

Funktioner hos metadata

Pass 3: Metadata

BAS Online 2021-01-20

Den här presentationen kommer att ta upp olika funktioner som metadata kan ha. Jag kommer också visa att det finns olika typer av metadata beroende på vilken funktion de är tänkta att fylla.

Ett sätt att kategorisera metadata är utifrån funktion. Vi ska titta på tre huvudsakliga kategorier lite mer ingående och ta upp deras funktioner. De tre huvudkategorierna av metadata typer är:

- *beskrivande metadata*
- *strukturella metadata*
- *administrativa metadata*.

Administrativa metadata kan delas upp vidare, och jag kommer att kort nämna något om så kallade tillgänglighetsmetadata och bevarandemetadata, som på engelska heter *rights management* respektive *preservation metadata*.

Beskrivande metadata används för att synliggöra en resurs så att man kan söka efter och hitta den i en metadata katalog. I FAIR-termer pratar vi här främst om F:et: dessa metadata gör dataset *findable*. Exempel på beskrivande metadata är resursens titel och skapare eller författare, en sammanfattning eller ett abstract, beskrivande nyckelord, vilket språk den finns på och geografisk information som till exempel beskriver var den kommer ifrån eller vilket område den rör. I en metadata katalog ligger tyngdpunkten ofta på beskrivande metadata just för att det är den här sortens information som behövs för att sälla fram intressanta resurser. I en metadata katalog över forskningsstudier hjälper sådana metadata

forskare att söka fram studier och forskningsdata som kan användas för ny forskning.

Strukturella metadata, å andra sidan, ger information om hur data är uppbyggda eller organiserade. Det kan handla om antalet variabler eller antalet undersökningsobjekt, om mappstrukturer eller kapitel i en bok, om hur filer organiseras i en databas eller hur ett numeriskt dataset är uppbyggt i kolumner och rader. Strukturella metadata har en viktig funktion för att beskriva data och i förlängningen behövs de för att data ska kunna undersökas, det vill säga för själva forskningsanalysen. Dessa metadata kan kopplas till FAIR-principernas I och R: de bidrar till att göra data interoperabla och återanvändningsbara.

Administrativa metadata ger information om hur data får användas, var de kommer ifrån och hur de underhålls. Exempel på administrativa metadata inkluderar filformat, teknisk information som till exempel om resursen kräver någon särskild programvara, vilka rättigheter som är kopplade till resursen och om det finns några embargon att ta hänsyn till. I korthet skulle man kunna säga att administrativa metadata omfattar information som behövs för att förstå hur en viss resurs används och hur man får tillgång till den, och vissa administrativa metadata är av värde för att göra data tillgängliga – FAIR:s *accessible*.

Ibland delas administrativa metadata in i undergrupper, och två sådana får stå som exempel här. *Tillgänglighetsmetadata* beskriver vilka regler som finns rörande åtkomst till data och vilka rättigheter som finns, medan *bevarandemetadata* ger information som är av betydelse för arkivering och lagring.

När man pratar om olika kategorier av metadata ska man också tänka på att gränserna mellan de olika typerna av metadata inte alltid är knivskarpa. Beroende på vilket perspektiv man anlägger så kan vissa metadata passa in i flera olika kategorier. Det innebär

även att kopplingarna mellan metadatakategorierna och de olika FAIR-principerna bör ses mer som en antydan eller tumregel, inte som lagar huggna i sten.

SND:s katalog är i förändring, men oavsett gränssnitt så är metadata i varje katalogpost i möjligaste mån strukturerade efter funktion och den kategori de främst tillhör. Bilden nedan (översta bilden s. 4) visar en post från SND:s katalog den 10 december 2020 för studien "Journalist 2011"¹. Under fliken Beskrivning kan man se att många fält kan kategoriseras som beskrivande metadata. Man får bland annat veta vem primärforskaren är, att studien ingår i en samling och vad den har för syfte, och det finns även en beskrivning.

Under fliken Översikt (nedersta bilden på s.4) finns ett par exempel på administrativa metadata, närmare bestämt vilken tillgänglighetsstatus data har och en länk till forskningsprojektets hemsida.

Om man går vidare och tittar under fliken Data för samma data-beskrivning (s. 5) ser vi några exempel på strukturella metadata. Vi får bland annat information om antalet variabler och individer och får därigenom veta hur omfattande datasetet är.

