

## FoU-ansökan

### Användbara auktoriteter för datadriven samlingsforskning

Centrala museer: Nytt forsknings- och utv.projekt Centrala museer

**Startår:** 2021

**Slutår:** 2023

**Söks hos RAÄ 2021:** 1 186 151

**Avses sökas hos RAÄ 2022:** 1 200 000

**Avses sökas hos RAÄ 2023:** 1 200 000

**Medelsförvaltare**

Nationalmuseum

Nationalmuseum

Box 16176

10324 Stockholm

**Kontaktperson ekonomi**

-

#### Projektledarens personuppgifter

*Namn, titel och personnummer:* -

*Examina, Disputerad år:* Fil dr. i historia

*Adress:* Nationalmuseum, Nationalmuseum, Forskningsavdelningen, Box 16176 , 103 2 Stockholm

*Telefon, ePost:*

#### Kortfattat syfte

Projektets syfte är att utreda om vokabulärer som länkad data stimulerar ny datadriven forskning på museisamlingar och om Wikidata är en lämplig plattform att använda sig av för att länka samman olika auktoriteter och vokabulärer. Detta genom att utveckla en metod för att göra svenska personauktoriteter kompatibla med varandra, dvs genom att öka museisamlingars semantiska interoperabilitet.

#### Samarbetspartner/tvärvetenskap

Nationalmuseum Statens historiska museer Wikimedia Sverige Nätverket digital humaniora, Uppsala universitet Centre for Image Analysis, Uppsala universitet Digisams arbetsgrupp för interoperabilitet fungerar som referensgrupp.

#### Eventuella andra projekt som projektmedlemmar medverkar i

NEO Collections - Digital museum collections in the 21st century (NM) Kartläggning av religion i vardagen: Medeltida helgonkulturer i Sverige och Finland (SHM) The digital Periegesis (UU) Strategisk inkludering av biblioteksdata på Wikidata (WMSE)

**Ev. ytterligare kontaktperson** -

**ePost-adress** -

## Populärvetenskaplig sammanfattning

Projektet underlättar för forskare, och andra, att hitta, förstå, analysera och använda relevant information på tvärs över de medverkande museernas samlingar. Den ökade samsökbarheten uppnås genom att länka ihop persondata från olika museisamlingar. Sammanlänkningen sker via Wikidata, en global och växande dataplattform, vilket gör det möjligt att länka vidare till alla museisamlingar den i sin tur länkar vidare till. I projektet ingår att med hjälp av visualiseringar av samband mellan historiska personer och föremål underlätta förståelse av annars ogreppbara och abstrakta datamängder. Ett ytterligare syfte är att höja kompetensen om länkad data och dess potential inom kulturarvssektorn.

## Summary

The project will make it easier for researchers and other users to find, understand and analyse relevant information distributed across different museum collections. The increased searchability will be achieved by linking data about persons from different museum collections. The data will be linked via Wikidata and thus offers an open, universally useful dataset for analysis and research. The project includes visualisations of relations between historical persons and/or objects to show the potential of otherwise abstract large amounts of data. Another aim is to increase the knowledge about Wikidata and it's possibilities in the cultural heritage sector.

## Bakgrund, syfte och mål

### Bakgrund

Det är idag mycket svårt för forskare och andra användare att med precision söka, hitta i, eller analysera information i samlingar från olika museer. De stora informationsmängder som finns digitaliserade i olika samlingar följer oftast institutionsspecifika dokumentationsstandarder, trots att många museisamlingar överlappar tematiskt. Det kräver att användaren måste sätta sig in i varje samlings dokumentationsmodell och dess tillhörande system för att få ut användbar information. Olikheterna mellan hur de olika samlingarna är registrerade vad gäller personposter, ämnes- och typlistor, och andra typer av kontrollerade listvärden gör det även svårt att automatiskt läsa, analysera och berika samlingsdatat. Samlingarna är inte semantiskt interoperabla. I detta projekt avser vi att utveckla en metod som, tillämpad av svenska museer, avhjälper denna problematik.

### Syfte

Syftet är att förbättra möjligheterna för forskare, användare och maskiner, att med precision och tillförlitlighet söka i och analysera museers samlingsdata, på tvärs över institutionerna.

Sammanlänkning av data gör det enklare att söka, analysera, visualisera, och förstå information som finns samlad om enskilda personer tvärs över olika museisamlingar och att använda information relaterad till dessa personer som källmaterial.

Genom att testa Wikidatas potential som "växel" mellan olika personauktoriteter kommer vi att skapa en metod för museer att jobba med länkad öppen data på befintlig infrastruktur, utan att vara begränsade till eller beroende av enskilda leverantörer. Att använda öppna kunskapsprojekt som Wikidata ger möjligheter för specialister utanför de deltagande museerna att bidra med information, samt tillgängliggör museernas data för en stor internationell användargrupp.

