

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

Teknisk Review av K-samsök

Rapport

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

Sammanfattning

Riksantikvarieämbetet har genomfört en oberoende utvärdering av systemet K-samsök som syftat till att identifiera styrkor och svagheter i systemet vad det gäller den tekniska lösningen, hur systemet löser de uppgifter det är tänkt att lösa, hur systemet förhåller sig till den semantiska webben samt hur väl dokumenterat systemet är.

Vad det gäller den tekniska lösningen i K-samsök har utvärderingen kommit fram till att arkitekturen är adekvat och skalbar, att koden är förvaltningsbar och flyttbar och att lösningen smälter väl in i öppen-källkods-samfundet om delar av systemet skulle licensieras som öppen källkod. Det fåtal mindre problem som finns med den tekniska lösningen bedöms rymmas inom normal förvaltning av systemet.

Avseende de uppgifter K-samsök är framtaget för att lösa är bedömningen att det finns ett behov av att se över vilka format som accepteras och används för de olika uppgifterna. Detta för att möjliggöra att de olika uppgifterna kan lösas var för sig och på det sätt som är optimalt för den enskilda uppgiften men även för att nå andra målgrupper.

För att fungera väl på den semantiska webben finns det ett behov av att formalisera det egna RDF-formatet, att arbeta för att innehållsleverantörerna länkar sin information till fler auktoriteter och om möjligt externa datakällor samt att se över identitetshandlingen. Det bedöms även finnas en outnyttjad potential i den semantifierade information som hanteras av K-samsök vilken skulle kunna användas för att inferera samband som inte finns uttryckligen i informationen och för att möjliggöra semantiska frågor i det externa API:et.

Slutligen har utvärderingen kommit fram till att dokumentationen av systemet är utförlig och förståelig med en viss avsaknad av information som riktar sig mot personer som har liten verksamhetskunskap.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

Innehåll

1	REVISIONER	4
2	INLEDNING	5
2.1	AVGRÄNSNINGAR	5
3	ÖVERSIKT	6
4	TEKNISK LÖSNING	8
4.1	WEBBAPPLIKATIONER	8
4.2	DATABASER	8
4.3	KOD	9
4.4	BEROENDEN	10
4.5	ÖPPEN KÄLLKOD	10
5	INFORMATION OCH TJÄNSTER	12
5.1	SKÖRDA	13
5.2	LEVERERA	14
5.3	TILLGÄNGLIGGÖRA	14
5.4	AVREFERERA	16
5.5	FÖRÄDLA	17
5.6	ÖVRIGT OM K-SAMSÖKS RDF-FORMAT	17
6	WEBBSEMANTIK	18
6.1	FORMAT	18
6.2	LÄNKADE DATA	18
6.3	IDENTITETER	19
6.4	INFERENS	19
6.5	SEMANTISKA SÖKNINGAR	19
7	DOKUMENTATION	21
7.1	EXTERN	21
7.2	INTERN	21
8	SLUTSATSER	22

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

1 Revisioner

Utg.	Datum	Utställare	Kommentar
PA1	2013-01-18	Johan Behrenfeldt	Preliminär första utgåva för leverans till Riksantikvarieämbetet.
PA2	2013-01-22	Johan Behrenfeldt	Korrigerat informationen om de lokala portarna och om formatet Europeana Data Model.
PA3	2013-01-25	Johan Behrenfeldt	Korrigerat informationen om leveransen av information till CARARE. Justerat formuleringarna rörande databasanvändare och formateringen av JSON i UGC-hubben. Inkluderat information om vilka som ställt upp på intervjuer.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

2 Inledning

Riksantikvarieämbetet har genomfört en oberoende utvärdering av systemet K-samsök. Syftet med utvärderingen kan sammanfattas till att söka svar på frågan om hur väl rustat K-samsök är för den framtid som väntar systemet. Mer specifikt så handlar utvärderingen om att besvara frågor om vilka styrkor och svagheter systemet har avseende dess tekniska lösning, de uppgifter systemet är tänkt att lösa, dess position på den semantiska webben samt den dokumentation som finns av systemet.

Utvärderingen genomfördes under december 2013 och informationsinsamlingen bestod huvudsakligen i ett antal möten med Johan Carlström, Henrik Summanen, Börje Lewin och Lars Lundqvist på Riksantikvarieämbetet. Dessutom inhämtades en del information under K-samsökmötet den 11 december där ett antal innehållsleverantörer och representanter för andra museer närvarade. Kod och dokumentation för K-samsök tillhandahölls av Johan Carlström och Börje Lewin och på webbsidan www.ksamsok.se.

2.1 Avgränsningar

Även om utvärderingen rör ett flertal aspekter av systemet och har ett brett perspektiv så finns det avgränsningar för vad som har inkluderats i utvärderingen och hur omfattande den gjorts.

Den utvärdering av arkitekturen och koden som genomförts har inte fokuserat på kvantitativa mått eller olika typer av metrik utan ingår som en del av det större kontext som söker svar på de mer allmängiltiga frågorna som presenterades tidigare.

Utvärderingen av arkitekturen har även enbart studerat den logiska arkitekturen av systemet och inte den fysiska. Samtidigt är det förstås så att en viktig egenskap hos en bra logisk arkitektur och systemdesign att den inte begränsar vilka möjligheter som finns att skapa en väl fungerande fysisk arkitektur eller möjligheten att skala ut eller upp systemet genom förändringar i den fysiska arkitekturen. En bra logisk arkitektur möjliggör alltså en bra fysisk arkitektur och det är den egenskapen som har utvärderats.

En viktig ingrediens för att nå ut med det externa API som ingår i K-samsök och för att skapa intresse hos applikationsutvecklare för API:et är att olika typer av marknadsföringsaktiviteter och liknande genomförs. I den här utvärderingen har inte den marknadsföring som genomförts hittills eller som pågår nu studerats. Inte heller innehåller den här rapporten några förslag på förbättringar inom det området.

De personer som är och har varit inblandade i utvecklingen och förvaltningen av K-samsök har heller inte utvärderats. I utvärderingen av hur förvaltningsbart systemet är så berörs dock problematiken kring personberoenden.