Under denna flik hittar vi också studiens DOI som är en permanent identifierare för studien. Förkortningen står för Digital Object Identifier. En DOI går in under beskrivande metadata. Vi kan även se att det bara finns en version av data.

Journalist 2011
Den här studien är en del av samlingen Journalistundersökningarna

Beställ samling
Beställ data

Översikt | **Beskrivning** | Data

Beskrivande metadata

Ansvärg Institution/enhet
Göteborgs universitet, Institutionen för journalistik, medier och kommunikation

Skapare/primärforskare
Kent Asp - Göteborgs universitet, Institutionen för journalistik, medier och kommunikation

Identifierare
SND-ID: SND 0926

Beskrivning
Journalist '11 är en vetenskaplig frågeundersökning som genomförs vid Institutionen för journalistik, medier och kommunikation, Göteborgs universitet (JMG). Formuläret går ut till 2 500 slumpmässigt valda journalister i hela Sverige.
Undersökningen syftar till att belysa hur svenska journalister ser på sig själva och samhället, sin yrkesroll, arbetssituation och medieutvecklingen.
Journalist '11 är den sjätte undersökningen som genomförs vid JMG. Den första gjordes 1989. Resultaten från den senaste undersökningen Journalist '05 presenterades bland annat i boken Den svenska journalistikaren. Undersökningsserien har även resulterat i ett flertal doktorsavhandlingar i universitetsämnet Journalistik.
Under de tjugotvå år som gått sedan 1989 har det svenska mediasystemet förändrats i grunden. Etermediemonopolet har avskaffats, kommersiell radio och TV har vuxit fram, internet och nya medier har kommit till, arbetssätt har förändrats och påverkats av nya tekniker. Och under senare år har de sociala mediernas roll i journalistiken fått särskild uppmärksamhet.
Det främsta syftet med Journalist '11 är
--
Visa mer...

Språk
Svenska

Tidsperiod(er) som undersökts
2011-09 -- 2012-02

Geografisk utbredning
Geografisk plats: Sverige

Lägsta geografiska enhet
Kommun

Högsta geografiska enhet
Land

Kontakt för frågor om data
Ulrika Andersson

Ingår i samling hos SND
Journalistundersökningarna
Visa relaterade forskningsdata

Tillhörande dokumentation
Enkät_L_011.pdf
Kodbook_Journalist 2011.pdf

Ladda ner metadata
• Datafile
• OAI-AP
• OAI 2.5
• JSON

Journalist 2011
Den här studien är en del av samlingen Journalistundersökningarna

Beställ samling
Beställ data

Översikt | **Beskrivning** | Data

Administrativa metadata

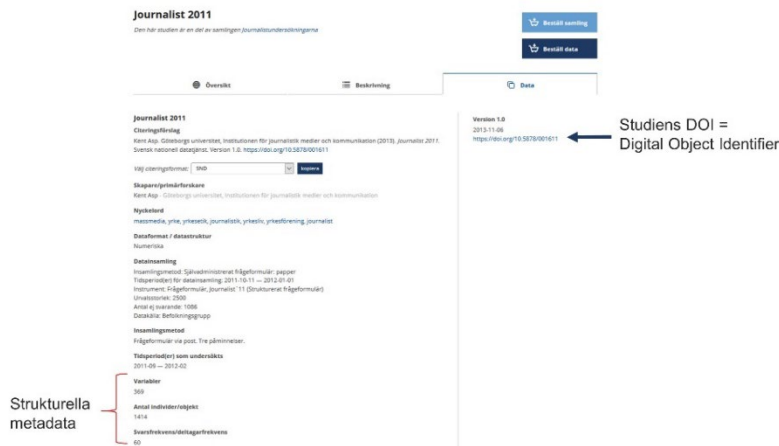
Skapare/primärforskare
Kent Asp - Göteborgs universitet, Institutionen för journalistik, medier och kommunikation

