



**D**igisam

# Vägledning till ett sammenlänkat kulturarv



# Vägledning till ett sammanlänkat kulturarv

## Innehåll

Vägledning till ett sammanlänkat kulturarv.....	1
Inledning .....	2
Inför uppstarten.....	4
Länkade data som metod .....	6
Strategier för att komma igång.....	9
Praktiska exempel på identifierare och beskrivningar.....	13
Vidare läsning.....	16

Digisam  
Riksantikvarieämbetet  
Box 1114, 621 22 Visby  
Tel 08-5191 80 00  
[www.digisam.se](http://www.digisam.se)  
[digisam@raa.se](mailto:digisam@raa.se)

Vägledning till ett sammanlänkat kulturarv

Diarienummer RAÄ-2021-1139

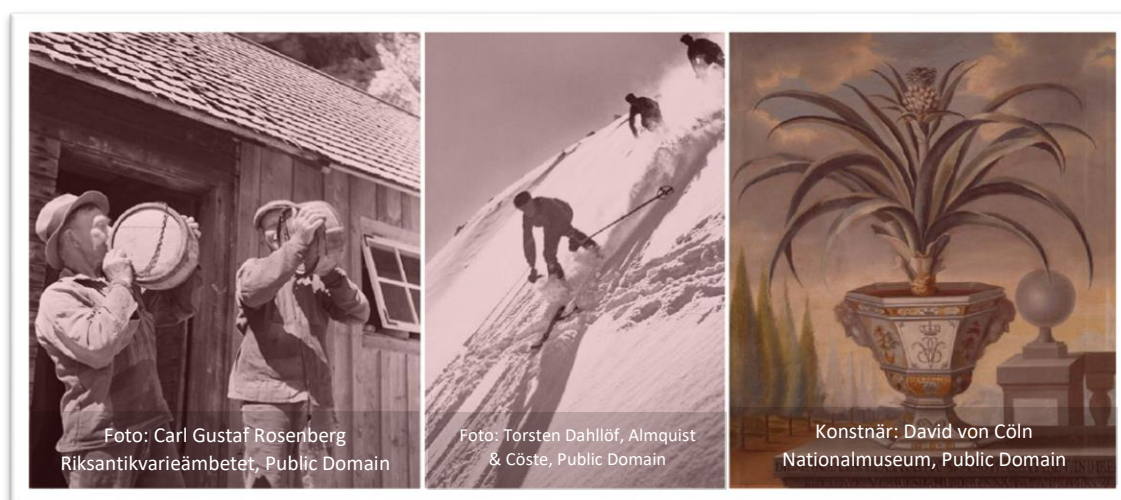
Text: CC0

Bilder och illustrationer: Creative Commons licens CC BY där inget annat anges

Villkor på <http://creativecommons.org/licenses/by/4.0/deed.sv>

# Inledning

Arkiv, bibliotek, museer och andra kulturarvsinstitutioner har tillsammans en enorm skattkammare med information av både stor bredd och stort djup. Denna information kan användas på många sätt inom exempelvis forskning, utbildning och näringsliv såväl som av allmänheten. Finns informationen i digital form når den långt fler, oavsett användarnas geografiska tillhörighet och tid på dygnet, och kan tillämpas inom områden såsom digital tjänsteutveckling och storskaliga dataanalyser.



Digitalt tillgängliggörande via internet behöver ske på ett sätt som möjliggör utbyte av data, harmonisering och sammanlänkningar mellan olika datamängder. När Tim Berners-Lee, skaparen av World Wide Web, såg hur internet utvecklades lanserade han idén om länkade data.<sup>1</sup> Han såg behovet av att inte bara skapa länkar mellan webbsidor utan också skapa länkar på datanivå.<sup>2</sup> När alla publicerar sina data online, gör dem maskinläsbara<sup>3</sup> och beskriver dem på ett standardiserat sätt, samt länkar dem till andra data kan internet i sig användas som en databas.

Vad betyder det i praktiken? Jo, att det blir möjligt att bygga digitala tjänster som baseras på data från en mängd olika aktörer, exempelvis kulturarvsinstitutioner. Det innebär också att användarna inte behöver leta runt på olika kulturarvsinstitutioners webbplatser för att hitta information om en specifik företeelse. Istället kan de söka fritt i och ställa frågor till den totala mängden publicerade data utan att behöva veta vad som förvaltas hos vilken institution. Då blir det till exempel möjligt att skapa specifika sökingångar för personer, platser eller

<sup>1</sup> W3C, *Linked Data*, <https://www.w3.org/standards/semanticweb/data>, hämtad 14 april 2021.

<sup>2</sup> Data är representation av fakta, instruktioner, idéer eller liknande i en form lämpad för överföring, tolkning eller bearbetning av människor eller maskiner. Termerna information och data betraktas i dagligt tal som synonymer men har olika betydelser. Data innebär fakta eller idéer i sina mindre beståndsdelar, medan information hänvisar till mer komplexa koncept som ofta består av flertalet data. Data blir information när någon har tolkat innebörden av dem.

<sup>3</sup> Maskinläsbara data är strukturerade data som är skapade i syfte att läsas och bearbetas av en maskin.

tidsperioder som hittar information från våra olika datakällor. Dessutom kan användarna få information presenterad för sig som de inte visste fanns men som relaterar till den företeelse de inledningsvis sökte efter. I förlängningen innebär det att kulturarvsinstitutionernas information kan få större räckvidd och ökad synlighet, samt bidra till en ännu bättre helhetsbild av vårt gemensamma kulturarv genom att kopplas till hela det nätverk av information och data som internet består av.

