmodellerna, utrustningen, syftet, vilka foton som visade faktisk verksamhet och vilka som var iscensatta.

Eftersom konceptet var framgångsrikt var planen att genomföra ytterligare två metadatataskrivsugor inom ramen för projektet varav ett med fokus på en annan samling. Tyvärr innebar kalenderkrockar och andra åtaganden att en av dessa inte kunde genomföras medan den andra slutligen genomfördes på SLU utan Wikimedia Sveriges eller Nordiska museets direkta inblandning.

## Veckans tävling

För att bjuda in den bredare Wikimediagemenskapen till engagemang med bilderna genomförde vi 27 november–3 december en *Veckans tävling*<sup>20</sup> på svenskspråkiga Wikipedia

<sup>19</sup> [https://commons.wikimedia.org/wiki/ISA\\_Tool](https://commons.wikimedia.org/wiki/ISA_Tool)

<sup>20</sup> [https://sv.wikipedia.org/wiki/Wikipedia:Veckans\\_tävling/Bildminnen](https://sv.wikipedia.org/wiki/Wikipedia:Veckans_tävling/Bildminnen)



med fokus på att både använda bilderna i Wikipediaartiklar samt att förbättra informationen om bilderna.

Veckans tävling är en fokuserad satsning med ett gamifieringsvinkel, där onlinegemenskapen arbetar med materialet under en vecka och man får poäng för olika insatser. Vinnaren, samt alla som får över ett visst antal poäng, får en digital utmärkelse. I detta fall fick pristagarna även priser vilka donerats av Nordiska museet. Även om tävlingen ägde rum på svenskspråkiga Wikipedia så omfattades lämpliga redigeringar på alla Wikimedia-plattformar. Vi marknadsförde tävlingen på svenskspråkiga Wikipedia, Wikimedia Commons samt sociala medier.

Ett viktigt betänkande när man sätter upp denna typ av tävling är att designa den på ett sätt som leder till konstruktiva redigeringar som är i linje med vad gemenskapen på det projektet skulle vilja se. Beroende på tävlingens form finns det olika taktiker för detta. I detta läge bedöms risken låg och var främst kopplad till att en användare skulle lägga in stora gallerier med bilder i artiklar, utan att ge bilderna någon kontext. För att motverka detta gavs extra poäng för bildtexter samt för bilder som lades till artiklar som tidigare saknade illustration.



Bildbeskrivning efter tävlingen (inte alla språk):

en: *The Brandenburg Gate at the Pariser Platz in Berlin, photographed from the West German sector. In the background, you see the Eastern part of the city.*  
sv: *En röd Volvo 144 framför Brandenburger Tor 1967.*

Bildbeskrivning innan tävlingen: *Tyskland Berlin, allmänt.*

Foto: K.W. Gullers / Nordiska museet, [CC BY-SA 4.0](#).

Uppslutningen i tävlingen var i linje<sup>21</sup> med andra *Veckans tävling* med totalt 8 deltagare varav 5 gjorde tillräckligt stora bidrag för att vinna pris. Inom ramen för tävlingen förbättrades 300 unika bilder (ca 6% av bilderna<sup>22</sup>). Utöver dessa deltagare såg vi även en ökad interaktion med bilderna av användare som inte skrev upp sig som deltagare i tävlingen. En mätning<sup>23</sup> av antalet förbättringar, under tävlingsperioden samt de efterföljande fyra dagarna, visar att fler än 900 unika bilder förbättrades, dvs. enbart 1/3 av förbättringarna rapporterades till tävlingen.<sup>24</sup>

Enbart 13 artiklar fick nya bilder inlagda vilket kan anses vara ett relativt lågt resultat. För att få en djupare förståelse för varför har vi ställt ett antal uppföljningsfrågor till några av deltagarna i tävlingen. Utöver denna återkoppling har vi även själva identifierat ett antal möjliga orsaker till detta:

- Det är svårare att hitta en lämplig artikel för en bild än att förbättra denna med beskrivningar eller kategorier. Detta hade delvis kunnat åtgärdas genom att ytterligare öka antalet poäng som illustration av en artikel gav.
- De tävlande utgick ofta från en eller en handfull batcher. Inom var batch är materialet ofta liknande så det kan vara svårt att identifiera mer än en bild som kan användas för att illustrera en artikel.
- Många av bilderna lutade mer åt konstnärlig vardagsfotografi vilket gör att de ofta inte är encyklopediskt relevanta.
- I flera batcher var det svårt att identifiera exakt vad som avbildades. Ju enklare något är att identifiera, desto troligare att det är känt nog att redan ha en illustration på Wikipedia.

Det deltagarna återkopplade som mest givande med tävlingen var att arbeta med bilderna och upptäcka detaljerna i dem. Det som var mest frustrerande var fel i metadata, antalet dubletter och snarlika bilder där samma information måste läggas till på flera bilder.

## Organiskt av gemenskapen

Genom att mer informellt engagera gemenskaperna på Wikimediaprojekten kring tematiskt avgränsat materialet har vi också sett goda resultat. Genom det rika mångfald av olika typer av gemenskaper på plattformarna har vi kunnat pröva detta både på geografiskt avgränsat material (Göteborg och Köpenhamn på två olika språkversioner av Wikipedia, se nedan) samt material fokuserat på flygplan (genom *WikiProject Aviation* på Wikimedia Commons<sup>25</sup>).

För batchen med bilder för [Göteborgsboken](#) frågade vi *Projekt Göteborg* på svenskspråkiga Wikipedia<sup>26</sup> om de kunde hjälpa till att identifiera bilderna. Bilderna inspirerade till en

---

<sup>21</sup> Snittet för de tio direkt föregående tävlingarna var 6,1 deltagare. Illustrationstävlingar har dock historiskt haft högre deltagande än snittet.

<sup>22</sup> Utifrån att 4832 bilder var uppladdade vid tävlingens start.

<sup>23</sup> Denna gjordes med hjälp av [Extraheringsverktyget](#).

<sup>24</sup> Statistiken för förbättringarna finns i [Appendix B](#).

<sup>25</sup>

[https://commons.wikimedia.org/w/index.php?title=Commons\\_talk:WikiProject\\_Aviation&oldid=825489795#Archival\\_Scandinavian\\_Airlines\\_photos\\_-\\_call\\_for\\_help](https://commons.wikimedia.org/w/index.php?title=Commons_talk:WikiProject_Aviation&oldid=825489795#Archival_Scandinavian_Airlines_photos_-_call_for_help)

<sup>26</sup>

[https://sv.wikipedia.org/w/index.php?title=Wikipediadiskussion:Projekt\\_Göteborg&oldid=52876257#TIPS:\\_nya\\_bilder\\_från\\_Nordiska\\_museet,\\_Göteborgsmotiv\\_att\\_identifiera](https://sv.wikipedia.org/w/index.php?title=Wikipediadiskussion:Projekt_Göteborg&oldid=52876257#TIPS:_nya_bilder_från_Nordiska_museet,_Göteborgsmotiv_att_identifiera)

fantastisk diskussion med mycket lokalkunskap, barndomsminnen av Göteborg på 40-talet och insikter som utmynnade i ett intensivt arbete för att försöka identifiera varje bild. Resultatet var att alla 110 bilder fick förbättrad bildinformation.

*Göteborgs hamn från Stenpiren till Fiskhamnen kände jag dels från söndagspromenader med min pappa, dels från egna turer tillsammans med en kamrat något eller några år senare.*

– Citat från en av deltagarna i Projekt Göteborg.

Från batchen med SAS-bilder identifierades en [delmängd av bilderna som avbildar Köpenhamn](#). För dessa frågade vi på danska Wikipedia om de kunde hjälpa till att identifiera vad de avbildar.<sup>27</sup> Utöver att identifiera vilken gata, torg etc. som avbildas på samtliga 149 bilder kunde gemenskapen även säga exakt när bilderna togs genom att identifiera att de sammanföll med Drottning Elizabeth II besök i Köpenhamn samma år. Ett fantastiskt exempel på lokalkunskap och engagemang.



