

## Exempel på metadastandarder

### Pass 4: Metadastandarder

*BAS Online 2021-01-20*

Nu är det dags att ta upp några exempel på metadastandarder. Jag ska börja med att titta lite snabbt på ett urval av metadastandarder som vi på SND har stött på i olika sammanhang. Vissa känner du kanske redan till: MARC är till exempel bekant för de flesta bibliotekarier och Dublin Core förekommer i många olika sammanhang.

Dublin Core är en metadastandard som upprätthålls av DCMI, *The Dublin Core Metadata Initiative*, och tanken med standarden är att man ska kunna hitta och identifiera resurser av väldigt skilda slag. Däremot finns det med Dublin Core inte möjlighet att ge detaljerad information om resurserna. Själva kärnan i standarden utgörs av de femton element som du kan se listade nedan. Det formella namnet för de här elementen är *The Dublin Core Metadata Element Set*.

contributor	format	rights
coverage	identifier	source
creator	language	subject
date	publisher	title
description	relation	type

Det finns även en utökad variant av Dublin Core-standarden som heter DCMI Metadata Terms. Förutom de femton elementen i den mindre varianten så innehåller den 40 så kallade *terms*. Inte alla de 55 elementen i DCMI Metadata Terms är obligatoriska utan det är ganska mycket ett smörgåsbord som man kan välja och vraka från efter behov. Dessutom är det i det närmaste obligatoriskt att man,

när man introducerar Dublin Core för någon, påpekar att den har fått sitt namn efter en workshop i Dublin, Ohio, och inte har något med Dublin på Irland att göra.

MARC är en standard från biblioteksvärlden. Den har funnits sedan 1960 och fungerar bra när man ska katalogisera böcker och andra publikationer.

METS är en standard för digitala publikationer som vid sidan av katalogisering också funkar bra för att överföra metadata från ett ställe till ett annat. METS innehåller främst administrativa metadata.

PREMIS används för digital arkivering och innefattar administrativa och tekniska metadata för digitala objekt.

DDI, *Data Documentation Initiative*, är den standard som SND använder mest. Det är en ganska omfattande standard med omkring 1 200 element som har sitt ursprung i samhällsvetenskaplig kvantitativ metod, och den är väldigt detaljerad när man ska beskriva tabulära surveydata. Bland annat kan man med DDI-element beskriva data ned på variabelnivå. Jag kommer att ta upp DDI i mer detalj alldeles strax.

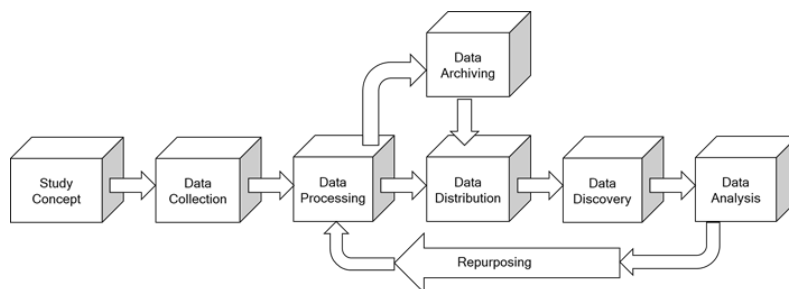
OLAC är en standard som bygger på och har avsevärda likheter med Dublin Core. Man har lagt till några kontrollerade vokabulärer som gör den mer lämplig för språkvetenskapliga data och mindre lämplig för allt annat. Ett exempel på ett sådant vokabulär är språkkoder.

TEI används för att göra extremt detaljerade uppmärkningar av text, till exempel grammatik, stilistik, ordförråd och handskriftsrelaterad information. Man kan också märka upp platser, personer, datum och föremål som nämns i texten och länka till andra ställen där det finns mer information om dem. Dokumentationen för TEI är på 1 800 sidor och standarden en väldigt bred tillämpning.