Beskrivning
Journalist '11 är en vetenskaplig frågeundersökning som genomförs vid Institutionen för journalistik, medier och kommunikation, Göteborgs universitet (JMG). Formuläret går ut till 2 500 slumpmässigt valda journalister i hela Sverige.
Undersökningen syftar till att belysa hur svenska journalister ser på sig själva och samhället, sin yrkesroll, arbetssituation och medieutvecklingen.
Journalist '11 är den sjätte undersökningen som genomförs vid JMG. Den första gjordes 1989. Resultaten från den senaste undersökningen Journalist '05 presenterades bland annat i boken Den svenska journalistikaren. Undersökningsserien har även resulterat i ett flertal doktorsavhandlingar i universitetsämnet Journalistik.
Under de tjugotvå år som gått sedan 1989 har det svenska mediasystemet förändrats i grunden. Etermediemonopolet har avskaffats, kommersiell radio och TV har vuxit fram, internet och nya medier har kommit till, arbetssätt har förändrats och påverkats av nya tekniker. Och under senare år har de sociala mediernas roll i journalistiken fått särskild uppmärksamhet.
Det främsta syftet med Journalist '11 är att följa upp tidigare undersökningar och ge en bred belysning av vilka förändringar som ägt rum i journalistikaren under de två senaste decennierna. Många frågor i Journalist '11 är därför desamma som ställts i tidigare undersökningar.
Likalydande frågor ställs också till allmänheten, och till riksdagsledamöter. Jämförelser görs med SOM-institutets undersökning 2011 och med Riksdagenskat 2010. Även i dessa fall görs jämförelser bakåt i tiden med Journalist '09 som startpunkt.
Den långa undersökningsserien är unik, även i ett internationellt perspektiv. En motsvarande undersökningsserie finns endast i USA.

Huvudman
Göteborgs universitet

Tillgänglighetsstatus
Åtkomst till data via SND
Tillgång till data är begränsad

Hemsida
Journalistundersökningen



Journalist 2011
Den här studien är en del av samlingen Journalistundersökningarna

Översikt | Beskrivning | Data

Journalist 2011
Självklaringsstudie
Kersti Ahl, Göteborgs universitet, Institutionen för journalistik, medier och kommunikation (2018), Journalist 2011
Svensk version av datagräns, version 1.0, <https://doi.org/10.5879/001611>

Välj dataformat:

Skapare/primärforskare
Kersti Ahl, Göteborgs universitet, Institutionen för journalistik, medier och kommunikation

Nyckelord
massmedia, yrke, yrkesetik, journalistik, yrkesetik, yrkesföring, journalist

Datamaterialet / datasetstruktur
Numeriska

Datansamling
Insamlingsmetod: Självadministrerat frågeformulär: papper
Tidsperiod(er) för datainsamling: 2011-10-11 – 2012-01-01
Instrument: Frågeformulär, Journalist 11 (Självadministrerat frågeformulär)
Urvalstorlek: 2500
Antal gånger insamlad: 1008
Datakälla: Beröringsgrupp

Insamlingsmetod
Frågeformulär via post, Tre administratörer
Tidsperiod(er) som undersöktes:
2011-08 – 2012-02

Variabler
369
Användningsobjekt:
1414
Svenskrekogningskategoriseringskod:
60

Strukturella metadata

Studien DOI = Digital Object Identifier
<https://doi.org/10.5879/001611>

Metadata på olika nivåer

Jag ska nu övergå från att prata om olika typer av metadata till att prata om metadata på olika nivåer. Data finns på olika nivåer och metadata kan beskrivas på liknande sätt.

På **projektnivån** eller **studienivån** finns metadata som ger övergripande information om ett forskningsprojekt. I SND:s katalog används benämningen "studier".

Nästa nivå kallas **fil-** eller **dataset-** eller **databasnivå** och handlar om data som har samlats in som ett delprojekt, eller en del av en studie, och där det alltså finns något som håller samman informationen i filen eller datasetet.

På **variabel-** eller **objektnivå** handlar det om dokumentation av specifika frågor eller mätningar. Det är kanske lättast att jämföra med SPSS eller Excel där varje variabel innehåller information om den specifika frågan och hur den är kodad.

Olika metadata är aktuella på de olika nivåerna. Projekt- eller studienivå omfattar metadata som beskriver sådant som är av övergripande karaktär för projektet. Typiska metadata på den här nivån inkluderar sådant som vilka frågeställningar forskaren vill

utforska, vilket syfte projektet har, hur det startade, vilket metodval forskaren har gjort och hur projektets geografiska täckning ser ut.

På filnivå hittar man exempelvis metadata som beskriver hur en enskild fil är uppbyggd, hur en samling filer hör ihop och hur tabeller i en databas förhåller sig till varandra och är uppbyggda. Det kan röra sig om olika versioner av datasetet, vilket eller vilka filformat det finns sparad i och eventuella checksummor som kan användas för att säkerställa att filerna inte har förvanskats. Om det finns flera filer inom ett dataset kan det också behövas en beskrivning av vad fil A innehåller och hur den skiljer sig från filerna B och C.

