

## Underlag för utvecklingsplan för K-samsök

Författare: Marie Gustafsson Friberger, Good Measure  
(med bilaga av Matthias Palmér, Metasolutions)

Mottagare: Matilda Karlsson, Riksantikvarieämbetet

Licens: Creative Commons BY-SA 4.0

15 september, 2014 (Version 2)

### **SAMMANFATTNING**

Rapporten presenterar ett underlag för utvecklingsplan för K-samsök, som tar upp både tekniska och icke-tekniska aspekter, de förstnämnda främst kopplade till länkade öppna data. Rapporten grundar sig på dokument samt intervjuer med personer internt på Riksantikvarieämbetet och externt hos innehållsleverantörer, relaterade organisationer och Europeana. Ett antal möjliga roller för K-samsök beskrivs: aggregator, samordnare eller ägare till auktoritetslistor, mötesplats för innehållsleverantörer, stöd för utvecklare samt skapare av kontaktytor mot allmänheten. Vägval på kort (1-2 år) och lång (3-5 år) beskrivs, både på ett nationellt plan och i förhållande till Europeana.

## Innehållsförteckning

<b>1 Inledning</b>	<b>3</b>
1.1 Rapportöversikt	3
<b>2 Bakgrund</b>	<b>3</b>
2.1 Länkade data	3
2.2 K-samsök idag	4
<b>3 K-samsöks roll och funktion i en framtida digital kulturarvsinfrastruktur</b>	<b>5</b>
3.1 Behövs en nationell aggregator?	6
3.2 Roll 1: Aggregator	7
3.3 Roll 2: Auktoritetslistor – samordnare och/eller ägare	8
3.4 Roll 3: Mötesplats för innehållsleverantörerna	9
3.5 Roll 4: Stöd för utvecklare	9
3.6 Roll 5: Skapare av kontaktytor för allmänheten	9
3.7 Vägval och förslag på kort och lång sikt	10
3.7.1 Förslag på kort sikt	10
3.7.2 Analys och förslag på lång sikt	11
<b>4 K-samsök i förhållande till Europeana</b>	<b>11</b>
4.1 På kort sikt	12
4.2 På längre sikt	13
<b>5 Summering av vägval</b>	<b>13</b>
<b>Bilaga 1: Intervjupersoner</b>	<b>14</b>
<b>Bilaga 2: MetaSolutions view on a linked data architecture for GLAM – applied to the Swedish Open Cultural Heritage project K-Samsök (author: Matthias Palmér)</b>	<b>15</b>

# 1 Inledning

K-samsök är en aggregator som samlar och tillgängliggör kulturarv från svenska museer och andra minnesinstitutioner, i dagsläget 45 institutioner. K-samsök har sitt ursprung i Museiutredningen 2007–2009, med motivet att skapa en samsökstjänst och samordna museisystem. Riksantikvarieämbetet (RAÄ) tog 2009 ansvaret att utveckla och förvalta denna lösning. Ungefär samtidigt lanseras Europeana, och K-samsök levererar data från sina institutioner vidare till det europeiska aggregatet.

K-samsök har varit en föregångare vad gäller länkade öppna data (LOD) i Sverige. Uppdraget som en nationell nod passar mycket väl med fördelarna med länkade data: att kunna länka samman spridda resurser med hjälp av gemensamma vokabulär och kopplingar mellan resurs-URI:er. Samtidigt är K-samsök i dagsläget endast delvis en plattform för länkade data, och på senare tid syns vissa brister i K-samsöks protokoll och funktionella struktur.

Syftet med denna rapport är att lämna ett förslag på:

- K-samsöks roll och funktion i en framtida digital kulturarvsinfrastruktur och som central nod för kulturmiljövården i Sverige, på kort sikt (1-2 år) och lång sikt (3-5 år).
- K-samsöks roll på kort och lång sikt i förhållande till Europeana

Rapporten baserar sig på dokument samt på intervjuer med personer både internt (på Riksantikvarieämbetet) och externt (museisystems- eller LOD-kunskap). Bilaga A listar intervjuade personer.

## 1.1 Rapportöversikt

Rapporten inleds med bakgrund kring länkade data och K-samsöks nuläge. Efter det presenteras möjliga roller och funktioner för K-samsök, samt vägval på kort (1–2 år) och lång (3–5 år) sikt. Sedan analyseras K-samsök i relation till Europeana, även här på kort och lång sikt. Slutligen läggs det nationella och Europeiska perspektiven samman i summerande vägval.