### Mål

- Utveckla en dokumenterad, testad, och till olika museers behov anpassningsbar metod för att länka samman personauktoriteter från olika museisystem.
- Tillgängliggöra hela SHM:s personauktoritet på webben, med persistenta identifierare för varje persondataobjekt.
- Via Wikidata länka samman alla NMs och SHMs auktoritetsposter för personer födda 1500 - 1800.
- Producera och dokumentera minst 5 visualiseringar av datamängden och som visar på möjligheterna med länkade öppna persondata.
- Producera en kvalitativ undersökning om sammanlänkning av auktoritetsdata i projektet är hjälpsam för forskning i de deltagande museer.
- Genomföra föreläsningar, workshops och andra kommunikationsinsatser för att öka förståelsen i museibranschen och forskarsamhället för fördelarna med att använda länkade data-kompatibla auktoriteter och vokabulärer samt samarbeta med Wikidata.

## Teori, metod samt nationell och internationell översikt av tidigare forskning

### Teori

Projektet bygger på visionen om den semantiska webben, en webb av meningsfull och maskinläsbar data. Den underliggande teknologin är länkade öppna data, data som möjliggör kopplingar mellan informationsobjekt från

olika källor. Målet är att förenkla återanvändning och interoperabilitet. För att uppnå detta är det avgörande att informationsobjekt, såsom personposter, har entydiga och beständiga identifikatorer så att de kan refereras till och från i olika sammanhang.

För att publicera länkad data måste institutioner förse sina poster om personer, platser och andra företeelser med entydiga identifikatorer (URI:er). När en företeelse - t.ex. en person - har olika identifikatorer i olika system behöver länkas samman till varandra så att maskiner kan "veta" att de identifierar samma företeelse och då även samla information om samma koncept från olika institutioner. Majoriteten av system som används i kulturarvssektorn har dock inte funktioner för att skapa länkad data av informationsobjekt eller, när de har, enkelt länka samman dem med andra auktoriteter och vokabulärer. I stället för att utveckla egna och dyra lösningar, kan institutioner med fördel använda och samarbeta med öppna och hållbara plattformar som tillhandahålls av andra.

Wikidata är en sådan plattform och har snabbt kommit att bli den mest använda av alla källor till länkad data. Den stora räckvidd och den öppna gemenskap som Wikidata har bidrar till en mångfald och tät sammanlänkning i datat som få andra, ofta smalare och institutionsbaserade, auktoritetsprojekt kan uppnå. Att referera till Wikidatas (auktoritetsdata)objekt från en museisamling är därför ett kraftfullt sätt att koppla samman museisamlingar via personer och andra företeelser. Museernas persondatabaser kan också, via Wikidata, kompletteras med data från externa betrodda vokabulärer.

Att bidra till öppna kunskapsbaser som Wikidata är också ett gott sätt att omsätta principer som OpenScience och FAIR-principerna i praktiken.

### Metod

I projektet använder vi ett urval av auktoritetsposter som vår testdatamängd. Urvalet omfattar personer och andra aktörer födda 1500 - 1800 eKr som beskrivs i Nationalmuseums- och Statens historiska museers samlingar. De deltagande museerna har dokumenterat personerna från olika perspektiv och i olika tekniska system. I och med att museerna skiljer sig i sin forskning och därför i hur och vilken information som sparas kring personerna är vår hypotes att en sammanlänkning av biografiska auktoriteter kommer att ge nya insikter i historiska sammanhang och nätverk mellan personer.

Vi kommer att, via Wikidata, länka samman de i projektet medverkande museernas auktoritetsposter. För att länka samman datat använder vi verktyg som OpenRefine. I de fall museerna har fullständig och väl strukturerad data kan matchningen mot Wikidata ske automatiskt. I andra fall behöver en människa avgöra om matchningen som skett automatiskt är korrekt eller avgöra vilken av ett antal förslag på Wikidataobjekt som är korrekt. I fall där museernas personposter är mycket bristfälliga, kommer manuell länkning krävas. I de fall där museets post inte har en motsvarighet i Wikidata skapar vi nya personobjekt där, med museets personpost som auktoritet.

Vi använder det kraftfulla frågespråket SPARQL i Wikidata Query Service för att ställa komplexa frågor till datat, visualisera svaren, och exportera ut svarsdatat (för vidare analys eller användning).

### Nationell och internationell översikt av tidigare forskning

I Sverige och Norge finns sedan 2013 projektet Kulturnav, som tillhandahåller auktoritetslistor över personer, händelser och termer för objekt i samlingarna. Kulturnav är ett värdefullt verktyg för de anslutna museerna, men är svår att använda för museer som inte använder sig av samlingsförvaltningssystemet Primus. Genom att använda Wikidata istället för en sluten plattform kommer berikningarna att vara tillgängliga oavsett intern förvaltningssystem.

