

Hur FAIR är svensk digitiserad kulturarvsdata idag?

ARON AMBROSIANI* OCH ÅSA M LARSSON**

Abstract

Historical research is increasingly dependent on digital methods to find materials and information online. This requires a very different form of source criticism in order to understand what is available or not through different services. In this paper we present an overview of how digitalised heritage data from Swedish museums and archives are made publicly available on various platforms today, and highlight the often hidden challenges researchers face when trying to make use of them. The study is grounded in the FAIR data principles: to which extent is digitalised heritage data findable, accessible, interoperable and reusable? Content available through aggregated web services by the National Heritage Board, the National Archives and the National Library, among several others, are examined and compared. The published collections data of the Nordic Museum are used to highlight the limits of FAIR. We identify areas where lack of national strategies and digital practices following the FAIR principles cause problems, and suggest ways forward. Digitalisation is much more than digital versions of analogue practices and materials. Trusted vocabularies combined with linked data and persistent identifiers will enable a richer and more multivocal heritage that is not restricted in space or to the information that fits onto a catalogue card. Researchers and heritage institutions need deeper collaboration to ensure digitalisation results in usable data, and they need to embrace good digital practices.

Keywords: digitalisation, collections, digital cultural heritage, FAIR principles, linked open data, controlled vocabularies

Inledning

Detta temanummer av *Lychnos* fokuserar på hur digitalisering har påverkat historisk forskningspraktik. Vi vill i denna artikel uppmärksamma historiker och andra kulturhistoriska forskare på källkritiska aspekter av det digitala tillgängliggjorda kulturarvet, beroende på vilken organisation som ansvarar för det och i vilket syfte det gjorts. Vi gör detta genom att utifrån FAIR-principerna analysera några av de största svenska sök-

* Aron Ambrosiani, fil mag i historia, intendent för digitala samlingar, ArkDes, aron.ambrosiani@arkdes.se

** Åsa M Larsson, fil dr i arkeologi, Riksantikvarieämbetet, asa.larsson@raa.se

plattformarna för digitaliserade samlingar och arkiv, samt genom fallstudier av digitaliserat material från Riksantikvarieämbetet och Nordiska museet. De som eftersöker källmaterial får olika resultat beroende på de val som gjorts av samlingsförvaltare och systemutvecklare. En annan viktig aspekt vi lyfter fram är hur analog praktik påverkat digitaliseringen, så att teknikens möjligheter ännu inte utnyttjas till sin fulla potential. Vi menar att det är av stor vikt att de som söker efter och beforskar digitaliserade material vinnlägger sig om att förstå digitaliseringens kontext, på samma sätt som de vinnlagt sig om att förstå den historiska kontexten. Utan denna förståelse finns en påtaglig risk att datadriven forskning som använder digitaliserat kulturarv blir undermålig eller rent av felaktig. Som historikern Ian Milligan skriver i *The Transformation of Historical Research in the Digital Age* (2022):

Digitized sources will continue to dominate and shape the historical profession. As historians get a deeper understanding of this shift, they will hopefully look back on the first quarter of the twenty-first century as an aberrant period of unreflective digital practice. With more training and attention paid to the mediating influence of platforms, historians need to go down a mental checklist and consider the roles of mediation, algorithmic bias, and context both when they write, read, and evaluate scholarship.¹

Allt mer forskningsfinansiering inom humaniora går till digitalisering. Samtidigt har forskare och minnesinstitutioner delvis olika syften med sina insatser.² Vi argumenterar därför för att den fortsatta digitaliseringen av kulturarvet kräver mycket närmare samarbete mellan minnesinstitutioner och akademi, där båda parter måste bli bättre på att förstå varandras förutsättningar och behov. Minnesinstitutionerna behöver engagera forskare och experter i digitaliseringsprocessen så att kvaliteten på metadata ökar. Forskare i sin tur måste involvera minnesinstitutionerna redan på planeringsstadiet, så att resulterande data blir återanvändbar. En bra utgångspunkt för båda dessa ansatser är att aktörerna förstår och följer de så kallade FAIR-principerna för data: riklig beskrivande metadata, specialanpassade vokabulärer, standardiserade terminologier, beständiga identifierare och tydlig licensinformation.³

Efter en kort bakgrund ges en genomgång av centrala begrepp och en presentation av det digitala materialet och tjänsterna som används för den aktuella studien. Utifrån detta analyseras sedan tillgänglighet (*Accessible*) först via aggregerade plattformar för kulturarvsdata och därefter utifrån Riksantikvarieämbetets respektive Nordiska museets samlingar. En viktig poäng är inte bara att en extremt liten andel av samlingarna är digital tillgängliga, utan att denna tillgänglighet är en glidande skala: från en kort katalogtext om ett helt personarkiv till digitala avbildningar av

enskilda objekt. Att information om digitaliserat material överhuvudtaget är tillgängligt är en förutsättning för att det ska vara sökbart (*Findable*), vilket nästa avsnitt går in på. Där ser vi närmare på framförallt publicering av beständiga identifierare för kulturarvsdata på de stora söktjänsterna, vilket krävs för att kunna hänvisa digitalt till specifika källmaterial.

För interoperabilitet (*Interoperable*) har vi i detta sammanhang valt att studera kompatibilitet mellan olika vokabulärer för begrepp och kategorier som är i användning i databaser och samlingssystem. Wikidatas roll som potentiell resurs lyfts fram men problematiseras. Ett lika stort problem som att inte kunna kombinera likartad information från olika samlingar, är risken att kombinera objektsdata som skiljer sig åt trots att det lätt kan framstå som att det är samma sak. Som exempel på detta lyfts begreppet ”socken” fram, en viktig kategori inom många historiska forskningsämnen, där det finns kronologiska och geografiska aspekter som kan påverka datadrivna analyser. Slutligen presenteras en analys av öppenhet och återanvändbarhet (*Reusable*) av digitaliserade samlingar och utvecklingen under det senaste decenniet. Finns tydliga villkor för återanvändning och i vilken mån sätter de upp hinder för användarna? Detta är avgörande för om forskare ska kunna publicera ny kunskap utifrån det digitaliserade kulturarvet. Vi ser en positiv utveckling på detta område, men mycket återstår att göra. Det är också viktigt att forskare säkerställer att den digitalisering av kulturarvet de genomför inom sina projekt blir återanvändbar för andra.

I den avslutande diskussionen summerar vi vad forskare bör tänka på vid användning av digital kulturarvsdata, och vad de och minnesinstitutionerna kan göra för att underlätta för framtida forskning på digitaliserat källmaterial.

Bakgrund

Under andra halvan av 1900-talet blev det allt vanligare att omvandla analoga material till digitala format (digitisering) och att övergå till digitala arbetssätt (digitalisering). På 2000-talet är detta så självklart att de flesta av oss knappt reflekterar över de digitala system vi är helt beroende av i stort och smått. Omställningen var så snabb och så genomgripande att dess konsekvenser först nu börjar uppmärksammas på bred front.⁴ Det handlar om att kvarlevande analog praxis inte drar full nytta av digitaliseringen eller än värre, att digitala verktyg och metoder vi använder inte är anpassade till källmaterialens beskaffenhet. Detta gäller i synnerhet för humanistisk och kulturhistorisk forskning som ofta får anpassa sig till digitalisering utvecklat utifrån tekniska och naturvetenskapliga premisser. Även om digital humaniora alltmer etablerats som ett eget kunskapsfält

med en rad olika metodologiska tillvägagångssätt, så är det knappast någon historiker, lingvist eller arkeolog verksam idag som inte kan sägas arbeta digitalt i någon mån. Vare sig det handlar om att leta efter och tillgodogöra sig digitaliserade material, eller om att producera egna resultat i digitala format (text, tabeller, illustrationer, databaser).

Till skillnad från naturvetenskaperna så varierar det humanistiska materialets utseende och beskaffenhet i mycket högre grad beroende av tid, plats och kontext – och de begrepp som används kan inte alltid sägas utgöra empiriska kategorier. Denna komplexitet är inte en felaktighet som ska rättas till, det är en grundläggande förutsättning för studiet av människors historia och kulturella uttryck. För att använda ett talesätt inom programmering: "it's a feature, not a bug". Till det kommer att humanioras källmaterial i hög utsträckning består av föremål och handlingar som finns utspridda på ett stort antal mer eller mindre kända samlingar och arkiv, både inom det egna landet och i andra länder. Denna fragmentering är just det som digitaliseringen kan bli ett botemedel för. För att det ska ske krävs det mer än pengar – det krävs tydligare styrning och samarbete på bred front som följer FAIR-principerna.⁵

FAIR-principerna och kulturarvsdata

FAIR-principerna för data fungerar som en vägledning för att data, till exempel den som framställs av forskning eller som skapas i samband med digitalisering, ska bli tillgänglig och användbar i bred bemärkelse.⁶ Akronymen står för sökbarhet (*Findable*), tillgänglighet (*Accessible*), interoperabilitet/kompatibilitet (*Interoperable*) och återanvändbarhet (*Reusable*). Principerna går framförallt ut på att göra data *maskinläsbar*, men det yttersta syftet med detta är att göra data *användbar för människor*. Alla sorters sökningar i digitala register och databaser, på internet eller i slutna system, kräver att maskinen kan tolka vad människan är ute efter och kan identifiera rätt material. Det är dessutom en stor skillnad på att kunna se information på en skärm, och att komma åt underliggande data för vidare bearbetning. Att kunna kombinera data från olika källor för att skapa nya synteser kräver att det finns gemensamma nycklar och begrepp. Slutligen: utan information om villkor för användning (licens) kan digitaliserat material och data inte återanvändas.