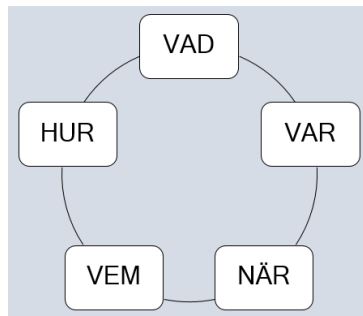
På variabel- eller objektnivå kan det för till exempel en surveyundersökning handla om hur en fråga är formulerad. Om svaren är kodade på något sätt behövs en kodlista som förklarar koderna. Ibland kan värdet på en variabel vara beräknat utifrån en eller flera andra variabler och därmed beroende av dessa, och det är då viktigt att det framgår tydligt för varje beroende variabel.

Med andra ord så varierar det från ämnesområde till ämnesområde och från studie till studie vad som behöver dokumenteras – det vill säga vilka metadata som ska användas – och i vilken omfattning. I regel kan man säga att ju mer detaljerad nivå som metadata har, desto större är skillnaderna. Syfte och abstract är exempel på metadata som används inom i princip alla forskningsområden, medan variabler och objekt kan se helt olika ut beroende på vetenskaplig disciplin.

Funktioner

Ett annat sätt att dela upp metadata är utifrån deras olika funktioner.

Ett exempel på en sådan funktion är sökbarhet. Då talar man ofta om de fem frågeorden: Vad, Var, När, Vem och Hur. Metadata som svarar på de här fem frågorna behövs för att det ska gå att hitta resursen, det vill säga för att kunna bygga en sökfunktion kring den.



Bra metadata för sökbarhet är viktiga också för spridning mellan olika system, men det är då viktigt att de följer en standard. Nästa pass kommer därför att handla mer om diverse metadata-standarder.

En viktig sak att tänka på i samband med metadata för sökbarhet är att det är bra att använda standardiserade kontrollerade vokabulärer i de fall det är möjligt. Ett kontrollerat vokabulär utgörs av en lista med förbestämda termer, och det är endast termer från listan som får användas för att fylla i metadataelementet som vokabulären hör till. En fördel med att använda kontrollerade vokabulär är att i den mån de finns översatta så blir det enklare att exponera forskningsdata på flera språk. Samma kontrollerade vokabulär kan användas i olika metadatakataloger, vilket gör att man kan använda samma söktermer i olika kataloger och hitta forskningsdata från olika källor.

En annan funktion hos metadata som vi inte har pratat om tidigare är att man ska kunna citera forskningsdata. Datacitering innebär att man kan referera till forskningsdata som använts i studien på samma sätt som man refererar till andra källor, som till exempel tidskriftsartiklar eller rapporter.

En citering till data bör innehålla information som gör att de data som använts går att hitta igen om man skulle vilja bygga vidare på

dem eller använda dem för reproduktion eller verifiering. Det kan till exempel gälla

- namnet på primärforskare och/eller organisationen som tagit fram data
- titeln på studien det rör sig om
- året som datasetet publicerades
- vilken version av datasetet som använts för analysen
- vilket repositorium eller dataarkiv som har förmedlat data; och
- en permanent identifierare så att man enkelt kan hitta rätt.

Ytterligare en viktig funktion hos metadata är att samla information som rör olika rättigheter för hur de får användas. Information om copyright, licens, huvudmannskap, tillgänglighetsnivå och förmedling är alla exempel på rättighetsmetadata.

Sammanfattning

Vi började med att titta på olika kategorier hos metadata, såsom beskrivande metadata, strukturella metadata och administrativa metadata. Gränserna mellan de här kategorierna är inte knivskarpa, men är ett sätt som kan hjälpa en att tänka kring metadata.

Metadata kan också indelas i olika nivåer, från en generell studienivå, via datasetsnivån och ned på objektsnivå. Oavsett hur man väljer att dela upp metadata så kräver olika typer av data olika typer av dokumentation och metadata.

Referenser

Svensk nationell datatjänst (2013). *Journalistundersökningen 2011*: <https://snd.gu.se/sv/catalogue/study/snd0926> (Hämtad 2020-12-10)