

Utvecklingsplan över webbplatserna

I denna rapport presenteras utvecklingsplanen över de webbplatser som Riksarkivet förvaltar och vidareutvecklar inom ramen för regeringsuppdraget om att främja myndigheternas arbete med data för vidareutnyttjande, projektet ”Öppna data och PSI”.

Diarienummer: 04-2016/6854

Version 1.0. Licens CC-0. En senare version kan finnas på vidareutnyttjande.se.

1	Inledning.....	4
2	Bakgrund.....	4
3	Inventering av webbplatserna	4
3.1	Dataportalen oppnadata.se	4
3.1.1	Bakgrund	4
3.1.2	Funktionalitet	4
3.1.3	Ett europeiskt perspektiv.....	5
3.2	Webbhandledningen vidareutnyttjande.se	6
3.2.1	Bakgrund.....	6
3.2.2	Nuvarande förvaltning.....	6
3.2.3	Funktionalitet	6
3.2.4	Ett europeiskt perspektiv.....	7
4	Förslag till vidareutveckling och förvaltning av webbplatserna	7
4.1	Dataportalen oppnadata.se.....	7
4.1.1.1	Endast CKAN	7
4.1.1.2	Innehållshanteringssystem där man implementerar CKAN.....	7
4.1.1.3	Innehållshanteringssystem som kallar på en specifik egenutvecklad datakatalog via API.....	8
4.1.2	Nya funktioner	8
4.2	Webbhandledning vidareutnyttjande.se	9
4.2.1	Vägledning om PSI-förteckningar och öppna data	9
4.2.2	Interaktiv kommunikations- och informationsyta för projektet (blogg, m m)	9
5	Riktlinjer för licensiering av innehållet på webbplatserna	10
6	Förvaltning av DCAT-AP.....	10
7	Ordlista	11
8	Akronymlista	11

1 Inledning

I denna rapport presenteras utvecklingsplanen över de webbplatser som Riksarkivet förvaltar och vidareutvecklar inom ramen för regeringsuppdraget om att främja myndigheternas arbete med data för vidareutnyttjande, projektet "Öppna data och PSI".

Utvecklingsplanen har gjorts i form av en inventering av webbplatserna, beskrivning av dessa och en konsekvensanalys av olika tekniska alternativ kopplade till vidareutveckling av dataportalen, oppnadata.se. Resultatet är ett förslag till vidarehantering av dessa under 2017.

2 Bakgrund

Som en del av regeringsuppdraget till Riksarkivet (Fi2015/02025/SFÖ, FI2016/01537/SFÖ)¹, ska Riksarkivet bland annat stimulera myndigheter att publicera öppna data enligt gemensamma riktlinjer, förvalta och utveckla webbportalen oppnadata.se samt ansvara för att förvalta webbhandledningen vidareutnyttjande.se.

3 Inventering av webbplatserna

Dataportalen oppnadata.se

Bakgrund

Dataportalen oppnadata.se skördar och visar metadatabeskrivningar av öppna data från offentliga organisationer. För att data ska kunna visas i dataportalen måste databeskrivningarna följa metadataspecifikationen DCAT-AP² och därmed vara uttryckta i maskinläsbart RDF/XML-format. I samband med vidareutvecklingen av oppnadata.se har portalen anpassats för att använda den senaste versionen av DCAT-AP, 1.1. Delar av DCAT-AP 1.1 har tidigare anpassats på beställning av Vinnova³ utifrån funktionaliteten i oppnadata.se.

Funktionalitet

Dataportalen oppnadata.se har idag vissa begränsningar gällande funktionaliteten. Detta beror på att det underliggande systemet, CKAN (The Comprehensive Knowledge Archive Network),⁴ fungerar främst som en datakatalog för metadatabeskrivningar. Systemet baseras på en öppen källkod. CKAN är byggd med Python (back-end) och JavaScript (front-end).