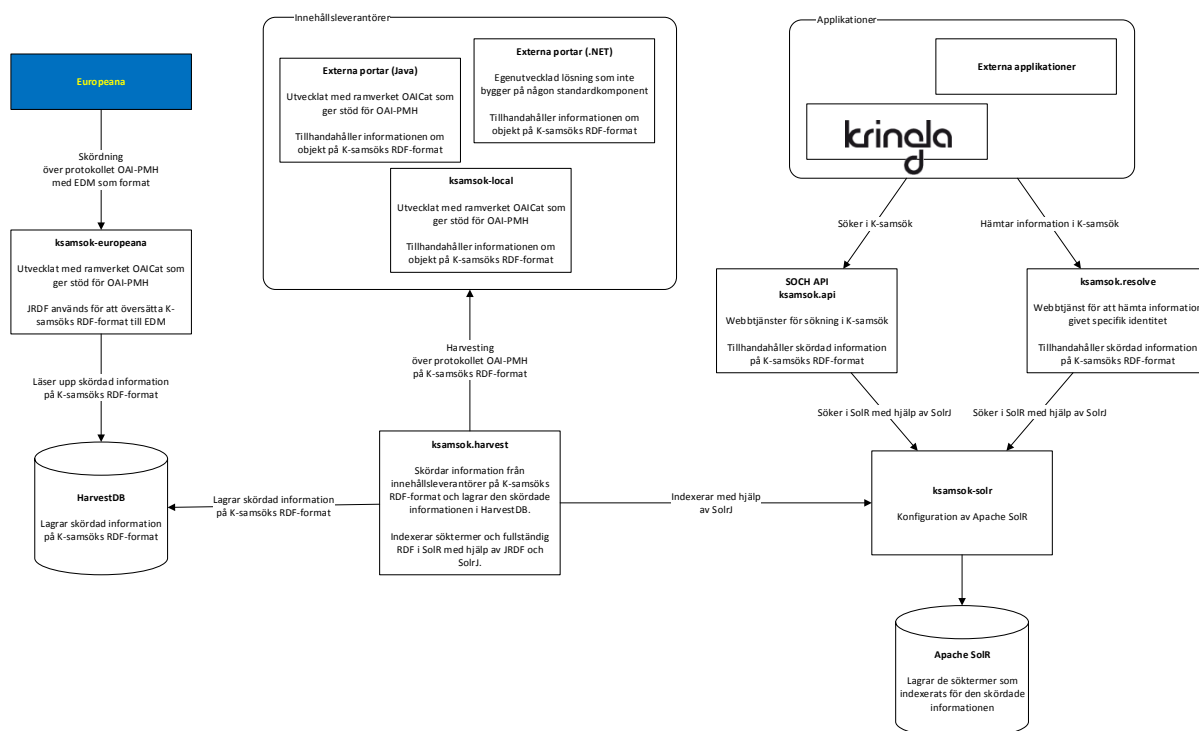
<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

3 Översikt

Systemet K-samsök har i dag fem huvudsakliga uppgifter, nämligen att:

- Skörda¹ information från innehållsleverantörerna.
- Leverera den skördade informationen till Europeana.
- Tillgängliggöra den skördade informationen till externa applikationer genom ett externt API med sök- och statistikfunktionalitet.
- Tillhandahålla den skördade informationen när de URI:er som identifierar informationen avrefereras².
- Förädla den skördade informationen med information från allmänheten och andra källor.

I översikt bilden återges de olika delar systemet som rör de fem huvuduppgifterna och deras inbördes beroenden samt deras beroenden till externa system.



Figur 1: Översikt av systemet K-samsök.

¹ Skörda är en, i det här sammanhanget, vedertagen och lämplig term för att samla in och aggregera information från en eller flera parter. Den kommer därför användas genomgående i den här rapporten och synonymer undvikas för att undgå osäkerhet om det är samma aktivitet som åsyftas.

² I den här rapporten används den inte helt etablerade översättningen avreferera av det engelska ordet *dereference*. Termen används uteslutande för att beskriva handlingen att tillhandahålla informationen för en viss URI.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

I bilden saknas UGC-hubben som är en självständig systemdel som saknar kopplingar till de andra delarna av K-samsök.

De fem systemdelar som har ingått i utvärderingen och deras syften är:

- **ksamsök:** Innehåller funktionalitet för att skörda information från innehållsleverantörerna på K-samsöks egna RDF-format över skörde-protokollet OAI-PMH, ett sök-API som tillgängliggör den skördade informationen på det egna RDF-formatet genom en mängd sök- och statistisk-webbtjänster samt en tjänst som tillhandahåller den skördade informationen givet specifika identiteter.
- **ksamsök-europeana:** Översätter den skördade informationen till formatet EDM och levererar detta till Europeana över skörde-protokollet OAI-PMH.
- **ksamsök-local:** Tillhandahåller information från Riksantikvarieämbetets egna system till K-samsök på K-samsöks egna RDF-format över skörde-protokollet OAI-PMH.
- **ksamsök-solr:** en konfiguration av indexeringsmotorn Apache SolR som indexerar den skördade informationen och tillgängliggör den genom sök- och statistik-funktioner.
- **UGC-hubben:** ett API för att lagra och söka fram användargenererat innehåll om objekt genom webbtjänster.

Även om systemet innehåller fler delar så täcker de fem delarna ovan in samtliga av de huvudsakliga uppgifterna systemet har i dag.

Utöver dessa delar är det värt att nämna de övriga lokala portar som levererar innehållsleverantörernas information till K-samsök ansvarar innehållsleverantörerna själva för. Riksantikvarieämbetet tillhandahåller två grundkodpaket, ett för Java-miljö och ett för Microsoft .NET-miljö. De går att anpassa för olika museisystem så att de kan leverera information till K-samsök. I dag finns det portar för museisystemen Primus, Carlotta och MuseumPlus³.

En genomgripande princip som ligger till grund för viktiga delar i designen av systemet är att K-samsök inte modifierar den information som skördas från innehållsleverantörerna. Det är således samma RDF-beskrivning som används i alla systemdelar utom UGC-hubben.

³ En port för CollectiveAccess är på gång men tidsplanen är oklar.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

4 Teknisk lösning

Den tekniska lösningen har utvärderats främst ur perspektivet av hur väl systemet klarar förändringar i vad systemet ska förmå, hur flyttbart systemet är och dessutom hur väl det är implementerat.

