



Strategisk inkludering av biblioteksdata på Wikidata 2019

Rapport

2021-01-29



Lund, Botanik, Särskilt fysco-
(Håkansson, /Karl/ Artur /Heribert/)
logi... 1944] (Br) Pedagog.
Skrivelse till matematisk-naturvetenskapli-
ga sektionen i Lund med anledning av sakkun-
nigutlåtandena angående professuren i botanik-
särskilt fysiologi och anatomi.
8:o Lund 194

Rapport: Öppen källkod och öppen data — Strategisk inkludering av biblioteksdata på Wikidata

Projektid: Från den 15 augusti 2019 till den 31 december 2020

Projektsamarbete: Kungliga biblioteket – Digitalt först med användaren i fokus och Wikimedia Sverige

Författare av rapport: Wikimedia Sverige (Alicia Fagerving, André Costa)

Nyckelord: Bibliotek, Wikipedia, Fri kunskap, Plattformer, Öppna data, Wikidata, Libris

Licens: [CC BY 4.0 Creative Commons — Attribution 4.0 International — CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

Bild på omslag: Foto: Alicia Fagerving, CC BY-SA 3.0 Bilden kan hämtas från:
<https://w.wiki/vnk>

Sammanfattning

Strategisk inkludering av biblioteksdata 2019 var ett projekt och samarbete som genomfördes av Kungliga biblioteket och Wikimedia Sverige under åren 2019–2020. Projektet var en fortsättning av **Strategisk inkludering av biblioteksdata 2018** vilket genomförts åren innan. Utöver att fortsätta och fördjupa en del av de satsningar som initierades av det tidigare projektet utforskades nya utvecklingsmöjligheter och datasamlingar.

Arbetet inom projektet pågick i flera parallella spår. Data om publikationer och författare laddades upp till Wikidata; tekniska verktyg för att förbättra och uppdatera Wikidataobjekt utvecklades; genom löpande omvärldsbevakning togs planer och förslag fram för att hålla svensk biblioteksrelaterad data fortsatt relevant för den globala gemenskapen; och synergier med andra plattformar utnyttjades, både inom och utanför Wikimedias ekosystem för fri kunskap. Trots att Covid 19-pandemin omöjliggjorde fysiska möten genomförde vi en uppskattad digital Wikidatautbildning för bibliotekspersonal och berättade om vårt arbete och våra lärdomar på flera internationella evenemang.

Vi hoppas att de resurser som vi utvecklat och de samtal vi haft med Kungliga bibliotekets och andra biblioteks personal kommer att fortsätta bära frukt och inspirera bibliotekarier och Wikimediagemenskapen till vidare satsningar och innovationer. Wikimedia Sveriges långsiktiga mål är att ta en ledande roll i samarbetet med kulturarvsinstitutioner och stötta de aktörer världen över som väljer att engagera sig i öppen kunskap-rörelsen; vårt arbete som beskrivs i den här rapporten har tagit oss en bra bit på vägen dit och pekar ut flera möjliga vägar framåt.

Innehåll

Inledning	5
Aktiviteter och resultat	5
Uppladdningar av bibliografiska data	5
Böcker som hänvisas till från Wikipedia	6
Svenska tidskrifter	8
Statens Offentliga Utredningar – demokratidokument i ett nätverk	9
Robusta auktoritetsnätverk med VIAF	9
Synergier med andra plattformar	10
Storskaliga uppladdningar – hur mycket får plats?	11
Sammanfattning	12
Infrastruktur för att arbeta med bibliografiska data	12
Att hantera parallella Libris-identifikatorer på Wikidata	13
Att använda VIAF för att stärka kopplingar mellan Wikidata och Libris	13
En möjlig väg mot halvautomatisk berikning av Wikidata med Librisposter	14
Bättre källhänvisningar på Wikipedia med Wikidata	17
Strukturerade data på Wikimedia Commons – ett nytt sätt att lagra information om digitaliserade publikationer	19
Utbildning	21
Wikidatadagarna	21
Wikipedia i Biblioteken	22
Informationsspridning och kommunikation	22
Wikimania 2020, Stockholm (16–19 augusti 2020)	22
Liber 2020, online (24 juni 2020)	23
Hack for Heritage, online (25 september 2020)	24
Libris användardag, Stockholm (10 november 2020)	25
Code4Lib, online (mars 2021)	25
Diskussion	25
Wikidata som ett centralt nav för kulturarvsdata	26
De fem effektmålen	28
Slutsatser	28
Framtidsmöjligheter	29
Ämnen och genrer	29
Öppna vetenskapliga tidskrifter	30
Kompletterande datamängder	31
Verk och utgåvor	32
Förbättrad länkning till faksimiler och andra digitala resurser	33

Inledning

Denna rapport redogör för aktiviteterna genomförda inom ramarna för projektet **Strategisk inkludering av biblioteksdata på Wikidata 2019**. Detta var ett större strategiskt projekt för att inkludera bibliotekens digitala information på Wikimedias plattformar och genomfördes av Wikimedia Sverige i samarbete med Kungliga biblioteket mellan den 15 augusti 2018 och den 31 december 2020.

Genom åren har KB och WMSE engagerat sig i flera gemensamma initiativ, fokuserade både på teknik och infrastruktur – som detta projekts direkta föregångare, *Strategisk inkludering av biblioteksdata på Wikidata 2018*¹ – och på utbildning, som *Wikipedia i biblioteken*². Ett nära och långsiktigt samarbete mellan bibliotekssektorn och Wikimediagemenskapen gynnar båda parter genom att bidra till vårt gemensamma mål: att göra mer kunskap tillgänglig för flera människor.

Projektet var designat för att uppnå följande effektmål:

1. Öka användningen av svensk biblioteksrelaterad data.
2. Nationellt och internationellt synliggöra det arbete som skett runt LOD i Sverige samt samarbetet med Wikimediarelsen.
3. Sporra bibliotekens tekniska personal som jobbar med LOD m.m. genom att ge konkreta exempel på det värde deras arbete skapar samt tillsammans identifiera framtida utvecklingsmöjligheter för fri kunskap.
4. Visa på bibliotekens värde för allmänheten genom att tydliggöra insatserna och det värdefulla material de förvaltar.
5. Öka kunskapen om hur Wikimediaprojekten kan användas av biblioteken för att prioritera sin digitalisering.

Aktiviteter och resultat

Uppladdningar av bibliografiska data

Vi fortsatte att ladda upp data om publikationer från Sveriges nationalbibliografi till Wikidata, ett arbete som påbörjades inom projektets första omgång 2018/19. En nyhet i år var att arbetet kunde utföras mer effektivt tack vare nya tekniska lösningar. Vi började att använda oss av applikationen OpenRefine³, som är ett verktyg för att analysera och bearbeta stora mängder

¹ Se slutrapport:

https://commons.wikimedia.org/wiki/File:Biblioteksdata_2018_Slutrapport_Wikimedia_Sverige.pdf

² Se slutrapport: https://se.wikimedia.org/wiki/Projekt:Wikipedia_i_biblioteken_2019/Rapport

³ <https://openrefine.org/>

data. En viktig funktionalitet i OpenRefine är att dataposter kan matchas mot Wikidataobjekt och laddas upp till Wikidata, antingen genom att lägga till ny information i befintliga objekt eller genom att skapa nya. Vi behövde alltså inte lägga ner lika mycket tid på teknisk utveckling som under 2018 års projekt, och kunde arbeta mer flexibelt med flera olika datasamlingar. Dessutom hade vi nytta av den modellering – dvs. den mall för att omvandla fälten i Librisposten till Wikidatas egenskaper – som vi hade tagit fram.

Intressant material att arbeta med var det knappast ont om. Sveriges nationalbibliografi, som är den del av Libris som omfattas av licensen CC0 och därför skulle kunna importeras till Wikidata, omfattar i skrivande stund⁴ över 761 000 objekt, varav 674 000 böcker. Det var därför en utmaning att välja ut några mindre datamängder för att utforska mer specifika aspekter av bibliografiska data på Wikidata.

Böcker som hänvisas till från Wikipedia

Inom 2018 års projekt genomförde vi första försök att massimportera data om böcker från Libris till Wikidata.⁵ Utfallet var positivt och vi fortsatte arbetet även i år. Precis som tidigare var diskussionen om vilka data som skulle importeras relevant, och vi valde att ta upp tråden med svenskspråkiga Wikipedias källhänvisningar – böcker som är relevanta för vår lokala gemenskap, med förhoppning att Wikidataobjekten kommer att ses och förbättras av volontärerna.

Vårt arbete bestod av följande steg:

1. Extrahera alla förekomster av mallen {{Bokref}}⁶ från en dump av svenskspråkiga Wikipedias innehåll. Mallen används för att ange källinformation på ett strukturerad sätt, med separata fält för titel, författare, Libris-ID m.m. Drygt 60 000 instanser av mallen hittades.
2. Begränsa urvalet till mallinstanser med fältet Libris ifyllt. Drygt 13 000 mallinstanser identifierades.
3. Begränsa urvalet till unika identifierade Libris-ID:n (eftersom samma bok kan hänvisas till på flera ställen på Wikipedia). 5 500 unika ID:n identifierades.

⁴ <https://libris.kb.se/hitlist?q=db:natbib>

⁵

[https://se.wikimedia.org/wiki/Projekt:Strategisk_inkludering_av_biblioteksdata_på_Wikidata_2018/Rapporter/Arbete#Verk_och_utgåvor](https://se.wikimedia.org/wiki/Projekt:Strategisk_inkludering_av_biblioteksdata_p%C3%A5_Wikidata_2018/Rapporter/Arbete#Verk_och_utg%C3%A5vor)

⁶ <https://sv.wikipedia.org/wiki/Template:Bokref>

```

{{bokref
| efternamn1      = Ahlin
| förnamn1        = Lars
| efternamn2      = Zander
| förnamn2        = Jens
| efternamn3      = Slimane
| förnamn3        = Ben
| titel           = Principles of wireless communications
| utgivare        = Studentlitteratur
| utgivningsort   = Lund
| år              = 2018
| ISBN            = 978-91-44-12653-1
| libris          = 22444889
}}

```

Ahlin, Lars; Zander, Jens; Slimane, Ben (2018). *Principles of wireless communications*. Lund: Studentlitteratur. [Libris länk](#). ISBN 978-91-44-12653-1

Fig. 1. En ifylld Bokref-mall och hur informationen visas för läsaren i en Wikipediaartikel.