Enligt *Digisams vägledande principer för arbetet med digitalt kulturarv*<sup>4</sup> ska länknings mellan egen och andras information eftersträvas. Denna vägledning är en vidareutveckling som ger dig som arbetar för en kulturarvsinstitution, till exempel som samlingsförvaltare, databaskoordinator, utvecklare eller i någon annan roll där förståelse för hur informationen kan hanteras och tillgängliggöras har stor betydelse, en introduktion till länkade data som metod och ett antal pragmatiska strategier för att realisera metoden. Olika organisationer har olika förutsättningar och ambitionsnivåer. Huvudsaken är att vi alla väljer någon strategi och kommer igång!

## Öppna data och länkade öppna data

Öppna data är data som vem som helst kan använda fritt, i princip till vad som helst utan hinder i form av kostnader eller licenser.<sup>5</sup> Länkade data tillför länkar på datanivå i ett format kallat Resource Description Framework (RDF).<sup>6</sup> När länkade data tillgängliggörs öppet benämns det länkade öppna data (LOD/Linked Open Data).

Denna vägledning berör länkade öppna data. För mer information om hur du kan arbeta med öppna data, se Digisams rekommendation om öppna data<sup>7</sup> eller de stöd och vägledningar kring öppna och delade data som Myndigheten för digital förvaltning (DIGG) har på sin webbplats.<sup>8</sup>

---

<sup>4</sup> Digisam, *Vägledande principer - Använda*, <http://www.digisam.se/vagledande-principer-for-arbetet-med-digitalt-kulturarv/vagledande-principer-anvanda/>, hämtad 14 april 2021.

<sup>5</sup> Internetstiftelsen, *Öppna data*, <https://internetstiftelsen.se/kunskap/for-samhallet/oppna-data/>, hämtad 14 april 2021.

<sup>6</sup> Vitbok länkade öppna data, *Introduktion till länkade data*, <https://lankadedata.se/vitbok/introduktion.html>, hämtad 14 april 2021.

<sup>7</sup> Digisam, *Rekommendation för öppna data*, [http://www.digisam.se/wp-content/uploads/2015/06/Rekommendation\\_for\\_oppna\\_data.pdf](http://www.digisam.se/wp-content/uploads/2015/06/Rekommendation_for_oppna_data.pdf), hämtad 14 april 2021.

<sup>8</sup> Myndigheten för digital förvaltning (DIGG), *Öppna och delade data*, <https://www.digg.se/utveckling-av-digital-forvaltning/oppna-och-delade-data/>, hämtad 14 april 2021.

# Inför uppstarten

Inför uppstarten av ett arbete med länkade data är det bra att ringa in din organisations förutsättningar. Nedan följer några viktiga aspekter som har bäring på arbetet och resultatet.

## Styrning och resurser

Inför ett arbete med länkade data krävs förmodligen kompetensutveckling på området såväl som tekniska lösningar som gör det möjligt att realisera metoden. För att komma igång behövs stöd från ledningen, och allra helst behöver arbetet med länkade data bli en självklar del av verksamhetens digitaliseringsstrategi.

Guiden *Linked Open Data – A Quick Start Guide For Decision Makers*<sup>9</sup> riktar sig till beslutsfattare och beskriver hur du kan forma en strategi för länkade öppna data, vilka investeringar ett arbete med länkade öppna data kräver och vilka fördelar det ger, samt en rad goda exempel från omvärlden.



## Samordning av metadata

Utgångspunkten för länkade data är att olika informationsmängder och enskilda dataposter ska vara möjliga att koppla samman via internet. En “typ” från en institution ska gå att koppla ihop med en “typ” från en annan institution. Till exempel kan metadata<sup>10</sup> om ett vikingasvärd hos ett museum kopplas ihop med metadata om ett vikingasvärd hos ett annat museum. Länkade data kan på detta sätt fungera som en förlängning av den egna institutionens metadata.

Metadata hos olika institutioner ser olika ut. För att möjliggöra länkade data behöver metadata om samma eller liknande objekt och företeelser samordnas. En sådan samordning rör vilka standarder, ontologier, vokabulärer och auktoritetsposter som exponeras på internet. Poängen med detta är att åstadkomma semantisk interoperabilitet,<sup>11</sup> det vill säga att olika

<sup>9</sup> REEEP, *Linked Open Data: The Essentials – A Quick Start Guide for Decision Makers*, <https://www.reeep.org/LOD-the-Essentials.pdf>, hämtad 14 april 2021.

<sup>10</sup> Metadata är strukturerad information som beskriver och identifierar annan data eller information för att underlätta förståelse, sökning och användning.

<sup>11</sup> Wikipedia, *Semantisk interoperabilitet*, [https://sv.wikipedia.org/wiki/Semantisk\\_interoperabilitet](https://sv.wikipedia.org/wiki/Semantisk_interoperabilitet), hämtad 14 april 2021.

datorsystem kan utväxla information med varandra på ett sätt där det mottagande systemet kan tolka informationens innebörd och producera användbara resultat för slutanvändaren.