Av de olika systemdelarna är det egentligen enbart ksamsök och UGC-hubben som har egna arkitekturer. Systemdelarna ksamsök-local och ksamsök-europeana är utvecklade med ramverket OAICat och följer det ramverkets arkitektur och ksamsök-solr är en konfiguration av Apache SolR som har sin egen arkitektur.

4.1 Webbapplikationer

Såväl ksamsök som UGC-hubben är implementerade som webbapplikationer vars gränssnitt består av så kallade servletter⁴. En sådan webbapplikation installeras i en servlet-motor som tar hand om kommunikationen till servlettarna och även tillhandahåller en mängd annan standardfunktionalitet. Alla servlet-motorer som följer Javas specifikationer kan användas som värd för webbapplikationer av den här typen och i K-samsöks fall har valet fallit på servlet-motorn Apache Tomcat. Andra möjliga val vore till exempel Jetty och Oracle WebLogic Server.

Den arkitektur som är vald för webbapplikationerna ksamsök och UGC-hubben är ändamålsenlig och tillåter att systemdelarna skalas ut⁵ och skalas upp⁶ genom förändringar i den fysiska arkitekturen. Genom att utveckla webbapplikationerna enligt Javas servlet-specifikation undviks beroenden till en specifik programvara och därigenom blir personberoendet mindre. Arkitekturen möjliggör även att webbapplikationerna utökas med exempelvis fristående behörighetsstyrning, loggning, med mera som följer samma specifikationer.

4.2 Databaser

Den skördade informationen lagras i en PostgreSQL-databas och även UGC-hubben nyttjar en PostgreSQL-databas för lagringen av förädlade data. PostgreSQL är en öppen-källkod-databas som har nått en hög mognad och är ett adekvat val för den här typen av system. Communityn runt PostgreSQL är levande och uppdateringar släpps regelbundet.

Att byta databas låter sig inte göras lika enkelt som ett byte av servlet-motor och eventuella negativa sidor hos den databas som är vald måste hanteras. I fallet med PostgreSQL är den mest negativa sidan att den har en, i förhållande till alternativen, relativt liten marknadsandel. Det har givetvis konsekvenser för hur väl spridd erfarenhet av och kunskap om databasen är. För utvecklingen och förvaltningen av K-samsök är det förmodligen ett relativt litet problem eftersom systemet enbart använder standardfunktionalitet hos databasen och inte sådan

⁴ Även OAICat-ramverket och därigenom ksamsök-local och ksamsök-europeana är implementerade på det sättet.

⁵ Att *skala ut* en del av ett system innebär att flera instanser av systemdelen installeras på olika maskiner.

⁶ Att *skala upp* en del av ett system innebär att miljön som systemdelen är installerad på förstärks med exempelvis ökad processorkraft eller ökat arbetsminne.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

funktionalitet som enbart återfinns hos PostgreSQL. Vad det gäller driften av databasen kan problemen eventuellt vara större när databasen behöver optimeras, felsökas eller administreras.

All access av databaserna för den skördade och förädlade informationen sker med hjälp av Java Database Connectivity (JDBC) och utan stöd av något ORM-ramverk som Java Persistence API (JPA), Hibernate eller liknande. Object-Relational Mapping (ORM) är en teknik för att översätta mellan de objektmodeller som används i Java och den relationella databasmodell som informationen i databasen är lagrad i.

Genom att använda ett ORM-ramverk kan man arbeta mer oberoende av den specifika databasmotor som lagrar informationen i systemet och implementationen av systemet kan dra nytta av styrkor hos programspråket och utvecklingsmiljön som inte finns hos databasen. Det omvända gäller, så klart, att vissa styrkor hos databasen och databasspråket SQL skulle gå förlorade i en ORM-mappning. Det bör även tilläggas att medan ORM-ramverk har kommit och gått med jämna mellanrum har tekniker som JDBC och SQL levt i 15 respektive 40 år.

I K-samsök är lagringsmodellerna väldigt små och behovet av ett ORM-ramverk är ganska litet i dagsläget om inte, till och med, direkt kontraproduktivt men om lagringsmodellerna skulle utökas kan det vara lämpligt att omvärdera nyttan med att gå över till en ORM-lösning.

Noterbart i accessen av databaserna är att samma databas användare används i alla sammanhang. Till exempel ansluter det externa API:et till databasen med samma databas användare som skörde-funktionaliteten. Det kan vara en svaghet ur säkerhetssynpunkt eftersom en databas användare som används av en externt exponerad del av systemet måste ha behörigheter som ger databas användaren rätt att skapa och ta bort information i databasen för att skörde-funktionaliteten ska fungera.

4.3 Kod

All kod i de utvärderade systemdelarna är implementerad med hjälp av utvecklingsramverket och programspråket Java och för närvarande används version 1.6 av Java.

Koden är föredömligt strukturerad och dokumenterad och genomgående används kända och lämpliga designmönster. Det finns även en hög grad av enhetlighet i hur koden implementerats och beroendena mellan olika delar i koden har hållits på en låg nivå. Allt detta tillsammans gör koden lättfattlig och förvaltningsbar.

Generellt finns det dock väldigt få enhetstest eller andra automatiserade test. Det kan göra koden svårare att underhålla, särskilt för nya utvecklare, då det kan vara svårt att avgöra om ett tillägg till eller förändring av koden påverkar någon annan del av systemet än den som utvecklaren arbetar med. Med hjälp av automatiserade tester kan den typen av så kallade regressionstester genomföras ögonblickligen och utan förkunskaper om hela systemet.

Även om det enbart rör en mycket liten del av koden är det även värt att nämna att den XML och JSON som tillhandahålls av API:et för UGC-hubben skapas som ren text. Det är oftast att föredra att använda specifika formatterare när data på ett visst format ska skapas för att inte behöva hantera alla olika egenheter och potentiella problem hos formatet. I det här fallet är det

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

förmodligen än viktigare då formatet JSON är sårbart för så kallade injektionsattacker när det genereras och behandlas utan korrekt hantering av specialtecken.