4. I OpenRefine, hämta strukturerad data om böcker som identifieras av dessa ID:n, via URL:en <https://libris.kb.se/resource/bib/{Libris-ID}>.
5. Extrahera information om vilken/vilka bibliografier boken tillhör.
6. Begränsa urvalet till böcker som återfinns i Sveriges Nationalbibliografi (bibliografibeteckning NB). Det är endast denna delmängd i katalogen som omfattas av licensen CC0 och därmed är upphovsrättsmässigt kompatibel med Wikidata.⁷ De flesta av de andra deldatabaserna förvaltas av andra organisationer, inte Kungliga biblioteket. 2 700 poster återstod i det slutgiltiga urvalet.
7. Skapa nya, eller förbättra existerande (identifierade med Libris-ID), Wikidataobjekt för böckerna. Objekten – till exempel [Q101093673](https://www.wikidata.org/wiki/Q101093673) – innehåller följande information, konverterad från Librisposterna:
 - Instans av → upplaga.
 - Titel och ev. undertitel.
 - Författare, antingen som länkat Wikidataobjekt (om Librisposten innehåller författarens auktoritetsidentifierare och ett Wikidataobjekt med samma ID kunde hittas) eller som textsträng.
 - Utgivningsort, om textsträngen i Libris med stor sannolikhet kunde matchas mot ett Wikidataobjekt.
 - Förlag, om textsträngen i Libris med stor sannolikhet kunde matchas mot ett Wikidataobjekt.
 - Språk.

⁷ <https://data.kb.se/>

- Utgivningsdatum.
- Libris-ID samt -URI.
- ISBN10 eller/och ISBN13.

Vi resonerar mer utförligt kring utmaningen att konvertera Librisdata till Wikidataobjekt i avsnittet [Diskussion](#).

Svenska tidskrifter

Vi importerade grundläggande data om 33 000 **tidskrifter** som finns i Sveriges nationalbibliografi. Den välvärdade datamängden var tacksam att arbeta med, i synnerhet med tanke på att det inte funnits så mycket data om tidskrifter, inte minst svenska, på Wikidata. Ett direkt resultat av det arbetet är att av de 53 000⁸ tidskrifter som har ett Wikidataobjekt har 33 000⁹ ett Libris-ID. Svenska är just nu det i särklass största språket för tidskrifter på Wikidata.

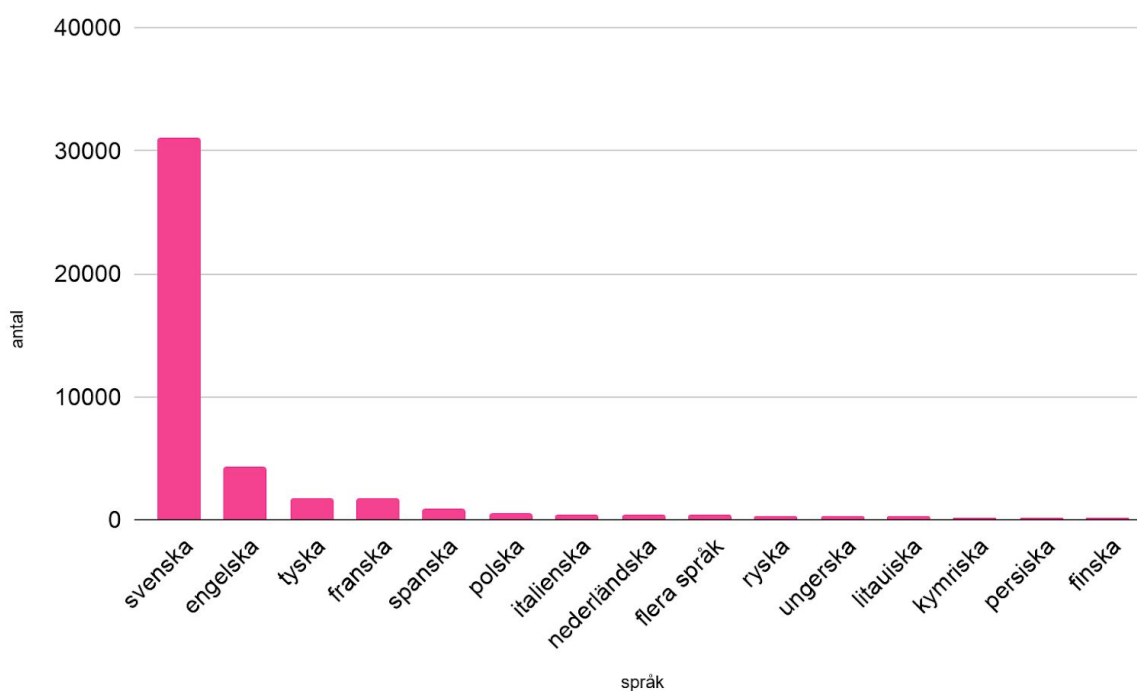


Fig. 2. Tidskrifter på Wikidata uppdelade efter språk i december 2020. Källa: <https://w.wiki/qmw>.

Bibliografisk data om tidskrifter är särskilt intressant då den utgör det nav som knyter samman olika källhänvisningar till artiklar i en tidskrift och bidrar med de grundobjekt som även

⁸ <https://w.wiki/qnc>

⁹ <https://w.wiki/qne>

möjliggör skapandet av Wikidataobjekt för individuella artiklar. Inom ramen för ett annat projekt gjorde vi precis detta för ca 2 000 artiklar i Nordiska Museets *Fataburen*.¹⁰

Statens Offentliga Utredningar – demokratidokument i ett nätverk

Vi valde även att arbeta med 6 000 **Statens Offentliga Utredningar** (SOU). Detta är en annan mindre, välavgränsad datamängd och dokumenten har ett högt informationsvärde för samhällsmedborgarna. Eftersom Kungliga biblioteket har digitaliserat och tillgängliggjort dokumenten¹¹ var det särskilt värdefullt att kunna belysa den satsningen genom att hänvisa till faksimilen i Wikidataobjekten. Se till exempel [Q80296336](#), *Betänkande med förslag angående avsättningen av smör och andra fettämnen av inhemskt ursprung (1934)*.

Genom att synliggöra SOU-samlingen riktade vi ljuset mot bibliotekets roll som en viktig aktör i det demokratiska samhället. Ökad tillgång till viktiga dokument, både historiska och samtida, gör den demokratiska processen mer transparent och bidrar till en mer medveten och engagerad allmänhet. Detta yttrar sig även i liknande insatser genomförda av volontärgemenskapen både i Sverige och internationellt; ett exempel värt att nämna är det arbete med svenska Riksdagens öppna data som genomfördes av frivilliga under sommaren och hösten 2020. Många tusen motioner, interpellationer, utskottsbetänkanden, förordningar m.m. laddades upp till Wikidata och gemenskapen uppmuntrades till att berika de enskilda dataposterna med ytterligare data så som ämnesord och orsakssamband.¹² Det finns alltså både mycket data om våra demokratiska processer och många engagerade användare på Wikimediaplattformarna som gärna tar del av och förbättrar det.

Robusta auktoritetsnätverk med VIAF

För att göra datat från Libris mer robust och sammanlänkat med andra databaser arbetade vi även med Virtual International Authority File, **VIAF**, det vill säga den globala auktoritetsdatabasen som många nationella bibliotek, däribland KB, använder sig av. En personauktoritet i Libris har alltså en motsvarande identifierare i VIAF. VIAF-ID ([P214](#)) är en av Wikidatas i särklass största egenskaper, som i skrivande stund används i över 2,5 miljoner objekt. Eftersom VIAF är en hubb som kopplar ihop många biblioteks identifierare är det en mycket användbar egenskap, som gör det möjligt att hitta data om personerna i andra databaser. Täckningen är däremot långt ifrån komplett – det händer ofta att en användare endast lägger till en Libris-identifierare i ett nyskapat personobjekt. Det är helt naturligt med tanke på att allt innehåll på Wikimediaplattformarna byggs upp steg för steg av frivilliga. Det är möjligt att fylla dessa luckor (personobjekt med en Libris-auktoritetsidentifierare som saknar

¹⁰

https://meta.wikimedia.org/wiki/FindingGLAMs/White_Paper/DOCS#A_digitized_ethnographic_library_reaches_Wikimedians

¹¹ <https://www.kb.se/hitta-och-bestall/hitta-i-samlingarna/digitala-souer.html>

¹² <https://wikimedia.se/2020/11/17/framtiden-for-riksdagens-data-i-wikidata/>

VIAF-identifierare) maskinellt genom att leta upp poster i VIAF med Libris-ID:t. Vi genomförde alltså en synkroniseringsinsats mellan Libris och VIAF med följande steg:

1. Identifiera personobjekt på Wikidata som har ett Libris-ID men inget VIAF-ID, och lägga till detsamma (130 objekt); samt
2. Identifiera personobjekt på Wikidata som har ett VIAF-ID men inget Libris-ID (trots att VIAF-objektet innehåller detta) och lägga till detsamma (1 300 objekt).

Genom att berika Wikidataobjekten har vi gjort dem lättare att hitta för alla som använder sig av VIAF-ID:n, som är en betydligt större grupp än bara de som intresserar sig för svenska bibliotekssamlingar. Eftersom VIAF även inkluderar Wikidata-ID:n i sina poster har den globala identifierarnätet gjorts tätare och mer användbart. En dylik synkroniseringsinsats skulle kunna göras regelbundet för att underlätta för mänskliga användare som inte skulle behöva tänka på leta upp och lägga till ett antal identifierare för hand.

Synergier med andra plattformar

Nytt för i år var att vi inkluderade ytterligare två plattformar i vårt arbete, Litteraturbanken och Wikisource. Eftersom vi hade jobbat med Projekt Runeberg i det tidigare projektet¹³ så kändes det naturligt att fortsätta kartlägga och utforska andra svenska öppna litteraturplattformar.

Litteraturbanken¹⁴ är ett projekt för att tillgängliggöra äldre svensk litteratur som kommit till stånd genom ett samarbete mellan Svenska Akademien, Kungliga biblioteket och Språkbanken vid Göteborgs universitet. Det digitala biblioteket omfattar över 4 800 verk. På Wikidata finns två egenskaper med anknytning till Litteraturbanken: Litteraturbanken utgåva ([P5123](#)) och Litteraturbanken författare ([P5101](#)). Vi har:

- Träffat Litteraturbankens utvecklare vid Språkbanken för att berätta om det arbete som vi gör kring länkade öppna biblioteksdata och demonstrera hur sådan data fungerar på Wikidata; samt
- Laddat upp data om flera tusen bokutgåvor på Litteraturbanken till Wikidata.

Uppladdningen var möjlig tack vare att Litteraturbanken inkluderar Libris identifikationsnummer i sin databas, vilket gjorde det lätt att identifiera motsvarande Librisposter. Vårt arbete omfattade drygt 4 000 Wikidataobjekt, de flesta av vilka skapades från början.

Wikisource¹⁵ är ett av Wikipedias systerprojekt. Det är ett digitalt bibliotek för upphovsrättsligt fria böcker och andra dokument där faksimil visas jämte korrekturlästa och sökbara texter. Korrekturläsningen görs av volontärer som oftast utgår från OCR:ade texter (OCR, *optical*

¹³

https://se.wikimedia.org/wiki/Projekt:Strategisk_inkludering_av_biblioteksdata_p%C3%A5_Wikidata_2018/Rapporter/Projekt_Runeberg

¹⁴ <https://litteraturbanken.se/>

¹⁵ <https://sv.wikisource.org/wiki/Wikisource:V%C3%A4lkommen>

character recognition, är en teknik där datorn omvandlar inskannade bilder till text och som oftast kräver manuell kontroll och korrigerings).