Så, nu ska vi gå in lite mer på DDI. Som jag sa tidigare så skapades DDI från början som ett hjälpmedel när man skulle dokumentera tabulära datafiler från enkätundersökningar. Standarden används på många dataarkiv och datarepositorier världen runt och det beror på att de som arbetar med bevarande av data ofta har sitt ursprung i ämnesområden där tabulära data är väldigt vanliga. Kontrollerade vokabulärer är en viktig ingrediens i DDI liksom i många metadata-standarder och vi kommer att prata om dem i nästa presentation.

Det finns två varianter av DDI som används parallellt: DDI Codebook och DDI Lifecycle. Codebook kom först och har som huvudsakligt syfte att man ska kunna förstå alla rader och kolumner i en tabulär datafil. Från en metadatabeskrivning i DDI Codebook kan man generera en kodbok som innehåller all nödvändig information för att använda filen. DDI Lifecycle bygger istället på idén att data har en livscykel där de förändras kontinuerligt, från sin tillkomst tills deras syfte har uppfyllts. Man vill beskriva hela livscykeln med DDI Lifecycle och det är en av anledningarna till att standarden är så omfattande som den är.

Här nedanför ser du datalivscykeln så som DDI-alliansen föreställer sig den. Tänk på att livscykeln beskriver data och inte metadata. Metadata används när man ska beskriva alla dessa olika steg.



#### *DDI Alliance [inget år]*

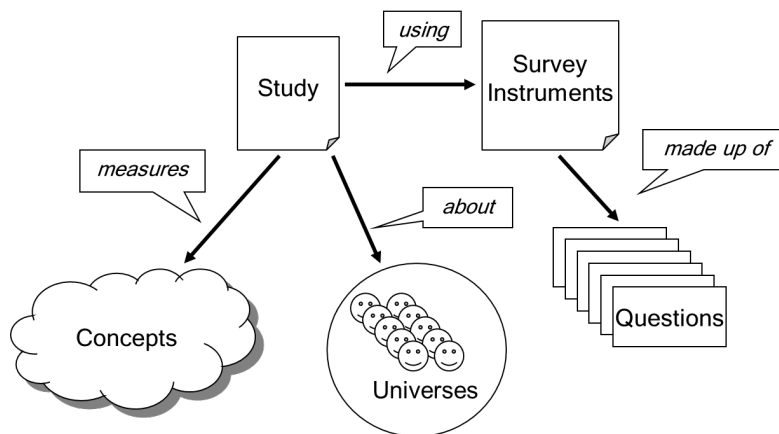
Längst till vänster börjar livscykeln med att man kommer på en projektidé och planerar hur man kan genomföra den. Sedan sker

själva datainsamlingen som i DDI-modellen vanligtvis är någon typ av enkät. När enkäterna kommit in så slås de ihop till en fil som bearbetas för att ge bästa möjliga underlag för analys. Med bearbetade data sker två saker. De distribueras eller lämnas över till den som ska utföra den vetenskapliga analysen men det är också den här bearbetade versionen av data som överlämnas till ett dataarkiv eller datarepositorium, som i sin tur kan distribuera data vidare. Till slut hittar eller får forskaren data, om vederbörande inte själv samlade in dem. I mindre projekt är det vanligtvis samma person som har genomfört datainsamlingen och bearbetningen som sedan analyserar data, vilket därefter resulterar i en eller flera publikationer. Eftersom det finns en mekanism för distribution kan någon annan hitta data och komma på ett nytt användningsområde för dem och så börjar processen om med ny bearbetning, till exempel genom att slå ihop variabler eller kombinera dataset.

För att den här återvinningen ska vara möjligt är det viktigt att alla de tidigare stegen i processen dokumenterats ordentligt, annars kan den nya forskaren inte veta på vilka sätt det är rimligt att använda de data som har hittats.