# 2 Bakgrund

Här beskrivs bakgrund för rapporten kring länkade data och K-samsöks nuläge. Syftet med Avsnitt 2.1 är inte att ge någon uttömmande beskrivning av länkade data<sup>1</sup>, snarare att förklara några grundläggande begrepp som återkommer i rapporten.

## 2.1 Länkade data

Länkade data är ett sätt att dela och organisera information för att kunna koppla samma informationsmängder från flera organisationer. Tankesättet baserar sig på att beskriva saker (resurser) snarare än dokument, och att dessa beskrivningar blir maskinläsbara. Genom att länka mellan resurser och dataset kopplar man samman det som tidigare bara varit sökbart separat. W3C har presenterat ett antal principer<sup>2</sup> för länkade data:

1. Använd URI:er för att namnge saker
2. Använd HTTP URI:er så att dessa namn kan användas för att få mer information

---

<sup>1</sup> På <http://www.linked-data.se/> finns en introduktion till länkade data på svenska.

<sup>2</sup> <http://www.w3.org/DesignIssues/LinkedData.html>

3. När någon kollar en URI, ge användbar information genom standarder (RDF\*, SPARQL)

4. Inkludera länkar till andra URI:er så att fler saker kan upptäckas

Resource Description Framework (RDF) beskriver ting genom subjekt-predikat-objekt triplar. SPARQL är ett frågespråk byggd på detta.

Ofta vill man kunna söka på specifika ting, snarare än enbart textsträngar. Genom URI-användning för att identifiera ting, kan man tydligare koppla resurser vilket möjliggör att man kan söka på exempelvis Domkyrkan i Uppsala och få information kopplad till denna byggnad på ett sammanhållet vis. Effekten kommer inte av *ett* dataset tillgängliggjort som länkade data, utan när information från flera institutioner kan länkas samman.

Om olika dataset använder olika URI:er för att namnge samma sak, kan man specificera det med speciell typ av länk: sameAs. För att slippa att använda sameAs-länkar i onödan kommer olika datakällor ofta överens om att använda specifika thesaurus eller auktoritetslistor, ofta framtagna i formatet SKOS<sup>3</sup>.

För systemutvecklare som vill bygga applikationer som tillhandahåller eller använder länkade data finns bl a kodbibliotek, triple stores (som lagrar triplar) och SPARQL-endpoints.

## 2.2 K-samsök idag

Flera intervjupersoner nämner att K-samsöks roll idag, utöver att vara en aggregator, är något vagt definierad. Nedan sammanfattas aspekter av hur K-samsök fungerar idag som kommit fram i intervjuer, som delvis informerar roller och vägval som tas upp i nästa avsnitt.

Om man ser till K-samsök som **system** och till dess drift, upplevs det som stabilt och driftsäkert (vilket även tas upp i en teknisk review<sup>4</sup> från 2013). Om man ska se systemet ur perspektivet länkade data, förefaller det vara en hybridlösning: en RDF-lösning i stor utsträckning byggd på traditionella tekniker som relationsdatabaser. Det finns ingen triple store, vilket innebär att man får bygga nya tabeller varje gång en ny relation läggs till. Vidare lagras alla länkar mellan objekt i K-samsök separat, så att objekten inte känner till dem, eftersom K-samsök inte äger objekten.

En aspekt av K-samsök är dess **protokoll**, där man kan se vissa brister. Ett exempel på detta är att mycket information finns ostrukturerad i ett presentationsblock, vilket gör att den kan vara svår att använda för andra syften (när man inte enbart vill visa en presentationstext). Det har tagits fram ett nytt protokoll med stöd för LOD, där exempelvis kopplingar till händelser och personer (agenter) kan anges. Dock är det än så länge endast en institution som levererar på detta format. En viktig aspekt kring protokolländringar kräver uppdateringar från portleverantören och kan således vara en kostnadsfråga.

Flera intervjupersoner tar upp problem med **datakvalitet** och **ägarbesiddelse**. Dataleveransen upplevs som spretig, delvis på grund av att det inte finns någon överenskommelse kring master data (som är ett sätt att se auktoritetslistor) och för att mycket data inte är strukturerad, vilket gör att man måste matcha i fritext. Detta kopplas till frågan om ägarbesiddelse, och att en konsekvens av skördningen är att data frikopplas från förvaltaren. Den varierande datakvaliteten beror också på underliggande datainsamling och museisystem.