Vårt arbete med Wikisource bestod av att koppla texterna till utgåvornas Wikidataobjekt vilka i sin tur – om Wikisourcesidan innehöll ett Libris-ID – fylldes med data från Libris. Vi upptäckte att en stor del av Wikisourceposterna var kopplade till helt tomma Wikidataobjekt, eller till inga objekt alls, vilket gjorde det svårt för Wikidataanvändarna att hitta till dem. På detta sätt redigerade vi 154 objekt.

Ett resultat av vårt arbete är att data från flera svenska litteraturplattformar samsas nu på Wikidata. Här ingår, förutom Libris, Litteraturbanken och svenska Wikisource, även Projekt Runeberg som vi arbetat med i förra projekt. Det finns nu ett antal objekt som innehåller länkar till alla fyra plattformarna, vilket demonstrerar Wikidatas roll som en central informationshubb.

Storskaliga uppladdningar – hur mycket får plats?

En återkommande tråd under vårt arbete med bibliografiska data har varit frågan om hur mycket data som egentligen borde laddas upp till Wikidata. Det är vi inte ensamma om – den internationella WikiCite-gemenskapen har också ägnat detta en del diskussion. Å ena sidan är Wikidata ett öppet projekt, med mycket generösa relevanskriterier¹⁶, som uppfylls av alla publikationer och personer med ett Libris-ID; vårt resonemang kring detta återfinns i förra projektrapporten.¹⁷

Å andra sidan är det inte optimalt, varken teknik- eller innehållsmässigt, att ladda upp hundratusentals svenska katalogposter till Wikidata. Wikidata är som sagt en allmän databas och att överbelamra det med data från en källa utan någon tanke på vilka som skulle kunna dra nytta av det hade inte uppskattats av gemenskapen. Wikidata växer fort, och redan nu upplever en del användare svårigheter med att hitta bland objekten eller exekvera komplexa sökningar mot databasen. WikiCite-gemenskapen har därför identifierat ett behov av tydligare riktlinjer att förhålla sig i framtiden i samband med massuppladdningar av bibliografiska poster. Flera scenarier har tagits fram: från att fortsätta ladda upp så mycket bibliografiska data som möjligt till att sätta upp en – eller flera! – separat plattform dedikerad till denna.¹⁸

Vårt tillvägagångssätt har definitivt påverkats av dessa diskussioner. I vårt urval fokuserade vi på datamängder som antingen 1) har en tydlig koppling till de andra Wikimediaplattformarna, för att öka sannolikheten att gemenskapen uppmärksammar och använder dem (se [Böcker som hänvisas till på Wikipedia](#), [Wikisource](#)), 2) har en koppling till andra plattformar (se

¹⁶ <https://www.wikidata.org/wiki/Wikidata:Notability>

¹⁷

[https://se.wikimedia.org/wiki/Projekt:Strategisk_inkludering_av_biblioteksdata_på_Wikidata_2018/Rapporter/Inklusionsguide](https://se.wikimedia.org/wiki/Projekt:Strategisk_inkludering_av_biblioteksdata_p%C3%A5_Wikidata_2018/Rapporter/Inklusionsguide)

¹⁸

https://www.wikidata.org/wiki/Wikidata:WikiCite/Roadmap#Four_scenarios_for_the_future_of_WikiCite

[Litteraturbanken](#)), eller 3) är av en mer sammanhållen och välavgränsad art (se [SOU:er](#), [tidskrifter](#)). Dessutom lade vi en del resurser på att förbättra befintliga objekt genom att lägga till fler länkar till VIAF och Libris, vilket beskrivs i avsnittet [Infrastruktur för att arbeta med bibliografiska data](#).

Sammanfattning

Knappt 50 tusen Wikidataobjekt redigerades inom projektet; tidskriftsdatauppladdningen – de flesta nyskapade objekt – stod för den största delen. I jämförelse med 2018 års projekt (500 böcker som hänvisas till på Wikipedia och vidare 5 000 med koppling till Projekt Runeberg) laddade vi upp betydligt mer bibliografiska data, vilket var möjligt dels tack vare att en grundkringmodellering och databearbetning redan var på plats och dels tack vare den anpassningsbarhet och flexibilitet som OpenRefine erbjuder vid bearbetning av datamängder som är uppbyggda på olika sätt.

Samling	Antal förbättrade eller nyskapade Wikidataobjekt
Statens Offentliga Utredningar	6 119
Tidskrifter	33 192
Böcker i Nationalbibliografin som hänvisas till från Wikipedia	2 534
Böcker i Litteraturbanken	4 048
Auktoritetsposter	1 585
Wikisource	154
Alla	47 632

Fig. 3 Sammanställning av antalet Wikidataobjekt som redigerades under projektets olika stadier.

Infrastruktur för att arbeta med bibliografiska data

Under projektets gång utvecklade vi flera tekniska lösningar för att arbeta med Librisdata och Wikidata, både när det gäller de huvudsakliga uppgifterna (uppladdningar av bibliografiska data) och mindre stödåtgärder. All mjukvara utvecklad av Wikimedia Sverige är tillgänglig under en öppen licens och kan därför fritt granskas, användas och vidareutvecklas av

gemenskapen; den kan även – förhoppningsvis! – anpassas till andra datakällor och -samlingar. Nedan listas de som vi ser ha störst återbruksvärde.

Att hantera parallella Libris-identifierare på Wikidata

Källkod på Github:

https://github.com/Vesihisi/Biblioteksdata/blob/master/Biblioteksdata2/add_uri.py

I samband med utvecklingen av Libris XL skapades en ny identifierare, en URI, som till skillnad mot de tidigare separata identifierarna för auktoriteter och publikationer används till alla sorters resurser i bibliotekets kataloger. Samtidigt som det underlättar både för dem som jobbar med bibliotekets system och för användarna så introducerar det en extra komplexitet, då både gamla och nya identifierare finns i omlopp. Nya resurser (Librisposter tillagda efter URI:ns introduktion) får endast en ny identifierare, medan äldre resurser kan refereras till med bådadera, vilket avspeglas på Wikidata. Se till exempel objektet [Q4769173](#) som innehåller de två egenskaperna Libris-URI och SELIBR-ID.

Under vårt arbete blev vi varse om att det fanns ett stort antal Wikidataobjekt som endast innehöll ett gammalt ID. Vidare skapas det fortfarande objekt med bara detta, eftersom det är lättare för många användare som inte är medvetna om Libris XL och hämtar informationen från Libris "gamla" plattform. För att göra det lättare att hålla Wikidata i synk med Libris XL skapade vi därför ett litet skript som identifierar sådana objekt och lägger till en URI.

Skriptet fungerar på följande sätt:

1. Hitta alla Wikidataobjekt som innehåller 1 gammalt Libris ID (antingen för auktoriteter eller publikationer). Begränsningen till exakt ett värde innebär att vi skippar objekt med potentiellt felaktigt tillagda multipla identifierare, där ytterligare tillägg riskerar att försämra datakvalitén.
2. Hämta motsvarande URI från Libris' API.
3. Lägg till URI:n med källhänvisning till Libris.

Knappt 40 000 objekt redigerades i den första körningen, en stor backlog har åtgärdats. Även om vi inte kan förutse ett behov av en lika omfattande insats i framtiden så håller vi på att implementera ett sätt att köra skriptet automatiskt på regelbunden basis, till exempel varje vecka eller varje månad, för att löpande åtgärda de brister som uppstår och avlasta användarna från att behöva hitta och lägga till både gamla och nya identifierare. Vi har identifierat detta som en enkel åtgärd för att förbättra kopplingarna mellan Wikidata och Libris.

Att använda VIAF för att stärka kopplingar mellan Wikidata och Libris

Källkod på Github:

<https://github.com/Vesihisi/Biblioteksdata/blob/master/Biblioteksdata2/viaf.py>

Virtual International Authority File (VIAF) är ett internationellt system för auktoritetsdata där identifierare från många bibliotek, däribland KB, samlas. VIAF är en omtyckt resurs för Wikidataanvändarna. VIAF ID ([P214](#)) är med sina 2,5 miljoner användningar en av Wikidatas största egenskaper. Det finns många Wikidataobjekt som innehåller ett VIAF-ID men saknar de identifierare som kan inhämtas från VIAF, däribland Libris-identifierare. Vi gjorde därför en insats för att berika ett antal personobjekt på Wikidata med data inhämtad via VIAF.

Uppdateringen genomförde vi i två steg:

1. Identifiera Wikidataobjekt som innehåller ett VIAF-ID som saknar Libris-ID trots att detsamma finns i VIAF-posten. Lägg till det Libris-ID som inhämtats från VIAF¹⁹.
2. Identifiera Wikidataobjekt som innehåller ett Libris-auktoritets-ID men saknar VIAF-ID (alla poster i Libris auktoritetsdatabas finns per definition med i VIAF). Lägg till motsvarande VIAF-ID som inhämtats från VIAF.

Sammanlagt uppdaterade vi över 1 500 objekt. I likhet med föregående skript lämpar sig även detta för regelbunden körning för att löpande åtgärda de luckor som uppstår när volontärer redigerar Wikidata.

En möjlig väg mot halvautomatisk berikning av Wikidata med Librisposter

En intressant möjlighet som vi undersökte inom projektet var hur processen att bearbeta Librisdata och lägga till det på Wikidata skulle kunna förenklas och tillgängliggöras för flera. Vi vill gärna se att användarna engageras av och arbetar med KB:s bibliografiska data även efter att vår satsning har avslutats. Detta skulle gynna såväl KB (genom en ökad synlighet av dess resurser) och Wikidata (genom sänkta trösklar för användare). I synnerhet skulle det gynna de med intresse för biblioteks- och informationsvetenskap, vars expertis är särskilt värdefull för plattformen.