Att konsekvent och utifrån ett helhetsgrepp arbeta med metadata som en del av det digitala tillgängliggörandet gör att din institution kan komma framåt snabbare. Viktigt att säga i sammanhanget är dock att arbetet med länkade data är möjligt, till och med önskvärt, att påbörja utan att alla metadata finns på plats eller håller högsta möjliga kvalitet från början.

## Rättighetsmärkning, licenser och känsliga uppgifter

Tydligt uppmärkt information som klargör för användaren hur informationen får användas är en förutsättning för ökad användning av den. Ha i åtanke att så länge de använda märkningarna eller licenserna<sup>12</sup> innehåller någon typ av begränsning innebär det att den möjliga återanvändningen av informationen är begränsad. Använd därför så kallade öppna märkningar i så lång utsträckning som möjligt.

Enligt *Digisams vägledande principer för arbetet med digitalt kulturarv*<sup>13</sup> kan metadata i allmänhet licensieras helt öppet, det vill säga med CC0, även om de refererar till skyddade digitala objekt.<sup>14</sup> I *Digisams vägledning om upphovsrätt och angränsande lagstiftning*<sup>15</sup> kan du lära mer om hur du kan bedöma och märka upp information. Se även rekommendationerna från Myndigheten för digital förvaltning (DIGG).<sup>16</sup>

Ett angränsande område är de etiska frågorna. Arkivmaterial kan innehålla känsliga personuppgifter och språkbruk eller formuleringar som idag anses vara kränkande. Det kan också finnas uppgifter som rent juridiskt är oproblematiska men där nu levande släktingar har synpunkter på att uppgifterna görs tillgängliga. För att komma igång med länkade data kan det därför vara klokt att börja med datamängder som är oproblematiska och inte innehåller begränsningar.

---

<sup>12</sup> Se t.ex. 1) Creative Commons, *Om licenserna*, <https://creativecommons.org/licenses/?lang=sv>, hämtad 14 april 2021, eller 2) Europeana, *Available rights statements*, <https://pro.europeana.eu/page/available-rights-statements>, hämtad 14 april 2021.

<sup>13</sup> Digisam, *Vägledande principer för arbetet med digitalt kulturarv*, <https://www.digisam.se/vagledande-principer-for-arbetet-med-digitalt-kulturarv/>, hämtad 14 april 2021.

<sup>14</sup> Ett digitalt objekt är en tydligt avgränsad enhet med digital information.

<sup>15</sup> Digisam, *Vägledning om upphovsrätt och angränsande lagstiftning*, [http://www.digisam.se/wp-content/uploads/2020/07/Vagledning\\_om\\_upphovsratt\\_3.1.pdf](http://www.digisam.se/wp-content/uploads/2020/07/Vagledning_om_upphovsratt_3.1.pdf), hämtad 14 april 2021.

<sup>16</sup> Myndigheten för digital förvaltning (DIGG), *Rekommendation om öppna licenser och immaterialrätt*, <https://www.digg.se/utveckling-av-digital-forvaltning/oppna-och-delade-data/offentliga-aktorer/oppna-licenser-och-immaterialratt>, hämtad 14 april 2021.

# Länkade data som metod

I detta avsnitt presenteras de moment som länkade data som metod bygger på. I nästa avsnitt går vi närmare in på hur sammanlänkningen sker i praktiken.

## Säkerställ att informationen i sin grundform kan förstås och användas av andra

För att exempelvis en databas eller ett register ska bli en resurs på internet som kan avläsas maskinellt behöver den vara begriplig. Till exempel bör strukturerad och ostrukturerad information (såsom fritextfält) hållas separat och fältnamn och förkortningar uttryckas på sätt som kan förstås även utanför institutionens gränser (se termlistor nedan). Information som är underförstådd behöver också göras explicit. Till exempel behöver en lista över författare som endast innehåller namn kompletteras med information om att personerna avser författare.



Fotograf: Peter Häll, Tekniska museet (CC BY)

## Publicera dataposterna som webblänkar

Vid ett arbete med länkade data publiceras dataposter som webblänkar där länkarna pekar ut resurser. En länk som uttrycker en företeelse, ett begrepp eller en term på internet kallas för URI,<sup>17</sup> Uniform Resource Identifier (se även IRI<sup>18</sup>), eftersom länken inte enbart är en webbadress utan även en identifierare för en viss sak. I och med att identifierarna är webbadresser kan information om en resurs (exempelvis en agent, ett objekt, en händelse eller ett koncept) levereras med ett vanligt webbanrop. Det gör det också möjligt för den som känner till identifieraren att länka till och göra påståenden om denna resurs.

Möjligheten att skapa länkar mellan objekt från flera institutioner påverkas av mängden exponerade identifierare, det vill säga antalet URI:er. Därför är det viktigt att institutionerna

<sup>17</sup> Wikipedia, *URI*, [https://sv.wikipedia.org/wiki/Uniform\\_Resource\\_Identifier](https://sv.wikipedia.org/wiki/Uniform_Resource_Identifier), hämtad 14 april 2021.

<sup>18</sup> The Internet Engineering Task Force (IETF), *Internationalized Resource identifiers (IRIs)*, <https://tools.ietf.org/html/rfc3987>, hämtad 14 april 2021.

ser till att antalet exponerade identifierare ökar. En annan viktig aspekt av identifierare är deras grad av persistens, det vill säga under hur lång tid framåt en identifierare "pekar" på samma sak och kan leverera rätt information.