¹ <http://www.regeringen.se/pressmeddelanden/2016/07/regeringen-oppnar-dorren-for-mer-oppna-data/>

² https://joinup.ec.europa.eu/asset/dcat_application_profile/description

³ https://docs.google.com/document/d/17-vEfZXlu9kykcmjXZo1_Z8QKkr7-Prgwd6YUKLRrjk/edit

⁴ <https://ckan.org/>

Sökningarna implementeras av Solr. CKAN-installationer tillhandahåller automatiska sökningar genom webb-API:er.⁵

CKAN-systemet ger tillgång till en grundläggande funktionalitet. Den har en modulär arkitektur som gör det möjligt att utveckla tillägg (plug-ins) för att tillhandahålla ytterligare funktioner. För en förbättrad redaktionell funktionalitet, layout, m m, behöver CKAN även kombineras med ett innehållshanteringssystem (Content Management System, CMS) och/eller vidareutvecklas på annat sätt. CKAN lagrar metadata om olika datamängder och presenterar dessa i ett webbgränssnitt.

Det metadataformat som används i CKAN är DCAT-AP, ett format rekommenderat av EU-kommissionen för dataportalen.

Riksarkivet har övertagit lagring, tekniskt support och drift för dataportalen. Detta sköts för närvarande delvis genom externa leverantörer.

Ett europeiskt perspektiv

De flesta dataportalen på europeisk nivå använder CKAN, eller varianter av den (exempelvis med implementering av ett innehållshanteringssystem). En genomgång av olika rekommendationer och goda exempel (best practices) från de andra europeiska länderna visar att CKAN i en del länder också innefattas i rekommendationerna kring öppna data som standard för dataportalen.

Bland andra system som används finns egenutvecklade, specialanpassade lösningar, t ex OpenDataSoft, DKAN, MicrosoftSharePoint.⁶ Dataportalen som använder olika tekniska lösningar försvårar utbyte av metadata. För närvarande finns det på europeisk nivå över 150 icke-kompatibla öppna dataportalen. Nackdelen med olika lösningar utan möjlighet till samordning är att allmänheten inte kan komma åt informationen genom en gemensam sökning utan behöver söka efter information som finns utspridd över olika webbsidor och portalen i alla europeiska länder. Detta fragmenterade landskap av dataportalen bidrar till att öppna data inte kan nå sin fulla potential.

Skördningen från Sverige till den europeiska dataportalen, European Data Portal (EDP), har hittills fungerat efter en modell där man skördade till EDP från olika svenska portalen där det fanns öppna data och inte endast genom en gemensam ingång, som oppnadata.se.

På nationell nivå publiceras datamängder i vissa fall genom myndigheternas webbsidor eller genom en sektorspecifik dataportal eller databas. Dessa datamängder syns inte alltid i den nationella dataportalen och är inte heller alltid helt öppna. Exempel på dessa dataportalen är exempelvis K-samsök, Geodataportalen, Nationella arkivdatabasen, Trafikdataportalen, m fl. En del portalen tillgängliggör data som är definierade som öppna data, andra gör det endast delvis eller inte alls. Även vad gäller kommuner och regionala sammanställningar av data så finns det exempel på portalen och sökfunktioner, exempelvis Dataportalen Stockholm,⁷

⁵ <https://en.wikipedia.org/wiki/CKAN>

⁶ <https://joinup.ec.europa.eu/node/161966/>

⁷ <http://dataportalen.stockholm.se/dataportalen/>

dataportalen Göteborg,⁸ Öppna data Helsingborg,⁹ Open North,¹⁰ Västra Götalands regionen,¹¹ med flera. Även Sveriges Kommuner och Landsting, SKL, tillgängliggör öppna data i en tjänst som visar sammanställda nyckeltal.¹²

Med denna bakgrund har Riksarkivet inom arbetet med regeringsuppdraget om öppna data inlett en process för att den nationella portalen ska kunna skörda metadata om öppna datakällor från andra dataportaler inom landet, vilket innebär att EDP på sikt behöver skörda data endast från den nationella portalen. Detta innebär att man kommer att undvika att skörda dubletter samt att det kommer att skapas en gemensam sökingång på nationell nivå, som effektiviserar informationshantering och ökar möjligheterna för slutanvändare att söka, hitta och använda öppna datamängder.

Webbhandledningen vidareutnyttjande.se

Bakgrund

Arbetet med handledningen påbörjades av Peter Krantz i sin anställning på E-delegationen och togs sedan över av Björn Hagström på E-delegationen. Björn slutförde arbetet med den första versionen av vägledningen och de juridiska bedömningarna stod till stor del Daniel Westman för.