För att data och digitaliserat material ska kunna hittas och återanvändas krävs att det finns en beständig identifierare som långsiktigt kan peka på källan. Sökbarheten är dessutom beroende av att digitala objekt (filer) har beskrivande metadata enligt standardiserade termer och vokabulärer. En term eller ett objekt som har en beständig identifierare som tillhandahåller en maskinläsbar definition och beskrivning, helst via en URL (webblänk),

Gustav IV Adolf of Sweden (Q52929)

King of Sweden between 1792-1809

roi de Suède Gustave IV | King of Sweden Gustav IV Adolph | kung av Sverige Gustav IV Adolf | King of Sweden Gustav IV Adolf | King of Sweden Gustaf IV Adolf | King Gustav Adolph of Sweden | Gustav IV Adolf | roi de Suede Gustave IV | Gustavus IV

- In more languages

Configure

Language	Label	Description	Also known as
English	Gustav IV Adolf of Sweden	King of Sweden between 1792-1809	roi de Suède Gustave IV King of Sweden Gustav IV Adolph kung av Sverige Gustav IV Adolf King of Sweden Gustav IV Adolf King of Sweden Gustaf IV Adolf King Gustav Adolph of Sweden Gustav IV Adolf Gustavus IV
Swedish	Gustaf IV Adolf	kung av Sverige 1792–1809	Överste Gustafsson Gustav IV Adolf Gustaf IV Adolph
Finnish	Kustaa IV Aadolf	Ruotsin kuningas 1792–1809	
Tornedalen Finnish	No label defined	No description defined	

All entered languages

edit

Wikipedia (50 entries) edit

arwiki	غوستاف الرابع أدولف
arzwiki	جوستاف الرابع أدولف
azbwiki	دۆرۆنچو قوستاو ادولف (سوئد)
bewiki	Густав IV Адольф
bgwiki	Густав IV Адолф
brwiki	Gustav IV Adolf
cawiki	Gustau IV Adolf de Suècia
cswiki	Gustav IV. Adolf
dawiki	Gustav 4. Adolf af Sverige
dewiki	Gustav IV. Adolf (Schweden)
elwiki	Γουσταύος Δ΄ Αδόλφος της Σουηδίας
enwiki	Gustav IV Adolf
eswiki	Gustavo I.a Adolfo (Svedio)
eswiki	Gustavo IV Adolfo de Suecia
etwiki	Gustav IV Adolf
euwiki	Gustavo IV.a Adolfo Suediakoa
fiwiki	گوستاو چهارم ادولف
fiwiki	Kustaa IV Aadolf
frwiki	Gustave IV Adolphe
glwiki	Gustavo IV Adolfo de Suecia

Fig. 1. Gustav IV Adolf i Wikidata

kan kallas för en *auktoritet*. Ett typiskt exempel på auktoriteter är de som Wikimedia-rörelsen publicerar via Wikidata och som låter användare hitta rätt artikel oavsett om de söker på ”Gustav IV Adolf” eller ”överste Gustafsson” (<http://www.wikidata.org/entity/Q52929>), eller låter dem skilja på staten Danmark (www.wikidata.org/entity/Q35) och orten Danmark i Uppland (www.wikidata.org/entity/Q5273924).

Detta är i korthet vad FAIR-principerna går ut på, vilket den enskilde forskaren själv kan applicera på sin forskningsdata: öppna dataformat för långtidsbevarande, standardiserad metadata och vokabulärer, beständiga identifierare och tydlig licensiering. Samt öppen publicering av åtminstone beskrivande metadata och länkar till källan. Det är helt i enlighet med FAIR-principerna att juridiskt eller etiskt känsligt digitalt material inte är fritt tillgängligt på nätet, men källdata för analyser och slutsatser måste bevaras och det måste gå att redovisa var källmaterialet förvaras. Detta gäller för kulturarvsdata oavsett om de tillkommit genom forskningsprojekt, publika projekt (crowdsourcing) eller löpande samlingsförvaltning.

Material och metod

Vi vill i vår artikel titta närmre på nuläget för kulturarvsdata i Sverige utifrån FAIR-principerna. Detta gör vi på tre nivåer:

1. Utifrån svenska plattformar för aggregerat digitaliserat kulturarv och hur de förhåller de sig till varandra, eftersom de belyser hur samma material kan presenteras och tillgängliggöras på olika sätt.
2. Utifrån innehållet i aggregatorn K-samsök och hur det har förändrats

över det senaste decenniet, vilket återspeglar utvecklingen vid både stora och små minnesinstitutioner.

3. Utifrån specifika exempel från Nordiska museet och Riksantikvarieämbetet på hur kulturarvsdata tillkommit och hur data och processer förhåller sig till FAIR-principerna på en konkret ”hands on”-nivå.

På så sätt hoppas vi erbjuda en värdefull ögonblicksbild av hur olika kulturarvsinstitutioner tillhandahåller digital kulturarvsdata. Studier av kulturarvsdata på övergripande nivå är värdefulla (se t.ex. Golub et al. (2022)), men innebär också att detaljerna försvinner i mängden; vi ser skogen men inte de olika träden. Det aggregerade kulturarvet problematiseras därför i denna artikel genom att presentera fallstudier som belyser institutionella variationer. Vi undersöker hur kulturarvsdata tillkommit och hur data och processer förhåller sig till FAIR-principerna på en konkret nivå vid Nordiska museet och Riksantikvarieämbetet. Att just dessa två institutionernas data är utvalda för fallstudierna beror på att artikelförfattarna har erfarenhet av att arbeta med digital kulturarvsdata vid dem. Syftet är inte att värdera olika sorters digitaliseringsstrategier, utan enbart att ge läsaren en bättre inblick i hur kulturarvsdata uppstår inom en specifik verksamhet.

Digital svensk kulturarvsdata – en överblick

Digitaliserad svensk kulturarvsinformation finns idag tillgänglig genom en mängd olika tjänster och plattformar. Dessa tillhandahålls av minnesinstitutioner, myndigheter, konsortier och privata företag men är i mycket liten utsträckning sammanlänkade med varandra, vilket skapar utmaningar för den som vill hitta liknande information på flera olika plattformar. Även för specialister är det svårt att förstå vad som kuraterats och filtrerats via de olika tjänsterna och vilka källkritiska aspekter som finns på innehållet – samt vad som saknas digitalt och endast finns i fysisk eller analog form. Vi kommer i denna artikel fokusera på innehållet i K-samsök (tillhandahållen av Riksantikvarieämbetet), men andra större plattformar för svensk kulturarvsdata kommer också omnämnas: Libris (Kungliga biblioteket), Alvin (Uppsala universitetsbibliotek och tillhörande konsortium), Nationell ArkivDatabas (NAD) (Riksarkivet), DigitaltMuseum (KulturIT) och KulturNav (KulturIT). På senare år har Wikidata (Wikimedia Foundation) etablerat sig som en slags sambandscentral där termer och objekt kan kompletteras med ”samma som”-länkar till andra plattformar med hjälp av i skrivande stund 6 920 olika externa identifierare.⁷

K-samsök är en nationell aggregator för kulturarvsdata som utvecklats och förvaltas av Riksantikvarieämbetet sedan 2008.⁸ K-samsök samlar in och tillgängliggör kulturarvsdata från svenska minnesinstitutioner genom

att de installerar en lokal port i sina databaser. Via porten levereras metadata från, och länkar till, innehållet i samlingsdatabaserna. Det är institutionerna själva som bestämmer vilka delar av innehållet i databaserna som ska tillgängliggöras och de mappas mot K-samsöks termer så att det går att söka i flera olika databaser samtidigt, till exempel på person, plats, material, typ av objekt eller tidsperiod. Kvaliteten på informationen och sökträffarna är helt beroende av den data som tillhandahålls av samlings-systemen och om de har mappats mot rätt termer. K-samsök skapar även beständiga identifierare i form av webblänkar till objekten.⁹

Förutom rent beskrivande metadata så kan även digitaliserade bilder och skannade dokument tillgängliggöras av K-samsök genom en teknik som kallas API (Application Programming Interface). Originalen finns kvar i institutionernas system, men genom att använda sig av API:et så kan vem som helst med rätt kunskaper utveckla nya appar eller webbgränssnitt som kan visa upp bilder och data från samlingarna. Söktjänsten *Kringla* som tagits fram av Riksantikvarieämbetet visar upp all data som finns tillgänglig via K-samsök, inklusive bilder i de fall det finns och med länkar till källorna. Ett annat exempel på användning av API:et är webbplatsen *Runor*¹⁰ framtagen av Uppsala universitet och Riksantikvarieäm-

Runor

< TILLBAKA

Sö 179

Gripsholm
Kärnbo socken

Även känd som:
L 927

Inskrift

Runtext

• tula : lit : raisa : stain : pinsa| at : sun : sin : haralt : brupur : inkuars : þain furu : trikkla : fiari : at : kull : auk : austarla| arni : kafu : tuu : sunarla : a sirklan-ti

Formvästnordiska

Töla lét reisa stein þenna at son sinn Harald, bróður Ingvars. Þeir fóru drengila fiarri at gulli ok austarla erni gáfu, dóu sunnarla á Serklandi.

Runsvenska

Töla lét reisa stein þenna at sun sinn Harald, bróður Ingvars. Þeir fóru drengila fiarri at gulli ok austarla ærri gáfu, dóu sunnarla á Serklandi.

Engelska

Töla had this stone raised in memory of her son Harald, Ingvar's brother. They travelled valiantly far for gold, and in the east gave (food) to the eagle. (They) died in the south in Serkland.

Svenska

Töla lét resa denna sten efter sin son Harald, Ingvars broder. De for manligen fjarran efter guld och österut gav örnen (mat), de dog söderut i Särkland.

Referera enligt följande:

Runinskrift Sö 179 i Samnordisk runtextdatabas 2020, Institutionen för nordiska språk, Uppsala universitet.
<http://kulturarvsdata.se/uu/srdb/4013365a-38f1-4ec2-9f39-2c897d5914a>

☞ Läs mer om hur du citerar innehåll i Runor

FAKTA

Objektuppgifter
Föremål: **sten, runsten**
Runsten
Material: **granit (sten)**
Granit
Ristare:
Äskil 2, Attribuerad
Runtyp: Okänd
Stil: Fp

Ingvarsten

Proveniens
Plats: Gripsholm
Socken: Kärnbo socken
Härad: Selebo härad
Landskap: Södermanland
Kommun: Strängnäs
Län: Södermanland
Land: Sverige

Ursprunglig plats
Nej

Datering
1010-1050
Vikingatid

ID i nationellt lämningsregister
Sverige
fmsid: 10034300330001
kmrId: c51d7a51-9583-40af-96f5-69ebe924f4b8
raÄnr: 33:1
raaParish: Kärnbo
parishCode: 0343

Objekt URI
<http://kulturarvsdata.se/uu/srdb/4013365a-38f1-4ec2-9f39-2c897d5f914a>

BILDER

K-samsök
Klicka på en bild för förstoring och information.



< 1 2 >

DKUMENTATION

Relaterade dokument (3)

 ☑ Delat inventeringsbokuppslag, Lokal 33/34 (2)

 ☑ Inventeringsbokuppslag (1)

 ☑ Inventeringsbokuppslag (2)

Fig. 2. På webbtjänsten Runor kan användare söka på runinskrifter och träffsidan visar uppgifter, arkivhandlingar och bilder aggregerade från flera olika källor med hjälp av K-samsöks API.¹¹

betet. I denna söktjänst kan användare söka på runinskrifter från alla perioder och hela världen och se information, bilder och dokument från flera olika källor samlat på en och samma sida.