Kungliga biblioteket genomförde 2018-2019 ett liknande projekt, där auktoritetsdata från Libris har laddats upp till Wikidata. Målen var att förenkla och utvidga möjligheterna till sökning av bibliografisk information, att ge biblioteksdata en större synlighet utanför sektorn och att göra det enklare att källbelägga information på Wikipedia.

Riksantikvarieämbetet utvärderade 2019 möjligheterna för kulturarvsinstitutioner att hämta tillbaka berikningar från Wikimedia Commons och Wikidata till de egna dokumentationssystemen. Nationalmuseum har deltagit i denna pilotstudien och importerat över 5000 Wikidata-identifikatorer till personposter. Genom att utvidga pilotstudien på hela Nationalmuseums persondatabas (inkl data om avbildade personer mm), ökas tillgängligheten ytterligare.

Internationellt har bl.a. USAs Library of Congress i ett jämförbart projekt konstaterat att de tack vare integrationen med Wikidata kan tillhandahålla mycket mer och mer relevant information.

### Förteckning över forskningsområdets viktigaste arbeten

Allison-Cassin, S., Scott, D. (2018), Wikidata: a platform for your library's linked open data, Code4Lib Journal, 40 (2018-05-04). (<https://journal.code4lib.org/articles/13424>)

Boyd Davis, S., Vane, O. & Kräutli, F. (2016) Using Data Visualisation to tell Stories about Collections. In: Proceedings of the Conference on Electronic Visualisation and the Arts, EVA 2016, July 2016, London, UK. British Computer Society. pp.221-228.

Brouwer, J., & Nijboer, H. T. (2017). Golden Agents. A web of linked biographical data for the Dutch Golden Age. (<http://ceur-ws.org/Vol-2119/paper6.pdf>)

Cook, S. (2019) The uses of Wikidata for galleries, libraries, archives and museums and its place in the digital humanities, *Comma*: Volume 2017, Issue 2 (doi.org/10.3828/comma.2017.2.12)

Ferriter, M. (2019) Integrating Wikidata at the Library of Congress, available at <https://blogs.loc.gov/thesignal/2019/05/integrating-wikidata-at-the-library-of-congress/>

Foka, A., and Gelfgren, S. (2017) 'Visualisering som verktyg och metod för historieforskning'. In Erixon, P.O. and Pennlert, J (eds) *Digital humaniora och om humaniora i en digital tid Daidalos*: Gothenburg: 147-164.

Goldfarb D. and Merkl, D., (2018) Visualizing Art Historical Developments Using the Getty ULAN, Wikipedia and Wikidata, 2018 22nd International Conference Information Visualisation (IV), Fisciano, 2018, pp. 459-466, (doi: 10.1109/iV.2018.00086.)

Golub, K., Göransson, E., Foka, A., Huvila, I.(2019). 'Digital humanities in Sweden and its Infrastructure: Status quo and the sine qua non'. *Digital Scholarship in the Humanities* Oxford: Oxford University Press.

Klein, M. and Kyrios, A. (2013). VIAFbot and the Integration of Library Data on Wikipedia. *Code4Lib Journal*, 22 (2013-10-14). (<https://journal.code4lib.org/articles/8964>)

Romeo, F. (2016), Bringing [art] knowledge to everyone who seeks it, available at <https://medium.com/digital-moma/bringing-art-knowledge-to-everyone-who-seeks-it-899ec257a55c>

Stinson, A., Wyatt, L, Fauconnier, S (2018), Stepping Beyond Libraries : the Changing Orientation in Global GLAM-Wiki, *Italian Journal of Library, Archives and Information Science = Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione* : 9, 3, 2018][Macerata : EUM-Edizioni Università di Macerata. (DOI:10.4403/jlis.it-12480)

Stinson, A. (2017), Wikidata in Collections: Building a Universal Language for Connecting GLAM Catalogs, available at <https://medium.com/freely-sharing-the-sum-of-all-knowledge/wikidata-in-collections-building-a-universal-language-for-connecting-glam-catalogs-59b14aa3214c>

van Veen, T. (2019). Wikidata. *Information Technology and Libraries*, 38(2), 72-81. (doi.org/10.6017/ital.v38i2.10886)

Vane, Olivia, (2020), Thesis, Timeline design for visualising cultural heritage data PhD thesis, Royal College of Art.

Voß, J. et al. (2014): Normdaten in Wikidata. *Lulu.com*

Whitelaw, M. (2015) Generous Interfaces for Digital Cultural Collections, *digital humanities quarterly*, Volume 9 Number 1. (<http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html>)

Windhager, F., et al. (2018) "Visualization of cultural heritage collection data: State of the art and future challenges." *IEEE transactions on visualization and computer graphics* 25.6 : 2311-2330.