## **Beskriv relationer mellan objekt, namn och företeelser**

Formatet för länkade data, RDF, beskriver relationer mellan olika ting. Det består av tre delar, en så kallad tripplett: subjekt, predikat och objekt. Ett namn på en plats kan till exempel uttryckas som att "en plats" (subjekt) "har ett namn" (predikat) "Stockholm" (objekt). Andra poster kan tala om att samma Stockholm har status som huvudstad, att staden ligger i Sverige, att en specifik person är född i denna stad, att en viss bok är utgiven där och så vidare.

Det är alltså viktigt att vara överens om hur saker och relationer ska namnges. Däremot behöver olika aktörer inte vara överens om påståendena. I ett visst sammanhang är Stockholm inte alls någon huvudstad; det finns till exempel flera andra orter med samma namn. Det kan även finnas olika uppfattningar gällande om staden ska räknas till Uppland eller Södermanland, som kan underbyggas med länkar till olika källor och andra påståenden. Idén är att vara explicit och basera påståenden på URI:er som definierar det som avses. En textsträng som i en databas talar om att ett föremål är kopplat till "Göteborg" bör alltså istället eller dessutom uttryckas med en URI som anger en datapost för denna plats. Denna datapost innehåller (förhoppningsvis) ytterligare metadata, till exempel geografiska koordinater och andra relationer. Genom att inkludera URI:er i din institutions datamängder berikas datamängderna. Dessutom är det ett sätt att tala om för omvärlden att ni tillhandahåller mer information som kan kopplas till samma plats.

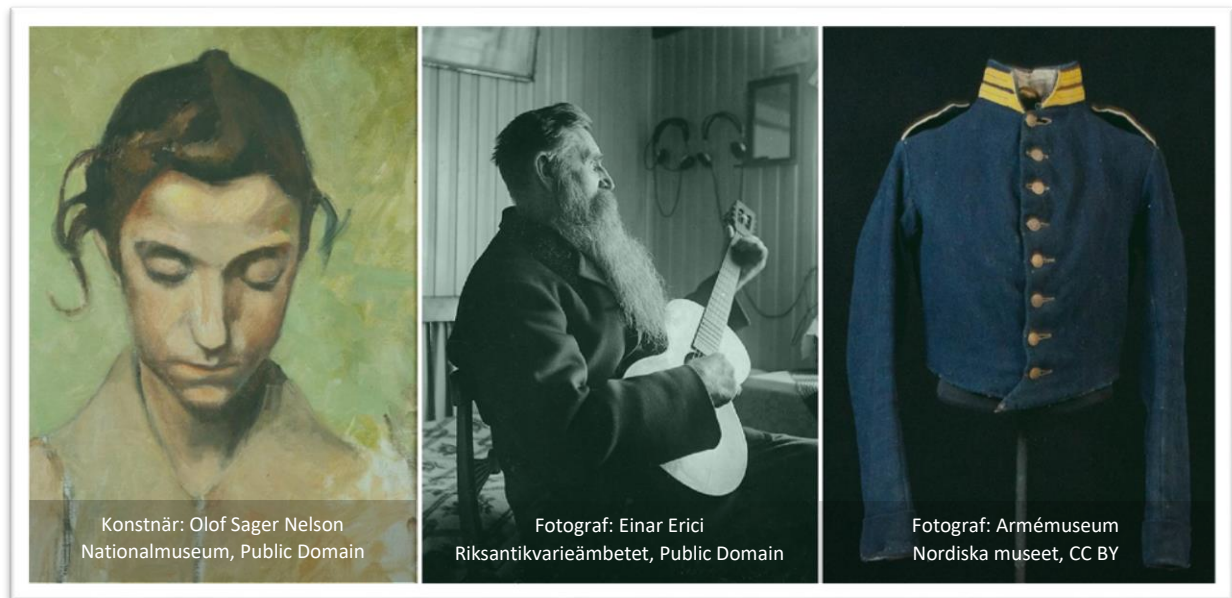
## **Använd termlistor för att uttrycka termer och begrepp**

Användande av ontologier och vokabulärer/terminologier (termlistor) är viktigt i sammanhanget för att främja den semantiska interoperabiliteten. I de fall standardiserade ontologier och vokabulärer inte har använts redan från början behöver informationen rättas efter sådana när den tillgängliggörs online. Både metadatabeskrivningarna och termlistorna bör då publiceras. Ett annat sätt att göra detta på, som oftast ökar interoperabiliteten, är att länka befintliga metadatabeskrivningar till existerande termlistor på internet. Bruket av termlistor avser här främst grundläggande information om exempelvis årtal eller tid, namn, upphovsperson, källa eller referenser. Även metadata om exempelvis bilder kan beskrivas med termlistor som specificerar saker såsom typ av bild och upplösning.

Information som av olika skäl inte är publicerad på internet, till exempel på grund av upphovsrättskäl eller integritetskäl, bör även den beskrivas med hjälp av termlistor på internet. Det gör det möjligt för exempelvis forskare, som kan få tillgång till materialet även om det inte är fritt, att hitta det.