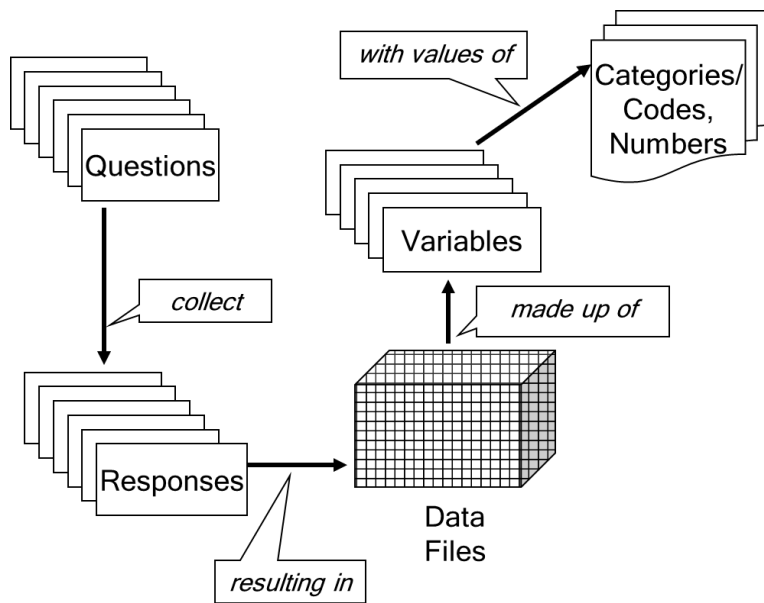
Tänk på att den här modellen beskriver en idealiserad bild av hur en viss typ av forskning går till. Mycket verklig forskning är betydligt rörigare än den här modellen visar och det finns andra forskningsmetoder som fungerar annorlunda.

Modellen i de två figurerna nedan visar hur en enkätundersökning genomförs. Det här är den typ av data som DDI är gjort för att beskriva. Kärnan är en studie som mäter koncept om en population. Koncept kan vara lite vad som helst, exempelvis kaffekonsumtion, och i studien undersöker man då kaffekonsumtionen hos en viss grupp individer. För att mäta kaffekonsumtionen använder man ett surveyinstrument, exempelvis en pappersenkät eller en nätenkät. Enkäten innehåller frågor, till exempel "Hur många koppar kaffe dricker du om dagen?" och "Vilken sorts kaffe brukar du dricka?".



Copyright © GESIS – Leibniz Institute for the Social Sciences, 2010.  
Published under Creative Commons Attribute-ShareAlike 3.0 Unported

När man ställt de här frågorna till sina deltagare får man ett antal svar som sedan sammanställs i en eller flera datafiler. Datafilerna består av variabler som motsvarar kolumner i en Excel-fil. En variabel är till exempel "Antal koppar kaffe om dagen". Varje rad i filen innehåller då en deltagares svar. Variabler kodas som kategorier eller numeriska värden. "Antal koppar kaffe om dagen" får rimligtvis ett numeriskt värde, exempelvis "3" medan "Typ av kaffedryck" får olika kategorier: bryggkaffe, espresso, latte osv. Av tradition brukar sådana kategorivariabler ersättas med siffror, så att "1" motsvarar "bryggkaffe", "2" motsvarar "espresso" och så vidare. De här koderna är en av anledningarna till att man behöver en kodbok för att kunna förstå en datafil.



Copyright © GESIS – Leibniz Institute for the Social Sciences, 2010.  
Published under Creative Commons Attribute-ShareAlike 3.0 Unported

## Sammanfattning

Jag har i den här presentationen tagit upp några av de standarder som du kan komma i kontakt med som DAU-medarbetare. Vi har tittat lite närmare på Dublin Core och DDI. Men det finns många andra standarder än de jag nämnt här.

Standarderna blir hela tiden fler och det är viktigt att inse att man som DAU-medarbetare inte kan arbeta med alla standarder. Man kan gradvis lära sig mer, men att verkligen lära sig en standard innebär att man måste arbeta med den i praktiken och inte bara läsa dokumentationen. Eftersom SND:s system bygger på DDI rekommenderar vi DDI som den första standard du verkligen tar itu med.

Nästa presentation handlar om hur man kan använda en metadata-standard mer i praktiken, utifrån sina egna behov.

## Referenser

DDI Alliance [inget år]. *DDI Data Lifecycle*.

<https://ddialliance.org/training/why-use-ddi> (Hämtad 2021-01-20).