Vi tog därför fram ett förslag till ett verktyg för att hämta bibliografiska uppgifter om en publikation från Libris och lägga till dem i Wikidata (antingen som ett nytt objekt eller som tillägg till ett existerande objekt). Vi utförde även en del arbete för att bygga den tekniska infrastrukturen för verktyget och ta fram en fungerande prototyp för att demonstrera processen. Arbetet bestod av följande steg:

En backend som hämtar och bearbetar Librisdata

Källkod på Github: <https://github.com/Vesihisi/booktool>

Libris erbjuder ett öppet API genom vilket data om en katalogpost med ett visst ID kan hämtas (t.ex. <https://libris.kb.se/resource/bib/11655186>). Den information som returneras är mycket omfattande; den inkluderar uppgifter som är av intresse för bibliotekets katalogisatörer men

¹⁹ Datat från VIAF hämtades genom <http://viaf.org/viaf/data/>.

inte för allmänheten. De relevanta uppgifterna, som bokens författare, titel, utgivningsår osv. utgör endast en liten del av den returnerade datamängden. Att förenkla datan och extrahera endast det som behövs för att fylla Wikidataobjektet var därför ett första steg i vårt arbete.

```

    @type:                "Instance"
    editionStatement:    "Ny, utök. och rev. uppl."
  ▼ extent:
    ▼ 0:
      @type:              "Extent"
      ▼ label:
        0:                "304 s."
    ▼ hasDimensions:
      @type:              "Dimensions"
      ▼ label:
        0:                "25 cm"
    ▼ hasTitle:
      ▼ 0:
        @type:            "Title"
        mainTitle:        "Svensk militärmakt"
        ▼ subtitle:
          0:              "strategi och operationer i svensk militärhistoria under 1 500 år"
    ▼ identifiedBy:
      ▼ 0:
        @type:            "ISBN"
        ▼ qualifier:
          0:              "inb."
        value:            "9789185789474"
    ▼ instanceOf:
      @type:              "Text"
      ▼ classification:
        ▼ 0:
          @type:          "ClassificationDdc"
          code:           "355.009485"
          edition:        "full"
          editionEnumeration: "23/swe"
        ▼ 1:

```

Fig. 4. En katalogpost i Libris (<https://libris.kb.se/vd677vd61b5wb9n/data.jsonld>) innehåller mycket metadata som kan vara svårnavigerad om man bara vill ha grundläggande uppgifter om publikationen.

Denna uppgift görs av ett Pythonskript som är driftsatt på Toolforge²⁰, en öppen plattform där alla kan implementera verktyg och tjänster som arbetar mot Wikimedias API:er. Skriptet anropas med en URL som inkluderar publikationens Libris-ID, t.ex.

<https://booktool.toolforge.org/api/11655186>. Det anropar i sin tur Libris API och tar fram följande uppgifter, som returneras i en enkel JSON-struktur:

- Libris-ID (det gamla systemet)
- Libris-URI (det nya systemet)
- Bidragsgivare och deras roller (författare, översättare, illustratör, redaktör)

²⁰ <https://wikitech.wikimedia.org/wiki/Portal:Toolforge>

- I en del Librisposter identifieras dessa med Libris-URI. I detta fall försöker verktyget hitta deras Wikidataobjekt och returnera dess ID. Om bidragsgivaren endast anges med textsträng returneras istället denna).
- Antal sidor
- Titel och undertitel
- Språk
- Utgivningsår
- Utgivningsort
- Förlag
- Antal sidor
- ISBN10 och ISBN13

Implementeringen av en backend som eftersänder datat, snarare än att hämta den direkt från Libris till Wikidatas gränssnitt, beror delvis på att Wikimedias projekt av säkerhetsskäl inte tillåter JavaScript-tillägg att hämta data från externa resurser (Cross-Origin Resource Sharing).

²¹ Toolforge ingår i Wikimedias ekosystem och är därför en tillåten resurs.

```

{
- response: {
- contributors: [
- {
  person: "https://libris.kb.se/20dggfhl2vcxf74#it",
  qid: "Q5714532",
  role: "author"
},
- {
  person: "https://libris.kb.se/jgvx0qz25l2pgc2#it",
  qid: "Q48526878",
  role: "illustrator"
}
],
- isbn: [
- {
  type: "ISBN13",
  value: "978-91-85789-47-4"
}
],
  language: "https://id.kb.se/language/swe",
  libris_id: "11655186",
  libris_uri: "vd677vd61b5wb9n",
  pages: "304",
- publication: {
  year: "2009"
},
  subtitle: "strategi och operationer i svensk militärhistoria under 1 500 år",
  title: "Svensk militärmakt"
}
}

```

Fig. 5. Samma katalogpost som i [Fig. 4](#) bearbetad av verktyget (<https://booktool.toolforge.org/api/11655186>). Uppgifter som kan användas på Wikidata har

²¹ https://en.wikipedia.org/wiki/Cross-origin_resource_sharing

extraherats.

Ett Wikidata-tillägg som lägger till uppgifterna

Källkod: [https://www.wikidata.org/wiki/User:Alicia_Fagerving_\(WMSE\)/booktool.js](https://www.wikidata.org/wiki/User:Alicia_Fagerving_(WMSE)/booktool.js)

MediaWiki, mjukvaran som används av Wikimediaplattformarna, erbjuder sina användare en möjlighet att förändra och utöka de befintliga funktionaliteterna med hjälp av egenutvecklade JavaScript-tillägg.²² Inloggade användare kan välja vilka av dessa som ska aktiveras, vilket är ett enkelt sätt att tillhandahålla verktyg som är av intresse för specifika, ofta smala, uppgifter. Vårt mål var att skapa en fungerande prototyp av ett tillägg till Wikidatagränssnittet för att hämta information från Libris om en publikation med ett känt Libris-ID och lägga till det i ett Wikidataobjekt.

Brister och vidare utveckling

Den prototyp som vi tagit fram visar ett fungerande tillvägagångssätt för att erbjuda Wikidatas användare en möjlighet att enkelt berika plattformen med data från Libris. Vidare utveckling är nödvändig innan verktyget når en stabil version som kan lanseras för

Wikimediagemenskapen. Bland annat måste en del tekniska buggar lösas och fler tester genomföras för att säkerställa att verktyget uppfyller användarnas förväntningar och skapar mervärde i deras arbete med bibliografiska data. De funktionaliteter som vi har identifierat som viktiga och som kommer att prioriteras inkluderar: att kunna skapa nya Wikidataobjekt, inte endast redigera befintliga; att kunna använda sig av Libris URI, inte bara gamla ID, för att identifiera motsvarande Librispost; att analysera den språkkod som returneras av Libris' API (t.ex. <https://id.kb.se/language/swe>) och matcha den mot rätt Wikidataobjekt för att kunna ange bokens språk; att implementera stöd för att kunna identifiera redigeringar gjorda med hjälp av verktyget samt samla in information om hur mycket det används.

Vidare hoppas vi att verktyget, som precis som alla andra resurser framtagna inom ramen för detta projekt är byggt på öppen källkod, kan tilldra sig intresse av volontärutvecklare som vill förbättra och bygga vidare på det. Med tanke på den stora och teknikkunniga gemenskap som finns samlad kring bibliografiska data på Wikidata är detta inte osannolikt.

Vi bedömer att det krävs ett par dagar arbete för att uppnå detta. Vi avser att genomföra arbetet senare under 2021. Det kommer att finansieras med det bidrag som vi erbjudits av LIBER Europe i samband med att vi har [tilldelats LIBER Award for Library Innovation](#).

Bättre källhänvisningar på Wikipedia med Wikidata

Källhänvisningar är en av Wikipedias grundstenar; de behövs för att bygga ett pålitligt och trovärdigt uppslagsverk. Både Wikimediautvecklarna och -gemenskapen ägnar därför mycket tanke åt hur man skulle kunna förenkla och effektivisera processen att lägga till och hantera

²² <https://www.mediawiki.org/wiki/Extension:Gadgets>

källhänvisningar, något som utgör en avsevärd tröskel för nya bidragsgivare. Detta har aktualiserats i samband med utvecklingen av bibliografiskt innehåll på Wikidata. Det är tekniskt enkelt att visa data från Wikidata på Wikipedia, och det finns därför en stor potential för att med en mindre arbetsinsats förbättra Wikipedia genom den stora mängd bibliografiska data som Wikidatas volontärer så engagerat arbetar med.

Som det ser ut på Wikipedia just nu infogas källhänvisningar genom att ange publikationens bibliografiska uppgifter i en mall. Det är upp till skribenten att välja vilka detaljer (som utgivningsort, förlag m.m.) som ska vara med, varför detaljrikedomen på källhänvisningarna kan variera (se övre vänstra delen i [Fig. 6](#)). Samma bok kan refereras till på olika sätt av olika skribenter och i olika artiklar. Det är svårt att få fram en standardisering på grund av det stora antalet skribenter.

Ett alternativt tillvägagångssätt är, förutsatt att boken har ett objekt på Wikidata, att endast ange objektets ID-nummer (se övre högra delen i [Fig. 6](#)) och låta mjukvaran hämta de bibliografiska detaljerna därifrån samt formatera dem på ett standardiserat sätt. Från en skribents synpunkt har det den stora fördelen att slippa mata in all information för hand. Å andra sidan kräver det att man är tillräckligt bekant med Wikidata för att hitta objektet man behöver, och kanske till och med förbättra objekten eller lägga till egna.

Under året har den internationella Wikimediagemenskapen gjort ett gediget arbete²³ med mallen **Cite Q**²⁴, vars syfte är just att visa bibliografisk information från Wikidata på Wikipedia – huvudsakligen den engelskspråkiga upplagan av uppslagsverket eftersom det är de mest aktiva utvecklarnas hemmaarena. Mallen är dock utvecklad med flerspråkighet i åtanke och ska kunna anpassas till andra språkversioner. Dess nyaste version har även implementerats på portugisiskspråkiga Wikipedia.²⁵

Vi har följt utvecklingen av mallen med stort intresse då den växande mängd data om svenskutgiven litteratur på Wikidata – som dessutom i mängd och mycket är utvald just på grund av att böckerna hänvisas till på Wikipedia – erbjuder en stor potential för uppslagsverket.

På svenskspråkiga Wikipedia har mallen implementerats rent tekniskt, dvs. den finns och kan användas. Den har infogats i några enstaka artiklar och några användare har diskuterat dess lämplighet. Det finns dock ingen konsensus om huruvida den ska rekommenderas på ett systematiskt sätt eller undvikas. Vi är därför intresserade av att ta reda på om intresset för mallen har ökat nu i samband med att flera av de verk som hänvisas till har importerats inom detta projekt, se Böcker som hänvisas till från Wikipedia.

²³ <https://diff.wikimedia.org/2021/01/14/automatically-maintained-citations-with-wikidata-and-cite-q/>

²⁴ https://en.wikipedia.org/wiki/Template:Cite_Q

²⁵ https://pt.wikipedia.org/wiki/Predefinição:Citar_Q

```

{{Bokref
| efternamn1=Ahlin
| förnamn1 = Lars
| efternamn2 = Zander
| förnamn2 = Jens
| efternamn3 = Slimane
| förnamn3 = Ben
| titel= Principles of wireless
communications
| utgivare= Studentlitteratur
| utgivningsort = Lund
| år=2018
| ISBN = 978-91-44-12653-1
}}

```

{{Cite Q|Q57415723}}

Lars Ahlin; Jens Zander; Ben Slimane (2018), *Principles of wireless communications*, Lund: Studentlitteratur, ISBN 978-91-44-12653-1, Wikidata Q5741572

Fig. 6. Två sätt att hänvisa till en bok på Wikipedia – med mallen *Bokref* (ö.v.) respektive *Cite Q* (ö.h.). I det senare fallet hämtas uppgiften från Wikidataobjektet. Den undre bilden visar det renderade resultatet av *Cite Q*-mallen, resultatet från *Bokref* skiljer sig inte väsentligt från detta.