4.4 Beroenden

K-samsök nyttjar externa komponenter flitigt vilket är en stor styrka hos systemet då det minskar beroendet till de personer som utvecklade systemet. Nackdelen med beroenden till externa komponenter är att de riskerar att sluta underhållas eller förändras på ett sätt som gör dem obrukbara.

I K-samsöks fall har valet av externa komponenter oftast hamnat på välkända och stabila komponenter. Som tidigare nämnt används OAICat som lösning för systemdelarna ksamsok-local och ksamsok-europeana och Apache SolR för systemdelen ksamsok-solr. Vidare används följande komponenter för att hantera specifika uppgifter hos systemdelarna:

- **Apache Commons:** Hjälpfunktioner för hantering av bland annat strängar.
- **Apache log4j:** Loggningsfunktionalitet.
- **CQL-Java:** Hantering av frågespråket CQL.
- **Java API for XML Processing:** Tolkning av XML-dokument.
- **Joda Time:** Förbättrad datum- och tidhantering för Java.
- **JRDF:** Funktionalitet för att arbeta med RDF-dokument.
- **JUnit:** Funktionalitet för att implementera enhetstester.
- **OAIHarvester2:** Funktionalitet för att skörda information från en OAI-PMH-tjänst.
- **Quartz Enterprise Job Scheduler:** Schemaläggning av skörde-jobb och indexerings-jobb.
- **SLF4J:** Abstraktionslager för olika typer av loggning, däribland Apache log4j.
- **Solrj:** Funktionalitet för att interagera med en Apache SolR-server.
- **Spring:** Ramverk för utveckling av Java-applikationer som används i K-samsök för att binda ihop olika systemdelar konfigurationsmässigt i stället för programmatiskt.

Samtliga av dessa komponenter är mogna och underhålls fortlöpande än i dag. Undantaget är komponenten JRDF vilken var ett relativt litet öppen-källkodsprojekt som blev officiellt inaktivt den 8 maj 2011. Ett lämpligt alternativ till JRDF vore kanske Jena som är ett Apache-projekt.

4.5 Öppen källkod

Samtliga av komponenterna ovan är licensierade som öppen källkod och de licenser som används tillåter att komponenterna integreras i ett annat system utan att det systemet måste licensieras som öppen källkod. K-samsök har inte heller några andra beroenden till proprietär programvara.

Det innebär att varje del av K-samsök eller systemet som helhet kan licensieras under godtycklig licens. Med andra ord står det Riksantikvarieämbetet fritt att välja om licensformen ska vara proprietär eller öppen källkod och om den är öppen källkod även välja huruvida det ska finnas ett krav på att härledda verk också måste vara licensierade med samma licens.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

Vilka delar av K-samsök som vore intressanta att licensiera som öppen källkod är en flerfacetterad fråga. Om syftet är att skapa en community kring en del av systemet som förvaltar och vidareutvecklar koden är det viktigt att det finns ett stort intresse av funktionaliteten och att funktionaliteten kan ge ett mervärde i något annat sammanhang. I dag är det svårt att se att det finns någon del som innehåller sådan funktionalitet. Mycket beror det på att det är verksamhetsspecifik information och ett egendefinierat format som behandlas i systemet.

I det fallet att systemet utvidgades till att hantera standardformat vore det dock tänkbart att det skulle finnas en större chans att lyckas skapa en aktiv community kring de delarna. Det vore även en möjlighet att flytta delar av utvecklingen till existerande öppen-källkods-projekt så att exempelvis ksamsok-europeana skulle kunna utvecklas som ett tillägg till OAICat.

Om syftet med att licensiera delar av K-samsök som öppen källkod är att sprida kunskap och stötta externa applikationsutvecklare kunde ett effektivt sätt vara att licensiera Kringla eller delar av Kringla som öppen källkod. Den funktionalitet som finns i Kringla är förmodligen till större hjälp för en utvecklare som vill utveckla en mobilapplikation eller webbapplikation som integrerar sig med K-samsök än vad de systemdelar som utgör själva K-samsök är.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

5 Information och tjänster

I K-samsök används huvudsakligen en och samma informationsmodell genomgående i systemet, nämligen det egna RDF-formatet. När den skördade informationen ska levereras till Europeana översätts den till antingen Europeana Data Model (EDM) eller CARARE metadata schema men i övrigt är det enbart det egna RDF-formatet som används.

Den RDF-beskrivning som skapas hos innehållsleverantören i någon av de lokala portarna och som skördas av K-samsök är även den som lagras i systemet och som tillhandahålls genom det externa API:et och av tjänsterna för att hämta skördad information givet en specifik identitet.

Att det är samma informationsmodell genomgående i systemet är, som nämnt tidigare, ett medvetet beslut där utgångspunkten är att innehållsleverantören äger och kontrollerar sin egen information och K-samsök enbart samlar in den för vidare spridning. Det är en vettig princip och som tydliggör ägandet av informationen och även ansvarsförhållandena för informationens kvalitet. Samtidigt finns det dock en del negativa konsekvenser av att nyttja samma informationsmodell genomgående och dessa måste beaktas.

I de fall som informationsmodellen fungerar som överföringsformat i integrationen mellan två system, till exempel i fallet med innehållsleverantörernas lokala portar och K-samsöks skördefunktionalitet, utgör modellen ett beroende mellan systemen. Ett sådant beroende medför att varje förändring av modellen innebär att båda systemen måste uppdateras och om samma modell används för flera integrationer att varje förändring innebär att alla de integrerade systemen måste uppdateras.

Så länge förändringarna i informationsmodellen syftar till att korrigera eller förbättra något som är av intresse för samtliga integrerade system som kommer bli tvungna att uppdateras är det inget bekymmer men då förändringsbehovet enbart finns hos något enskilt system kommer samtliga system ändå att behöva uppdateras utan att något mervärde erhålls för dem. I det fall att K-samsök skulle få till uppgift att lagra annan information än i dag kan de förändringar som måste ske bli onödigt stora och påverka existerande innehållsleverantörer negativt.

I tjänsteorienterade system som K-samsök brukar man vanligtvis sträva efter ett antal principer:

- Tjänsterna har tydliga gränser mellan varandra.
- Tjänsterna är självständiga.
- Tjänsterna delar schema och kontrakt men inte klasser.
- Kompatibilitet mellan tjänster baseras på policy.