Från de brittiska flaggorna framför stadshuset kunde man avgöra när bilderna togs.

[Foto](#): K.W. Gullers / Nordiska museet, [CC BY-SA 4.0](#).



Från samma bild kunde man även utvinna en illustration av en tidstypisk dansk buss av känd modell.

[Foto](#): K.W. Gullers / Nordiska museet, [CC BY-SA 4.0](#) (beskuren).

Vi har även sett hur gemenskapen själva organiskt har upptäckt materialet och gjort olika insatser kring att förbättra det. Utöver de typer av förbättringar vi redan belyst bör även särskilt nämnas insatserna kring att beskära och retuschera bilder. Oftast är syftet att ta bort skador samt den svarta ram som ligger runt bilderna för att de ska passa bättre in i en Wikipedia-artikel men det görs även mer krävande retuscheringar för att höja bilderna till statusen *Utvald bild* vilket markerar dem som en av de finaste bilderna på Wikimedia Commons.

<sup>27</sup>

[https://da.wikipedia.org/wiki/Wikipedia:Landsbybrønden/Historiska\\_foton\\_från\\_Köpenhamn\\_på\\_Commons](https://da.wikipedia.org/wiki/Wikipedia:Landsbybrønden/Historiska_foton_från_Köpenhamn_på_Commons)



Ett exempel på en bild av en SAS DC7 i luften som den först såg ut, samt till höger den version som tre användare tillsammans restaurerat och som därefter utsetts till *Utvald bild*.

Foto [v, h]: K.W. Gullers / Nordiska museet, [CC BY-SA 4.0](#) (bilden till höger är modifierad).

## Återförrel av information från Crowdsourcing

Även om det finns ett stort värde av att gemenskapen via olika insatser berikar det uppladdade materialet på Wikimedia's plattformar så realiseras det fulla värdet först när museet kan återföra detta material till sina egna plattformar (så kallad *roundtripping*). Samtidigt är det viktigt att ämnesexperterna vid museet själva äger sitt data och kan godkänna eventuella uppdateringar, alternativt att det material som inte kvalitetkontrollerats internt tydligt kan märkas upp. Detta är viktigt för museet trovärdighet, men också för Wikimedia-projekten för att förebygga så kallad *citogenesis*<sup>28</sup>.

## Svårigheter kopplade till att ta ut och förstå ändringar

Roundtripping har tittats på tidigare, bland annat i en rapport från Riksantikvarieämbetet från 2019.<sup>29</sup> Fokus här är därför på följande svårigheter.

**Ändringarna finns både som wikitext och länkad data (SDC).** Sedan 2019 finns två olika format för att lagra informationen som gemenskapen lägger till på filer på Wikimedia Commons. Även om de visuellt presenteras på samma plats utgör de två olika datamängder som både är olika i sin natur och i hur man kommer åt dem. En stor del förändringar som sker med wikitexten är dessutom kring kategorier som utgör en form av strukturerad, men olänkad data. En stor del av den länkade data som läggs till utgör information som maskinellt utvunnits ur wikitexten, så som upphovsman, tillkomstår med mera. Däremot finns det ingen mekanism som säkerställer att de två datamängderna (SDC och wikitext) inte börjar divergera eller rent av motsäger varandra.

**Ändringarna är inte statiska.** Förändringar av bildinformationen sker löpande; det är därför viktigt att kunna särskilja mellan vilka ändringar man redan tagit ställning till och vilka som tillkommit därefter. Eftersom användare ofta redigerar bara en av datamängderna kan detta resultera i att SDC och wikitexten påstår olika saker, även i de fall där samma SDC-data från början extraherades ur wikitexten.

<sup>28</sup> <https://xkcd.com/978/>

<sup>29</sup> [https://meta.wikimedia.org/wiki/Wikimedia\\_Commons\\_Data\\_Roundtripping](https://meta.wikimedia.org/wiki/Wikimedia_Commons_Data_Roundtripping)

**Information kan både läggas till och tas bort.** Ändringarna innebär inte bara att information läggs till, utan även att information kan tas bort. Detta sker om till exempel felaktig information avlägsnas men också (särskilt för kategorier och SDC) om information görs mer specifik/fördjupas. Ett typiskt fall är när en väldigt generell kategori (*Göteborg*) tas bort och ersätts med en mer informativ (*Eriksberg, Göteborg*). När denna fördjupning görs i en redigering är den enklare att identifiera som sådan, men ofta (alltid för SDC) sker fördjupningen genom två redigeringar där den fördjupande informationen läggs till och den gamla tas bort separat. Samma mönster ser vi även på vandalism som sedan återställs. En sista typ av informationsfördjupning som sakta tilltar är att wikitext raderas till förmån för SDC-data. Visuellt så tas ingen information bort, då SDC-data presenteras på den plats där wikitexten tidigare syntes, men för den som gör ett uttag av bara wikitexten saknas nu denna information.

**Informationsförädling är inte avgränsat till enbart Wikimedia Commons.** Bilderna som laddas upp utgör en resurs för alla Wikimediaprojekten och bildernas användning på dessa kan i sin tur utgöra informationsförädling. Exempel på detta är bildtexter som läggs till i samband med att bilden används i en Wikipediaartikel, eller att bilden används som illustration på ett väldigt specifikt Wikidataobjekt (t.ex. en person eller fordonsmodell). Detta är information som inte finns tillgänglig om man bara ser till bildens filbeskrivningssida på Wikimedia Commons.

**Vissa förädlingar är grafiska.** Utöver de mer vanligt förekommande förädlingarna finns det även förädlingar som är grafiska. Exempel på detta är restaurering/retuschering av bildfilen (för exempel se [Gemenskapen / Organiskt av gemenskapen](#)), uppladdningar av nya bilder härledda från originalbilden samt modern fotografering av samma objekt/vy, så kallad re-photography. Dessa kan vara svåra att identifiera, då resultatet oftast laddas upp som en separat fil, och även när länken mellan originalfilen och den nya filen är explicit kan det vara svårt att maskinellt avgöra relationen mellan dem.

## Extraheringsverktyg

Inom ramen för detta projekt tog Wikimedia Sverige fram en första version av ett verktyg, *Commons-diff*<sup>30</sup>, avsett att extrahera de förändringar som skett kopplat till en fil på Wikimedia Commons. Verktyget är byggt för att kunna användas av flera aktörer/institutioner och tillåter därför ett antal val kring vilka typer av förändringar som ska hämtas ut och levererar sedan denna data i ett standardiserat format. Förväntningen är att man från en viss institution dels ställer in vilka ändringar man vill ta ut, dels bygger ett tillägg som översätter standardiserade utdata till det format och struktur som är bäst lämpat för återbruk i det egna systemet.

Verktyget kräver att man åtminstone anger en kategori<sup>31</sup> samt ett startdatum. Det hämtar sedan ut alla förändringar sedan det datumet för alla filer som idag ligger i kategorin (dock ej i eventuella underkategorier). Om filen laddats upp efter det datumet så hämtas alla ändringar efter den första (där filen laddades upp)<sup>32</sup>. Detta kommer åt svårigheten med att

---

<sup>30</sup> <https://github.com/Wikimedia-Sverige/commons-diff>

<sup>31</sup> Man kan istället för en kategori välja att förse verktyget med en lista på de filnamn som ska behandlas.

<sup>32</sup> Då bilduppladdningen i praktiken kan bestå av flera separata redigeringar sätts startpunkten till den första redigering gjord av ett annat konto än det som laddat upp filen.

berikade data ej är statistiskt då en institution kan plocka ut förändringarna periodvis, och då bara de ändringar som uppstått sedan det förra uttaget.

Som svar på problemet med olika datamängder kopplade till filen låter verktyget användaren ange både vilka egenskaper (SDC) samt vilka fält i informationsmallen<sup>33</sup> (wikitext) som hen vill ta ut. Detta gör det även lättare att exkludera maskinellt utvunnen data om denna inte är önskvärd. Kategorier och *captions* tas alltid med och presenteras separat.