Strukturerade data på Wikimedia Commons – ett nytt sätt att lagra information om digitaliserade publikationer

Structured data on Commons²⁶ (i fortsättningen benämnt **SDC**) är ett nytt sätt att beskriva filerna i Wikimediaprojektens gemensamma mediadatabas Wikimedia Commons. Istället för att enbart lagra uppgifter som skapandedatum, skaparens namn och upphovsrättsligt status i ett stort fritextfält – den klassiska strukturen som har använts på Wikimedia Commons sedan den startades – kan de delas upp i separata datafält. Genom att gå över från fritext till semantiskt meningsfulla data blir informationen betydligt mer tillgänglig för exempelvis sökmotorer. SDCs möjliggör dessutom kopplingar till Wikidata, detta gör informationen flerspråkig/språkneutral vilket gör att stora delar av Wikimedia Commons, som är flerspråkigt projekt i sin natur, äntligen blir tillgängligt för helt nya publikationer. Kopplingarna till Wikidata innebär också att det, när det har uppnått större mognadsgrad (verktyg för att arbeta med och söka i SDC är fortfarande under utveckling), kommer man kunna genomföra betydligt mer nyanserade sökningar och analys än vad som är möjligt i dag samt att resurser på Wikimedia Commons kan länkas samman med resurser utanför Wikimedias projekt.

²⁶ https://commons.wikimedia.org/wiki/Commons:Structured_data

Även om merparten av Wikimedia Commons består av fotografier och digitaliserade konstverk så innehåller plattformen även en stor mängd digitaliserade böcker, artiklar och andra dokument²⁷. En del har skannats in av gemenskapen på Wikimediprojekten och andra har importerats från öppet tillgängliga källor, som Internet Archive²⁸ eller Kungliga bibliotekets digitaliserade böcker. Vi valde därför att undersöka hur den nya teknologin SDC kan tillämpas på den sortens dokument.

Vi undersökte 3 typer av faksimiler, och ett exempel på en extraherad bild, på Wikimedia Commons och försåg dem med strukturerade data²⁹ för att jämföra hur detta sätt att beskriva mediafiler skiljer sig mot vanliga fritextbeskrivningar. Följande uppgifter har vi bland annat kunnat uttrycka med strukturerade påståenden:

- Länk till Wikidataobjektet för den motsvarande bokutgåvan ([Q100272846](https://www.wikidata.org/wiki/Q100272846)).
- Länk till sidan på Wikisource där boken omvandlas till e-bok (https://sv.wikisource.org/wiki/Index:Adolfsfors_1920.djvu).
- Vem som har digitaliserat boken (Internet Archive).
- Var faksimilen kommer ifrån (en webbadress och intern identifierare hos Internet Archive).
- Faksimilens upphovsrättsliga status (Public Domain eftersom bokens författare har varit död i över 70 år).
- När faksimilen skapades (2011)

En viktig fråga att beakta vid arbetet med SDC är att den strukturerade datan på Commons endast beskriver *filen*, inte *boken* den representerar. Det är viktigt med tanke på hur många digitaliserade konstverk som finns på Wikimedia Commons; i varje fall måste man skilja på konstverket och den digitala avbildningen.³⁰

Resultatet av vår undersökning blev att det finns stora möjligheter att beskriva digitala faksimiler med strukturerade data för att tydligt och maskinläsbart koppla den direkt till Wikidataobjektet för motsvarande utgåva av boken, och därmed även till Libris. För att undvika att blanda ihop data om filen och själva boken – vilket är lätt hänt om man inte är insatt i de av gemenskapens framtagna arbetsflöden – krävs att det finns (eller skapas) ett Wikidataobjekt för den aktuella bokutgåvan.

Vi planerar att fortsätta vårt arbete med SDC, både med digitala faksimiler och andra mediafiler, i föreningens andra projekt. Det förberedande arbete som lagts ner inom det här projektet har gett oss en stabil grund att stå på.

²⁷ En grov uppskattning baserat på förekommande filtyper visar att det rör sig om ungefär 5% av Wikimedia Commons bestånd, dvs. 3 miljoner filer.

²⁸ <https://archive.org/>

²⁹ Till exempel: https://commons.wikimedia.org/wiki/File:Adolfsfors_1920.djvu

³⁰ Se *Advice from the cultural sector* i

<https://diff.wikimedia.org/2020/02/28/structured-data-on-commons-and-glam-open-questions-and-fresh-challenges/>

Utbildning

Wikidatadagarna

I oktober 2020 organiserade vi ett **webbinarium** om Wikidata för bibliotekspersonal i Sverige. Vi hade hoppats på att kunna bjuda in till en fysisk träff, men det var omöjligt på grund av den rådande Covid 19-pandemin. Att genomföra kursen online hade sina utmaningar men även en fördel: deltagare från olika håll i landet kunde vara med utan långväga resor. Kungliga biblioteket stod för marknadsföringen, och som resultat hade vi 50 deltagare från Ystad i söder till Piteå i norr. De representerade olika delar inom bibliotekssektorn, från bibliotekarie vid folkbibliotek till utvecklare vid högskolebibliotek, vilket vittnar om ett stort intresse för Wikimediaplattformarna.

En av anledningarna bakom utbildningen var att det saknades välstrukturerat informationsmaterial på svenska för personer som av olika anledningar var intresserade av Wikidata men saknade erfarenhet av Wikimediaplattformarna. En stor del aktiva Wikidataanvändare är insatta i Wikimediarörelsen sedan tidigare, och den dokumentation som finns antar oftast en viss grad av vana av att redigera på Wikimediaplattformarna och orientera sig på deras hjälpsidor. Detta är en barriär för nya användare som skulle vilja aktivera sig på Wikidata för att bidra med sina ämneskunskaper. Experter inom biblioteks- och informationsvetenskap är en sådan grupp, och under vårt samarbete – inte minst i samband med det hackathon som vi organiserade som en del av projektets första omgång 2018–2019 – hade vi ofta fått höra att det var just en enkel introduktion som efterfrågades.

Under det heldagslånga webinariet gick vi igenom de viktigaste aspekterna av Wikidata, från plattformens syfte och koppling till Wikipedia till grunderna i frågespråket SPARQL som används för att ta fram information från Wikidata. En modul handlade specifikt om bibliografisk data på Wikidata, dess struktur och hur den internationella gemenskapen jobbar med det. Under dagen varvades förinspelade presentationer och fria diskussioner där deltagarna kunde ställa frågor och lära både av oss och varandra. Deltagarnas frågor och våra förklaringar har lagts upp på evenemangets sida på föreningens wiki och ingår, tillsammans med presentationerna och övrig information, i det utbildningsmaterial som både deltagarna och andra intresserade kan ta del av.³¹ Allt detta material är tillgängligt under en öppen licens och kan därmed användas och vidareutvecklas av alla.

Ett misslyckande blev däremot det **hackathon** som vi också ville få av stapeln i samband med utbildningen. Det skulle ha blivit ett praktiskt inriktat event där personal vid Kungliga biblioteket som till vardags jobbar med Libris och länkade data skulle kunna utforska konkreta problem och ta fram lösningar och prototyp. Denna typ av evenemang, som bygger på aktivt arbete och samarbete i små grupper, lämpar sig betydligt mindre för en digital form, vilket vi

³¹

https://se.wikimedia.org/wiki/Projekt:Strategisk_inkludering_av_biblioteksdata_p%C3%A5_Wikidata_2019/Wiki_datadagarna

bedömer var anledningen till att endast två intresserade anmälde sig. Eftersom vi inte kunde få med oss en kritisk massa deltagare valde vi att ställa in hackathonet.

Wikipedia i Biblioteken

Under året genomförde vi också ett annat projekt i samarbete med Kungliga biblioteket, *Wikipedia i Biblioteken*.³² Vi bidrog till KB:s nationella satsning *Digitalt först med användaren i fokus* med utbildningstillfällen om Wikimediaplattformarna och informationsmaterial som publicerades på den digitala lärplattformen Digiteket. Vår satsning var fokuserad på Wikipedia, men även Wikidata inkluderades för att ge deltagarna en samlad bild av Wikimedias ekosystem och inspiration till framtida aktiviteter.³³

Informations spridning och kommunikation

Vi deltog i flera evenemang, där vi spred kunskap om biblioteksdata och om vårt samarbete med Kungliga biblioteket. Dessutom arrangerade vi ett utbildningstillfälle med fokus på Wikidata för intresserad personal vid bibliotek i Sverige. Ett urval utav de evenemang vi deltog på finns nedan.

Wikimania 2020, Stockholm (16–19 augusti 2020)

Wikimania är Wikimediarörelsens årliga globala konferens. Detta gav oss en värdefull arena att kommunicera om arbetet och knyta kontakter med både andra Wikimedianer och ämnesexperter intresserade av fri kunskap. Det finns en stor internationell gemenskap samlad kring öppna biblioteksdata, **WikiCite**, och konferensen hade ett särskilt spår tillägnat just detta. Vi kunde alltså berätta om vårt samarbete med KB för en engagerad och insatt publik. Eftersom konferensen var öppen för allmänheten var det också ett tillfälle att belysa bibliotekens roll inom fri kunskap-landskapet och inspirera fler till att engagera sig på Wikimediaplattformarna.

Antal åhörare: 38

Presentation:

- https://commons.wikimedia.org/wiki/File:Strategic_inclusion_of_library_data_on_Wiki_data.webm
- https://commons.wikimedia.org/wiki/File:Strategic_inclusion_of_library_data_on_Wiki_data_Wikimania_2019.pdf

I samband med Wikimania organiserades även ett heldagsmöte med fokus på Wikidata och Wikibase för nationella bibliotek, **Wikidata & Wikibase for National Libraries Meeting**. Mötet anordnades av Europeana i samarbete med Wikimedia Sverige, Wikimedia Deutschland och

³² Slutrapport: https://se.wikimedia.org/wiki/Projekt:Wikipedia_i_biblioteken_2019/Rapport

³³ <https://digiteket.se/kurs/wikidata-en-vag-mot-mer-oppen-och-lankad-data/>

Kungliga biblioteket. Deltog gjorde, förutom oss, representanter för bland annat National Library of Wales, Finlands nationalbibliotek, Bibliothèque nationale de France, Deutsche Nationalbibliothek och Linked Data for Production Phase 2, ett samarbetsprojekt mellan flera högskole- och forskningsbibliotek i USA. Det är alltså flera betydande institutioner runt om i världen som på något sätt arbetar med Wikidata och Wikibase (den öppna mjukvara som Wikidata använder och som också kan användas för att bygga egna tjänster och plattformar) och som fick en möjlighet till diskussion och erfarenhetsutbyte.