Dessa principer syftar bland annat till att lösa just den typ av problem med förändringar som sprider sig till flera integrerade system. Särskilt den tredje principen om att tjänster delar schema och kontrakt men inte klasser grundar sig på att man i varje integration av två tjänster i möjligaste mån aldrig vill förändra protokoll eller format i deras interaktion med varandra.

Önskvärt är alltså att för varje integration specificera ett protokoll och ett format som bäst löser den uppgift integration syftar till att lösa och att det protokollet och det formatet sedan hålls

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

stabil så länge uppgiften förblir densamma. I ett system med flera integrationer så blir även konsekvensen att varje integration bör hanteras var för sig om de inte löser samma uppgift. Det innebär att det kan bli flera protokoll och flera format om det krävs för att lösa uppgifterna på bästa sätt.

Även om det ofta kan vara enklast att skraddarsy ett protokoll eller ett format för en viss integration är ett bra alternativ att förlita sig på ett standardiserat protokoll eller format. Även standardiserade protokoll och format uppfyller oftast de behov av stabilitet som en integration kräver och har fördelen av att oftast vara väl dokumenterade och kända.

Genom att se över K-samsöks huvudsakliga uppgifter var för sig framträder en möjlighet att förenkla och förbättra vissa tjänster om rätt protokoll och format används.

5.1 Skörda

Vid skördandet av information från innehållsleverantörerna används i dag skördeprotokollet OAI-PMH vilket är en specifikation framtagen av Open Archives för just syftet att skörda information. Genom att ett spritt och välkänt protokoll används vid skördandet så finns det även färdiga komponenter att stödja sig på. Det är något K-samsök har dragit nytta av och OAIHarvester2 används för att skörda information och OAICat används i såväl Riksantikvarieämbetets egen port ksamsok-local som i de lokala Java-portar som används hos flertalet innehållsleverantörer. Dessutom används samma protokoll i leveransen av information till Europeana vilket beskrivs i nästa avsnitt.

Informationen som skördas är på K-samsöks RDF-format och för att skapa sådana RDF-beskrivningar innehåller de lokala portarna översättningar och tolkningar som kompletterar funktionaliteten i OAICat. Att ta fram dessa översättningar och tolkningar har av en del innehållsleverantörer upplevts som svårt och tidskrävande och bättre stöd för detta har efterfrågats. Varje förändring av K-samsöks RDF-format riskerar även att innebära att samtliga portar behöver anpassas.

För att komma bort från beroendet mellan skördandet av information och andra uppgifter hos K-samsök som kan behöva förändra RDF-formatet och för att underlätta för innehållsleverantörerna att leverera information till K-samsök vore en möjlighet att tillåta fler överföringsformat än enbart det egna RDF-formatet.

Det kanske mest intressanta överföringsformatet i sammanhanget är Lightweight Information Describing Objects (LIDO). LIDO är en de facto standard för att överföra metadata om museiföremål och används flitigt i såväl Europeana-projekten som i andra projekt och av andra system. Bland annat finns det stöd att exportera information på LIDO-formatet i museisystemen MuseumPlus och CollectiveAccess vilka används av flera innehållsleverantörer.

Nackdelen med att tillåta leveranser av information på LIDO-formatet vore att K-samsök skulle bli tvunget att utifrån LIDO-beskrivningen skapa den RDF-beskrivning som behövs i andra delar av systemet. För att hantera det problemet skulle det vara tänkbart att den översättningen hanteras av en port som skulle kunna förvaltas gemensamt av alla innehållsleverantörer som levererar på LIDO-formatet, kanske till och med som öppen källkod.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

5.2 Leverera

Vid leveransen av information till Europeana är det inte K-samsök som valt varken protokoll eller format.

Även här används dock OAI-PMH som protokoll för skördandet. Valet av protokoll är positivt för K-samsök både ur aspekten att det är ett spritt och välkänt format för vilket det finns dokumentation och komponenter att stödja sig mot och ur aspekten att det är ett protokoll som används i andra delar av K-samsök.

De format som används vid leveranserna är Europeana Data Model (EDM) och CARARE metadata schema. EDM är framtaget för att utgöra en fullständig och välutvecklad modell för digitala kulturarvsdata vilket kan göra formatet instabilt om det behöver uppdateras ofta för att förbli fullständigt. Samtidigt är det ett flitigt använt format vilket bör innebära att Europeana strävar efter stabilitet och i de fall där det inte är möjligt erbjuder bakåtkompatibilitet.

CARARE metadata schema är skräddarsytt för uppgiften att skörda information från lokala system. I dag skapas ett interim-beskrivning av den information som ska levereras till CARARE genom en översättning av K-samsöks RDF-format. Den interim-beskrivningen översätts sedan av CARARE till deras metadata schema genom en översättning som tagits fram med hjälp av verktyget MINT. Av samma anledning som för EDM borde inte CARARE metadata schema uppdateras i onödan och eventuella förändringar av K-samsöks RDF-format kan i viss mån hanteras i översättningen till interim-beskrivningen.

Vidare stöder såväl Europeana som CARARE leveranser på andra format än de egna, däribland formatet LIDO som nämnades i förra avsnittet. I det fall att K-samsök skulle utökas med möjligheten att skörda information på LIDO-formatet så vore det eventuellt möjligt att använda de LIDO-beskrivningarna som de är vid leverans av information till Europeana. För övrig information som ska levereras till Europeana skulle det kunna vara intressant att titta på version 1.1 av OAICat-tillägget OAICatMuseum som ska innehålla stöd för LIDO.

5.3 Tillgängliggöra

För att tillgängliggöra den skördade informationen genom ett externt API med sökfunktionalitet användes initialt protokollet *Search/Retrieve via URL* (SRU) i K-samsök. Det protokollet överges dock successivt till förmån för det egna SOCH API:et. Skillnaden är inte så stor och i båda protokollen används i grunden frågespråket CQL för att ange sökvillkor. I SOCH API:et finns det dock en större möjlighet att påverka sökningarna och resultatet genom ytterligare parametrar.