Att informationsförädling kan vara uppdelat i flera redigeringar angriper verktyget genom att den bara jämför informationen i början och slutet av tidsspannet. Detta medför en informationsförlust i metadata om bl.a. vem som gjort en viss ändring men erbjuder istället ett mer lättkonsumerat datauttag med fokus på den data som ändrats. Detta innebär även att vandalism som återställs exkluderas från uttaget.

Verktyget har en del kända begränsningar, några är ett direkt resultat av den information som finns tillgänglig via Wikimedia Commons.<sup>34</sup> Främst bland dessa är att om man utgår från en kategori så är det de filer som idag är medlemmar i den som behandlas, inte de som var medlemmar vid ett givet start-/slutdatum. Det går därför inte att säga om avsaknaden av en fil mellan två uttag beror på att filen har raderats eller bara tagits ur kategorin.

Informationsförädling från andra Wikimediaplattformar inkluderas ej i den data verktyget levererar, detta då det inte går att automatiskt få fram hur filanvändningen såg ut vid ett tidigare datum. Det går däremot att använda ett annat verktyg vi utvecklat (se [nedan](#)), som tar ut en ögonblicksbild av filanvändning samt de bildtexter som finns på dessa plattformar.

Verktyget ger inte heller någon information om grafiska förändringar, antingen i form av nya separata filer eller uppdatering av den aktuella filen. Till en senare version av verktyget planeras ett tillägg som åtminstone kommer att flagga om filen har skrivits över sedan sist.

## Återbruk

Tack vare *Commons-diff*, och de manuella insatser som gjordes med informationsförädling av pilotuppladdningen (se [Inledande pilotuppladdning](#)) kunde Wikimedia Sverige ta fram ett konkret exempel med verklig data och presentera detta för vår Kulturlt-kontakt på museet samt genom dem även för innehållsexperterna på Arkivet. Detta, samt tillhörande diskussioner, gav museet en möjlighet att bättre ta ställning till vilka informationsmängder som kan vara intressanta att återföra, hur dessa förhåller sig till de strukturer som idag används samt vilka närliggande informationsmängder (t.ex. etiketter och objektclasser från Wikidata) som behövs för att ge informationen tillräcklig kontext.

---

<sup>33</sup> I inställningarna kan man även ange vilken informationsmall som förväntas.

<sup>34</sup> Andra begränsningar som kan åtgärdas i senare versioner av verktyget inkluderar: Att välja bort kategorier och SDC-beskrivningar ur utdata; att begära ut all SDC-data; stöd för att flera olika informationsmallar används parallellt; att även inkludera filer i underkategorier (detta kräver idag att man använder andra verktyg för att få fram alla filer och sedan skickar listan med alla dessa till Commons-diff).

Något som blev tydligt var att Wikidata-koncepten (via SDC) var lättare att hantera än exempelvis Wikimedia-kategorier<sup>35</sup> då de senare ofta kombinerar olika typer av koncept samt inte är flerspråkiga (oftast, men inte nödvändigtvis, är de på engelska). För att konsumera Wikidata-koncepten behövs dock även tillgång till deras etiketter på Wikidata. Då Nordiska museet i sin egna metadata skiljer mellan avbildade platser, personer och övriga koncept är det också viktigt att kunna identifiera vad för objektclass<sup>36</sup> ett Wikidata-koncept är.

Återkoppling var nyttig för att förbättra Commons-diff och resulterade i en del direkta ändringar som bl.a. kraftigt minskade antalet flaggade ändringar som i själva fallet bestod av information som redan var känd av museet. Återkopplingen är även helt central för att ta fram det tillägg som omvandlar utdata till ett format som Nordiska museet själva kan konsumera. På grund av att kritisk personal hade flera andra parallella åtaganden gavs det inte möjlighet att, inom ramen för projektet, helt färdigställa detta tillägg.

Flödet för hur förbättringarna i någon större skala ska kunna återföras är dock inte tydligt. För att den ska integreras i museets samlingsdatabas, Primus, krävs att personal manuellt stämmer av ändringen, vilket i sin tur kräver dedikerade personalresurser. Ett förslag för att ändå göra förbättringarna tillgängliga är att de istället visas i den publika plattformen för samlingarna, DigitaltMuseum, och där det då tydligt märks upp som förbättringar från Wikimedia Commons. Genom att göra informationen tillgänglig här skulle den även bli tillgänglig för den personal som arbetar direkt i samlingsdatabasen. Det skulle dessutom öka allmänhetens kunskap om Wikimedia Commons och om att vem som helst kan bidra med egna förbättringar.

## Statistiska mått

Det finns en uppsjö med statistik vi kan ta fram kring de bilder som har laddats upp. I detta stycke har vi fokuserat på vilka mått som redan finns, samt dess styrkor och svagheter. Vi tittar även på saknade mått och belyser några mått vi tagit fram inom ramen för projektet. Några av resultaten för projektet kommer att belysas här, övrig statistik återfinns i [Appendix B](#).

## Bild- och metadatauppladdning

*6014 bilder laddades upp i 6 batcher med totalt 32 teman.*

Genom att samla alla bilder som har laddats upp inom projektet i en metakategori<sup>37</sup> kan vi direkt från plattformen läsa ut antalet bilder som har laddats upp. Genom att ytterligare märka upp bilderna med metakategorier utifrån vilken batch de har laddats upp i, eller vilket tema de avbildar går det lättare att i ett senare steg dels segmentera statistik utifrån vilket material det innehåller samt vilka insatser som genomförts. Det blir även enklare att ta höjd för skillnaderna i antalet visningar (se [Filvisningar](#)) utifrån hur länge bilderna har legat uppe.

---

<sup>35</sup> Även om vissa kategorier kan kopplas till Wikidata-koncept är det ofta så att förbättringar av filens kategorisering kan leda till att den hamnar i en mer specifik kategori vilken inte längre har en koppling mot Wikidata.

<sup>36</sup> Dvs. att plocka ut konceptets *instans av*-påstående ([P31](#)) och därigenom klassificera det.

<sup>37</sup> [https://commons.wikimedia.org/wiki/Category:100\\_000\\_Bildminnen](https://commons.wikimedia.org/wiki/Category:100_000_Bildminnen)

Svagheten i systemet med att utgå från kategorier är att det inte finns något hinder för att någon plockar bort en sådan kategori från någon av filerna, och inte heller ett lätt sätt att i efterhand upptäcka att detta har gjorts. Det samma gäller om en fil läggs till i kategorin eller om en fil raderas helt. Denna typ av ändringar tillhör dock ovanligheten och vi borde kunna fränse effekten av detta på statistiken.

Något som däremot sker, men infrekvent är att ett derivat av en fil, till exempel en beskuren version, laddas upp och då (vanligen) ges samma kategorier som ursprungsfilen. Den negativa påverkan på statistiken från detta, samt om filen i sig uppdateras, (t.ex. att en fil inte ingick i kategorin under en tidsperiod eller en insats som man undersöker) kan dock ses som försumbar. Värdet är då större derivaten ändå ingår i samma urval då de ofta skapas just för ett vidareanvändningsbehov. För mätning av förekomsten av derivat se stycket nedan om statistik för [Interaktioner, förbättringar och engagemang](#).

Filuppladdningen kombineras med en metadatauppladdning både som wikitext och länkade data. Att få ett mått över denna informationsmängd är viktigt både för att kunna särskilja denna informationsmängd från den som tillkommit genom senare insatser samt för att få en bättre förståelse för hur stor del av den totala informationen som utgörs av ursprunglig data. Ett förslag på ett enkelt mått per bild skulle vara att ta antalet SDC-påståenden, antalet kategorier samt antalet tecken wikitext på filens beskrivningssida. En brist med det måttet är att information kan läggas till utan att de nödvändigtvis ändras.