Antal åhörare: 29

Presentation:

[https://commons.wikimedia.org/wiki/File:Wikidata & Wikibase for National Libraries WMS E.pdf](https://commons.wikimedia.org/wiki/File:Wikidata_%26_Wikibase_for_National_Libraries_WMS_E.pdf)

Liber 2020, online (24 juni 2020)


På denna konferens, som riktar sig till verksamma vid forskningsbibliotek, presenterade vi en poster och ett abstract om projektet.³⁴ Abstraktet var ett av tre som tilldelades årets LIBER Award for Library Innovation.³⁵ Priset, ett bidrag på upp till € 1 000 för ett biblioteksrelaterat projekt, tänker vi använda för att vidareutveckla den prototyp som beskrivs i avsnittet [En möjlig väg mot halvautomatisk berikning av Wikidata med Librisposter](#) och lansera den för Wikidatagemenskapen.

Antal åhörare: 262

Presentation: https://www.youtube.com/watch?v=_Gz6SfZfvrM

³⁴ <https://zenodo.org/record/3907332#.XxbILGP7Q5k>


³⁵ <https://liberconference.eu/winners-of-the-liber-award-for-library-innovation/>



Wikidata as an open, collaborative bibliographic database

The case of the National Library of Sweden


ALICIA FAGERVING (WIKIMEDIA SVERIGE) <alicia.fagerving@wikimedia.se>



Wikidata: a Linked Open Data Platform open to everyone

Wikidata ([wikidata.org](https://www.wikidata.org)) is a platform for Linked Open Data. It's one of the sister projects of Wikipedia. Just like Wikipedia, Wikidata can be edited by anyone, and the content is publicly accessible under an open license (CC0) – meaning it can be freely re-used! Wikidata editors are volunteers, and have imported many open data datasets.

Wikidata contains over **86 million data items**. Thanks to its flexible structure, it encompasses all domains of knowledge: people, geographical objects, historical events, books and much more, can be described in Wikidata. The volunteer community works together to develop modelling standards and best practices for each domain. A dedicated community project for books (wikidata.org/wiki/Wikidata:WikiProject_Books) has developed a data model based on FRBR and its work/edition distinction.



Open data at the National Library of Sweden

Wikimedia Sverige is a non-profit association based in Sweden, working towards making knowledge freely accessible to as many people as possible. We have an ongoing collaboration with the **National Library of Sweden**, where we work on the Wikimedia platforms (Wikidata and Wikipedia) with the open data released by the library. We also support the library in its endeavors to become an active and knowledgeable participant in the open knowledge movement.

In 2018, the National Library of Sweden became the first national library in the world to transition to linked data in its catalog. The new **Libris** – a national union catalog containing data from over 500 libraries in Sweden – is based on **Bibframe 2.0**, making the data easier to understand, analyze and re-use. Most importantly, it makes Swedish bibliographic data more compatible with the rest of the web, including free and open platforms such as Wikipedia and Wikidata.

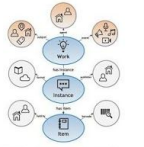
The **Swedish National Bibliography** contains metadata of over 700,000 books, periodicals and other materials published in Sweden. As part of its open data strategy, the National Library of Sweden has released this data under an open license, **CC0**. The same applies to the Libris Authority File. The CC0 marking is the most permissive Creative Commons license: it means the data can be re-used by anyone, without having to credit the source. Removing the attribution requirement is important in order to be able to combine and visualize multiple datasets when crediting every single source becomes an impossibility. This is the data we have been working with.

Working with Swedish bibliographic data on the Wikimedia platforms

We have been working on developing tools and workflows to import **bibliographic metadata** from the National Library's catalog to Wikidata. As a pilot project, we created ca. 500 items to represent book editions. We selected the books that are most often used as references on Swedish Wikipedia to make our work more relevant and interesting to the community. We also added Libris authority identifiers to over 60,000 items of authors on Wikidata.

One of the biggest **challenges** of mass-importing bibliographic data to Wikidata is that items representing authors, publishers etc. need to be uniquely identified in order to be linked from the book items. A prerequisite for doing this automatically is that the author's identity in the source dataset is stated using an authority ID and that a Wikidata item containing the same ID can be found – or created, if we can be sure it does not yet exist. We have found that many author statements in Libris are still stated as strings, making it impossible to automatically and correctly match them to Wikidata items. This problem is even more acute for publishers, where no authority IDs are used at all.

We are also exploring how bibliographic data on Wikidata can be continuously **improved** and kept up-to-date. Wikidata is a living platform, and when community members interact with the data we imported, it proves that our work is engaging and relevant. In particular, we are planning to build tools to enable Wikidata contributors to create new items based on Libris data when they are needed. We see this as an important contribution to the community which will enable the Swedish bibliographic commons to grow even after the conclusion of our project.



The **Bibframe 2.0** data model, used in the National Library's new catalog, uses three levels of abstraction to organize bibliographic data.

References

- Fischer, Barbara and Jens Ohlig. (2020). *Could you wikify an authority file? Wikibase has been evaluated for the Integrated Authority File (IAF)*. <https://blogs.wikimedia.org/?2020/03/04/wikibase-iaf-and/>
- Feritiz, Meghan. (2019). *Integrating Wikidata at the Library of Congress*. <https://blogs.loc.gov/thesignal/2019/06/integrating-wikidata-at-the-libr-06-02-2019/>
- Association of Research Libraries Task Force. (2019). *ARL White Paper on Wikidata: Opportunities and Recommendations*. <https://www.arl.org/resources/arl-whitepaper-on-wikidata/>

Wikidata and bibliographic metadata around the world

Libraries and other cultural heritage institutions around the world are implementing Wikidata in their work.

- Wikibase**, the software behind Wikidata, has been evaluated by the **German National Library** as a possible platform for the Integrated Authority File (IAF).
- The **Library of Congress** has added Wikidata identifiers to over one million entries in its authority file.
- The **Association of Research Libraries'** white paper on Wikidata gives more examples of how Wikidata and Linked Open Data can benefit cultural heritage organizations and scholarly communication.
- WikiCite** (<https://meta.wikimedia.org/wiki/WikiCite>) is an international initiative to develop open citations and linked bibliographic data to serve free knowledge, using Wikidata and Wikibase as its platform.

Our work with the National Library of Sweden is not unique; it is part of a global movement to make bibliographic data more robust, accessible and useful.

Fig. 7. Poster om Wikidata för bibliografiska data och samarbetet med Kungliga biblioteket som presenterades på konferensen Liber 2020 (CC BY-SA-4.0, tillgänglig på [Wikimedia Commons](https://commons.wikimedia.org/wiki/File:Hack_for_Heritage_2020_-_presentation_da_g_1.pdf)).

Hack for Heritage, online (25 september 2020)

Vi deltog i online-hackathonet som organiserades av Riksantikvarieämbetet som partner och support. På första dagen hade vi en kort presentation om vår verksamhet och på andra dagen en workshop på en halvtimme om Wikimediaprojekten och det kulturarvsrelaterade material som finns där. Kungliga biblioteket nämndes som ett exempel på en GLAM-institution³⁶ som aktivt bidrar med sina resurser till Wikimediaprojekten. Därtill fanns det tillfällen för handledning för teamen som trots att de lyftes flitigt inte kom att utnyttjas. Inte heller använde sig något av projekten av data från Wikimediaplattformarna.

Antal åhörare: 25

Presentationer:

- https://commons.wikimedia.org/wiki/File:Hack_for_Heritage_2020_-_presentation_da_g_1.pdf

³⁶ GLAM står för Galleries, Libraries, Archives and Museums.

- https://commons.wikimedia.org/wiki/File:Hack_for_Heritage_2020_-_workshop_dag_2.pdf

Libris användardag, Stockholm (10 november 2020)

Vi blev inbjudna att delta med en presentation om projektet. Vi berättade om vad Wikidata är och varför det är relevant inom bibliotekssektorn, samt vilka typer bibliografiska data som finns där. Vi gav även en överblick över svenska bibliografiska data på Wikidata, d.v.s. objekt med koppling till Libris, Litteraturbanken, Projekt Runeberg och svenska Wikisource.

Antal åhörare: 410

Presentation:

<https://www.kb.se/samverkan-och-utveckling/nytt-fran-kb/nyheter-samverkan-och-utveckling/2020-11-24-presentationer-fran-libris-anvandardag-2020.html>

Code4Lib, online (mars 2021)

Vi har skickat in ett presentationsförslag till konferensen, som äger rum våren 2021. Konferensens syfte är att belysa innovativa tillämpningar av öppna teknologier inom ABM-sektorn, ett naturligt sammanhang för att berätta om Kungliga bibliotekets satsning på Wikidata och våra erfarenheter från det nyss avslutade projektet.

Diskussion

När man laddar upp data från externa källor till Wikidata utgörs arbetet till stor del av att identifiera vilka Wikidataobjekt som motsvarar värdena i indatan. Det är just detta som är kärnan i Wikidata i synnerhet och Länkade Öppna Data i allmänhet: att dataobjekten länkar till varandra. Detta arbete underlättas väsentligt om entiteterna i indatan identifieras med unika identifierare. Att arbeta med textsträngar kräver betydligt mer arbete, även om verktyg som OpenRefine erbjuder (halv-)automatisk matchning av strängar mot Wikidataobjekt. Det är inte alltid som sådan matchning kan göras med tillräckligt stor säkerhet, ett typiskt exempel är författare med vanliga namn.

Många författare i Libris saknar fortfarande auktoritetsposter, vilket gör det svårare för Wikidatagemenskapen att arbeta med KB:s kataloger. Problemet är ännu tydligare när det gäller utgivningsorter och förlag. Ett och samma förlag kan förekomma under flera olika namnformer eller förkortningar i bibliotekskatalogen; flera förlag med samma eller liknande namn kan ha existerat vid olika tidpunkter i historien; förlag byter namn, slås samman, läggs ner och återuppstår. Dessutom är Wikipedias och Wikidatas täckning av svenska förlag långt ifrån komplett. Allt detta gör det svårt att förlita sig på automatisk matchning av textsträngar mot Wikidataobjekt.

Vi vill därför understryka att aktivt auktoritetsarbete, med såväl personer (författare, illustratörer, redaktörer m.m.) som förlag, orter och annan viktig information nog är det viktigaste steg bibliotek kan ta för att främja återanvändning av sina data och stärka sin position inom Länkade Öppna Data. Här skulle biblioteken kunna dra nytta av det arbete som Wikidatagemenskapen redan gjort och använda Wikidatas identifierare – till exempel för orter och förlag – direkt i sina kataloger.

En återkommande tråd i våra samtal med biblioteksverksamma har varit både den stora nyfikenheten på Wikidata och behovet av mer information. Vår [nybörjarutbildning](#) fyllde till viss del denna luckan, men av både tids- och tillgänglighets-skäl kunde vi inte fördjupa oss i mer tekniska aspekter av Wikidata, som exempelvis frågespråket SPARQL. Vi har hört från flera håll att just kunskaper inom SPARQL efterfrågas, vilket är mycket positivt, då detta är nästa steg till att verkligen realisera Wikidatas möjligheter. Det finns alltså många potentiella aktiva Wikidataanvändare bland biblioteksverksamma i Sverige, som i sin tur kan föra vidare kunskaperna till både sina kollegor och biblioteksanvändarna.