Resultatet av en sökning skickas alltid tillbaka som ett XML-dokument. För statistik- och fasettsökningar är det enkla listliknande XML-format som används. I fallet med vanliga sökningar beror innehållet i XML-dokumentet på parametern *recordSchema* vilken kan sättas till RDF, presentation eller XML där följderna blir att varje post i sökresultatet representeras av:

- **RDF:** Dess fullständiga RDF-beskrivning så som den skördats från innehållsleverantören.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

- **presentation:** Enbart värdet på egenskapen presentation i RDF-beskrivningen vilket är en förenklad beskrivning som lämpar sig för presentationsammanhang.
- **XML:** Enbart de egenskaper i RDF-beskrivningen som anges i sökparametern *fields*.

Syftet med att tillhandahålla tre utdata-format för vanliga sökningar är att tillgodose de olika behov och förutsättningar som finns hos de applikationsutvecklare som integrerar sig med det externa API:et. Det är en klok ambition som sannolikt skulle kunna drivas ännu längre för att tillgodose än fler behov och förutsättningar.

Trots XML-baserade överföringsformats många förtjänster är de ofta relativt bandbreddskrävande och otympliga att arbeta med. I många utvecklingsramverk, som exempelvis .NET och Java, finns därför funktionalitet för att tolka XML-format och översätta dem till objektstrukturer som är väsentligt enklare och stabilare att utveckla med.

För en webbapplikation som ska integrera sig med K-samsöks externa API kommer integrationen med allra största sannolikhet skötas på något av följande sätt.

- Webbapplikationen skapar webbsidor innehållande skript som exekveras i webbläsaren och som anropar det externa API:et direkt.
- Webbapplikationen skapar webbsidor innehållande skript som exekveras i webbläsaren och som anropar ett server-API hos webbapplikationen. Server-API:et sköter i sin tur anropet av K-samsöks externa API.
- Webbapplikationen använder det externa API:et för att skapa de webbsidor som visas i webbläsaren.

I de två senare fallen kommer integrationen sannolikt ske med hjälp av ett utvecklingsramverk som har bra stöd för att tolka och översätta XML till objektstrukturer medan samma möjligheter inte finns när anropen sker från skript som exekveras i webbläsaren. Det är inte så att det är omöjligt att åstadkomma detta och för många utvecklare är det inte ens svårt att göra det men för många utgör det antagligen en tröskel. Att tolka XML är även något som kan vara resurskrävande och det kan vara ett bekymmer när en webbapplikation lägger över det ansvaret på webbläsaren.

I webbapplikationen Kringla har man, som exempel, valt att gå på den tredje lösningsvägen, det vill säga att webbapplikationen sköter anropen på servern för att skapa webbsidorna som visas i webbläsaren. Motiven till det kan vara flera men ett skulle kunna vara att man vill undvika att arbeta med XML i webbläsaren.

I dag är det vanligt i API:er som ska användas av webbapplikationer att resultatet kan levereras på formatet JSON. JSON är mer kompakt än XML och har även fördelen att det finns inbyggt stöd för att tolka formatet i webbläsaren och är enkelt att arbeta med i skript som exekveras av webbläsaren.

Att kunna begränsa vilka egenskaper som ska returneras är en bra möjlighet och känns motiverad oavsett vilket utdataformat som används. Det förändras inte även om stöd för JSON införs.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

Behovet av att kunna få ut en förenklad beskrivning genom valet presentation tror jag är stort hos de flesta applikationsutvecklare som vill integrera sig med det externa API:et. Även här vore förmodligen möjligheten att få ut den beskrivningen som JSON önskvärt. En av de största styrkorna i presentations-beskrivningen är att genom att den utgör en, på sätt och vis, fristående del av RDF-formatet kan det formatet hållas stabilt även när RDF-formatet behöver förändras. På så sätt kan K-samsök erbjuda ett stabilt kontrakt mot integrerade system även under faser då förändringar av RDF-formatet sker.

Ett annat bekymmer i dag med RDF-formatet är att de lokala portarna ger upphov till RDF-beskrivningar som ur ett RDF-perspektiv är identiska med som ur ett XML-perspektiv ser väldigt olika ut. Det medför att applikationsutvecklare som integrerar sig med det externa API:et måste förstå eller kunna arbeta med RDF och att enbart kunskaper i XML inte räcker till. Det skulle kunna lösas genom att RDF-beskrivningarna normaliseras så att de får beskrivningar som liknar varandra.

Dokumentationen av K-samsöks RDF-format är utförlig och bra⁷ men begränsar sig i dag till den dokumentation som är framtagen av Riksantikvarieämbetet. Det finns heller inget färdigt stöd för att arbeta med formatet. Samma sak skulle gälla även med en JSON-version av formatet. Att kunna erbjuda utdata på ett format som har större spridning skulle ge applikationsutvecklare en större erfarenhetsbank att hämta stöd ur. Formatet LIDO skulle kunna vara ett lämpligt sådant format men kanske finns det även än mer generella format som går att använda i det externa API:et.

Ett sådant format och tillika protokoll som skulle vara användbart för exempelvis fornlämningar och andra historiska platser är Open GIS Consortiums standarder Web Feature Service (WFS) och Web Map Service (WMS). Spridningen av dessa är mycket stor och de allra flesta webbkartor och kartverktyg har stöd för att visa information som tillhandahålls via WFS eller WMS. Det krävs dock att informationen har ett geografiskt läge för att WFS eller WMS ska vara användbart.

I dag erbjuder det externa API:et en mängd metoder för att söka fram skördad information och även ta fram statistik om den skördade informationen. Något som inte finns är dock möjligheten att ställa semantiska frågor. Vad som skulle krävas för att göra det möjligt och vad en sådan funktionalitet skulle ge diskuteras närmare i avsnittet *Semantiska sökningar*.

5.4 Avreferera

De URI:er som används som identiteter för den skördade informationen hänvisar till domänen kulturarvsdata.se och anger HTTP som protokoll. När en klient avrefererar en sådan URI via ett HTTP-anrop hanteras det av en servlet som ingår i K-samsök.

Protokollet som används är alltså HTTP där skördad information hämtas genom anrop med identiteten på den efterfrågade informationen som URI. K-samsök avrefererar URI:en och besvarar anropet med den RDF-beskrivning som skördats från innehållsleverantörerna.