I praktiken kompliceras processen dock av att alla metadata inte läggs till i samma redigering (p.g.a. begränsningar i mjukvaran) samt att delar av den data som laddas upp kan bli tillgängligt först i efterhand<sup>38</sup>. I detta projekt har alla filuppladdningar, samt uppladdningar av ursprunglig data, genomförts av kontot *Alicia Fagervig (WMSE)* och vi klassificerar därför alla redigeringar från detta konto som ursprunglig data. Vad detta upplägg missar är om vi i efterhand genomför någon större åtgärd utifrån inkommen återkoppling från gemenskapen eller om vi med det kontot genomför en redigering på begäran av en deltagare på ett event som inte själv har möjlighet att redigera (se [Metadataskrivstuga](#)).

Inget verktyg för att ta fram detta mätetal finns idag och det bedömdes inte som prioriterat att ta fram inom ramen för projektet, även om det är förhållandevis okomplicerat.

## Filanvändning

*180 unika bilder används 238 gånger på 57 sidor över 7 plattformar.*

Wikimediaplattformarna ger ett gott stöd för att ta fram hur en fil just nu används inom Wikimedia's ekosystem av plattformar. Vill man göra detta för en större uppsättning filer är det dock rekommenderat att använda ett verktyg såsom *GLAMorous2*<sup>39</sup> vilket kan ta fram en överblick av antalet unika filer som används, det totala antalet filanvändningar samt antalet sidor och plattformar som detta skett över.<sup>40</sup> Att få en överblick av vilka plattformar som används gör det möjligt att separera användning avsedd för en extern publik (så som i artiklar på Wikipedia) från användning i t.ex. interna projektrapporter (ytterligare filtrering kan

---

<sup>38</sup> I vårt fall gjordes inledningsvis DigitaltMuseum-kopplingen först senare då filerna först publicerades på Wikimedia Commons.

<sup>39</sup> <https://glamtools.toolforge.org/glamorous/>

<sup>40</sup> Verktyget *wp\_captions* (se [nedan](#)) tar fram samma statistik i samband med att den körs.



göras genom att sortera ut bildanvändning utanför huvudnamnrymden, t.ex. i ofärdigt arbetsmaterial eller på diskussionssidor).

Vad Wikimedia Commons inte ger något stöd för är att kunna spåra filanvändning utanför Wikimediaplattformarna.

Det går inte heller att idag få fram historisk filanvändning. Detta innebär att man löpande måste göra datauttag för att kunna följa hur användningen utvecklas. Detta är extra viktigt inför en insats såsom den [Veckans tävling](#) som genomfördes med syfte att öka filanvändningen.

## Filvisningar

*100 000 Bildminnen fick 100 000 visningar på en enda dag och över 800 000 visningar under 2023.*

Antalet gånger var fil har visats är den statistik som ofta rapporteras då synlighet oftast är ett av de främsta målen för att dela material på Wikimedia Commons och inkludering av en fil i en populär artikel på Wikipedia snabbt kan leda till väldigt stor synlighet.

De vanligaste verktygen för att ta fram statistik för detta är BaGLAMa2<sup>41</sup>, GLAMorgan<sup>42</sup> och GLAMorous2<sup>43</sup>. Alla tre utgår från vilka sidor filerna används på (se [Filanvändning](#) ovan) vid ett givet tillfälle och summerar månadens visningar för den sidan. Siffran inkluderar enbart besök från människor och exkluderar botten från t.ex. sökmotorer.

För GLAMorgan och GLAMorous2 utgår den från filanvändningen vid det tillfälle då statistiken begärs. Detta innebär att den visningssiffran man får tillbaka inte är kopplad till om filen används på sidan under den månad som statistiken efterfrågas och det går till och med att få "filvisningar" för en månad långt före det att bilderna laddades upp<sup>44</sup>. För BaGLAMa2 utgår den från filanvändningen vid ett slumpvis valt tillfälle under den månad för vilken statistiken efterfrågas, men alla sidvisningar tillskrivs fortfarande filen, oavsett om den var med på sidan hela månaden eller ej vilket är relevant om den t.ex. har lyfts fram på Wikipedias startsida under en kortare period. Utöver detta ger statistiken inte heller en indikation på om bilden verkligen sågs av den som besökte sidan (om den t.ex. finns långt ner i artikeln) och den missar även visningar från andra webbplatser som bäddat in filen från Wikimedia Commons samt nedladdning av filen direkt från Wikimedia Commons.

Wikimedia tillhandahåller funktionalitet för att tillhandahålla det faktiska antalet filvisningar för en given dag eller månad genom att utgå från det faktiska antalet gånger som filen<sup>45</sup> begärts från Wikimedia Commons servrar. Denna statistik, kallad *media-views* eller *mediarequests* går dock enbart att begära per fil och används ännu inte i några öppna verktyg<sup>46</sup>.

<sup>41</sup> <https://glamtools.toolforge.org/baglama2/>

<sup>42</sup> <https://glamtools.toolforge.org/glamorgan.html>

<sup>43</sup> <https://glamtools.toolforge.org/glamorous/>

<sup>44</sup> Se exempelvis [denna sökning för visningar under augusti 2019 av de bilder som laddats upp inom projektet under 2023](#).

<sup>45</sup> inkl. nedskalade versioner av filen.

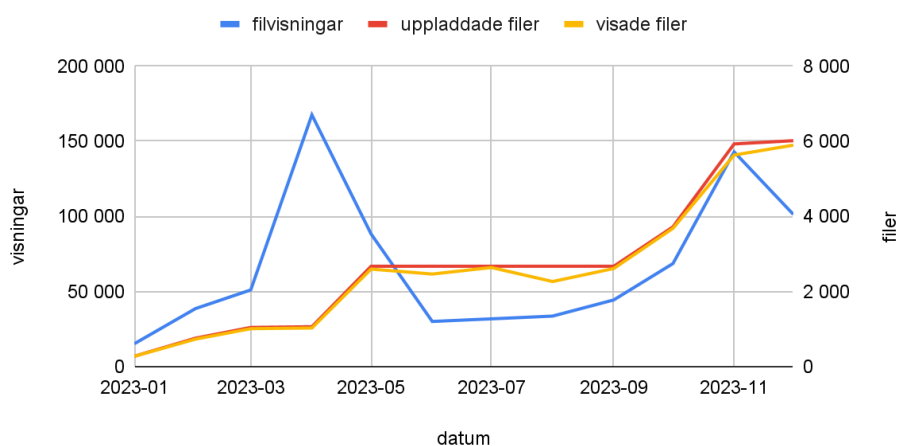
<sup>46</sup> *GLAM Wiki Dashboard* använder sig av denna statistik men kräver idag att man som GLAM-institution först registrerar sig. Fokus för verktyget är på institutionens samtliga bilder, och det går inte att få statistik för en specifik insats eller projekt. Se till exempel <https://glamwikidashboard.wmcloud.org/NDM> för Nordiska museet.

I november 2023 visades 93% av projektets bilder minst en gång.

Inom ramen för projektet har vi därför tagit fram ett litet verktyg, *mediaviews*<sup>47</sup>, som kan ta ut *mediarequests*-data för bilderna i en kategori på Commons för ett givet datumspann. Verktöget är flexibelt i att den även stödjer underkategorier samt att man kan välja att segmentera data per fil, dag eller månad. Verktöget har visat sig vara stabil även för att ta fram statistik för alla Nordiskas 20 000 bilder.

## Filvisningar 2023

För bilder i 100 000 Bildminnen



*Antalet filvisningar, visade filer och uppladdade filer<sup>48</sup> per månad under 2023 för filerna i kategorin [100 000 Bildminnen](#). Spiken för visningar i april syns tydligt.*

När verktöget användes för att ta fram visningarna för projektets bilder upptäcktes en tydlig spik i statistiken för april. Genom att segmentera först per dag, och sedan även per bild, kunde vi se att det den 13 april var över 100 000 visningar, och att dessa var spridd över ca 1 000 bilder. Detta kunde kopplas direkt till ett pressomnämmande (se [Synligheten](#)) som uppmanade läsarna att utforska bildskatten på Wikimedia Commons. Detta är således visningar som de traditionella verktygen inte skulle ha plockat upp.