Det är frivilligt och kostnadsfritt för kulturarvsinstitutioner att ansluta sina samlingsförvaltningssystem till K-samsök. För att institutionens kulturarvsdata ska kunna skördas av K-samsök krävs dock särskild teknisk anpassning som kostar en del i tid och resurser; K-samsök skördar med hjälp av det standardiserade protokollet OAI-PMH men med ett eget metadataformat/datamodell.¹² Flera etablerade samlingsförvaltningssystem stödjer detta men inte alla. Sammantaget innebär det att informationen i K-samsök är omfattande men inte komplett. Vid utgången av 2021 hade 80 institutioner anslutit sig. Det motsvarar cirka 64 procent av läns- museerna och 60 procent av centralmuseerna. Utöver det ingår vissa regionala museer, stadsmuseer, militärhistoriska museer, kommunarkiv, universitetsbibliotek och stiftelser samt Riksantikvarieämbetet självt som levererar metadata och bilder från sina arkiv (Arkivsök) och kulturhistoriska register (Fornsök, Bebyggelseregistret).

Vokabulärer – en grundläggande förutsättning för FAIR kulturarvsdata

En grundläggande förutsättning för FAIR är att det finns kontrollerade vokabulärer för begrepp och termer, vilka bör förses med beständiga identifierare och allra helst vara både människo- och maskinläsbara.¹³ Några av de digitalt tillgängliga vokabulärer som är relevanta för svenskt kulturarv och historia är de som publiceras av Riksarkivet för de termer (exempelvis orter) som ingår i Nationell ArkivDatabas, av Kungliga biblioteket för termer och ämnesord i Libris och av KulturIT i systemet KulturNav.¹⁴ Det finns också flera internationella vokabulärer som i hög grad är användbara för svenska forskare: Getty Research Institute publicerar AAT för konst och arkitektur, IA för ikonografiska begrepp och ULAN för konstnärer, Nomisma är en vokabulär för numismatiska termer och GeoNames samlar geografiska platser, för att bara nämna några.¹⁵ Dessutom finns auktoriteter som identifierar enskilda objekt i samlingar, arkiv och bibliotek, till exempel de som publiceras av Riksantikvarieämbetet i K-samsök och av KulturIT på DigitaltMuseum. Vare sig de är generella för begrepp eller specifika för objekt är dessa identifierare av stor betydelse för att göra samlings- och forskningsdata sökbar och interoperabel.

Detta kan verka som en betryggande mängd vokabulärer, men det finns fortfarande stora luckor för den som är verksam inom svensk historisk och förhistorisk forskning. Skapandet av dem har nämligen i första hand styrts utifrån systemutvecklingsbehov, inte behoven hos fältet i dess helhet. Det

saknas till exempel en digitalt tillgänglig vokabulär med auktoriteter för kulturhistoriska lämningstyper. Att auktoriteterna dessutom skulle kunna användas av människor i skapandet av forskningsdata är en insikt som kommit ganska sent. Även när det tekniskt sett finns auktoriteter är det inte alltid lätt för den enskilde användaren att hitta dessa, eftersom de ofta publiceras främst för maskinläsbarhet av andra system. Auktoriteter för svenska tidsperioder publiceras till exempel av Statens historiska museer, men för att hitta den beständiga länken för en specifik period utan att kunna tyda XML-kod behöver man i nuläget använda söktjänsten för den internationella aggregatorn PeriodO, som inte är helt lättnavigerad.¹⁶

Lite enklare att hitta auktoriteter och länkar är det via tjänsten KulturNav. En egenhet med plattformen är dock att den ibland innehåller flera olika auktoritetsposter för samma begrepp. Detta beror på att KulturNav innehåller flera överlappande vokabulärer, skapade av olika användare av samlingssystemet Primus. Till exempel har textilforskaren Gunnel Hazelius-Berg två auktoritetsposter; en från Nordiska museet och en från Världskulturmuseerna.¹⁷ Även icke museianknytna listor i KulturNav uppvisar dubletter, till exempel tidsperioden ”medeltid”.¹⁸

Wikimedia Foundations plattform Wikidata har auktoriteter som är enkla att söka upp, men dess universella anspråk innebär svårigheter inte minst för humanistisk forskning, till exempel när det kommer till tidsperioder som har skilda kronologiska spann i olika regioner (återigen: ”medeltid”). Vem som helst kan skapa en ny wiki-auktoritet eller redigera en existerande och den öppenheten är en faktor att ta med i bedömningen av om man ska förlita sig på dess definitioner.

Det är med andra ord inte lätt för forskare att veta vilka auktoriteter de ska använda sig av. Problemet för forskare, och för minnesinstitutioner som inte vill behöva uppfinna hjulet i samband med digitalisering av samlingar, är att det saknas ett nationellt helhetsgrepp kring att både identifiera och möta de behov som finns kring kontrollerade vokabulärer. Det finns sammanställningar av vokabulärer som är relevanta för kulturarvet, men det är svårt för den enskilde att sätta sig in i för- och nackdelar med olika alternativ.¹⁹ Ska den som forskar om Selma Lagerlöf använda sig av Libris, K-samsöks, KulturNavs eller Wikidatas person-auktoritet? Gör det någon skillnad? Finns det auktoriteter som kan hjälpa forskaren att specificera att det ”Sverige” som förekommer i dennes forskningsdata motsvaras av en specifik kronologi och geografi? Digitaliseringen har potential att revolutionera humanistisk forskningsmetodik, men för att detta ska bli verklighet så krävs att kulturarvsdata lever upp till FAIR-principerna.

Tillgänglighet (A)

Först ett förtydligande: med ”tillgänglighet” menas i detta sammanhang publik digital tillgång till data, på engelska benämnt ”access”. Tillgänglighet för personer med funktionsnedsättning (”accessibility”) är inte i fokus för resonemangen som följer, även om detta såklart också är relevant, särskilt för de plattformar som omfattas av det så kallade webbtillgänglighetsdirektivet.²⁰ I denna artikel fokuserar vi alltså på huruvida kulturarvsdata är digitalt tillgänglig i bemärkelsen att den går att ta del av via webben.

Forskare och studenter använder i allt högre grad digitalt tillgängligt material i sina studier och mycket av den inledande informationssökningen utgår från digitala plattformar, från generella som Google till specifika för en arkivsamling.²¹ Det är viktigt att digitalisering av arkiv och samlingar tar hänsyn till forskarnas perspektiv och behov. Dels eftersom de har ett professionellt behov av digitaliserat material, dels eftersom de genom sitt arbete skapar ny kunskap, vilket i mångt och mycket är det yttersta syftet med samlingarna.

Att fråga sig hur mycket som återstår att digitalisera av kulturarvet är som att fråga sig hur långt borta horisonten befinner sig. Tillgänglighet handlar om mer än att ha överfört informationen från ett analogt registerkort till ett digitalt samlingssystem, eller att ha skapat en digital avbildning med kamera eller skanner. För att bli tillgängligt måste kulturarvsdata beskrivas med metadata och framförallt måste det publiceras. Digital tillgänglighet för kulturarvssamlingar, vare sig det rör sig om pappershandlingar, föremål, fotografier, konstverk eller andra typer av objekt (ljudinspelningar, filmer, berättelser) kan definieras på olika nivåer. Det mest grundläggande är om det ens finns ett digitalt register över samlingarna (metadata). Registreringsnivån skiljer sig ofta åt beroende på institutionell tradition och typ av material; arkivmaterial registreras ofta på arkiv- och volymnivå (även om allt fler arkiv digitaliseras ända ned till enskilda ark sökbara i fulltext). Föremålssamlingar som digitaliseras registrerar ofta enskilda föremål och föremålsgrupper, men här kan det ske en medveten prioritering av de objekt som anses estetiskt tilltalande. Gamla arkeologiska undersökningar kan ha stora mängder fynd som enbart digitaliseras utifrån materialtyp (keramik, sten, ben), och där bara enstaka föremål får en egen post i systemet i samband med digitalisering.

Även om de allra flesta minnesinstitutioner numera har övergått till digitala samlingssystem i någon form, så betyder det inte att informationen i dem vare sig är i närheten av att vara komplett, eller att den är tillgänglig för utomstående. Bildfångst på materialet är dessutom än mer sällsynt än rena kataloguppgifter. Riksarkivet har gjort en grov uppskatt-

ning av att fem procent av deras material har bilder – och endast två procent är tillgängligt online.²² Det finns en överhängande risk att studenter och forskare som nu har tillgång till olika sorters sökplattformar in i arkiv och samlingar överskattar hur mycket av det analoga materialet som är digitalt tillgängligt och att detta kan snedvrída forskningen.

Tillgänglighet i aggregerade plattformar för kulturarvsdata

En utmaning för den som söker information är att kulturarvet är utspritt på ett mycket stort antal minnesinstitutioner, från nationell till lokal nivå. Denna fragmentering kan motverkas av plattformar som aggregerar kulturarvsdata från flera olika källor.

K-samsök har sett en stadig ökning av antalet medverkande institutioner sedan lanseringen 2008. Konsekvent statistik över innehållet finns framförallt från och med år 2013, då 42 institutioner redan var anslutna.²³ Drygt 5,3 miljoner digitala objekt var då indexerade och tillgängliga via API samt söktjänsten Kringla, varav 40 procent hade någon sorts bildmedia och 29 procent hade en ”geotagg”, det vill säga en uppgift om en geografisk position som relaterar till objektet, vilket kan handla om allt från fyndort till avbildad plats. År 2021 hade K-samsök ökat i omfång till 80 anslutna institutioner, vilket motsvarar en majoritet av central- och läns-museerna, men inte samtliga. Riksarkivet och Kungliga biblioteket levererar heller inte metadata om sina samlingar till K-samsök. Tjänsten innehåller information om över 9,3 miljoner digitala objekt, varav 58 procent har en avbildning. Andelen objekt med geotagg minskade under perioden från 29 procent till 19 procent, vilket reflekterar att de flesta nytillkomna digitala samlingar saknar den typen av metadata för exakt geografisk position. Totalt har antalet digitala objekt ökat med 74 procent och antalet digitala bildfiler ökat med 150 procent sedan 2013.²⁴ Även om det är en betydande mängd tillgänglig kulturarvsdata motsvarar det fortfarande bara en bråkdel av allt som finns bevarat i museisamlingar och arkiv.