⁷ För en mer utförlig kommentar om dokumentationen av K-samsöks RDF-format, se avsnittet *Dokumentation*.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

För den här huvuduppgiften hos K-samsök känns såväl protokoll som format som det naturliga valet. Genom att använda HTTP som protokoll blir den skördade informationens identiteter avrefererbara över webben och RDF är det format som rekommenderas för länkade öppna data.

5.5 Förädla

UGC-hubben exponerar ett servlet-baserat API genom vilket förädlande information om den skördade informationen kan skapas, förändras, tas bort, läsas och sökas fram. Indata anges som URL-parametrar eller som formulär i kroppen av ett POST-meddelande och utdata levereras på formattyperna XML eller JSON.

Såväl URL-parametrar som formattyperna XML och JSON är lätta att använda för en extern utvecklare och med adekvat dokumentation av API:et och innehållet i utdata finns inga svårigheter att integrera ett externt system med UGC-hubben.

UGC-hubben löser just den uppgift den är tänkt att lösa och använder för detta ett protokoll och ett format som löser uppgiften på ett effektivt och korrekt sätt. Lösningen är även självständig och har ett minimalt beroende till och gränsyta mot övriga delar av K-samsök. I det hänseendet är UGC-hubben ett typexempel på hur principerna för ett tjänstebaserat system kan uppfyllas.

5.6 Övrigt om K-samsöks RDF-format

Det RDF-format som används i K-samsök har även ett par generella brister som är oberoende av för vilka tjänster det används, nämligen att:

- Det saknas en formell definition av formatet (till exempel ett RDF-schema) även om det finns en "mall" tillgänglig för formatet.
- Formatet refererar till andra RDF-scheman såsom OWL, CIDOC CRM och BIO men formatet uppfyller inte deras specifikation av definitions mängd och värdemängd. Det är troligen inget problem för någon som har behov av att arbeta med just K-samsöks data men för system som länkar ihop en mängd olika datakällor och inte har någon specialhantering för K-samsök kommer inte formatet att kunna tolkas och förstås som det är tänkt.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

6 Webbsemantik

K-samsök verkar på den semantiska webben genom att tillhandahålla information från de olika innehållsleverantörerna på ett format som är semantiskt strukturerat.

6.1 Format

Formatet som används för att strukturera informationen är det egna RDF-formatet och att använda ett RDF-format i det här sammanhanget är det mest naturliga och det som vanligtvis rekommenderas.

De RDF-beskrivningar som skördas och tillgängliggörs av K-samsök skapas, som nämnt tidigare, hos de enskilda innehållsleverantörerna men alla beskrivningar följer samma RDF-format.

Initialt beskrev formatet enbart Fysiska ting men sedan version 1.1 finns även möjlighet att beskriva Agenter (personer, organisationer och grupper), Event (historiska händelser och utställningar) och Koncept (företeelser som inte är avgränsningsbara i rum och tid).

Några av de brister som finns hos formatet och hur formatet används beskrevs tidigare i avsnittet *Övrigt om K-samsöks RDF-format* och inbegrep bland annat att det saknades en formell definition och att refererade RDF-scheman inte nyttjades korrekt. En formell och korrekt definition av RDF-formatet är centralt för att informationen ska vara maskinellt tolkningsbar och användbar på den semantiska webben. Som det är i dag krävs det en hel del handpåläggning för att tolka informationen som tillgängliggörs av K-samsök.

6.2 Länkade data

K-samsök länkar inte ihop den skördade informationen med andra auktoriteter utan det är något som varje enskild innehållsleverantör måste göra. Mycket av arbetet med semantifieringen av informationen är alltså ett ansvar som faller på museisystemen och faller utanför K-samsöks ansvarsområde.

I dag så refererar den skördade informationen väldigt sällan eller aldrig till auktoriteter andra än de som tillhandahålls av K-samsök och de auktoriteter som tillhandahålls av K-samsök är inte länkade till externa auktoriteter. Det finns alltså inga länkar ut från K-samsök i dagsläget. Trots att det inte är K-samsöks ansvar att skapa dessa länkar skulle ett sådant arbete förmodligen kunna drivas eller stöttas från K-samsöks håll.

Avsaknaden av en formell specifikation av RDF-formatet i K-samsök försvårar även länkningen med andra datakällor genom att kopplingen mellan begrepp i K-samsöks RDF-format och andra format och ontologier blir oklar. Det innebär att det är svårt för externa datakällor att förstå hur deras information länkar in i den information som finns i K-samsök.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

6.3 Identiteter

De URI:er som används som identiteter för informationen i K-samsök skapas av innehållsleverantörerna. De byggs oftast upp genom ett kombinerat protokollet HTTP, domänen kulturarvsdata.se samt en sökväg som är unik för auktoriteten med formatet institution/tjänst/id. Institutionen pekar oftast ut vilken innehållsleverantör informationen är skördad från, tjänsten vilket system informationen är hämtad ur och id den identitet informationen har i det systemet hos den institutionen.

Ett problem med identiteter som innehåller information om det de identifierar är att om den informationen förändras blir identiteten antingen missvisande eller behöver bytas. Inget av detta är önskvärt generellt och i synnerhet inte på den semantiska webben.

I de flesta av portarna är det den interna identiteten i innehållsleverantörernas system som oftast valts som identitet i sökvägen för URI:en. Det medför att identiteten förändras om en innehållsleverantör väljer att byta system.

För att hantera stabila identiteter i distribuerade system väljer man ofta att använda globalt unika identiteter eller en centralt styrd löpnummerserie. En centralt styrd löpnummerserie känns dock inte applicerbart i det här fallet.

En globalt unik identitet skulle exempelvis kunna vara en GUID eller UUID men oavsett lösning måste identiteten vara en egenskap hos objekten i innehållsleverantörernas system och inte en identitet för att möjliggöra ett byte av system eller proveniens. Om det är görligt beror på de olika museisystemen och behöver undersökas.

6.4 Inferens

En möjlighet som uppstår genom att semantifiera information är att det går att härleda samband som inte finns uttryckligen i informationen. I K-samsök sker i dag enbart en typ av inferens och det är att enkelriktade relationer mellan två objekt i K-samsök utvidgas till att bli dubbelriktade, det vill säga relationen görs även tillgänglig som en egenskap hos det refererade objektet och inte enbart hos det refererande objektet.