Eftersom verktöget även kan användas för att ta fram historisk data kan vi jämföra det värde som *mediaviews* ger mot motsvarande siffra från BaGLAMa2 för att få en indikation på hur missvisande siffror från de sidvisningsbaserade verktygen kan vara. När vi gjorde denna jämförelse för samtliga av Nordiska museets bilder under 2022<sup>49</sup> gav den sidvisningsbaserade statistiken visningssiffror som var 35-50% högre än antalet faktiska mediavisningar.

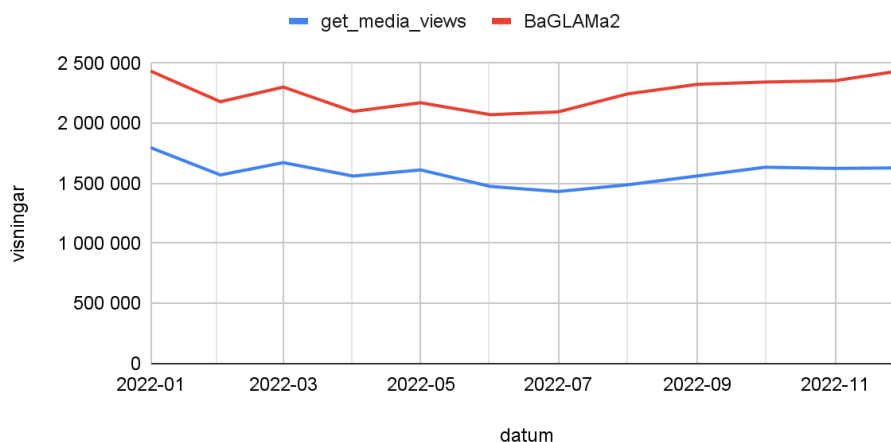
<sup>47</sup> Se *mediaviews* på <https://github.com/Wikimedia-Sverige/nordiska-2022>.

<sup>48</sup> Denna data togs fram med verktöget *file\_count*, se <https://github.com/Wikimedia-Sverige/nordiska-2022>.

<sup>49</sup> 2022 valdes snarare än 2023 då BaGLAMa2 tappat data under 2023.

## Filvisningar 2022

För Images from Nordiska museet



Antalet filvisningar per månad under 2022 för filerna i kategorin [Images from Nordiska museet](#). Mätt med verktygen *mediaviews* samt *BaGLAMa2*.

## Interaktioner, förbättringar och engagemang

Ett av målen med projektet, och särskilt med insatserna, är att skapa ett ökat engagemang med filerna. Detta kan dels ta formen av att filerna används på Wikipedia och övriga plattformar (se [Filanvändning](#)), dels genom att användare lägger till information om bilden på dess filbeskrivningssida på Wikimedia Commons (antingen som captions eller bildbeskrivningar), dels genom att det skapas derivat av filen eller att filen retuscheras. Vi skiljer på vanliga redigeringar och bot-redigeringar, hel- eller halvautomatiserade redigeringar som genomförs av gemenskapen men från särskilt uppmärkta konton. Olika plattformar, och ibland olika informationsmängder på samma plattform, har olika regler om var gränsen går för när halvautomatiserade redigeringar måste märkas som bot-redigering.

*68 användare har totalt redigerat filerna 3492 gånger.<sup>50</sup>*

Det finns idag ett verktyg, *Catanalysis*<sup>51</sup>, som utifrån en kategori (i vårt fall [Category:100 000 Bildminnen](#)) kan ta fram statistik för både antalet användare som har redigerat materialet, samt det totala antalet redigeringar som gjorts samt om dessa utförts av en bot eller ej. Att kunna exkludera botar är viktigt då de huvudsakligen maskinellt bryter ut informationsmängder från wikitexten till SDC. Från denna statistik kan vi även exkludera de användarkonton som härstammar från Wikimedia Sverige (märkta med *WMSE*), däribland kontot som används för den initiala uppladdningen. Verktöget tittar dock enbart på interaktioner på Wikimedia Commons och missar därför de användare som interagerat med materialet på andra plattformar genom t.ex. filanvändning eller genom att lägga till bildtexter direkt på de plattformarna. Informationen från *Catanalysis* går även att få segmenterad per månad och kan därför jämföras mot genomförda initiativ.

<sup>50</sup> Exklusive bot-redigeringar, WMSE-redigeringar samt redigeringar av icke-inloggade användare.

<sup>51</sup> <https://meta.toolforge.org/catanalysis/>

*2158 filer (35%) har redigerats och totalt fått 2990 nya kategorier, 348 nya captions på 8 olika språk, 309 nya depicts-påståenden och 166 förändrade bildbeskrivningar.*

Det extraheringsverktyg för förbättringar<sup>52</sup> som togs fram (se [Extraheringsverktyg](#)) ger en sammanställning av all information som lagts till sedan filen först laddades upp. Utifrån denna data kan vi ta fram statistik över både mängden information som har lagts till samt vilken typ av information som har tillkommit. Detta inkl. antalet kategorier som har lagts till, antalet depicts-påståenden, antalet bildbeskrivningar och captions samt vilka språk dessa är på. Vad extraheringsverktyget däremot inte gör är att särskilja mellan vanliga användare och bot-konton. Genom att köra verktyget direkt efter en insats<sup>53</sup>, och sätta startdatum till direkt innan insatsen får man en god inblick i utfallet. Verktyget användes exempelvis för att få en bättre inblick i de som engagerats genom [Veckans tävling](#) men valt att inte anmäla sig till tävlingen. Statistik är dock inte detta verktygs huvudsakliga användningsområde.<sup>54</sup>

*252 filer har fått minst ett depicts-påstående.<sup>55</sup>*

För att specifikt se till den länkade data som har lagts till på filerna kan man använda sig av SPARQL<sup>56</sup>-frågor för Wikimedia Commons<sup>57</sup>. Utifrån detta kan man t.ex. titta på andelen bilder som har minst ett depicts-påstående eller vilka bilder som är geositionerade. Verktyget är dock i beta och kräver därför att man har ett Wikimediakonto, både för att skapa frågor och ta del av resultaten.

*11 nya filer som skapats utifrån det uppladdade materialet kunde identifieras.*

Ett mätetal som vi idag inte har tillgång till är antalet härledda filer som skapats samt antalet filer som har retuscherats. Retuscherade filer och derivat som ligger i huvudkategorin för projektet borde vara relativt enkla att få fram utifrån vilket konto som har laddat upp dem. För derivat som inte ligger i kategorin blir det betydligt svårare. Det finns stöd i SDC för att ange dessa relationer men i dagsläget används det i praktiken inte då gränssnittet ej stödjer tillägg eller visning av detta. Det finns även en möjlighet att man utifrån en analys av länkarna mellan filbeskrivningssidorna kan härleda derivat. I vårt fall, där ursprungliga data inte innehåller länkar mellan filerna, fungerade denna analys och kunde ta fram dessa länkar utifrån vilka derivaten lätt kunde identifieras.<sup>58</sup>

*33 bilder fick bildtexter på 5 olika språkversioner av Wikipedia.*

Slutligen har vi tagit fram ett verktyg, *wp\_captions*<sup>59</sup>, som tar fram de bildtexter som har lagts till direkt på plattformarna där filerna används. Antalet bildtexter kan sättas i relation till

<sup>52</sup> Notera att verktyget exkluderar precis den "ursprungliga data" som diskuterats i [Bild- och metadatauppladdning](#).

<sup>53</sup> I en framtida version ska verktyget även få stöd för "slutdatum" så att statistik även kan tas fram för tidigare insatser.

<sup>54</sup> Ett litet verktyg för att ta fram statistiken från extraheringsverktygets resultatfil finns som *diff\_stats* på <https://github.com/Wikimedia-Sverige/nordiska-2022>.

<sup>55</sup> <https://w.wiki/7eZM> (kräver Wikimedia-konto)

<sup>56</sup> SPARQL är ett frågespråk som används för länkade datamängder.