## Skapa relationer till andra datamängder

Databaser har traditionellt sett skapats utifrån ett tankesätt som bygger på att databaserna antingen har eller saknar det som användarna efterfrågar, och svarar därmed oftast på sökningar utifrån ett sådant antagande. Men inom den semantiska webben,<sup>19</sup> det vill säga metoder och tekniker för att möjliggöra för maskiner att förstå innebörden i informationen på webben, består tankesättet i att det användarna efterfrågar kan finnas någonstans i den semantiska webbens distribuerade databas även om en enskild databas saknar det som efterfrågas.



När flera aktörer bygger kopplingar utifrån existerande data hänvisas användarna vidare till resurser som ger mer information om samma sak. Den som publicerar länkade data bör därför ta som vana att skapa relationer till andra befintliga datamängder. När information länkas samman med extern information kan den dessutom upptäckas av fler och eventuellt dubbelarbete minskas. Behovet av egenpåhittade definitioner och strängar eller fritextfält, som kan innehålla fel och vara svåra för maskiner att läsa, minskar också. Det blir till exempel möjligt att hänvisa till en befintlig datapost om en specifik källa istället för att själv hålla information om var, när och hur källan är skapad.

Antalet länkar ska dock inte överdrivas. Det är lämpligt att hänvisa till och återanvända det som är relevant för den egna datamängden och det egna sammanhanget.

<sup>19</sup> Wikipedia, *Semantiska webben*, [https://sv.wikipedia.org/wiki/Semantiska\\_webben](https://sv.wikipedia.org/wiki/Semantiska_webben), hämtad 14 april 2021.

# Strategier för att komma igång

Eftersom kulturarvsinstitutionernas förutsättningar och ambitionsnivåer ser olika ut finns det inget enda “rätt sätt” att arbeta med länkade data på. Vissa institutioner har stora utvecklingsavdelningar och sköter all utveckling internt medan andra använder sig av externa leverantörer för exempelvis sina webbplatser och samlingssystem. Inom vissa sektorer finns en aggregator med ett utpekat nationellt ansvar för aggregering och publicering av kulturarvsinstitutionernas information, inom andra saknas detta. Detta innebär att det krävs flera separata strategier som tillsammans kan användas för att uppnå ett mer sammanlänkat kulturarv. Den gemensamma ansatsen bör dock bestå i att öka både mängden maskinläsbara beskrivningar och mängden länkar inom och mellan organisationer.

Länkade data kräver alltså hittills två saker: en någorlunda stabil identifierare i form av en URI och en maskinläsbar beskrivning av det som identifieras. De kvarvarande frågorna är var och hur data ska exponeras. Detta kan göras på ett flertal olika sätt. Gemensamt för de alternativ som inte innefattar att data skickas till en extern aggregator, men i praktiken även de, är att ändringar kan behöva göras på den egna webbplatsen eller att en separat webbplats för länkade data sätts upp. Därför är en av de viktigaste åtgärderna att se till att din organisation har kontroll över er webbnärvaro oavsett om det gäller framsidan (“skyltfönstret”) eller samlingarna. Detta kan ske antingen genom egen utvecklarkompetens eller genom en leverantör. Detta gör det möjligt att, över tid, förvalta de exponerade datamängderna vilket är avgörande för att länkarna ska bli så persistenta som möjligt.

Så länge organisationer har någon sorts kontroll över sin webbnärvaro eller möjlighet att sätta upp en separat lösning som är åtskild från webbplatsen är tröskeln låg för att komma igång med länkade data. Oavsett vilken strategi som väljs (se nedan) behöver beskrivningarna vara maskinläsbara. En beskrivning behöver dock inte vara uttömmande. Att enbart exponera exempelvis resursens typ och en etikett kan framstå som magert, men gör det samtidigt möjligt för maskiner att upptäcka vilka resurser din organisation har och för människor att länka till dem vilket är en stor vinst. Tillsammans skapar dessa identifierare, beskrivningar och länkar en graf som gör det möjligt att hoppa mellan olika delar av kulturarvet utan att behöva aggregera allt i förväg.



## Strategi 1: Publicering integrerad på den egna webbplatsen

Finns en befintlig webbplats med sidor som beskriver exempelvis föremål eller personer är det möjligt att infoga metadata i den HTML-kod som gör sidan maskinläsbar. Det finns fyra huvudsakliga sätt att arbeta utifrån denna strategi:

1. Använd standarden RDFa<sup>20</sup> för att göra en redan befintlig vy mer semantisk genom att "strössla" semantik över den. Exempel på detta finns på [id.kb.se](https://id.kb.se/).<sup>21</sup>
2. Inkludera JSON-LD<sup>22</sup> som en separat beskrivning i sidan. Exempel på detta finns hos Google Search Central.<sup>23</sup>
3. Använd HTTP headers<sup>24</sup> för att leverera en maskinläsbar beskrivning när en klient frågar efter det explicit, så kallad *content negotiation*.<sup>25</sup> Detta är en avancerad metod som kräver kontroll över servermiljön.
4. Länka till och bädda in information från en extern beskrivning som någon annan aktör är värd för, exempelvis Europeana,<sup>26</sup> Archives Portal Europe<sup>27</sup> och K-Samsök.<sup>28</sup>

## Strategi 2: Publicering på en separat domän

Ett annat sätt att publicera data på är att skapa en separat domän för identifierare där metoden beskriven ovan integreras. Denna kan exempelvis ta formen **id.domän.se**. Fördelen är att inga ändringar behöver ske i din befintliga webbplats. Istället byggs något nytt upp vid sidan om vilket ger möjlighet till full kontroll över ditt skapande av en URI. En annan vinst är att det blir möjligt att byta underliggande system utan att det behöver anpassas till gamla URI:er. Se även nästa avsnitt *Praktiska exempel på identifierare och beskrivningar*.