En annan stor plattform för kulturarvsdata är DigitaltMuseum, som är den gemensamma söktjänsten för de institutioner i Norge och Sverige som använder sig av samlingssystemet Primus. I augusti 2022 fanns knappt 3,7 miljoner digitala objekt från cirka 92 museer på den svenska delen av DigitaltMuseum.²⁵ Det är viktigt att notera att många av de minnesinstitutioner som använder sig av Primus också levererar metadata och länkar till K-samsök. Innehållet på DigitaltMuseum är därför delvis detsamma som i söktjänsten Kringla och vice versa. Detta är inget negativt; tvärtom är det digitaliseringens främsta potential att kulturarvet kan tillgängliggöras på olika plattformar, i olika kombinationer och utifrån olika sökkriterier.

Nationell ArkivDatabas (NAD) är en plattform som tillgängliggör arkivinformation från Riksarkivet samt drygt 650 andra anslutna minnesinstitutioner.²⁶ Från och med april 2022 finns metadata om cirka 85 000 arkiv från Riksarkivet, Malmö stadsarkiv, Stockholms stadsarkiv och Värmlandsarkiv samt cirka 80 000 skannade arkivhandlingar ur statliga arkivbestånd, också tillgängligt via ett öppet API.²⁷ I söktjänsten redovisas kulturarvsdata som arkivbestånd och volymer och varje enskild post i NAD kan därför innehålla hundratals eller tusentals brev, fotografier, dokument och handlingar. NAD visar dessutom bildfångst enbart för de fyra institutionerna ovan. Digitaliserade bilder och dokument samt detaljerad objektsinformation kan alltså mycket väl finnas hos minnesinstitutionen, men det framgår inte av träfflistan, inte heller finns det länkar till källorna. Med undantag för de fyra institutionerna uppdateras inte informationen i NAD automatiskt, utan det sker genom manuell export från respektive institution, där vissa institutioner uppdaterar flera gånger i månaden medan det för andra kan gå flera år mellan uppdateringarna.²⁸ Mycket mer detaljerad digitaliserad kulturarvsdata kan alltså finnas än vad som framgår av denna söktjänst, men för att ta reda på det måste användaren söka vidare på andra plattformar, som Kringla eller respektive institutions egna söktjänster. Samtidigt finns det ett värde i att kunna söka på hela bestånd av arkivsamlingar, för den forskare som vill ta del av allt som en viss person lämnat efter sig eller som berör en organisation eller lokalitet.

Utmaningen ligger i att som användare förstå vad som har filterats ut och tillgängliggjorts – eller inte – på de många olika plattformarna som finns tillhanda. Kungliga biblioteket har portalen Libris för den som söker i de nationella biblioteksbestånden, men KB:s bestånd är inte sökbart via plattformarna NAD eller Kringla, endast via Libris samt deras söktjänster för samlingarna (Regina, SMDDB, Svenska dagstidningar och så vidare). Ytterligare en plattform som tillgängliggör kulturarvsdata är portalen Alvin som utvecklats av ett konsortium bestående av Uppsala, Lunds och Göteborgs universitetsbibliotek för deras biblioteks-, arkiv- och museisamlingar, men som numera även tillgängliggör data från flera andra lärosäten och minnesinstitutioner. I augusti 2022 fanns knappt 380 000 poster i söktjänsten.²⁹ Alvin-konsortiet har utvecklat en port för att leverera metadata och länkar till K-samsök, så att innehållet kan tillgängliggöras även via Kringla, men i dagsläget återstår en del arbete innan detta blir verklighet. Alvins resurstyper måste till exempel mappas mot K-samsöks objekttyper och de är inte helt synonyma.

Tillgänglighet – Riksantikvarieämbetet

Riksantikvarieämbetet har digital information om fler än en miljon objekt från sina arkiv och en stor andel av dessa har någon sorts bildrepresentation: 90–100 % av dokument, fotografier, teckningar, litteratur och kartor. Undantaget utgörs av ritningar där bara 10 % har bildfångst utöver kataloginformation.³⁰ Hur stor andel det digitaliserade utgör av allt analogt material går dock inte att säga. Mycket av innehållet i arkivet är inte katalogiserat i detalj ens analogt: det handlar om pärmar med allt från rapporter och brev till fotografier och teckningar. Gamla digitaliseringar kan dessutom behöva göras om eller kompletteras med ny metadata för att nå upp till dagens nivå.

Digitalt fött arkivmaterial utgör ytterligare en utmaning, till exempel digital dokumentation från arkeologiska undersökningar som kan finnas på servrar och på CD- och DVD-skivor – inte sällan i proprietära format som riskerar att bli oanvändbara med tiden. Den som känner till exakt vilken mätfil de behöver kan kontakta arkivet och få en kopia skickad till sig, men de är inte tillgängliga i ordets rätta bemärkelse. Åren 2020–2023 pågår ett samarbetsprojekt mellan Uppsala universitet och Riksantikvarieämbetet, finansierat av Riksbankens Jubileumsfond, i syfte att rädda en del av den digitalt födda dokumentationen från Riksantikvarieämbetets arkeologiska utgrävningar och göra den FAIR.³¹

Digitisering och digitalisering betyder alltså inte nödvändigtvis tillgänglighet, och även i de fall data görs tillgänglig för allmänheten kan detta ske med begränsningar. Att all information i ett samlingssystem inte kan vara öppet är självklart (till exempel den exakta placeringen av objektet), men riklig samlingsdata är av stort intresse för forskare och kan användas som underlag för datadrivna analyser. En positiv utveckling de senaste åren är därför att allt fler minnesinstitutioner tillgängliggör sin samlingsdata på nätet, men det sker fortfarande främst utifrån andra användarbehov än forskningens.³² Framför allt tillgängliggörs samlingarna via webbsök med ett fåtal förvalda kriterier och resultatet redovisas som träfflistor där informationen i bästa fall kan laddas ned som textfiler, ibland enbart objekt för objekt. Att få sökresultaten som strukturerad data underlättar vidare bearbetning, men för det krävs idag ofta kompetens inom programmering. Även när den kompetensen finns kan mycket av samlingarnas data vara oåtkomlig, trots att det inte rör sig om känsliga uppgifter, eftersom man inte bedömt det som viktigt för den förment primära användargruppen – den intresserade allmänheten.

För att kunna tala om sann digital tillgänglighet enligt FAIR-principerna behöver forskare kunna göra riktiga datauttag i strukturerade format med så mycket information som möjligt. Sådana datauttag skulle inte bara spara tid, utan även kunna användas som grund för kvalitetshöjning av

samlingsinformationen om forskningsprojekten levererar data tillbaka till minnesinstitutionerna.

Tillgänglighet – Nordiska museet

De aggregerade mätetalen från NAD och K-samsök ger en övergripande bild över digitaliserat kulturarv, men ytterligare nyanser blir synliga om situationen beskrivs på institutionsnivå. Nordiska museets samlingar består av ”cirka 1,5 miljoner föremål, 6 800 arkiv, 6 miljoner fotografier och uppskattningsvis 260 000 volymer böcker, tidskrifter och annat tryckt material”.³³ Digitiseringen av samlingarna redovisas löpande i museets årsredovisningar. Förutsättningarna för digitisering skiljer sig åt främst utifrån materialtyp; olika kulturarvsdata digitiseras på olika sätt.

När det gäller föremålssamlingarna uppgår antalet poster i samlingsförvaltningssystemet Primus till drygt 550 000 poster vilket på grund av databasens utformning motsvarar samtliga cirka 1,5 miljoner föremål. Katalogposterna är digitaliserade sedan lång tid tillbaka och fungerar kvalitativt annorlunda än den äldre analoga hanteringen i form av huvudliggare och kortkataloger. Viss information från katalogkorten följde dock inte med i digitaliseringen; därför är även själva katalogkorten digitaliserade och finns sökbara i en separat databas, endast åtkomlig för intern personal. Drygt 200 000 (37 procent) av föremålsposterna inkluderar minst en digital bildfil. Inte alla poster är publikt tillgängliga; vid årsskiftet 2021 var cirka 356 000 föremålsposter (65 procent) publicerade på DigitaltMuseum varav cirka 136 000 hade bildfil.³⁴

Alla användare, vare sig det är allmänhet eller specialister, uppskattar naturligtvis på att få en bildrepresentation av föremålen i samlingarna, vilket i Nordiska museets fall är möjligt för knappt en fjärdedel av det totala antalet föremålsposter; en såvitt vi kan bedöma relativt hög andel bland jämförbara institutioner. Externa forskare kan efter personlig kontakt få tillgång till opublicerade poster, men det är en mer omständlig process än för publicerade kulturarvsdata och uppfyller varken F eller A i FAIR-principerna.

Nordiska museets arkiv omfattar runt 6 000 hyllmeter handlingar.³⁵ Merparten av materialet är inte digitalt registrerat. Ungefär 21 % av arkiven är digitalt förtecknade och 10 % är publicerade i NAD.³⁶ Liksom för arkivsektorn i stort är de digitala sökningarna mer att betrakta som stickprov eller skyltfönster för de betydligt mer omfattande fysiska samlingarna. Utvalt innehåll, i första hand maskinavskrifter av frågelistsvar, har digitaliserats och finns tillgängliga (och sökbara i maskintolkad fulltext) genom en separat tjänst utvecklad av KulturIT.³⁷ Av museets cirka 6 miljoner fotografier är 106 534 (knappt 2 %) förtecknade som enskilda fotografier,

varav 72 % har digital bildfil.³⁸ Resterande fotografiskt material finns endast förtecknat på volymnivå eller inte alls.

Aggregerade söktjänster som NAD, DigitaltMuseum, Kringla, Alvin och Libris, precis som specifika söktjänster för enskilda minnesinstitutioner, är självklart en enorm hjälp för både allmänhet, studenter och forskare. Men det är viktigt att förstå på vilka premisser det digitala materialet förbereds och presenteras och varför det kan skilja sig åt mellan dem. Att det går att söka i samlingar är inte samma sak som att det går att hitta det man letar efter, vilket leder oss vidare till nästa FAIR-princip.

Sökbarhet (F)

Att data finns publikt tillgängliga via API-anrop betyder inte att användare kan *hitta* innehållet. För det krävs antingen kunskap inom programmering eller, vilket är långt vanligare, utveckling av webbgrensnitt som tillåter sökningar på ett användarvänligt sätt. Söktjänsterna levererar dock sina resultat utifrån på hur de designats och hur de tolkar användarens fråga – vilket kan variera även med sökningar i samma sorts datamängder. Ett exempel: eftersom Primus och K-samsöks datamodeller inte är identiska kan det leda till att samma objekt definieras olika i respektive söktjänst. I Kringla kan användare filtrera på typen ”dokument”, medan sådana handlingar oftast hittas under ”föremål” i DigitaltMuseum.