<sup>57</sup> <https://commons-query.wikimedia.org/>

<sup>58</sup> Detta gjordes med verktyget *deriv\_detector*, se <https://github.com/Wikimedia-Sverige/nordiska-2022>.

<sup>59</sup> Se *wp\_captions* på <https://github.com/Wikimedia-Sverige/nordiska-2022>.

mängden filbeskrivningar och captions som lagts till på Wikimedia Commons, samt kompletterar bilden över vilka språk materialet har interagerat med. En begränsning i verktyget är att den underliggande tjänsten<sup>60</sup> för att ta fram bildtexter enbart är aktiverad på en delmängd av Wikimedias plattformar. Detta inkluderar alla språkversioner av Wikipedia, där vi förväntar oss att användningen av bildtexter är störst. Man bör dock ha denna begränsning i åtanke om man jämför antalet bildtexter med siffrorna från [Filanvändning](#).

## Övrig statistik

Från de event och insatser vi har genomfört har vi samlat ytterligare statistik. Denna omfattar antalet event, antalet deltagare på eventet, typen av event samt ett mått på de förbättringar som genomfördes under själva eventet.

Vi har även fört en logg över de tillfällen då projektet (eller filerna) fått synlighet genom omnämmande i media, publicering av blogginlägg eller i nyhetsbrev (se [Appendix C](#)). Där detta har skett via våra plattformar har vi tillgång till visningsstatistik. Där detta skett via andras plattformar har vi sällan sådan statistik. Vi kan dock jämföra dessa insatser med visningsstatistiken för bilderna för att se om den har gett något utslag där.

## Segmentering

Rubrikerna ovan tittar på en del av den statistik och statistiska mått som redan fanns tillgänglig eller som har tagits fram genom projektet. För att bättre förstå *varför* vi ser vissa resultat skulle det vara intressant att segmentera statistiken utifrån batch eller insats för att få en bättre förståelse över vilken effekt olika insatser har haft. Här kan vi både utgå från vilka filer som omfattades av en viss insats, samt se till när i tiden insatsen låg och titta på visningar/interaktioner i samband med detta.

Detta är både intressant för att få en bättre förståelse för vilka insatser som har gett störst effekt men det kan även vara intressant att se om olika typer av insatser har resulterat i olika typer av interaktion (SDC vs. bildbeskrivningar vs. bildanvändning). Detta ger ett bättre underlag för att i framtiden välja rätt insats utifrån önskat resultat.

I en mindre utsträckning testades en sådan segmentering för [Veckans tävling](#) samt för att koppla spiken i filvisningar till mediaomnämmande inför [Metadataskrivstuga](#). För att dra några verkliga slutsatser skulle det dock fler individuella insatser behöva genomföras och kombineras med bättre dokumentation kring vilket material som var i fokus för insatsen.

---

<sup>60</sup> Den underliggande tjänsten levererar inte heller de bildtexter som förekommer i bildgallerier. Verktyget har därför utökats med funktionalitet för att även även extrahera dessa.

## Synlighet och kommunikation



Bilden till blogginlägget [En get är en get är en Q2934](#) som beskriver hur bildmaterialet kan berikas genom koncept från Wikidata.

**Foto:** K.W. Gullers / Nordiska museet, [CC BY-SA 4.0](#) (modifierad av Alicia Fagerving).

Utöver det arbete som nämnts ovan har vi även genomfört kommunikativa insatser för att öka synligheten av materialet och projektet. En komplett lista över dessa återfinns i [Appendix C](#).

En självklar målgrupp för vår kommunikation har varit Wikimediagemenskapen, och då särskilt den del som arbetar med GLAM (ABM-sektorn). Dessa har primärt hållits informerade genom det internationella nyhetsbrevet *This Month in GLAM*<sup>61</sup> där vi i 10 av de 17 månadsuppdateringar som skapats under projektets gång löpande har delat med oss av uppdateringar kring projektet. På den årliga världskonferensen för Wikimediarelsen, Wikimania, gav vi också en presentation om projektet.<sup>62</sup> Vårt experiment med en Metadataskrivstuga beskrev vi även på Diff<sup>63</sup>, Wikimedia-gemenskapens blogg.

Ytterligare en målgrupp har varit den allmänhet som dels kan vara intresserad av materialet och i bästa fall skulle kunna aktiveras till att dela med sig om sin kunskap av det. Här visade

<sup>61</sup> Outreach Wiki – [GLAM/Newsletter](#)

<sup>62</sup> Wikimania 2023 – [100 000 Photo Memories – museums and Wikimedians in synergy.webm](#)

<sup>63</sup> Diff – [SAAB Veterans Shared Their Knowledge at Metadata Workshop](#)

sig bildmaterialet vara extra tacksamt för ett *Bildextra* som Corren satte ihop<sup>64</sup> inför Metadaskrivstugan på Flygvapenmuseum, efter en pressrelease från oss. Efter publiceringen hörde Corren av sig för att meddela oss att det var en av deras mest välbesökta artiklar. Parallellt med detta kunde vi även se hur denna artikel fick en enorm påverkan för antalet visningar av bilderna (se [Filvisningar](#)). Detta är extra intressant då det blir ett bevis på att insatsen inte bara resulterat i synlighet bland Correns läsare, utan även har fått dem att ta sig vidare till Wikimedia's plattformar för att ytterligare utforska materialet.

## Sammanfattning, slutsatser och framtida arbete

I detta projektet tittade vi på flera olika steg av arbetsflödet vid ett tillgängliggörande av innehåll på Wikimedia Commons:

- urval av material att ladda upp;
- bearbetning och sammanjämkning av metadata;
- uppladdning;
- berikning och återanvändning;
- återförsl av berikad information;
- analys av resultatet.

Några ovanliga omständigheter har särskilt påverkat projektet.

- Att Wikimedia Sverige deltog från projektets start, innan digitaliseringen ens startat.
- Att materialet enbart hade en väldigt begränsad mängd metadata om vad som avbildats och det sammanhang i vilket det skapats.
- Att detta utvecklingsprojekt var en del av ett räddningsprojekt, där materialet ibland varit skadat.

Projektet visade tydligt att insatsen blir mer tidskrävande när det finns mindre metadata om filerna. Detta var särskilt framträdande om man önskar att använda en delmängder av filerna i riktade initiativ där de t.ex. bör vara tematiskt eller geografiskt avgränsade. I projektet testades ett antal olika angreppssätt för att balansera tidsåtgång och hur välintegrerat materialet blir på Wikimedia Commons. De hade alla olika styrkor utifrån t.ex. antal bilder och behov av avgränsningar men ingen lösning med universal passform hittades.

Genom projektet utvecklades och testades ett nytt koncept för skrivstugor med speciellt fokus på bilder. Konceptet är väldigt lovande och det finns redan planer på att upprepa det igen utanför projektet. Vi genomförde även olika typer av initiativ från interaktion med dedikerade delar av gemenskapen till en bredare tävling. Vi utökade samtidigt de möjligheter vi har till att utvärdera utfallet från dessa insatser. Projektet var lyckat i att pröva dessa initiativ samt ge dem en inledande utvärdering, det skulle dock ha krävts att samtliga insatser upprepas flera gånger för att vi skulle kunna dra en klar slutsats om vilken som är mest lämpad för ett önskat utfall.

I projektet togs ett verktyg fram för att extrahera den berikning av metadata som skett hos bilderna för att kunna göra denna tillgängliga för GLAM-institutionen. Initiala resultat visar att verktyget kan utvinna information som skulle vara värdefull, och möjlig, att integrera i GLAM-institutionens egna samlingssystem. För att kunna dra full nytta av detta, utanför ramen av ett utvecklingsprojekt, kräver det dock att institutionen har avsatt dedikerade

---

<sup>64</sup> Corren – [BILDDEXTRA: Bildskatt från Saab i Linköping görs tillgänglig](#)

personalresurser till att gå igenom och bekräfta den inkomna informationen. Alternativet är att titta på automatiserade sätt att göra informationen tillgänglig för brukarna av samlingssystemets data, men separat från övrig data och med en tydligt uppmärkt härkomst. Flera förslag kring vidareutveckling av extraheringsverktyget presenteras i rapporten och då verktyget inte är knutet till ett specifikt projekt eller GLAM-institution finns goda möjligheter för att verktyget ska kunna fortleva och utvecklas efter projektets slut.