Denna strategi kan kombineras med strategi 1 genom att de två sajterna länkar till varandra, det vill säga att beskrivningen av objektet länkar till sidan om objektet och vice versa.

Exempel på detta tillvägagångssätt finns hos Library of Congress<sup>29</sup> och Kungliga biblioteket.<sup>30</sup>

---

<sup>20</sup> W3C, *RDFa 1.1 Primer – Third Edition*, <https://www.w3.org/TR/rdfa-primer/>, hämtad 14 april 2021.

<sup>21</sup> Kungliga biblioteket, *Allmänt ämnesord - Fysik*, <https://id.kb.se/term/sao/Fysik>, hämtad 14 april 2021.

<sup>22</sup> W3C, *JSON-LD 1.1*, <https://www.w3.org/TR/json-ld11/>, hämtad 14 april 2021.

<sup>23</sup> Google Search Central, *Understand how structured data works*, <https://developers.google.com/search/docs/guides/intro-structured-data>, hämtad 14 april 2021.

<sup>24</sup> Wikipedia, *Header (datorterm)*, [https://sv.wikipedia.org/wiki/Header\\_\(datorterm\)](https://sv.wikipedia.org/wiki/Header_(datorterm)), hämtad 14 april 2021.

<sup>25</sup> Wikipedia, *Content negotiation*, [https://en.wikipedia.org/wiki/Content\\_negotiation](https://en.wikipedia.org/wiki/Content_negotiation), hämtad 14 april 2021.

<sup>26</sup> Europeana, <https://www.europeana.eu/sv>, hämtad 14 april 2021.

<sup>27</sup> Archives Portal Europe, <https://www.archivesportaleurope.net/sv>, hämtad 14 april 2021.

<sup>28</sup> K-Samsök, <https://www.raa.se/hitta-information/k-samsok/>, hämtad 14 april 2021.

<sup>29</sup> Library of Congress, *ID.LOC.GOV – Linked Data Service*, <https://id.loc.gov/>, hämtad 14 april 2021.

<sup>30</sup> Kungliga biblioteket, *id.kb.se*, <https://id.kb.se/>, hämtad 14 april 2021.

### Strategi 3: Publicering via extern leverantör

Strategin följer samma upplägg som ovan med tillägget att det är viktigt att ha en exitstrategi som gör det möjligt att byta leverantör utan att behöva byta alla URI:er. I praktiken betyder det att en URI måste skapas på ett sätt som gör att organisationen kan ha kontroll över den. En URI enligt formen **kulturarvsinstitution.leverantör.org** är vanskelig eftersom den alltid kommer att vara knuten till leverantören. En bättre lösning är att de data som leverantören tillhandahåller publiceras via organisationens egen domän, exempelvis **id.kulturarvsinstitution.se**. Länkarna till de specifika objekten "ägs" och förvaltas därmed av institutionen.

Exempel på detta tillvägagångssätt finns hos Riksarkivet som tillhandahåller en LOD-tjänst om historiska platser.<sup>31</sup>

### Strategi 4: Publicering via domängemensam aggregator eller central infrastruktur

Vissa domäner har en utpekad aggregator till vilken metadata skickas för publicering och spridning, alternativt så används en central infrastruktur av alla medlemmar. De tydligaste exemplen i Sverige är K-samsök<sup>32</sup> som driftas av Riksantikvarieämbetet, Libris<sup>33</sup> som driftas av Kungliga biblioteket och Nationella arkivdatabasen (NAD)<sup>34</sup> som driftas av Riksarkivet.

K-samsök fungerar som en kopplingsdosa mellan institutionernas databaser och de aktörer som vill använda informationen i egna tillämpningar, till exempel på webbplatser eller i mobila lösningar. Institutionernas data skördas upp till K-samsök och sparas ner i ett index. Via ett öppet API<sup>35</sup> kan utvecklare bygga e-tjänster mot olika målgrupper. K-samsök är också nationell aggregator för leverans av data till Europeana.<sup>36</sup>

Libris är en nationell infrastruktur som syftar till att effektivisera informationshanteringen för Sveriges bibliotek. Libris innehåller information om titlar, personer, ämnesord, etc. som finns eller används på svenska bibliotek. Libris har stöd för en rad olika sätt att komma åt och använda Libris data.<sup>37</sup>

NAD innehåller information om ca 230 000 arkivbestånd från en mängd arkivförvarande institutioner. Metadata om de förtecknade arkiven publiceras även i Archives Portal Europe

---

<sup>31</sup> Riksarkivet, *Tora*, <https://data.riksarkivet.se/tora>, hämtad 14 april 2021.

<sup>32</sup> Riksantikvarieämbetet, *K-Samsök*, <https://www.raa.se/hitta-information/k-samsok/>, hämtad 14 april 2021.

<sup>33</sup> Kungliga biblioteket, *Libris*, <http://libris.kb.se/>, hämtad 15 april 2021.

<sup>34</sup> Riksarkivet, *Nationell arkivdatabas*, <http://sok.riksarkivet.se/nad>, hämtad 15 april 2021.

<sup>35</sup> Wikipedia, *API*, <https://en.wikipedia.org/wiki/API>, hämtad 15 april 2021.