Sökbarheten försvåras ytterligare av att olika institutioner kan ha registrerat informationen enligt den logik som bäst passar deras existerande system. Detta blir extra tydligt i den begreppsförvirring som inte sällan uppstår över information om typ av objekt och upphovsperson när det handlar om digitala reproduktioner av fysiska objekt. En karta kan vara registrerad som ett fotografi, eftersom själva avbildningen är ett foto, och den angivna upphovspersonen kan vara fotografen snarare än den person som skapade kartan. Sökbarhet är alltså i hög grad beroende på de mänskliga faktorerna i samband med registrering av metadata, att det finns en gemensam förståelse för vad de olika attributen refererar till. En digital bildfångst av ett kulturarvsobjekt, vare sig det är en byggnad, ett dokument eller ett föremål, ger upphov till en egen uppsättning metadata som måste särskiljas från det som refererar till originalet.

Beständiga identifierare för kulturarvsdata

Enligt FAIR-principerna bör varje unik datapost ha en *beständig identifierare* (också kallat PID, ”Persistent Identifier”): en utpekad unik identifierare, gärna i form av en internetadress (URL), med åtminstone lika lång livstid som den utfärdande institutionen.³⁹ En beständig identifierare som innehåller information om i vilket sammanhang just detta ID-nummer

hör hemma, vilken organisation som är ansvarig för publiceringen, kan användas för att hänvisa till en källa på ett användbart och trovärdigt sätt för både maskiner och människor. De kan också användas för att redovisa vad som ingår i forskningsdata på ett sådant sätt att dessa dataset i sin tur blir sökbara. Maskinell sökbarhet via beständiga identifierare fungerar alltså åt *båda hållen*: att hitta källmaterial som ingår i studier och att hitta studier utförda på vissa källmaterial.

Det som krävs är att forskare vinnlägger sig om att få tillgång till och att redovisa dessa identifierare på ett korrekt sätt, vilket tyvärr inte är helt enkelt i dagsläget. De länkar som syns i adressfälten på webbgränssnitten vid en träff på en sökning är endast i undantagsfall beständiga och/eller människoläsbara. Istället måste användaren söka i den övriga informationen på träffsidan i hopp om att hitta något som kan fungera som en beständig identifierare, helst också människoläsbar.

Så hur ligger de stora plattformarna för kulturarvsdata till när det gäller redovisning av beständiga identifierare? Alvin tillhandahåller en ”permanent länk” till varje post med hjälp av URN-resolvern urn.kb.se. DigitaltMuseum i sin tur anger två identifierare på varje post: ”DigitaltMuseum ID” och ”DigitaltMuseum UUID”. Det framgår dock inte explicit att de ska betraktas som beständiga identifierare. Kringla tillhandahåller en ”fast länk” samt ytterligare fem identifierare dolda under den utfällbara fliken ”Teknisk data”; tre pekar till den ursprungliga posten och två är datorläsbara presentationer av den bakomliggande K-samsökposten. Det är alltså inte helt uppenbart för en forskare vilken av dessa identifierare som bör användas. Nationell ArkivDatabas tillhandahåller en ”kopiera länk”-knapp, medan Libris tillhandahåller en permanent länk kallad ”Länka till posten”. Under fliken Detaljer och rubriken ”Länkade data-URI:er (test)” återfinns länkar till relaterade poster i Libris XL (såväl auktoritetsposter för författare som termer för ämnesord). Tyvärr återfinns inte den permanenta länken i någon av de föreslagna referensformaten under ”Skapa referens”.

Alla plattformarna erbjuder alltså någon slags beständig identifierare (oftast kallad något i stil med ”permanent länk”) som åtminstone i teorin inte kommer att ändras. Utbudet skiljer sig något åt, liksom var i data-posterna identifierarna kan hittas och vad de kallas.

Beständiga identifierare – Nordiska museet

Nordiska museets cirka 700 000 databasposter är mestadels föremål och fotografier, vilka registreras i varsin modul i samlingsförvaltningssystemet Primus. Inventarienumren följer av tradition varsin serie: NM för föremålssamlingarna, NMA för Nordiska museets arkiv. Inventarienummer tilldelas numera vid registrering i Primus och varje nummer ska vara unikt



	<i>Identifierare</i>	<i>Publikt tillgänglig URL</i>
Nordiska museet inventarienummer	NMA.0044859 (registreras i Primus)	
DigitaltMuseum ID (KulturIT)	021018510873 (visas i adressfält)	https://digitaltmuseum.org/021018510873/
DigitaltMuseum UUID (KulturIT)	DFAA55E7-D1DF-4804-8385-BC13A56BEBC4	https://digitaltmuseum.org/DFAA55E7-D1DF-4804-8385-BC13A56BEBC4
DMS mediaserver ID (KulturIT)	03349TDWtQKZ (visas endast i kod)	https://mm.dimu.org/image/03349TDWtQKZ?dimension=max
K-samsök/Kringla ID (Riksantikvarieämbetet)	http://kulturarvsdata.se/nomu/photo/NMA0044859	https://www.kringla.nu/kringla/objekt?referens=nomu/photo/NMA0044859 http://kulturarvsdata.se/nomu/photo/NMA0044859

Fig. 3. Beständiga identifierare för ett objekt ur Nordiska museets samlingar: "Ritningar. Förslag till nybyggnad för Nordiska museet" av Evert Strömqvist 1887. Notera att det är ett fotografi av ritningen som ligger till grund för databasposten, snarare än själva ritningen.

i sin serie. De objekt som publiceras på DigitaltMuseum tilldelas därutöver dels en 10- eller 12-siffrig identifierare ("dimukod", även synlig i webbläsarens adressfält), dels en universell identifierare (UUID) på 128 bitar. Varje publicerad digital bildfil har därtill en egen identifierare på 12 tecken (versaler, gemener och siffror) som endast syns i kod. När infor-

mationen från DigitaltMuseum skördas till K-samsök följer uppdelningen i föremåls- och fotografiska samlingar med; de benämns ”nomu/object” respektive ”nomu/photo”.

För att hänvisa till en samlingspost behöver alltså någon av de beständiga identifierarna (DigitaltMuseum ID/UUID eller K-samsök ID) användas i hänvisningen (fig. 3).

Beständiga identifierare har i första hand utvecklats utifrån systembehov: att göra det möjligt för maskiner att söka fram och visa upp digitala objekt. Framväxten av digitala arbetssätt för forskare samt de ökade kraven på att tillgängliggöra forskningsdata och digitala källor medför dock att det finns ett påtagligt behov att tillgängliggöra dessa identifierare på ett tydligt, för människor förståeligt sätt via sökplattformarna. Utan detta blir det svårt för forskarna själva att nå upp till FAIR-principerna för sin forskningsdata. En annan viktig aspekt är att data från olika källor ska kunna kombineras samt att det ska gå att avgöra om en term är identisk med en annan term eller ej; med andra ord, om data har ett mått av interoperabilitet.

Interoperabilitet (I)

Interoperabilitet i FAIR-sammanhang handlar i hög grad om att använda väletablerade metadatastandarder, samt att använda kontrollerade vokabulärer. ”Metadatastandarder” innebär att metadata publiceras i maskinläsbar form. På språknivå betyder detta att XML eller Javascript ska användas (några av de format som pekas ut i ursprungliga FAIR-definitionen är RDF, OWL och JSON-LD).⁴⁰ Mer specialiserade metadataformat/datamodeller (som kan uttryckas i de ovan nämnda formaten) är t ex Europeana Data Model eller CIDOC-CRM.⁴¹

Posterna i K-samsök är tillgängliga i standarderna RDF och OWL och den egna datamodellen K-samsök-RDF version 1.1.⁴² Olika plattformar använder olika standarder; Libris kontrollerade vokabulärer publiceras som JSON-LD, RDF och Turtle och använder Kungliga bibliotekets egen datamodell KBV baserad på Bibframe.⁴³

Wikidata som sambandscentral

Wikidata innehåller åtminstone ett hundratal olika egenskaper för att länka wikidataposter till externa Sverigerelaterade databaser.⁴⁴ Den mest enskilt använda egenskapen, för att identifiera motsvarande post i K-samsök, används i skrivande stund i 171 568 wikidataposter. Den höga siffran för K-samsök beror främst på samarbeten med wikimediarörelsen där objekt skapats eller länkats till fornminnes-ID för tävlingen Wiki

Loves Monuments. Kungliga Bibliotekets databaser förekommer i drygt 130 000 wikidataposter (främst Libris-URI men även SELIBR-ID). De tre egenskaperna knutna till Nationell ArkivDatabas är dock sparsamt använda, mindre än 5000 poster sammanlagt.⁴⁵

I ett FoU-projekt mellan Nationalmuseum, Statens historiska museer och Wikimedia Sverige, under ledning av Karin Glasemann, kontrollerades och rättades personposterna i respektive samlingsförvaltningssystem, för att därefter länkas mot poster i Wikidata.⁴⁶ Projektet kunde på detta sätt identifiera drygt 1500 personer som förekom i båda museimyndigheternas samlingar. Projektet gör Nationalmuseums samlingar i högre grad interoperabla med samlingarna från Historiska museet, Ekonomiska museet, Hallwylska museet, Skoklosters slott, Livrustkammaren och Tumba bruksmuseum.⁴⁷ Interoperabiliteten med Wikidata medför dessutom att dessa museers samlingsposter kan berikas med mycket mer riklig information från personernas liv, vilket i sin tur kan användas till att skapa nya sätt att studera och visualisera kulturarvet.⁴⁸

Kontrollerade vokabulärer – Nordiska museet

Posterna i Primus och DigitaltMuseum använder ett flertal kontrollerade vokabulärer, bland annat ämnesord, material och teknik. Därtill kan poster relateras till personer och platser. Respektive vokabulär hanteras i Primus. Presentationen på DigitaltMuseum sker dock oftast i form av fritext, vilket omöjliggör den typ av semantiska sökningar vi tidigare nämnt (till exempel att särskilja landet Danmark från orten Danmark i Uppland).