Projektet kunde för första gången tydligt påvisa en koppling mellan extern kommunikation och faktiskt visningsresultat för bilderna. Detta öppnar upp för förbättrad analys av effektiviteten i framtida kommunikationsinsatser. Det verktyg som utvecklades för detta tar fram visningsstatistik som är mer rättvisande än den som vanligen redovisas. Genom att applicera verktyget på historisk data ges en indikation på i vilken utsträckning tidigare redovisad statistik har över- eller underskattats. Verktyget skulle också kunna integreras i en mekanism för att regelbundet ta fram ny visningsstatistik för en GLAM-institution eller ett projekt.

Projektet var en unik möjlighet för Wikimedia Sverige att kunna vara med redan i startgroparna för ett digitaliseringsprojekt. Det tidiga utbytet, utbildningarna och besöken gav båda parter en ökad förståelse för varandras möjligheter och omständigheter vilket kraftigt underlättade i senare diskussioner. Gemensamma diskussioner om hur metadatastrukturerna för digitaliseringen skulle se ut säkerställde att data, när det senare började produceras, fungerade för båda parter. Att förändra dessa strukturer senare hade varit väldigt kostsamt och en struktur som hade fungerat sämre för Wikimedia Sverige hade troligen krävt ytterligare manuella insatser med ett slutligen sämre resultat. I framtida initiativ, där digitaliseringen löper under en längre period, och där det finns mer kunskap om vad materialet kommer att innehålla, skulle man kunna dra ytterligare nytta av samarbetet genom att även låta återkoppling från Wikimediaplattformarna påverka digitaliseringsprocessen, exempelvis genom att väga in i prioriteringen av i vilken ordning materialet digitaliseras.

Avslutningsvis har samarbetet och projektet varit väldigt givande och lagt goda grunder för framtida samarbeten som ytterligare kan utforska några av de resultat som ges här.



Ett urval av bilder som digitaliserats i projektet, här från Saab i Linköping, Finspång i vinterskrud samt Lantbrukshögskolan i Ultuna.



# Appendix

## Appendix A: Uppladdningsbatcher

2022, juni

- [Pilotuppladdning](#)
  - Skidstaffett och Howard Frank

2023, januari

- [Nordiska Kompaniet](#)
  - Butiksinteriörer, kunder, föremål.

2023, februari-mars

- [Borås](#)
  - Arbete på Borås Väveri och Borås Yllefabriks AB. Några stadsvyer.
- [Finspång](#)
  - Stadsbilder i snöskrud. Några få okända personporträtt.
- [Helsingborg](#)
  - Serie *Hos Runar Karlströmer i Hälsingborg*, personer, hemmiljö.
- [Norrköping](#)
  - Skörde- och jordbruksmaskiner. Journalist P. Wange, troligtvis från reportage, oklart om det är om eller av honom. Några stadsbilder.
- [SAAB](#)
  - Bilder från fabriken, både personer och maskiner. Några bilar. Flygplan i luften.
  - I maj redigerades dessa under [skrivstugan i Linköping](#).
- [Trollhättan](#)
  - Industrimiljöer: AB Tacksfabriken och konf-fabr. AB Viking, SAAB Verkstäder, Stridsberg & Biörcks industri.
  - Generellt [blogginlägg om projektet](#).
- [Ulltuna](#)
  - Bilder från Lantbruksuniversitetet. Vetenskapliga maskiner, djur, odlingar. Många okända personer.
- [Vårgårda](#)
  - Arbete på Vårgårda Havregryn. Industrimaskiner, havre och havreprodukter, en del okända arbetare.

2023, maj

- [Göteborgsboken](#)
  - Byggnader och gator, sevärdheter. Hamnbilder med båtar och arbetare.
  - Notis på swwp, [Projekt Göteborg](#)
- [Ingrid Bergman](#)
  - En svensk skådespelare i arbete, medeltida utklädnad.
- [Mölnlycke](#)

- Industrimiljöer. En del beskrivna som *Mölnlycke Väverier*, andra utan specifikation så oklart om de också är därifrån.
- [SAS Scandinavian Airlines](#)
  - Flygplan i luften. Interiörbilder från flygplan. Flygplatser (på marken och i vänthallen). Resebilder från Köpenhamn.
  - Vi [bad den danskspråkiga](#) gemenskapen att förbättra [Köpenhamnsbilderna](#).
  - Vi nådde ut till [Wikiproject Aviation](#) på Commons. En del hjälp med att kategorisera bilderna. Anmärkning att det är väldigt mycket material.

2023, september-oktober

- [Flygflottiljen Tullinge](#)
  - Flygplan och flygplatser. Många personbilder. Några bilder inomhus, kontorsmiljö.
- [Flygvapnet Västerås](#)
  - Flygplan. Många personbilder, både porträtt och gruppbilder. Aktiviteter i skogen. Några interiörbilder i militär miljö.
- [Drottningholm](#)
  - Arkitektur och vattendrag. Några bilder på en folksamling.

2023, november

- 27 november inleddes Veckans tävling med tillhörande publicitet.
- [Astra Göteborg](#)
  - Arbete i kontors- och laboratoriemiljö. Några bilder på läkemedel. Astra var ett läkemedelsföretag.
- [Australien](#)
  - Bilder av turistkaraktär från oidentifierade ställen: arkitektur, landskap, stadsbilder, natur, stränder.
- [Keflavíks flygplats](#)
  - Människor i ett väntrum på flygplatsen.
- [Lionel Hampton](#)
  - Ett uppträdande på ett konserthus, okänt var. Många oidentifierade musiker.
- [Moskva](#)
  - Bilder av turistkaraktär: kyrkor, statyer och gator. Gatubilder med folksamlingar.
- [Bilar](#)
  - Bilar (Förenade Bil, BMW, Svenska Bil AB, Renault) i naturen. Plats identifierad på landskapsnivå: Roslagen, Skåne, Västkusten, Småland.
- [Flygbilder](#)
  - Svartvita bilder på kustområden, med grovt identifierat motiv (Marstrand, Falkenberg, Lysekil m.m.).
- [Skansen](#)
  - Mestadels barn ståendes i olika (oidentifierade) folkdräkter. Några gruppbilder.
- [Tunisien](#)
  - Ruiner och andra sevärdheter.
- [Berlin](#)
  - Arkitektur och transportinfrastruktur.
- [Östermalm](#)

- En lägenhetsinteriör hos en viss fru Thulstrup.
- [Sibirien](#)
  - Transportinfrastruktur och arkitektur. Några bilder från en arbetsplats samt lekande barn.
- [Pharmacia Uppsala](#)
  - Arbete i laboratoriemiljö. Pharmacia var ett läkemedelsföretag.
- [Leningrad](#)
  - Ett fåtal bilder på en stor båt i Leningrads hamn.
- [Jordanien](#)
  - Ruiner och andra sevärdheter.

## Appendix B: Statistik

### Bild- och metadatauppladdning

Idag har 6 014<sup>65</sup> bilder laddats upp. Dessa har laddats upp i 6 omgångar med totalt 32 identifierade teman<sup>66</sup>. Ytterligare 10 teman har identifierats efter uppladdningarna<sup>67</sup>.