Version två av vägledningen togs fram av E-delegationen och SKL. SKLs ramverk för öppna data slogs ihop med E-delegationens vägledning så att det bara skulle finnas en nationell vägledning inom området. Arbetet genomfördes av en arbetsgrupp bestående av representanter från E-delegationen, SKL, Kommunförbundet Stockholms län och PwC, med stöd av Daniel Westman och Jeanna Thorslund som stod för juridiska bedömningar.

Nuvarande förvaltning

Regeringen har gett Riksarkivet i uppdrag att främja arbetet med att tillgängliggöra information och öppna data från statliga myndigheter under regeringen. I uppdraget ingår att ansvara för att förvalta webbhandledningen vidareutnyttjande.se. Riksarkivet övertog ansvaret från och med 2016-07-01 och arbetar för närvarande med att ta fram en version tre av vägledningen.

Funktionalitet

Webbhandledningen är byggd i innehållshanteringssystemet WordPress och fungerar som en webbsida med redaktionell möjlighet.

⁸ <http://www25.goteborg.se/psidata/Default.aspx>

⁹ <https://oppna.helsingborg.se/>

¹⁰ <http://opendata.opennorth.se/sv/>

¹¹ <http://www.vgregion.se/om-vgr/om-webbplatsen/oppna-data/>

¹² <https://oppnadata.skl.se/page/firstpage>

Webbhandledningen består av två delar, "Juridik & rekommendationer" och "Skapa och publicera öppna data".

Under januari 2017 har bloggposterna fått en mer framträdande roll på förstasidan för att ge en bättre kommunikationsyta för att sprida nyheter om projektet, omvärldsbevakning, fråga efter och få in kommentarer och synpunkter kring vissa frågor och involvera öppna data communities och andra typer av användare, som t ex företag som inte behöver följa vägledningen men gärna vill använda och/eller leverera öppna data.

Ett europeiskt perspektiv

På en europeisk nivå har ett tematiskt nätverk "Share PSI, the network for innovation in European public sector information"¹³ kopplat till W3Cs "Data on the web Best Practices working group" och finansierat av EU-kommissionen, arbetat med att bland annat sammanställa "best practices" kring PSI och öppna data-frågor. En översiktlig genomgång av flera andra europeiska länders webbhandledningar kring öppna data ger en bild av att allmän praxis är att publicera webbhandledningen i dataportalen för öppna data.

4 Förslag till vidareutveckling och förvaltning av webbplatserna

Dataportalen opnadata.se

4.1.1 Teknisk plattform

Det finns tre möjliga lösningar för att skapa tekniska möjligheter till vidareutveckling. Alla kräver mer eller mindre kontinuerliga insatser av utvecklare för att få kopplingen från metadatarregistret att fungera med vad som visas på webbsidan, vare sig det är CKAN eller en specialkodad lösning.

CKAN är en de facto-standard för dataportaler som följer vad som kan bedömas vara praxis inom området (se 3.1.3). CKAN är framtaget av Open Knowledge Foundation och utvecklas kontinuerligt. Det finns olika varianter på CKAN-installationer:

Endast CKAN

Nackdelen med att ha en ren CKAN-portal är att det krävs många plug-ins om man vill ha en mer funktionell och attraktiv webbsajt. I sitt grundutförande är CKAN begränsad i vad man kan göra layout- och funktionsmässigt.

Innehållshanteringssystem där man implementerar CKAN

Innehållshanteringssystemet står för webbsidorna och CKAN för datat. Ett tillägg i innehållshanteringssystemet möjliggör API-anrop till CKAN. . Funktionerna finns i

13

<http://www.w3.org/2013/share-psi/>

innehållshanteringssystemet och information från datakatalogen hämtas med kodsträngar. Fördelen med denna lösning är att den bedöms vara relativt lättadministrerad.

Även en variant där CKAN står för data samt webbsidorna och ett innehållshanteringssystem hanterar autentisering, användare, etc.

CKAN tillsammans med innehållshanteringssystemet Drupal är den CKAN-lösning som används mest internationellt, t ex i Irland och Finland.

CKAN tillsammans med Wordpress används av t ex datapress.com samt den amerikanska dataportalen data.gov.

Innehållshanteringssystem som kallar på en specifik egenutvecklad datakatalog via API

Funktionerna finns i innehållshanteringssystemet och information från datakatalogen hämtas med kodsträngar. Nackdelen med detta alternativ är att det är en specialbyggd lösning som frångår praxis och som kan försvåra utbyte av metadata.