Fler och fler vokabulärer har sedan 2011 flyttat till KulturNav, KulturITs egna plattform för att hantera vokabulärer som länkade öppna data. Att en vokabulär finns tillgänglig i KulturNav innebär dock inte att den används av samtliga museer som använder Primus; det är nämligen upp till respektive museum att koppla ihop sina befintliga, lokala vokabulärer med de gemensamma i KulturNav.⁴⁹ Detta sker manuellt och kräver rejäla datastädningssatsningar vilket inneburit att de flesta listor på många museer (inklusive Nordiska museet) fortfarande endast är lokala. DigitaltMuseum i sin tur kan bara presentera de KulturNav-kopplade termerna som semantisk information, vilket förklarar varför så många vokabulärer presenteras som fritext. Övergången från lokala till gemensamma vokabulärer har kommit längst för personposter, där DigitaltMuseum både länkar och presenterar KulturNavs auktoriteter. Det framgår dock inte på DigitaltMuseum att personposten kommer från KulturNav, vilket kan komplicera för den som vill hänvisa på rätt sätt.⁵⁰ Till exempel kan det konstateras att åtminstone två andra museer i skrivande stund använder

separata, ej sammanlänkade personposter för Nordiska museets grundare Artur Hazelius.⁵¹ Golub et al. (2022) konstaterar att det inte går att använda kontrollerade vokabulärer för att ämnessöka i svenska museisamlingar.⁵² Vid första anblick stämmer detta. Men det som framstår som fritextfält för slutanvändaren är i själva verket ofta kontrollerade vokabulärer på institutionsnivå.

Svenska socknar – ett exempel på interoperabel kulturarvsdata

Ett talande exempel på både bristerna i tillgängliga vokabulärer för svenskt kulturarv, och potentialen de skulle kunna ha, är enheten ”socken”. De svenska socknarna är en grundpelare både för kulturhistoriska samlingar och för forskningen. För Riksantikvarieämbetet är socknarna en grundläggande enhet i Antikvariska-topografiska arkivet (ATA), och de kom också att användas som utgångspunkt i myndighetens inventeringar av kulturhistoriska lämningar under 1900-talet. Även många andra kulturarvsinstitutioner (till exempel Nordiska museet) har använt socknarna vid registrering av bruksningsort, givarens ort och så vidare.

Att åtskilliga socknar har exakt samma namn är inget problem i ett analogt arkiv där hyllorna sorteras först utifrån landskap och därefter socken. I samband med den första administrativa digitaliseringen så krävdes dock något som gjorde att systemet kunde skilja dem åt, varvid sockennummer skapades. Dessa enkla fyrsiffriga koder fungerar i ett slutet system, men är långt ifrån de beständiga identifierare som krävs för FAIR data. Sifferkoden refererar dessutom bara till ett *namn*, det finns ingen exakt definition av vilken geografiska utsträckning och kronologisk period som menas, trots att gränserna i många fall har ändrats i samband med olika historiska skeenden.

Än mer komplext blev det när Fornminnesregistret skulle digitaliseras. Under 1900-talet hade vissa socknar införlivats med stadsförsamlingar och ”upphört”. När det analoga Fornminnesregistret ersattes av Fornminnesinformationssystemet (FMIS)⁵³ i början av 2000-talet så uppstod en diskrepans mellan det digitala kartlager för sockengränserna som skapades utifrån de analoga kartorna från Lantmäteriet och den digitaliserade arkivinformatjonen som skapades utifrån inventeringsböckerna.⁵⁴ I böckerna hade ibland, men inte alltid, de gamla sockennamnen behållits vid ominventeringarna på 1970-talet och framåt, även om området formellt uppgått i en stad vid det laget. Detta för att slippa ändra namnen på ett stort antal fornlämningar som i det gamla systemet bestod av socken+löpnummer. Bristen på samordning mellan de som digitaliserade arkivet och de som digitaliserade kartlager fick konsekvenser för webb-sökgränssnittet Fornsök som utvecklades för FMIS. Det gick till exempel

inte att filtrera ut lämningar inom ”Hyllie socken” i Fornsök, trots att ett stort antal lämningar hade kvar det namnet, eftersom de tekniskt sett numera låg inom Malmö socken.

Ett problem för digital humaniora är att ATA:s socknar inte är publicerade som öppen länkad data. Inte heller de digitala sockenbeteckningarna i Fornsök finns tillgängliga som auktoriteter. K-samsök har publicerat en RDF-fil⁵⁵ i vilken ingår samtliga socknar relaterade till ATA-sockennummer, men det går inte att använda detta som auktoriteter för enskilda socknar i forskningsdata. Riksarkivet å andra sidan har faktiskt auktoritetsposter för topografiska beteckningar, som till och med inkluderar information om vilken tidsperiod som avses. Riksarkivets socknar är som sagt inte alltid definierade på samma sätt som i ATA eller i Kulturmiljöregistret (Fornsök), eller för den delen i Lantmäteriets egna digitala sockenlager som skapades på 2000-talet. Sammantaget leder detta till att den som samlar in kulturarvsdata inom en socken får data från *olika* geografiska områden om man kombinerar information från ATA, Kulturmiljöregistret, Riksarkivet respektive Lantmäteriet. Med rent analoga metoder var detta inte så stort problem, eftersom urvalet skedde manuellt, men med datadrivna analyser så finns det potentiella felkällor som lätt missas. Det räcker inte med kompetens inom historiska källmaterial och kompetens inom dataanalys – det krävs dessutom kompetens om det digitala källmaterialets beskaffenhet.

Komplexiteten med digital data från olika källor är inte nödvändigtvis ett problem. Rätt använd tillhandahåller digitaliseringen en stor potential för humanistisk forskning att skraddarsy analyser utifrån rumsliga, tidsmässiga och administrativa variabler, med hjälp av auktoriteter och geodata. Det ställer dock höga krav på användarnas förståelse av det digitala källmaterialet: hur digitala enheter är beskaffade, hur de skapades, när och av vilken anledning. Det finns en överhängande risk att användning av digitala hjälpmedel och resurser som utgår från en analog förståelse resulterar i missförstånd och i värsta fall felaktigheter.

Det är av största vikt att datadrivna analyser inom digital historia tydligt redovisar vilka digitala källor som använts för att resultatet ska bli reproducerbart och återanvändbart. Med andra ord, forskare måste vara minst lika noggranna med att redovisa vilka digitala källor de använt för datauttag och som underlag för datadrivna analyser, som de är att med att redovisa analoga källor. Detta blir extra viktigt vid kvantitativa analyser där större mängder kulturarvsdata hämtas eller analyseras. Eftersom nytt material tillkommer behöver nästa forskare kunna ta del av data-mängden vid det tillfälle analysen gjordes.

Vissa plattformar (till exempel Wikidata) erbjuder versionshanterade länkar, men i de flesta fall behöver forskarna fortfarande själva ordna en

permanent länk till just den version av en datapost som de använt. Förutom att själva spara ned data som webbarkiv kan forskare underlätta för läsare och granskare genom att spara sidor i Internet Archives Wayback Machine.⁵⁶ Forskare vid svenska universitet kan dessutom, efter bearbetning, spara och dela forskningsdata genom Svensk nationell datatjänst.⁵⁷ Men för att kunna publicera sina resultat och redovisa källor så behöver användarna få tydlig och begriplig information om de villkor som gäller vid återanvändning och bearbetning av materialet.

Återanvändbarhet (R)

För att tredje part (till exempel forskare) ska kunna återanvända data behövs information om villkor för detta, ofta kallad licensinformation. Med utgångspunkt i svensk upphovsrättslagstiftning kan respektive institution bedöma vilka kulturarvsdata (till exempel fotografier, konstverk och texter) vars upphovsrätt löpt ut. För upphovsrättskyddat material är det upp till personen eller organisationen som skapar materialet (eller den som förvärvat den ekonomiska upphovsrätten) att välja hur pass öppet det ska vara. I Digisams vägledande principer för digitalt kulturarv anges att ”Det digitaliserade materialet ska tillgängliggöras så snart som möjligt och på ett sätt som möjliggör en så bred användning som möjligt”.⁵⁸

Det dominerande systemet för licensiering av kulturarvsdata är Creative Commons. Licenserna CC0, CC BY och CC BY-SA är att betrakta som fria (”Approved for Free Cultural Works”) enligt Creative Commons, tillsammans med material markerat som ”in the public domain” (PDM).⁵⁹

Riksantikvarieämbetet har tagit ett policybeslut att myndighetens egenhändigt producerade material (publikationer, nya fotografier etc) ska licensieras som CC BY (fri användning men med attribution). Textuell metadata och geodata i Kulturmiljö- och Bebyggelseregistret är helt fri att spara, bearbeta och återpublicera (CC0), även om attribution självklart uppskattas.⁶⁰ All metadata som publiceras i K-samsök är märkt med CC0 och därmed återanvändbart utan några krav på attribution. Detta gäller alltså den *beskrivande* informationen för kulturarvsobjekten. Eventuella avbildningar och mediefiler kan ha en annan form av licens. Här kan det finnas skilda bedömningar hos olika minnesinstitutioner. Det beror inte nödvändigtvis på en ovilja att materialet kommer till användning, även om det ibland uttrycks oro för att ”tappa kontrollen”. Det är helt enkelt inte lätt att veta vilka rättsliga regler som gäller för olika sorters digitala reproduktioner av äldre material. Ett fotografi i samlingarna kan till exempel ha olika sorters upphovsrättsliga förbehåll beroende på vilket år det togs, i vilket sammanhang (porträtt?), av vem (amatör eller professionell?) och när fotografen ifråga avled. Inte sällan saknas många av dessa

uppgifter i samlingskatalogen, vilket medför att det krävs individuella bedömningar som tar mycket resurser i anspråk.⁶¹ Det kan också råda oklarheter och uppstå missförstånd om huruvida en digital avbildning av till exempel en tavla, en ritning eller ett föremål kan sägas ha eget upphovsrättsligt skydd eller ej. Med andra ord krävs både god juridisk kompetens och god kunskap om kulturarvsobjektets historik för att kunna avgöra villkoren för återanvändning. Detta är ett område där minnesinstitutionerna och forskarna genom att samarbeta och ta del av varandras kompetens kan komma avsevärt mycket längre än genom att arbeta på varsitt håll.