Fördelningen på filuppladdningarna per månad<sup>68</sup>:

månad	nya filer
2022-06	20
2023-01	256
2023-02	481
2023-03	286
2023-04	19
2023-05	1 609
2023-10	1 049
2023-11	2 207
2023-12	87

Filanvändning och filvisning:

### GLAMorgan/GLAMorous

GLAMorgan 2023-12 <sup>69</sup>	GLAMorous 2023-12 <sup>70</sup>	GLAMorous 2023-12 (enbart huvudnamnrymd) <sup>71</sup>
6,066 files in category tree. 68 files were viewed, out of 68 used. 42 pages on 7 wikis use those files. 22,635 file views in 2023-12.	# wikis 7 Distinct files used 175 Pages using files 55 Total file usages 233 Page views for 2023-12 32,245	# wikis 7 Distinct files used 68 Pages using files 42 Total file usages 102 Page views for 2023-12 22,635

<sup>65</sup> Antal filer i [https://commons.wikimedia.org/wiki/Category:100\\_000\\_Bildminnen](https://commons.wikimedia.org/wiki/Category:100_000_Bildminnen)

<sup>66</sup> [Appendix A: Uppladdningsbatcher](#)

<sup>67</sup> Antal kategorier i [https://commons.wikimedia.org/wiki/Category:100\\_000\\_Bildminnen](https://commons.wikimedia.org/wiki/Category:100_000_Bildminnen)

<sup>68</sup> *file\_count* på <https://github.com/Wikimedia-Sverige/nordiska-2022>.

<sup>69</sup>

<https://glamtools.toolforge.org/glamorgan.html?&category=100%20000%20Bildminnen&depth=0&year=2023&month=12>

<sup>70</sup>

[https://glamtools.toolforge.org/glamorous/?mode=category&text=100%20000%20Bildminnen&depth=0&pageviews=1&month=2023-12doit=1#gu\\_results](https://glamtools.toolforge.org/glamorous/?mode=category&text=100%20000%20Bildminnen&depth=0&pageviews=1&month=2023-12doit=1#gu_results)

<sup>71</sup>

[https://glamtools.toolforge.org/glamorous/?mode=category&text=100%20000%20Bildminnen&depth=0&ns0=1&pageviews=1&month=2023-12doit=1#gu\\_results](https://glamtools.toolforge.org/glamorous/?mode=category&text=100%20000%20Bildminnen&depth=0&ns0=1&pageviews=1&month=2023-12doit=1#gu_results)

## Mediaviews<sup>72</sup>

date	views	viewed_files
2023-01	15 217	274
2023-02	38 608	732
2023-03	51 022	1 010
2023-04	167 506	1 028
2023-05	88 270	2 596
2023-06	30 054	2 464
2023-07	31 787	2 638
2023-08	33 619	2 265
2023-09	44 373	2 612
2023-10	68 577	3 681
2023-11	142 892	5 631
2023-12	101 308	5 895
<b>total</b>	<b>813 233</b>	

## Interaktioner, förbättringar och engagemang

### Catalysis

Från Catalyst-verktyget ser vi att 68 unika användare (exkl. WMSE och botar) har totalt bidragit med 3 492 redigeringar på filer (och övriga sidor) i projektkategorin.<sup>73</sup> Om vi inkluderar botar och WMSE användare får vi istället 86 unika användare som bidragit med 49 234 redigeringar.

Nedan finns även interaktionsstatistik per månad. För antalet användare exkluderas botar samt användare med totalt färre än 10 redigeringar i kategorin.

	Interaktionsstatistik per månad			
	edits per month	New pages per month	Bytes added per month	editors per month
2022-06	151	25	58 223	1
2022-07	6	1	2 722	0
2022-10	42	0	6 427	2
2022-11	1	0	16	0
2023-01	1 568	259	540 652	2
2023-02	3 515	487	1 469 869	1
2023-03	4 338	289	1 850 558	2

<sup>72</sup> Verktöget *mediaviews* på <https://github.com/Wikimedia-Sverige/nordiska-2022> med `start=20230101` och `end=20231201`.

<sup>73</sup> <https://meta.toolforge.org/catalysis/?title=100+000+Bildminnen&cat=1&wiki=commonswiki>

2023-04	354	20	250 749	4
2023-05	10 845	1 613	5 759 977	7
2023-06	616	1	525 925	1
2023-07	301	0	106 908	0
2023-08	151	5	55 915	1
2023-09	68	0	21 695	0
2023-10	4 635	1 056	4 372 673	3
2023-11	21 244	2 228	11 964 860	19
2023-12	1 338	91	592 355	14
Total	<b>49 173</b>	<b>6 075</b>	<b>27 579 524</b>	<b>57</b>

### Commons-diff

2158 filer (35%) har redigerats och totalt fått 2990 nya kategorier, 348 nya captions på 8 olika språk, 309 nya depicts-påståenden och 166 förändrade bildbeskrivningar.

Notera att antalet som anges är nytillagda objekt av den typen minus borttagna objekt av samma typ.

Rådata<sup>74</sup>:

```
{
  "meta": {
    "cutoff": "2022-06-01",
    "files": 6014,
    "source": "Category:100 000 Bildminnen",
    "timestamp": "2024-01-10T11:12:13"
  },
  "stats": {
    "all": 6014,
    "caption_languages": 8,
    "captions": 348,
    "categories": 2990,
    "changed": 2158,
    "descriptions": 166,
    "statements": 309
  }
}
```

En körning av commons-diff verktyget för *100 000 Bildminnen*-kategorin tar ca 2,5 timmar.

wp\_captions:

34 bilder fick totalt 57 bildtexter på 7 Wikimediaprojekt<sup>75</sup>.

<sup>74</sup> <https://github.com/Wikimedia-Sverige/commons-diff/> varpå resultatfilen har använts i *diff\_stats* på <https://github.com/Wikimedia-Sverige/nordiska-2022>.

<sup>75</sup> *wp\_captions* på <https://github.com/Wikimedia-Sverige/nordiska-2022>.

Från denna data eliminerades de bildtexter som inte var beskrivande (t.ex. i samband med Veckans tävling) för att få fram resultatet:

33 bilder fick bildtexter på 5 olika språkversioner av Wikipedia.

## Veckans tävling

För veckans tävling [statistik från commons-diff för perioden 2023-11-27–2023-12-07]. Totalt redigerades 929 bilder. Nedan ges antalet ändringar per typ, antalet som anges är nytillagda objekt av den typen minus borttagna objekt av samma typ.

- captions: +315,
- categories: +846,
- changed: +929,
- statements: +144 och
- ändrade beskrivningar: 75

Det verkliga antalet förbättringar är troligen högre då fördjupningar, t.ex. en bred kategori som ersatts av en mer specifik kategori, inte plockas upp.

## Appendix C: Mediaomnämmande och övrig kommunikation

- Projekt-presentationer:
  - Wikimania 2023 – [100 000 Photo Memories – museums and Wikimedians in synergy.webm](#)
  - Commons – [Projektgalleriet](#)
- Fokus på [metadataskrivstugan](#):
  - Corren – [BILDEXTRA: Bildskatt från Saab i Linköping görs tillgänglig](#)
  - Wikipedia – [Konceptbeskrivning av metadataskrivstuga](#)
  - Diff – [SAAB Veterans Shared Their Knowledge at Metadata Workshop](#)
  - Wikimedia Sverige – [Saab-veteraner delade med sig av sin kunskap under metadata-stuga](#)
- Våra blogginlägg på [wikimedia.se](#):
  - [En get är en get är en Q2934](#)
  - [Bildminnen från hela Sverige – på museet och Wikimediaplattformarna](#)
  - [Saab-veteraner delade med sig av sin kunskap under metadata-stuga](#)
- Wikimedia Sveriges nyhetsbrev:
  - [April 2023](#)
  - [November 2023](#)
- Återrapportering av det [gemensamma pressmeddelandet](#):
  - Kamera&Bild – [Nordiska museet öppnar upp bildminnen för allmänheten](#)
- Rapporter på det internationella månadsbrevet [This Month in GLAM](#):
  - [November 2023 - Sweden report](#)
  - [October 2023 - Sweden report](#)
  - [August 2023 - Sweden report](#)
  - [May 2023 - Sweden report](#)
  - [April 2023 - Sweden report](#)
  - [March 2023 - Sweden report](#)
  - [February 2023 - Sweden report](#)
  - [January 2023 - Sweden report](#)
  - [August 2022 - Sweden report](#)
  - [June 2022 - Sweden report](#)