## Plan för genomförandet - tid, aktiviteter och rollfördelning för ansökningsåret

Nedan beskrivs enbart planen för år 1. År 2 och 3 beskrivs vid förnyad ansökan.

### Förberedelser och utbildning, M1-3

I denna aktivitet förbereder vi och planerar projektets genomförande. Vi modellerar även data, tar fram datakvalitetskriterier, samt utbildar personal i användandet av Wikidata, OpenRefine, samt andra verktyg för databearbetning. NM ansvarar, samtliga deltar.

### Katalogisering och databearbetning, M4-9

I denna aktivitet katalogiserar vi persondata i museernas respektive system samt laddar upp och länkar samman dessa i Wikidata. Vi gör initiala analyser och visualiseringar. NM ansvarar, samtliga deltar.

### Analys, visualisering, och felrättning, M9-11

I denna aktivitet skapar och dokumenterar vi komplexa sökningar och datauttagskriterier baserat på konkreta forskningsfrågor. Vi skapar även flera visualiseringar av datat. Identifierade fel i data korrigeras. NM ansvarar, samtliga deltar.

### Utvärdering och resultatspridning M11-12

I denna aktivitet utvärderar vi projektets resultat, kvalitativt och kvantitativt, med fokus på effektmålsanalys, användartillfredsställelse, och datakvalitet. Vi intensifierar vår redan tidigare påbörjade resultatspridning. NM ansvarar, samtliga deltar.

## Förväntade resultat och effekter

Det främsta resultatet är att de medverkande museernas persondata länkas samman till varandra - samt till alla andra museer som också är sammanlänkade med Wikidata. Forskare kan därmed utforska, visualisera, och ladda ner samlingsdata, på tvärs över institutionsgränser, med en frågekomplexitet och precision som innan projektet var omöjlig.

Baserat på detta arbete kommer vi att generalisera, dokumentera och sprida en metod som gör det möjligt för andra institutioner att göra samma typ av sammanlänkningar och visualiseringar. Visualiseringarna tar formen av t.ex. tidslinjer, kartor, och entitets-relationendiagram och tillgängliggörs på ett sådant sätt att andra kan anpassa och använda dem. Spridningen av den utvecklade metoden bidrar till att datakompetensen på Sveriges museer stärks generellt och specifikt vad gäller länkade öppna data, auktoriteter och vokabulärer.

På lång sikt blir effekten att allt fler samlingar blir mer "beforskningsbara" och därmed också mer beforskade. Vidare kommer de museer som implementerar metoden att få bättre översikt och insikt i sina egna samlingar och dess relationer till andra samlingar. Denna effekt blir än större om projektet år 2 och 3 utökar metoden till andra vokabulärer än persondata.

## Resultatspridning

Digisam är en grupp aktörer med stort intresse av att utveckla och förbättra metoder för förvaltning och publicering av samlingar. De kommer att vara en intressent i projektet och bidra med sin expertis. Vidare kommer vi att inom ramen för seminarier och andra events arrangerade av Digisam och dess medlemmar kunna nå ut med våra projektresultat, framförallt till den svenska museisektorn.

GLAMWikirörelsen är en global gemenskap i vilken Wikimedianer och professionella inom kulturarv samarbetar för att stödja varandras uppdrag. Genom Wikimedia's plattformar kan kulturarv få större och internationell användning. Wikimediarerelsens volontärer kan ofta stödja institutioner i deras digitaliseringsarbete. Genom samarbete med denna gemenskap kan vi sprida våra projektresultat nationellt och internationellt.

Uppsala universitets Nätverk för digital humaniora kommer att vara en intressent i projektet och bidra med sin expertis. Vidare kommer seminarier och andra events arrangerade av dem kunna nå ut med projektresultat till forskare och studenter.

Vi kommer att proaktivt bjuda in till samtal om vår metod och sprida projektresultat i flera kanaler och format. Tillsammans har projektdeltagarna stor räckvidd och förmåga att nå ut till intressenterna, nationellt och internationellt.

## **Etiska överväganden**

Datamängden från Nationalmuseum och Statens historiska museer inkluderar endast data som redan är fritt tillgänglig för allmänheten. Datamängden omfattas inte av GDPR då datan beskriver personer som varit döda sedan länge och inte innehåller några känsliga personuppgifter. Med tanke på detta ser vi inga stora risker gällande etiska överväganden för projektet.

Projektet kommer att tillämpa FAIR-principerna för datahantering, det vill säga att uppgifterna ska vara sökbara, tillgängliga, kompatibla och återanvändbara. Projektet kommer att tillämpa god forskningssed så som den beskrivs i t ex ALLEAs Den europeiska kodexen för forskningens integritet.