Nya funktioner

De förändringar som ligger närmast för oppnadata.se har till stor del baserats på en omvärldsbevakning genom bland annat dialog med den europeiska dataportalen och erfarenhetsutbyte med kollegor från andra europeiska länder. Riksarkivet har även tagit in synpunkter från användarna, bland annat inom ramen för den kartläggning av användarbehov som genomfördes under våren 2017.

Till hösten är följande funktioner planerade:

- möjlighet att ladda ned alla dataset på en gång (finns redan idag, men behöver bli tydligare för användarna).
- ladda ned dataset på metadatanivå (finns redan idag, men behöver bli tydligare för användarna).
- tydligare visning av datakategorier (finns flera kategorier än det visas idag, vi behöver bli tydligare kring vikten av att "tagga" dataset i beskrivningar av metadata).
- RSS-flöde för att användarna lätt ska kunna se vad som är nytt i portalen

Andra förändringar som är planerade men som kräver mer utvecklingsinsatser är:

- möjlighet för användarna att kommentera dataset
- användarna ska kunna lämna önskemål om vilka data de vill se på portalen
- uppmärkning av dataset enligt femstjärne-modellen för öppna data
- verktyg för DCAT-AP (att skapa och validera)
- kataloginformation i DCAT-AP visas i portalen

För att skapa kopplingar mellan metadata från olika organisationer och sektorer ligger mycket av fokus idag på att sammanlänka informationen genom så kallade auktoritetsfiler. Det innebär att datakällor som kan vara referenspunkter för beskrivning av informationen från

flera institutioner tillgängliggörs online. Dessa kan t.ex. vara personer, organisationer, platser, händelser, ämnen och liknande. Samtidigt är det viktigt att länkar som pekar på dessa datakällor online följer ett visst format.

Därför kommer även möjligheten att kunna hantera beständiga identifierare och auktoritetsfiler i dataportalen att undersökas.

Webbhandledning vidareutnyttjande.se

Vägledning om PSI-förteckningar och öppna data

Då genomgången av andra europeiska länders webbhandledningar kring öppna data gav en bild av att allmän praxis är att publicera webbhandledningen/vägledningarna i dataportalen för öppna data så kommer vi under 2017 att arbeta med att implementera webbhandledningen i dataportalen opnadata.se. Det blir tydligare rent kommunikativt att ha instruktionerna på den webbsida som instruktionerna gäller istället för att ha två olika webbsidor. Valet av CKAN med Wordpress-alternativet för vidareutveckling av dataportalen ger goda möjligheter till att implementera webbhandledningen i dataportalen.

Interaktiv kommunikations- och informationsyta för projektet (blogg, m m)

Riksarkivets arbete med regeringsuppdraget om data för vidareutnyttjande är ett projekt som handlar om öppenhet och proaktivitet i tillgängliggörande av information. Därför är också transparens en självklar del i projektets arbetssätt. Regeringsuppdraget handlar om att både publicera öppna data men även om att motivera, informera och främja myndigheternas arbete med öppna data samt förvalta och utveckla portalen för öppna data på ett sätt som gör att den är attraktiv för utvecklare, forskare och allmänheten i stort. En viktig del i uppdraget är därmed också den tekniska utvecklingen som befinner sig under ständig förändring och där det automatiskt ingår en stor bit omvärldsbevakning och kommunikation. Inom ramen för detta behöver Riksarkivet också ta in synpunkter från omvärlden, dra slutsatser, samt kommunicera tillbaka kunskap och det vi producerar ut till användarna. Vårt mål är att alla som är intresserade ska kunna följa vårt arbete på webben.

Via bloggen och sociala medier når vi ut till olika öppna data-communities, i många fall även internationellt. Effekten av detta har redan påvisats under projektets första månader, då det har varit bidragande orsak till informationsspridning och etablering av nya kontakter. Webbstatistiken för vidareutnyttjande.se visar att bloggposter är minst lika intressanta för användare som innehållet i vägledningen. Vi skickar även ut ett nyhetsbrev för att nå ut till de målgrupper som endast kan nås via mail och/eller som inte hittar till bloggen.

Bloggposterna ger också möjlighet att sätta fokus på olika delar i vägledningen, sprida nyheter om uppdateringar, omvärldsbevaka och skapa sammanhang.