Oavsett dessa utmaningar så har det under det senaste decenniet skett en mycket positiv utveckling för att göra digitaliserat kulturarv återanvändbart. 2014 var mindre än en tredjedel (31 procent) av mediafilerna som indexerades av K-samsök märkta med en fri licens och 15 procent saknade helt rättighetsmärkning. Från och med 2020 har samtliga mediefiler som kan sökas fram i Kringla någon sorts rättighetsmärkning (i annat fall har de avindexerats) och två tredjedelar (66 procent) har licenser som tillåter mycket fri återanvändning.⁶²

Betyder denna utveckling mot alltmer fria och öppna licenser att minnesinstitutionernas omfattande och resurskrävande arbete med digitalisering och tillgängliggörande kommer att osynliggöras? Inte alls. Bara för att en minnesinstitution inte alltid kan ställa villkor på användningen av det digitaliserade kulturarvet, så kan naturligtvis sådana krav uppstå i andra sammanhang. Forskare måste precis som tidigare kunna uppge källa för de arkivhandlingar, bilder och källmaterial de använt sig av när de publicerar sig vetenskapligt. Om de dessutom gör datadrivna analyser på strukturerad metadata från kulturarvssamlingar så behöver de redovisa källa för datauttaget, även om det tekniskt sett har licensen CCo. Det är alltså en skillnad på vad som krävs *i lag*, och vad som krävs enligt god *vetenskaplig praxis*. Den oro som en del företrädare för minnesinstitutioner uttrycker över att tillgängliggöra digitalt kulturarv är därför delvis överdriven – åtminstone i relation till akademiker och medborgarforskare. Användarna själva har ett intresse i att kunna ange var de hittat informationen och att kunna peka direkt på källan.

Slutligen bör det nämnas att en ändring av upphovsrättslagstiftningen är på väg. Preliminärt kommer kulturarvsinstitutionerna att få något större spelrum kring digitalt tillgängliggörande. Regeringen föreslog i juli 2022 bland annat att ”Ensamrätten för framställare av fotografiska bilder ska inte omfatta fotografier vars huvudsakliga motiv utgörs av ett konstverk för vilket upphovsrätt inte längre gäller”.⁶³

När det gäller återanvändbarhet kan vi identifiera några övergripande trender bland svenska kulturarvsinstitutioner under 2010-talet. Fler institutioner väljer att inte hävda ny upphovsrätt till digitala avbildningar

av verk där upphovsrätten löpt ut, särskilt tvådimensionella verk. Fler institutioner väljer dessutom att licensiera fotografier ur den egna verksamheten med CC-licenser. Från att mestadels ha använt den restriktiva licensen CC BY-NC-ND (som endast tillåter icke kommersiell användning och förbjuder bearbetning) använder institutionerna numera ofta licenserna CC BY (återanvändning tillåten med attribution) eller CC BY-SA (återanvändning tillåten med attribution; bearbetningar måste licensieras med samma villkor). Färre institutioner väljer att begränsa bildupplösningen i publicerat material. Sammantaget innebär dessa förändringar en betydligt högre grad av återanvändbarhet.

Slutdiskussion

Vad är viktigt för forskare att ta hänsyn till vid användning av digitala kulturarvsdata, och vad kan institutionerna göra för att underlätta för framtida forskning? Här summerar vi våra viktigaste slutsatser.

Det behövs ett nationellt helhetsgrepp kring behov av vokabulärer för begrepp och terminologier inom kulturhistorisk forskning. Riksantikvarieämbetet, Riksarkivet och Kungliga biblioteket behöver ta ledande roller för att förverkliga detta eftersom de har i uppdrag att stödja museer, arkiv och bibliotek i den digitala transformationen, så att de kan bidra till kunskapstillväxten.⁶⁴ Myndigheterna måste genomföra detta arbete i nära samarbete med varandra och i öppen dialog med minnesinstitutionerna som ofta har liknande behov. Framförallt behövs riktade insatser för att utveckla kontrollerade vokabulärer för historiskt signifikanta termer, till exempel geografiska och administrativa enheter.

För att institutionerna ska kunna förbättra och länka samman digitala kulturarvsdata behövs särskild finansiering, eftersom det digitala kulturarvet inte är en ersättning för det analoga som fortfarande måste omhändertas. Det räcker inte att finansiera aggregatorer och gemensamma plattformar – resurserna behövs även hos de institutioner som ansvarar för databasernas innehåll. Utan god kvalitet på data så uppfyller plattformarna inte sin funktion oavsett hur bra utvecklade de är och för att uppnå denna kvalitet krävs omfattande arbetsinsatser av personer med god kunskap om kulturhistoria. Med andra ord krävs fördjupade och långsiktiga samarbeten mellan akademien och minnesinstitutionerna, för att säkerställa att det digitaliserade kulturarvet blir sökbart och interoperabelt. Inte bara forskare utan även studenter och medborgarforskare är resurser som grovt underutnyttjas i dagsläget. Långsiktigt samarbete av detta slag kräver resurser för att etableras och drivas, men är helt avgörande för att massdigitaliseringen ska bli mer än en rent kvantitativ process där gamla felaktigheter och brister reproduceras.

Kraven på forskare att tillgängliggöra den data som de baserar sina analyser på ökar hela tiden, men det krävs att plattformarna hjälper dem på vägen. Söktjänsterna behöver göra det enklare och tydligare att hitta beständiga identifierare, så att det blir självklart för forskare hur de ska hänvisa till digital kulturarvsdata och objekt i arkiv och samlingar. På så vis blir det i förlängningen enklare för andra att hitta och bygga vidare på den forskning som genomförts. För att minnesinstitutioner, aggregatorer och forskningsinfrastrukturer ska kunna hitta och länka till forskning bedriven på arkiv och samlingar så krävs att den data som analyserna bygger på:

- Beskrivs med riklig, standardiserad metadata och får beständig identifierare (Findable)
- Bevaras på betrodda repositorer, även om åtkomsten kan vara begränsad (Accessible)
- Innehåller tillräcklig med information, helst med hjälp av kontrollerade vokabulärer, så att det blir begripligt och jämförbart (Interoperable), samt
- Har tydligt angivna villkor för vidare användning (Reusable)

Sist men inte minst: finansiärer och forskningsledare behöver ta höjd för att projekt som kräver digitaliserat kulturarv ofta behöver resurskrävande insatser för att kvalitetssäkra och kvalitetshöja informationen i samlingsförvaltningssystemen. Projekten behöver planeras i nära samarbete med minnesinstitutionerna i ett tidigt skede. Detta arbete kan i vissa fall likställas med grundforskning: det är inte lika spektakulärt som de resultat som utlovas i slutändan, men det är helt avgörande för att resultaten ska bli trovärdiga och användbara.

Noter

1. Ian Milligan: *The transformation of historical research in the digital age* (Cambridge, 2022), 29.
2. Joacim Hansson: ”How to catch an antelope. On authenticity in digitised cultural heritage documents” i Joacim Hansson & Jonas Svensson: *Doing digital humanities. Concepts, approaches, cases* (Växjö, 2020), 110.
3. Se diskussion i Henrik Summanen: *Kulturarvets digitalisering* (Stockholm, 2021).
4. Några exempel: Pelle Snickars: *Kulturarvets mediehistoria. Dokumentation och representation 1750–1950* (Lund, 2020); Koraljka Golub, Pawel Michal Ziolkowski & Goran Zlodi: ”Organizing subject access to cultural heritage in Swedish online museums” i *Journal of Documentation* 78:7 (2022), doi:10.1108/JD-05-2021-0094; Milligan: *The transformation of historical research in the digital age*.
5. Henrik Summanen: ”Debatt: Kulturarvets digitalisering” i *Fornvännen* 116:4 (2021).

6. GO FAIR: *FAIR Principles*, <https://www.go-fair.org/fair-principles/> [hämtad 2021-11-05]; Mark D. Wilkinson m.fl.: "The FAIR guiding principles for scientific data management and stewardship" i *Scientific Data* 3:1 (2016), doi:10.1038/sdata.2016.18.
7. Siffran framtagen genom att välja "show all properties" följt av "filter by data-type: external-id" i Hay's tools: *Wikidata Propbrowse*, <https://hay.toolforge.org/propbrowse> [hämtad 2022-04-13].
8. Riksantikvarieämbetet: *Bidrag till en nationell strategi för digitalisering, elektronisk tillgång och digitalt bevarande, samt förslag avseende framtiden för K-samsök. Redovisning av regeringsuppdrag* (Stockholm, 2010), <http://urn.kb.se/resolve?urn=urn:nbn:se:raa:di-va-3887>.
9. Marcus Smith: "Linked open data and aggregation infrastructure in the cultural heritage sector: A case study of SOCH, a linked data aggregator for Swedish open cultural heritage" i Koraljka Golub & Ying-Hsang Liu: *Information and knowledge organisation in digital humanities* (London, 2021).
10. *Runor*, <https://app.raa.se/open/runor/search> [hämtad 2022-04-17].
11. Institutionen för nordiska språk, Uppsala universitet: *Runinskrift Sö 179 i Samnordisk runtextdatabas* 2020, <http://kulturarvsdata.se/uu/srdb/4013365a-38f1-4ec2-9f39-2c897d5f914a> [hämtad 2022-08-27].
12. Riksantikvarieämbetet: *Att leverera data*, <https://www.raa.se/hitta-information/k-samsok/du-som-vill-leverera/att-leverera-data/> [hämtad 2022-04-14].
13. Catherine Smith: "Controlled vocabularies. Past, present and future of subject access" i *Cataloging & Classification Quarterly* 59:2-3 (2021), doi:10.1080/01639374.2021.1881007.
14. Riksarkivet: *Nationell ArkivDatabas – Sök orter*, <https://sok.riksarkivet.se/topografier> [hämtad 2022-04-17]; Kungliga biblioteket: *id.kb.se*, <https://id.kb.se/> [hämtad 2022-04-17]; Sara Kayser & Ulf Bodin: "KulturNav. A collaborative approach to the management of shared vocabularies" i CIDOC Conference: *University Heritage* (Genève, 2021).
15. *Getty Vocabularies (Getty Research Institute)*, <https://www.getty.edu/research/tools/vocabularies/index.html> [hämtad 2022-04-17]; *nomisma.org*, <http://nomisma.org/> [hämtad 2022-04-17]; *About GeoNames*, <https://www.geonames.org/about.html> [hämtad 2022-04-17].
16. Statens historiska museers periodlista finns maskinläsbar på adressen <https://mis.historiska.se/rdf/period/period.xml>. Samma periodlista finns tillgänglig i PeriodO på adressen <http://n2t.net/ark:/99152/povn2fr>.
17. *Hazelius-Berg, Gunnel (1905–1997) – KulturNav*, <https://kulturnav.org/87b2efce0e07d-4fcd-bd86-8412a971c927> [hämtad 2022-04-17]; *Hazelius-Berg, Gunnel Gunnarsdotter (1905–) – KulturNav*, <https://kulturnav.org/54ada828-8a09-4a56-af5c-bf2af4b09123> [hämtad 2022-04-17].
18. *Medeltid – KulturNav*, <https://kulturnav.org/0453a745-3ff0-49b9-b69e-e8bd25eb635f> [hämtad 2022-04-17]; *Medeltid (1100 e. Kr–1500 e. Kr) – KulturNav*, <https://kulturnav.org/8186a38a-1c8d-4d43-8c60-5ebfbdad5c4> [hämtad 2022-04-17].
19. Digisam: *Vägledning till ett sammanlänkat kulturarv*, RAÄ-2021-1139 (Stockholm, 2021), <http://urn.kb.se/resolve?urn=urn:nbn:se:raa:diva-6391> [hämtad 2022-08-27].
20. Vägledning för webbutveckling: *Webbdirektivet – översikt*, <https://webbriktlinjer.se/lagkrav/webbdirektivet/> [hämtad 2022-04-17].
21. Malin Lennartsson et al: "Arkiven och historikerna", *Historisk Tidskrift* 140:3 (2020).
22. *Ibid.*, 544.