Wikidata som ett centralt nav för kulturarvsdata

En central tanke inom både det här projektet och dess föregångare har varit att undersöka och involvera flera svenska litteraturplattformar. Förutom Libris har vi arbetat med Projekt Runeberg, Litteraturbanken och Wikisource. Förutom stärkta kontakter med både kulturarvsaktörer och öppen kunskap-gemenskapen, som sträcker sig utanför Wikimediaplattformarna, har det bidragit till att befästa Wikidatas roll som ett centralt nav för litteraturdata. Wikidata har över åren tagit plats som ett sådant i andra kulturarvssammanhang och för andra auktoriteter, inte minst inom museisektorn. Det växande intresset för öppna bibliografiska data som vi ser inom både Wikimediagemenskapen och bibliotekssektorn bådär gott för fortsatt utveckling i den riktningen.

Det måste påpekas att dataimporterna från Projekt Runeberg och Litteraturbanken har varit möjliga tack vare att bägge plattformarna inkluderar Libris-ID i sina databasposter. Den data som importerats kommer alltså ursprungligen från Libris. Detta är ett exempel på hur man kan undvika duplikation av data genom att hänvisa till andra, bättre resurser – Libris är den de facto kanoniska källan av bibliografiska data om svenskutgivna publikationer. På samma sätt hade det inte varit otänkbart för plattformarna att även inkludera Wikidatas identifierare i och med att all mer information om svenska verk och författare nu finns på Wikidata. Detta är något som vi har framfört i våra samtal med kulturarvsorganisationer. Det finns redan goda internationella exempel: Library of Congress inkluderar Wikidataidentifierare i sina auktoritetsposter.³⁷ Litteraturbanken gör likaså och hämtar bland annat information om författarnas födelse- och dödsdatum samt länk till dennes Wikipediaartikel från Wikidata.³⁸

³⁷ <https://blogs.loc.gov/thesignal/2019/05/integrating-wikidata-at-the-library-of-congress/>

³⁸ Se API-respons där uppgifter hämtade från Wikidata finns utbrutna: https://litteraturbanken.se/api/get_author/BremerF



Fig. 8. Ett Wikidataobjekt för en bokupplaga – [En naturforskarens resa omkring jorden](#) (1872) – är en ingångsport till flera externa resurser.

Ett centralt datanav gör det enklare för användarna att upptäcka den data som de kopplade aktörerna har om objekten. Detta är särskilt intressant för utvecklare som bygger applikationer och tjänster som återanvänder datan, eftersom de enkelt kan komma åt informationen via Wikidatas API:er. För kulturarvsaktörerna själva innebär det att de nu har en möjlighet att koppla sina resurser mot Wikidata. De kan komma åt de resurser som övriga aktörer sitter på och till exempel automatiskt visa komplementerande uppgifter utan att behöva lägga till dem själva. Med väldigt enkla medel kan de på det sättet erbjuda en rikare upplevelse för sina användare. Dessutom kan den datamängd som finns både hos aktören och på Wikidata få engagemang och feedback från gemenskapen. En del kulturarvsinstitutioner har börjat experimentera med att använda gemenskapens tillägg och ändringar för att berika sina egna resurser, en process som kallas för *roundtripping*.³⁹

Via Wikidata kan man bland annat upptäcka de digitaliserade resurser (faksimiler och e-böcker) som är kopplade till en bokutgåva, även om Libris inte känner till att de finns. Främst har detta möjliggjorts genom direkta kopplingar till plattformar utanför Wikimediaplattformarna, som Litteraturbanken och Projekt Runeberg. Med [Strukturerad data på Commons](#) kommer det även bli lättare att automatiskt kunna upptäcka faksimiler som finns tillgängliga på Wikimediaplattformarna, något som förhoppningsvis kommer att göra flera medvetna om Wikisource som en yta där dessa kan tillgängliggöras som e-böcker. Ett större engagemang i Wikisource innebär flera intressanta, men tidigare osynliga, publikationer som tillgängliggörs för allmänheten i ett attraktivt och lätthanterligt format. Det gör det helt enkelt lättare för flera att bekanta sig med den svenska litteraturskatten.

³⁹ https://meta.wikimedia.org/wiki/Wikimedia_Commons_Data_Roundtripping

Eftersom data på Wikidata är strukturerad öppnar det för avancerade sökningar och analyser. Man kan till exempel hitta böcker med utgången upphovsrätt som saknar digital faksimil. Om boken används som källa på Wikipedia tyder det på intresse från allmänheten – om boken tillgängliggörs digitalt kan läsarna ta del av den för att lära sig mer om ämnet, och fler volontärer kan använda den i artikelskrivandet. Denna data kan användas av bibliotek och arkiv som har boken i sina samlingar och vill planera sina digitaliseringsinsatser utifrån faktiska behov.

De fem effektmålen

Genom att ladda upp data om böcker, tidskrifter och dokument från Libris ökade vi **synligheten och användningen av svensk biblioteksrelaterad data**. Detta är särskilt värdefullt med tanke på att Wikidatagemenskapen är huvudsakligen engelskspråkig och innehållet på plattformen till övervägande del grundar sig på resurser och datamängder från den anglosaxiska delen av världen. Tack vare våra insatser har Wikimediamenskapen, både i Sverige och internationellt, blivit **medvetna om det arbete som skett runt Länkade Öppna Data i vårt land**, och förhoppningsvis nyfikna på liknande satsningar på olika håll i världen. Genom deltagande i konferenser nådde vi även ut till yrkesverksamma inom biblioteks- och informationsvetenskap utanför Wikimediamenskapen. På Wikimania, som var öppet för allmänheten, introducerade vi KB:s och våra satsningar på ett lättillgängligt sätt för att vidare **synliggöra bibliotekens roll i öppen kunskap-landskapet** och det värdefulla material de förvaltar. Löpande under projektåret kommunicerade vi även om bibliotekens och vårt arbete i våra sociala medier.

Vår utbildning om Wikidata var ett fantastiskt tillfälle att inte bara förmedla konkreta kunskaper utan även **inspirera bibliotekens tekniska personal** till att själva utforska och bidra till Wikimedias plattformar. Vårt mål har varit att ge deltagarna en bättre förståelse av och en stabil grund till att aktivt delta i de plattformar för fri kunskap som vi arbetar med.

Slutligen har vårt arbete med källor på svenskspråkiga Wikipedia haft för syfte att belysa vilka publikationer som är särskilt intressanta för Wikimediamenskapen. Den grund som vi byggt genom att skapa Wikidataobjekt för dem är intressant inte bara för uppslagsverkets frivilliga, utan även för biblioteken som **vägledning till hur eventuella digitaliseringsinsatser skulle kunna prioriteras** – allt för att bättre kunna svara på det behov av litteratur som finns på Wikimedias plattformar. Med fördel kan en sådan prioritering även dra nytta av vår insats att koppla andra plattformar mot Wikidata så att det blir lättare att upptäcka om digitaliserade resurser redan finns på annan plats.

Slutsatser

I det stora hela är vi nöjda med projektets utförande. Trots Covid-19-pandemin, som försvårade planeringen, uppfyllde vi de mål som vi hade satt upp. Tack vare omställningen från en fysisk träff till ett webinarium lyckades vi till och med nå fler deltagare än förväntat med vår utbildningsdag. Det låga intresset för hackathonet visade däremot att vissa typer av evenemang är betydligt svårare att anpassa.

Mycket grundarbete har gjorts i föregående projekt. Med en stabil grund att stå på valde vi att utforska flera olika områden, både inom Wikimediaplattformarna och utanför, istället för att fokusera på att ladda upp så mycket bokdata som möjligt till Wikidata. På detta sätt belyste vi de många vinklar och arbetsområden som finns för dem som vill jobba med biblioteksrelaterad data – både institutionellt och inom volontärgemenskapen – på Wikimediaplattformarna.

Projektet låg väl i tiden och gav oss möjlighet att följa den utveckling som gjorts både i Sverige och internationellt och fundera på Wikimedias roll i den. Inte minst den nationella plattformen för öppna vetenskapliga tidskrifter är mycket intressant från den synvinkeln. Samtidigt har intresset av att använda Wikimediaprojekten, i synnerhet Wikidata, för att synliggöra det demokratiska samhällets processer tagit fart; även här fyller biblioteken och den data de har en central roll.

Allt eftersom innehållet i Libris XL fortsätter att mogna ser vi stora möjligheter till att ytterligare stärka banden mellan bibliotekets resurser och Wikimediaplattformarna, till nytta för båda.

Framtidsmöjligheter

Som en del av Wikimediagemenskapens arbete med sin 2030-strategi⁴⁰ finns planer på att ge utökat ansvar till delar av rörelsen för specifika ämnesområden. Inom ramen för detta arbetar Wikimedia Sverige för att ta på sig en ledande roll för rörelsens arbete kring innehållspartnerskap, ett område där kulturarvsorganisationer länge haft en central roll. Som en del av detta arbete räknar vi med samarbeten med bibliotekssektorn, både inom Sverige och internationellt, kommer att få en betydande roll.

Ämnen och genrer

Även om vi har gjort mycket så har vi hittills blott skrapat på ytan när det gäller bibliografiska data på Wikidata. Det finns flera områden där gemenskapen gjort värdefulla insatser som

⁴⁰ https://meta.wikimedia.org/wiki/Wikimedia_2030

skulle kunna berikas med data från svenska bibliotek. Ett sådant exempel är **ämnen** och **genrer**.

Svenska ämnesord är en vokabulär om över 30 000 termer från många användningsområden som används av de flesta Librisanslutna bibliotek.⁴¹ På Wikidata finns egenskapen *huvudtema / main subject* ([P921](#)) med motsvarande funktion. Den används just nu av över 12 miljoner objekt, men om man begränsar urvalet till objekt med angivet språk svenska (det vill säga alla svenskspråkiga filmer, böcker, artiklar m.m. på Wikidata) så är det bara knappt 8 400 objekt som använder sig av den.⁴² Samtidigt finns det över 16 000 bokutgåvor på Wikidata som innehåller Libris-ID men saknar egenskapen *huvudtema*.⁴³ Dessa objekt skulle alltså kunna berikas med ämnesord importerade från Libris.

En förutsättning för att detta skulle kunna göras automatiskt är att ämnesorden är synkroniserade med Wikidata. I praktiken görs detta genom att en ny Wikidataegenskap skapas (jämför t.ex. egenskapen [P8647](#), *Library of Congress Children's Subject Headings ID*) och sedan används för att koppla respektive ämnes objekt med ämnesordet. Till exempel skulle Wikidataobjektet om kattrasen Turkisk Van, [Q42724](#), länka till <https://id.kb.se/term/sao/Turkisk%20van>. Detsamma skulle man kunna göra med andra metadatadatabaser som KB förvaltar, som **SAOGF - Genre och form**, **Barnämnesord** m.m.