<sup>36</sup> Europeana, <https://www.europeana.eu/en>, hämtad 15 april 2021.

<sup>37</sup> Kungliga biblioteket, *Libris - Teknik & format*, [http://librishelp.libris.kb.se/help/tech\\_swe.jsp?open=tech](http://librishelp.libris.kb.se/help/tech_swe.jsp?open=tech), hämtad 15 april 2021.

och därifrån vidare till Europeana.

## Strategi 5: Publicering via extern kooperativ aggregator

Ett exempel på en extern kooperativ aggregator som används på global nivå är Wikidata<sup>38</sup> som tillhör Wikimedia Foundations. Wikidata lagrar strukturerade data huvudsakligen från de artiklar som finns på Wikipedia. Wikidata är således ett komplement till Wikipedia och Wikimedia Commons som lagrar multimediala filer. Wikidata syftar till att utgöra en fri flerspråkig kunskapsdatabas som ska kunna läsas, redigeras, återanvändas och tolkas av såväl människor som datorprogram. Data lagras därför i en databas på ett strukturerat sätt i form av semantiska relationer. Exempelvis kan användare ställa en komplex sökfråga med hjälp av sökspråket SPARQL<sup>39</sup> i sökverktyget Wikidata Query.<sup>40</sup>

Notera att användning av Wikidata som plattform medför att ändringar kan göras av fler än den som laddat upp data. Detta kan vara en stor fördel, exempelvis genom att andra kan skapa länkar till objekten, men institutionen tappar samtidigt den totala kontrollen över sin egen information på just den plattformen. Därför bör inte institutionens egna tjänster bygga enbart på Wikidata utan plattformen ska ses som ett komplement, och strategin kombineras med andra strategier. Wikibase,<sup>41</sup> som består av en serie programvaror som används av Wikidata, kan när den används som teknisk basplatta för en egen instans fungera tillsammans med andra nämnda strategier.

---

<sup>38</sup> Wikidata, *Welcome to Wikidata*, [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page), hämtad 15 april 2021.

<sup>39</sup> Wikipedia, *SPARQL*, <https://en.wikipedia.org/wiki/SPARQL>, hämtad 15 april 2021.

<sup>40</sup> Wikidata, *Wikidata Query Service*, <https://query.wikidata.org/>, hämtad 15 april 2021.

<sup>41</sup> Wikibase, <https://wikiba.se/>, hämtad 15 april 2021.

# Praktiska exempel på identifierare och beskrivningar

Nedan följer praktiska exempel på hur identifierare kan göras mer stabila och beskrivningar utformas. För mer information om identifierare, se även *Digisams checklista för beständiga identifierare*.<sup>42</sup>

## Persistenta identifierare

Det finns inget perfekt tekniskt recept för hur en URI kan göras “persistent” eftersom ingen identifierare kommer att hålla för alltid. Detta beror på att det i grund och botten handlar om vilken process som har kopplats till förvaltningen av identifieraren. En specifik teknik kommer aldrig att göra en identifierare persistent så länge den inte också backas upp av en medveten förvaltning. Med detta sagt finns det ett antal tumregler som kan hjälpa till att göra identifieraren så stabil som möjligt. För att exemplifiera kommer en URI nedan att i flera steg transformeras till en både kortare och mer stabil URI.

### Exempel-URI:

`https://collection.example.org/MV/museumViewservice=ExtInt&module=artist&objectId=4128&shoesize=42`

### Tumregel 1: Inkludera inte tjänstespecifika detaljer

I exemplet ovan är strängarna *MV*, *museumView*, och parametern *service* interna angelägenheter för systemet som inte har med resursen att göra. Dessa strängar kommer, utöver att de gör denna URI väldigt lång, göra det svårare att byta system utan att alla som har skapat ett bokmärke till en viss resurs får en bruten länk. Tar vi bort strängarna har vi kvar:

`https://collection.example.org/?module=artist&objectId=4128&shoesize=42`

### Tumregel 2: Inkludera för resursen relevant information

Information som är inneboende i resursen, exempelvis id-nummer och typ av resurs, passar bättre som en del av en URI istället för som parametrar. En URI ska peka ut resursen, inte tjänsten som visar upp resursen. I det här fallet är värdet i parametern *module* resursens typ (“artist”) och *objectId* är ett löpnummer som har skapats av systemet och identifierar en person.

`https://collection.example.org/artist/4128?shoesize=42`

---

<sup>42</sup> Digisam, *Checklista: Beständiga identifierare*, [http://www.digisam.se/wp-content/uploads/2015/06/Checklista\\_bestandiga\\_identifierare.pdf](http://www.digisam.se/wp-content/uploads/2015/06/Checklista_bestandiga_identifierare.pdf), hämtad 15 april 2021.

### **Tumregel 3 - Var återhållsam med metadata**

Vissa metadata, såsom id-nummer och typ, kan fungera bra i en URI, speciellt om det finns risk för krock. Men med tanke på att nästan alla metadata kan komma att ändras är det viktigt att vara återhållsam med att inkludera dem. I exemplet ovan är också *shoesize* snarare ett attribut hos resursen än något som behöver finnas i en URI. Den informationen kan med fördel finnas i beskrivningen istället.