Genom ontologier som CIDOC CRM, SKOS, FOAF med flera skulle en mängd relationer och klassifikationer kunna härledas genom inferens. Det skulle även vara möjligt att definiera egna regler och ontologier för att härleda ytterligare relationer.

6.5 Semantiska sökningar

Ytterligare en möjlighet som uppstår genom att information är semantifierad är att den då går att lagra i system som kan erbjuda effektiva semantiska sökningar. De flesta sådana system utgår från lagring av RDF-beskrivningar i en så kallad *triple store* och stöder sökningar med hjälp av frågespråket SPARQL.

Genom att lagra den skördade information från innehållsleverantörerna i ett sådant system skulle K-samsök kunna erbjuda semantisk sökfunktionalitet i det externa API:et.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

Som triple store finns det en mängd lösningar att välja bland såsom Apache Jena och Sesame. Det finns även ett tillägg till Apache SolR som heter Semantic Information Retrieval Engine (SIREn) och som utökar funktionaliteten i Apache SolR med stöd för semantiska sökningar.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

7 Dokumentation

K-samsök är dokumenterat för såväl externa som interna intressenter. Till de externa intressenterna hör de som ska leverera information till K-samsök eller nyttja det externa API:et och till de interna hör främst de som ska förvalta systemet.

7.1 Extern

Det externa API:et och K-samsök generellt är väl dokumenterat på webbplatsen www.ksamsok.se. Den mesta dokumentationen är dessutom tillgänglig direkt på webbplatsen och drar fördel av länkar både inom webbplatsen och till andra webbplatser.

Däremot är inte informationen helt enkel att överblicka och den kan ibland vara utspridd på flera sidor, särskilt den information som rör applikationsutveckling. Mycket skulle troligen vara vunnet om sidan <http://www.ksamsok.se/om-k-samsok/att-utveckla/> utökades med en kort beskrivning av vilken information en utvecklare behöver ta till sig för att kunna utveckla en applikation. I dag finns en länk till information om det externa API:et men även en länk till beskrivningen av RDF-formatet vore på sin plats.

Beroende på vilka man vill nå ut med API:et till kan det också vara nödvändigt att tydliggöra vilken information som faktiskt finns tillgänglig genom K-samsök. För en runstensintresserad utvecklare som vill integrera möjligheten att söka fram runstenar i sin applikation är det kanske inte en självklarhet att den informationen finns att tillgå genom K-samsöks API.

Ett förslag som diskuterades på K-samsöksmötet i december var att applikationsutvecklare som kontaktar K-samsök för att få en API-nyckel även skulle få kontaktuppgifter till personer som arbetat med det externa API:et eller som kan hjälpa till och stötta på andra sätt.

7.2 Intern

I den här utvärderingen har inte en fullständig genomgång av K-samsöks dokumentation för förvaltning, drift och andra interna behov genomförts. Däremot så har ett antal sådana dokument studerats för att förstå K-samsök tillräckligt för att kunna utvärdera exempelvis systemarkitekturen.

Upplevelsen av dokumentationen har då varit att den är informativ och utförlig ur ett tekniskt perspektiv och på en detaljerad nivå. I arbetet med utvärderingen hade dock en tydlig skiss över hur systemet hänger ihop på en mer detaljerad logisk nivå och hur informationen flödar genom systemet varit ovärderlig och eventuellt finns redan en sådan dokumentation att tillgå.

<i>Utställare</i> Johan Behrenfeldt	<i>Utg.</i> PA3	<i>Datum</i> 2013-01-25
<i>Referens</i> Teknisk Review av K-samsök – Rapport	<i>Filnamn</i> Teknisk Review Rapport PA3.docx	

8 Slutsatser

Utvärderingen har syftat till att besvara frågor rörande vilka styrkor och svagheter systemet har sett till dess tekniska lösning, de uppgifter systemet är tänkt att lösa, dess position på den semantiska webben samt den dokumentation som finns av systemet.

Den samlade bedömningen av K-samsöks tekniska lösning är att systemet är utvecklat med en adekvat och skalbar arkitektur, att koden är förvaltningsbar och flyttbar och att lösningen smälter väl in i öppen-källkods-samfundet om delar av systemet skulle komma att licensieras som öppen källkod.

Det fåtal problem som finns med den tekniska lösningen–två möjliga avvikelser och en osupporterad tredjepartskomponent–ryms inom normal förvaltning av systemet.

Vad det gäller de uppgifter K-samsök är framtaget för att lösa är resultatet av utvärderingen att det finns ett behov av att se över vilka format som accepteras och används för de olika uppgifterna.

I dag sker de flesta av integrationerna med hjälp av K-samsöks egna RDF-format och det gör formatet väldigt kritiskt för systemet. Varje förändring av någon integration i systemet riskerar påverka alla andra integrationer även när de integrationerna inte är beroende av förändringen.

Det är sannolikt även önskvärt att kunna nå ut med informationen i K-samsök på mer spridda format för att nå en större målgrupp. Att se över främst utdataformaten i det externa API:et bedöms vara viktigt för att öka nyttjande av API:et.

Som en aktör på den semantiska webben är K-samsök och innehållsleverantörerna tidigt ute med att semantifiera information och tillgängliggöra informationen öppet.

Det finns dock ett behov av att formalisera det egna RDF-formatet för att tydliggöra de kopplingar som finns till format och data utanför K-samsök. Vidare bör K-samsök arbeta för att innehållsleverantörerna ska länka sin information till fler auktoriteter och externa datakällor om så är möjligt.

Hanteringen av identiteter i K-samsök behöver även ses över för att möjliggöra byte av museisystem hos innehållsleverantörerna och byte av proveniens hos föremålen.

Den information som skördas av K-samsök i dag bearbetas enbart för att möjliggöra textuella sökningar. Genom att informationen är strukturerad som RDF finns möjligheten att göra mycket mer med informationen som att inferera samband som inte finns där uttryckligen och att möjliggöra semantiska sökningar i det externa API:et.

Slutligen har utvärderingen kommit fram till att dokumentationen av systemet är utförlig och förståelig. Såväl den interna som den externa dokumentationen skulle dock kunna kompletteras med information som riktar sig mot personer som har inte har så stor verksamhetskunskap.