Att synkronisera sina vokabulärer med Wikidata är ett stort steg mot att göra den data man förvaltar lättare att dela, återanvända och utforska, vilket har utnyttjats bland annat av Europeana⁴⁴ och de nordiska kulturarvsinstitutioner som är anslutna till KulturNav.⁴⁵ Här kan Wikimediagemenskapen göra en stor insats och hjälpa till att matcha ämnesorden med verktyget Mix'n'Match.⁴⁶ Med kopplingen till Wikidata får man även flerspråkighet för ämnesorden, något som kan underlätta för den som vill söka i ett bestånd men ej har svenska som primärt språk eller för att hitta lämpligast ämnesord för böcker som ej är skrivna på svenska.

När en Wikipediaartikel använder en bok som källa så säger detta i sin tur något om bokens ämne, denna information kan på många sätt vara ett mer organiskt alternativ till det ämne som finns direkt på motsvarande Wikidataobjekt, som möjligen hämtats från ämnesorden i Libris. När ämnesorden i Libris har matchats mot Wikidata kan man därför undersöka hur källhänvisningarna kan användas för att förfina eller validera de ämnesord som inledningsvis satts. Arbetet med att ställa om till att [använda Wikidata i källhänvisningarna](#) skulle underlätta en sådan undersökning, men den är inget krav för att detta ska vara möjligt.

⁴¹ <https://metadatabyran.kb.se/amnesord-och-genre-form/svenska-amnesord>

⁴² <https://w.wiki/uDZ>

⁴³ <https://w.wiki/uDd>

⁴⁴ <https://pro.europeana.eu/page/get-your-vocabularies-in-wikidata>

⁴⁵ <https://wikimedia.se/2020/06/10/kulturnav-wikidata-sant/>

⁴⁶ <https://mix-n-match.toolforge.org/>

Öppna vetenskapliga tidskrifter

Kungliga Biblioteket har fått i uppdrag av regeringen att "inrätta och förvalta en nationell digital plattform för öppet tillgängliga svenska vetenskapliga tidskrifter".⁴⁷ Arbetet med plattformen inleddes under 2020 och har följts av oss med stort intresse. Den nationella plattformen skall göra det möjligt för offentligt finansierad forskning att vara öppet tillgänglig, mer användbar och synlig både i Sverige och utomlands.

Den nationella plattformen kommer att innehålla metadata om ett stort antal vetenskapliga artiklar. Ifall metadatan licensieras under den fria licensen CC0 blir den kompatibel med Wikidata; vi har från vår sida påpekat hur värdefullt det skulle bli för Wikimediagemenskapen. Vetenskapliga artiklar, huvudsakligen från engelskspråkiga publikationer inom medicin och naturvetenskap, utgör redan nu en betydande del av Wikidata (deras antal har vuxit från 2,3 miljoner i augusti 2017⁴⁸ till 22,5 miljoner i februari 2020⁴⁹), tack vare importen från öppna källor av vilka bland annat PubMed kan nämnas. Svensk forskning är knappt representerad på Wikidata, likaså forskning inom humaniora och samhällsvetenskap. Här skulle ett färskt tillskott av öppna bibliografiska data kunna göra en märkbar skillnad samtidigt som det skulle öka medvetenheten om och förhoppningsvis engagemanget i Wikidata bland svenska forskare och högskole- och forskningsbibliotekarier.

Vi fortsätter därför att följa utvecklingen av den nationella plattformen med stort intresse.

Kompletterande datamängder

Under vårt arbete med datan i nationalbiografien upptäckte vi att det fanns relaterade datamängder som saknades eller var ofullständiga i Libris.

I flera fall upptäckte vi Librisobjekt som saknar identifierare för författare, istället anges författarnamnet som en sträng. För dessa skulle det vara intressant att identifiera de fall där Wikimediaprojekten har länkat till författaren. Detta kan antingen ske genom att Wikidataobjektet länkar till ett författarobjekt; att källhänvisningen, när boken använts som källa på Wikipedia, länkar författarnamnet till en annan Wikipediaartikel eller att faksimilen på Wikisource har länkat till en författarsida. Utifrån dessa länkar skulle man kunna ta fram förslagslistor som kan ligga till grund för att KB själva ska kunna verifiera kopplingen och berika sina datamängder.

En annan datamängd som helt saknas är länkade data om förlag, tryckerier etc. Denna datamängd kompliceras ytterligare av de många olika sätt som samma organisation kan

⁴⁷

<https://www.kb.se/samverkan-och-utveckling/oppet-tillgang-och-bibsamkonsortiet/oppet-tillgang/plattform-for-oppet-tillgangliga-tidskrifter.html>

⁴⁸ Finn Årup Nielsen; Daniel Mietchen; Egon Willighagen (oktober 2017), "Scholia, Scientometrics and Wikidata", The Semantic Web: ESWC 2017 Satellite Events: 237-259, doi:[10.1007/978-3-319-70407-4_36](https://doi.org/10.1007/978-3-319-70407-4_36)

⁴⁹ <https://www.wikidata.org/wiki/Wikidata:Statistics>

hänvisas till namnbyten av organisationerna samt sammanslagningar av dem (med eller utan namnbyte).

För att ta fram denna datamängd skulle sett samarbete kunna genomföras med andra kommersiella- eller civilsammahällesaktörer för att ta fram ett material som sedan kan struktureras upp på Wikidata. Om källbeläggningen och kvalitén är hög nog skulle denna datamängd sedan kunna adopteras av KB. Potentiella samarbetspartners skulle kunna vara branschorganisationer⁵⁰, företagsarkiv⁵¹ samt KB (för eventuell information som mottagits i samband med pliktleveranser).

Verk och utgåvor

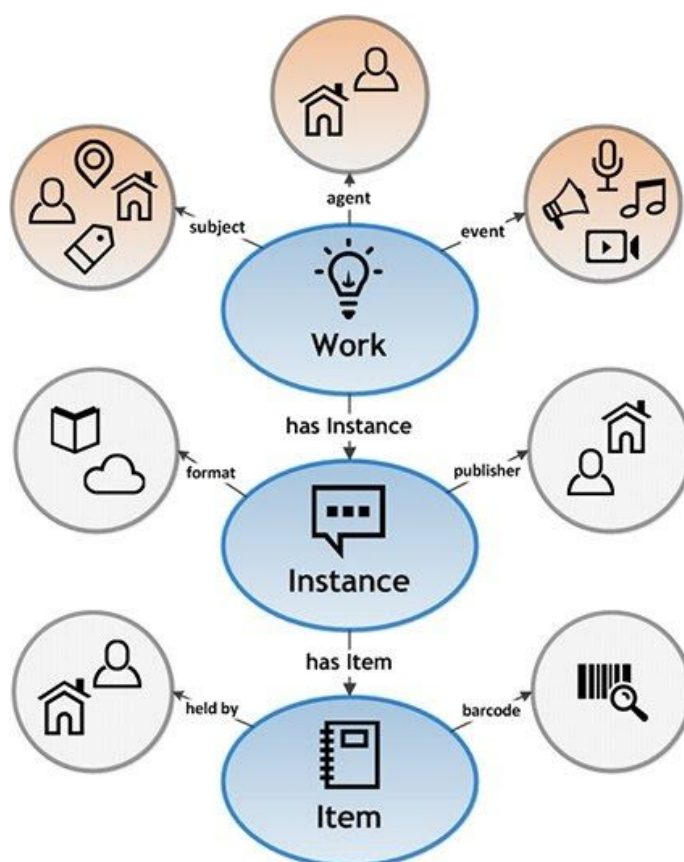


Fig. 8. Datamodellen BIBFRAME 2.0 som används bl.a. i Libris XL. Den av Wikidatagemenskapen framtagna datamodell för böcker är baserad på denna. Public Domain, via [Wikimedia Commons](https://commons.wikimedia.org/wiki/File:BIBFRAME_2.0.png).

Både Libris XL och Wikidata⁵² använder sig av en datamodell där unika exemplar (fysiska objekt) kopplas till instanser/utgåvor som i sin tur knyts samman via verk. Som ett exempel

⁵⁰ T.ex. Svenska Förläggareföreningen (<https://www.forlaggare.se>)

⁵¹ Förslagsvis via Centrum för Näringslivshistoria (<https://www.naringslivshistoria.se/>)

⁵² https://www.wikidata.org/wiki/Wikidata:WikiProject_Books

kan vi se det litterära verket *The Duc de L'Omelette* ([Q3793881](#)) som i sin tur är kopplat till två utgåvor, en på Engelska från 1832 ([Q81507681](#)) och en på Franska från 1887 ([Q81507373](#)).

Tyvärr är denna typen av länknings fortfarande relativt ovanliga på Wikidata⁵³ eftersom de underliggande fria datakällorna ofta själva saknar detta lager. Här skulle KB:s öppna datamängd kunna fungera som ett startskott för att strukturera upp informationen på Wikidata och erbjuda en stor mängd verk till vilket existerande objekt om utgåvor kan knytas.

Samtidigt kan den data som skapas på Wikidata, speciellt när det kommer till att knyta olika översättningar till samma verk vara av nytta för KB. Det beskriver en angränsande datamängd som tillför ytterligare kunskap om den kontext i vilka böckerna i Libris ingår, och kan användas exempelvis till forskning. Den är även intressant för användarna som kan hitta andra upplagor eller översättningar av en bok som t.ex. citeras på Wikipedia.

Förbättrad länkning till faksimiler och andra digitala resurser

Den centrala roll som Wikidata och Wikimedia Commons har fått inom den digitala kulturarvssfären internationellt gör att svenskspråkiga digitaliseringar hittar dit även från källor som inte traditionellt fokuserar på svenskspråkiga böcker. Det är därför en källa till digitala resurser som kanske inte finns på KB:s radar eller där källan själv inte är medveten om Libris och dess identifierare.

[Strukturerad data på Commons](#) gör det möjligt att koppla dessa digitaliseringar till motsvarande Wikidataobjekt och via detta till Libris. Initiativet är dock fortfarande relativt nytt och i dagsläget är det få⁵⁴ faksimiler som är kopplade till Wikidataobjekt. En insats kring att befolka Wikimedia Commons med strukturerad data om redan existerande digitaliseringar⁵⁵, samt vid behov skapande motsvarande Wikidataobjekt, behövs för att möjliggöra ytterligare nyttor.

Kopplat till detta ser vi en vinst med att ytterligare fokusera på länknings direkt från Wikidataobjektet för en utgåva till faksimiler och andra digitala resurser på externa plattformar. Detta ger dels en klarare bild över hur digitaliseringslandskapet ser ut – vilket kan informera framtida digitaliseringsinsatser – och gör Wikidata till ett nav för den som söker att hitta upphovsrättsligt fri svensk litteratur.

⁵³ <https://w.wiki/vIv>

⁵⁴ <http://bit.ly/CommonsLibrisFaksimiler>

⁵⁵ Idag rör det sig om ca 1 000 sådana filer <https://petscan.wmflabs.org/?psid=18324853>