<https://collections.example.org/artist/4128>

### **Tumregel 4 - Var precis, men inte för detaljerad**

Strängen *artist* indikerar att denna URI pekar ut en konstnär, men individer kan ha flera roller och vi vill inte skapa flera URI:er för samma person i dess olika roller. Konstnär kan i det här fallet med fördel generaliseras till *agent*, en term som ofta används för personer och organisationer. Det kan också vara värt att fundera på om *collection* är nödvändig information att ha med. Tar vi bort dessa får vi följande resultat:

<https://example.org/agent/4128>

Exempel-URI:n är nu mycket kortare och mer begriplig utan att någon information har gått förlorad. Men är den persistent? Det beror på. Den är onekligen mer stabil i och med att allt tjänstespecifikt är borta och skulle därför kunna överleva en flytt till ett nytt system. Om den är persistent kommer dock bero på om organisationen som står bakom den kan upprätthålla den över tid och om systemet kan garantera att id-numret 4128 inte raderas, återanvänds eller ersätts med något annat vid exempelvis en omindexering.

Kom dock ihåg att det går utmärkt att publicera länkade data även om URI:erna inte är lika rena som den ovan. Det är svårare att upprätthålla "stökiga" länkar, men det går.

## **Maskinläsbara metadata**

Metadata är aldrig perfekta och en informationsmodell blir aldrig klar. Det är därför viktigt att arbetet med länkade data drivs utifrån ett bredden-först-perspektiv där fokus läggs på att publicera data snarare än riskminimering där enbart "perfekta" data släpps, åtminstone till en början. Risker att ingenting blir publicerat på grund av för höga krav är långt större än risken att något som publiceras visar sig vara oanvändbart. Det är därför lämpligt att börja med grovkorniga typer såsom agent (personer, institutioner och dyl.), objekt, händelse, koncept och platser och efter det gå vidare med mer detaljerade beskrivningar.

Nedan ett förslag på prioritering.

#### **Nivå 1:**

- Typ (objekt, agent, händelse, koncept, geografi / topografi)
- Titel / namn

## Nivå 2:

- Interna länkar till relaterade resurser - exempelvis mellan agenter och objekt
- Utåtgående länkar
- Mer finkornig beskrivning

## Nivå 3:

- Än mer finkornig beskrivning

Det finns en uppsjö termlistor/vokabulärer som kan användas för att skapa beskrivningar. Många domäner har en eller flera utpekade och det kan vara svårt att välja. I linje med bredden-först-strategin ovan kan det finnas en poäng med att först utvärdera exempelvis Schema.org,<sup>43</sup> en vokabulär framtagen av Google tillsammans med experter från flera domäner. Det är inte en perfekt vokabulär, men den har de grundläggande typerna och har fördelen att sökmotorer lättare förstår beskrivningen jämfört med mer komplicerade vokabulärer såsom BIBFRAME.<sup>44</sup> En annan vanligt förekommande vokabulär inom länkade data är Dublin Core<sup>45</sup> som i likhet med Schema.org är generell och uttrycker många vanliga objekt och relationer.

Det är dock viktigt att påpeka att det inte är problematiskt att börja med Schema.org eller Dublin Core och sedan successivt göra beskrivningen mer finkornig med hjälp av en annan mer domänanpassad vokabulär. Det är viktigare att börja än att det blir perfekt från början.

---

<sup>43</sup> Schema.org, <https://schema.org/>, hämtad 15 april 2021.

<sup>44</sup> Library of Congress, *Bibliographic Framework Initiative*, <https://www.loc.gov/bibframe/>, hämtad 15 april 2021.

<sup>45</sup> Dublin Core Metadata Initiative, *DCMI Metadata Terms*, <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>, hämtad 15 april 2021.

## Vidare läsning

Bauer, Florian och Kaltenböck, Martin. *Linked Open Data: The Essentials - A Quick Start Guide for Decision Makers*. REEEP & Semantic Web Company.

<https://www.reeep.org/LOD-the-Essentials.pdf> (Hämtad 2021-04-15).

Digisam. *Checklista: Beständiga identifierare*. [https://digisam.se/wp-content/uploads/2015/06/Checklista\\_bestandiga\\_identifierare.pdf](https://digisam.se/wp-content/uploads/2015/06/Checklista_bestandiga_identifierare.pdf) (Hämtad 2021-04-15).

Digisam. *Rekommendation för öppna data*. [https://digisam.se/wp-content/uploads/2015/06/Rekommendation\\_for\\_oppna\\_data.pdf](https://digisam.se/wp-content/uploads/2015/06/Rekommendation_for_oppna_data.pdf) (Hämtad 2021-04-15).

Digisam. *Vägledande principer för arbetet med digitalt kulturarv*. [https://digisam.se/images/docs/rapporter/Vagledande\\_principer\\_for\\_arbetet\\_med\\_digitalt\\_kulturarv.pdf](https://digisam.se/images/docs/rapporter/Vagledande_principer_for_arbetet_med_digitalt_kulturarv.pdf) (Hämtad 2021-04-15).

Myndigheten för digital förvaltning (DIGG). *Öppna och delade data från offentliga aktörer*. <https://www.digg.se/utveckling-av-digital-forvaltning/oppna-och-delade-data/offentliga-aktorer> (Hämtad 2021-04-15).

Palmér, Matthias och Ebner, Hannes. *Vitbok länkade öppna data*. MetaSolutions AB. <https://lankadedata.se/vitbok/print/> (Hämtad 2021-04-15).