Vägledningen

Vad gäller vägledningen finns det parallellt med layouten också ett pågående arbete med att uppdatera innehållet. Fokus under våren 2017 ligger på att ta fram en vägledning för att

skapa och publicera PSI-förteckningar och öppna data. Inom ramen för detta görs en inventering av befintliga sidor kring att skapa och publicera öppna data.

Under hösten 2017 kommer fokus att ligga på en inventering och uppdatering av juridiska frågor samt en rekommendation kring licensiering och beständiga identifierare.

Det finns också möjlighet att implementera en utbildning om öppna data (via Europeiska dataportalen) på webbsidor, vilket är efterfrågat av användarna. Möjligheter att implementera en sådan utbildning i dataportalen oppnadata.se kommer att undersökas.

Showcase – goda exempel

Lades till under maj 2017. Där visas goda exempel som Riksarkivet kontinuerligt kommer att fylla på.

Om oss

Menyval som ger grundläggande information om Riksarkivets regeringsuppdrag, (d v s alla delar, inte bara om vägledningen), hur projektet drivs och vilka vi är som jobbar i projektet (kontaktuppgifter, m m).

5 Riktlinjer för licensiering av innehållet på webbplatserna

Metadatat på portalen är CC0. Detta gäller även för de rapporterna som projektet publicerar.

6 Förvaltning av DCAT-AP

Delar av den senaste DCAT-AP1.1-versionen har tidigare anpassats på beställning av Vinnova utifrån funktionaliteten i oppnadata.se och förändringar är dokumenterade på engelska.¹⁴ I samband med Riksarkivets vägledning om PSI-förteckningar kommer Riksarkivet att ansvara för översättning, förvaltning av och utveckling av en nationell DCAT-AP-anpassning. Detta arbete kommer att påbörjas i september 2017 och kommer troligtvis att påverka utvecklingen av dataportalen.

¹⁴https://docs.google.com/document/d/17-vEfZXlu9kykcmjXZo1_Z8QKkr7-Prgwd6YUKLRjk/edit

7 Ordlista

API (Application Programming Interface) är en regeluppsättning för hur en viss programvara kan kommunicera med annan programvara.¹⁵

Beständiga identifierare - Unika kodsträngar kopplade till varje objekt/post. PID – Persistent Identifier är en förutsättning för att man sedan med fasta länkar ska kunna skapa referenser mellan olika objekt/poster i ett nätverk som t.ex. Internet.

CKAN – ett verktyg med öppen källkod för att skapa dataportaler för öppna data.¹⁶

Innehållshanteringssystem - ¹⁷

Licensiering - En markering av materialets upphovsrättsliga status.

Metadata – Data om data (Anmärkning: Metadata är knutna endera till ett informationssystem eller till ett informationsobjekt med syftet att beskriva, administrera eller använda detta. Termen används vanligen för strukturerande uppgifter om digitala (och analoga) system. Metadata delas ofta upp i olika typer, exempelvis tekniska metadata som skapas för eller genereras av ett datasystem, administrativa metadata som beskriver hur ett system fungerar eller hur ett objekt skapats och metadata som rör objekten i systemet och deras egenskaper).¹⁸

Skördning – inom ramen för denna rapport avses automatiskt hämtning av metadataelement.¹⁹

Vidareutnyttjande: användning av handlingar för andra ändamål än det ursprungliga ändamål för vilket handlingarna behandlas av en myndighet. Lag (2015:289).²⁰

Öppna data - Avser digital information som man kan vidareutnyttja utan att betala avgift eller acceptera annat villkor än att man kan behöva ange källan.

8 Akronymlista

API - (Application Programming Interface)

CMS – Content Management System, se Innehållshanteringssystem.

DCAT-AP – DCAT Application Profile

¹⁵ http://sv.wikipedia.org/wiki/Application_Programming_Interface

¹⁶ <https://ckan.org/about/>

¹⁷ <https://sv.wikipedia.org/wiki/Inneh%C3%A5llshanteringssystem>

¹⁸ <http://digisam-ra.blogspot.se/p/begreppslista-bilddatabaser-och.html>

¹⁹ https://en.wikipedia.org/wiki/Metadata_discovery

²⁰ <http://rkrattsbaser.gov.se/sfst?bet=2010:566>

PSI - offentlig information (Public Sector Information)

RDF – Resource Description Framework

XML - Extensible Markup Language