23. Tillhandahållen av Maria Carlsson och Marcus Smith, Riksantikvarieämbetet.
24. Se Tabell 1 i Aron Ambrosiani & Åsa M. Larsson: *Hur FAIR är svensk digitaliserad kulturarvsdata idag Appendix Tabeller*, <https://zenodo.org/record/7140171>
25. Observera att antalet varierar beroende på om man räknar enskilda museer för sig (t.ex Sjöhistoriska museet) eller museiorganisation (Statens maritima och transporthistoriska museer).
26. Riksarkivet: *Sökning: *-NAD*, [https://sok.riksarkivet.se/arkivinstitutioner?InfoSida=&Sokord=*](https://sok.riksarkivet.se/arkivinstitutioner?InfoSida=&Sokord=) [hämtad 2022-09-12]
27. *Arkivinformation från Riksarkivet – nu som öppen data via API:er*, <https://sok.riksarkivet.se/nyhetsarkiv?item=117740&sida=nad> [hämtad 2022-04-18]
28. E-post från Katarina Ahlberg vid Riksarkivet, 2022-09-09.
29. *Alvin*, <https://www.alvin-portal.org/alvin/resultList.jsf?faces-redirect=true&includeViewParams=true&query=&searchType=EXTENDED&dswid=-2622> [hämtad 2022-08-27].
30. Se Tabell 2 i Aron Ambrosiani & Åsa M. Larsson: *Hur FAIR är svensk digitaliserad kulturarvsdata idag Appendix Tabeller*, <https://zenodo.org/record/7140171>
31. Daniel Löwenborg, Maria Jonsson, Åsa Larsson & Johan Nordinge: "A turn towards the digital. An overview of Swedish heritage information management today" i *Internet Archaeology* 58 (2021), doi:10.11141/ia.58.19.
32. Koraljka Golub et al.: "Organizing subject access to cultural heritage in Swedish online museums".
33. Nordiska museet: Årsredovisning. Stiftelsen Nordiska museet 2021 (Stockholm, 2022), https://www.nordiskamuseet.se/sites/default/files/public/arsredovisning_2021-final.pdf, 33.
34. Se Tabell 3 i Aron Ambrosiani & Åsa M. Larsson: *Hur FAIR är svensk digitaliserad kulturarvsdata idag Appendix Tabeller*, <https://zenodo.org/record/7140171>
35. Nordiska museet: Årsredovisning. Stiftelsen Nordiska museet 2021, 36, 56.
36. Se Tabell 4 i Aron Ambrosiani & Åsa M. Larsson: *Hur FAIR är svensk digitaliserad kulturarvsdata idag Appendix Tabeller*, <https://zenodo.org/record/7140171>
37. *Nordiska museets digitala arkiv*, <https://dokument.nordiskamuseet.se/> [hämtad 2022-08-28].
38. Nordiska museet: Årsredovisning. *Stiftelsen Nordiska museet* 2021, 36.
39. Hans-Werner Hilde & Jochen Kothe: *Implementing persistent identifiers. Overview of concepts, guidelines and recommendations* (London, 2006).
40. GO FAIR: *FAIR Principles*.
41. Lukas Koster & Saskia Woutersen-Windhouwer: "FAIR Principles for Library, archive and museum Collections: A proposal for standards for reusable collections" i *The Code4Lib Journal* 40 (2018).
42. *Protokoll och parametrar* (2015), <https://www.raa.se/hitta-information/k-samsok/att-anvanda-k-samsok/protokoll-och-parametrar/> [hämtad 2022-03-21].
43. GitHub: *Data Model · libris/definitions Wiki*, <https://github.com/libris/definitions> [hämtad 2022-04-18].
44. *Template: Sweden properties – Wikidata*, https://www.wikidata.org/wiki/Template:Sweden_properties [hämtad 2022-04-15].
45. Se Tabell 5 i Aron Ambrosiani & Åsa M. Larsson: *Hur FAIR är svensk digitaliserad kulturarvsdata idag Appendix Tabeller*, <https://zenodo.org/record/7140171>
46. *Användbara auktoriteter för datadriven samlingsforskning*, diarienummer RAÄ-2020-1762.

47. Linnéa Lundin: *Användbara auktoriteter för datadriven samlingsforskning* | SHM (2021), <https://shm.se/blog-article/anvandbara-auktoriteter-for-datadriven-samlingsforskning/> [hämtad 2022-04-13].

48. *Projekt:Wikidata för auktoritetskontroll 2021/Översikt – Wikimedia*, https://se.wikimedia.org/wiki/Projekt:Wikidata_f%C3%B6r_auktoritetskontroll_2021/%C3%96versikt [hämtad 2022-04-18]; Albin Larsson: *Visualiseringar för datadriven samlingsforskning* (2021), <https://byabbe.se/datadriven-samlingsforskning/> [hämtad 2022-01-10].

49. Sara Kayser & Ulf Bodin: "KulturNav. A collaborative approach to the management of shared vocabularies".

50. Notera att KulturNav-UUID 2de1784e-b458-4752-af85-624d3e0e2614 inte är synligt på denna sida: *Hazelius, Artur (1833-1901)*, <https://digitaltmuseum.org/021036011730/hazelius-artur-1833-1901> [hämtad 2022-04-18].

51. *Sök: name:"9A0F3191-02D5-4D1B-A708-C88A6B368EF9"*, <https://digitaltmuseum.se/search/?aq=name%3A%229A0F3191-02D5-4D1B-A708-C88A6B368EF9%22> [hämtad 2022-04-18]; *Sök: name:"CDC5A5B6-402D-48B3-B9F8-C7E96FFEDC1"*, <https://digitaltmuseum.se/search/?aq=name%3A%22CDC5A5B6-402D-48B3-B9F8-C7E96FFEDC1%22> [hämtad 2022-04-18].

52. Golub et al., "Organizing subject access to cultural heritage in Swedish online museums", 231-232.

53. Malin Blomqvist: *Informationssystemet för fornminnen. Lista med lämningstyper och antikvarisk praxis* (Stockholm, 2004).

54. Muntlig uppgift, Johan Andersson, Riksantikvarieämbetet.

55. Riksantikvarieämbetet: *Socknar som OWL*, <https://kulturarvsdata.se/resurser/aukt/geo/parish/parish.owl> [hämtad 2022-04-18].

56. *Wayback Machine*, <https://web.archive.org/> [hämtad 2022-08-27].

57. *Dela data på ett FAIR sätt | Svensk Nationell Datatjänst*, <https://snd.gu.se/sv/beskriv-och-dela-data/dela-data-pa-ett-fair-satt> [hämtad 2022-08-27]; *Förbered data för tillgängliggörande | Svensk Nationell Datatjänst*, <https://snd.gu.se/sv/forbered-data-tillgangliggorande> [hämtad 2022-08-27].

58. *Vägledande principer – Använda – Digisam*, <https://www.digisam.se/vagledande-principer-for-arbetet-med-digitalt-kulturarv/vagledande-principer-anvanda/> [hämtad 2022-04-18].

59. Creative Commons: *Understanding Free Cultural Works*, <https://creativecommons.org/share-your-work/public-domain/freeworks/> [hämtad 2022-04-18].

60. *Licenser och upphovsrätt*, <https://www.raa.se/om-webbplatsen/licenser/> [uppdaterad 2021-12-01]; *Öppna data från Kulturmiljöregistret*, <https://www.raa.se/hitta-information/oppna-data-fran-kulturmiljoregistret/> [uppdaterad 2022-03-18]

61. Digisam: *Vägledning om upphovsrätt och angränsande lagstiftning Version 3.1.* (Stockholm, 2018)

62. Uppgifter via Maria Carlsson och Marcus Smith, Riksantikvarieämbetet. Se även Tabell 5 i Aron Ambrosiani & Åsa M. Larsson: *Hur FAIR är svensk digitaliserad kulturarvsdata idag Appendix Tabeller*, <https://zenodo.org/record/7140171>

63. Regeringen och Regeringskansliet: *Upphovsrätten på den digitala inre marknaden*, 2022, <https://www.regeringen.se/rattsliga-dokument/proposition/2022/07/prop-202122278/> [hämtad 2022-08-27].

64. Riksdagsförvaltningen: *Förordning (2014:1585) med instruktion för Riksantikvarieämbetet Svensk författningssamling 2014:2014:1585 t.o.m. SFS 2020:870 – Riksdagen*, <https://>

www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/forordning-20141585-med-instruktion-for_sfs-2014-1585 [hämtad 2022-08-27]; Riksdagsförvaltningen: *Förordning (2009:1593) med instruktion för Riksarkivet Svensk författningssamling 2009:2009:1593 t.o.m. SFS 2020:869 – Riksdagen*, https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/forordning-20091593-med-instruktion-for_sfs-2009-1593 [hämtad 2022-08-27]; Riksdagsförvaltningen: *Förordning (2008:1421) med instruktion för Kungl. biblioteket Svensk författningssamling 2008:2008:1421 t.o.m. SFS 2022:1311 – Riksdagen*, https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/forordning-20081421-med-instruktion-for-kungl_sfs-2008-1421 [hämtad 2022-08-27].

Tack

Tack till Maria ”Mia” Carlsson och Marcus Smith för hjälp med information om K-samsök, Johan Andersson för frågor rörande digitaliseringen av Fornminnesregistret samt Maria Jonsson och Johan Nordinge för allt möjligt rörande ATA – och mycket mycket mer. Det är en ynnest att ha tillgång till så mångkunniga medarbetare på Riksantikvarieämbetet.