---

<sup>3</sup> Simple Knowledge Organization System: <http://www.w3.org/2004/02/skos/>

<sup>4</sup> <http://www.ksamsok.se/wp-content/uploads/2013/02/Teknisk-Review-Rapport-PA3.pdf>

Utöver rollen som nationell aggregator kan K-samsök även sägas ha rollen som **plattform** för kunskapsutbyte mellan museerna, särskilt de som jobbar med system och digital samlingsförvaltning, där olika museisystem möts. K-samsök lyfter frågor kring exempelvis licenser och vokabulär. Intervjupersonerna upplever att det fortfarande finns variation mellan museer i informationshantering, men att det är bättre än innan K-samsök och mindre personberoende.

För utvecklare som vill använda data från K-samsök kan ett **API** användas, som kan ge svar i XML, RDF och ett presentationsformat. Det finns en hel del dokumentation kring API:et, men den är svår att överblicka. Det finns tydliga ingångar för utvecklare och innehållsleverantörer längst upp på sidan, men det är inte tydligt hur detta mappar till huvudmenyn, t ex om utvecklare har nytta av materialet under ”Resurser och protokoll”.

Det väsentliga värdet av K-samsök kommer när informationen används, vare sig det är av experter eller allmänheten (även om där finns en spänning i att tillgodose behoven för dessa ytterligheter av användare). Kringla, en samsökstjänst framtagen av Riksantikvarieämbetet som bl a använder K-samsök, är en del i att visa upp det material som finns i K-samsök. Några museer har byggt egna webb- och app-lösningar baserat på K-samsök. Ett antal appar som bygger på data från K-samsök har skapats av tredje part, exempelvis Fornfynd. Dessa visas upp på K-samsöks blogg, allra helst genom gästlägg.

Det upplevs att K-samsök inte utnyttjas till sin fulla potential vad gäller användning av det insamlade materialet. En del av detta handlar om outreach-arbete, men en del om att relevant information saknas för många föremål: geografisk information, kunna placera på tidslinje, knyta ihop resurser kring ex en person (auktoritet).

Flera intervjupersoner tar upp att K-samsök har haft en viktig roll i att **driva utvecklingen framåt** (vad gäller tillhandahållande och likagörande av information) och att utvecklingen på många vis gått fort. En möjlig syn är att K-samsök ska ses som ”proof of concept”, snarare än en slutgiltig lösning. K-Samsök kan då ses som ett läroprojekt, både för institutioner och K-samsök som organisation.

### 3 K-samsöks roll och funktion i en framtida digital kulturarvsinfrastruktur

En aggregator bör utgå från att nyttan med data uppstår när den används – man behöver således utgå från användarbehov. En utmaning med K-samsök är att det finns ett flertal avnämare, vars behov bara i viss mån överlappar: domänexperter, utvecklare och allmänhet (som i sig är en heterogen grupp). Man behöver därför beakta den insamlade informationens användbarhet baserat på deras respektive behov, där domänexperten sannolikt efterfrågar att modellen ska fånga mer komplexa sammanhang än vad t ex utvecklaren behöver. Om man prioriterar domänexpertens behov finns en risk att det är på de andra användargruppernas bekostnad. Å andra sidan, domänexperten är en nyckel i att tillhandahålla data av god kvalitet, så denne måste se värdet i det.

I ett större sammanhang är tillgängliggörandet av information kring institutionernas samlingar en del i en större process kring att förändra tankesättet kring det digitalas roll för museisektorn, där det sannolikt kommer att behövas nya måtvärden. Det är mindre väsentligt att samlingarna syns i Europeana än vad detta kan få för effekt: exempelvis ökad synlighet i

nya sammanhang eller möjligheten att berika informationen kring samlingen med information från andra källor.

Det är lätt att kategorisera en förändring av K-samsök som hörandes hemma i teknikkategorin, men det är bara en komponent. Ur ledningars förståelse för det digitala möjligheter och förutsättningar, skapas organisatoriska och samarbetsmässiga prioriteringar, vilket i sin tur kopplar in verksamhetskompetens om det digitala behovet.

I rapporten syftar ordet roll på vilka normer och förväntningar som kan finnas på K-samsök, medan funktion syftar på en uppgift med på förhand beskrivet resultat som man vill ska finnas och fungera. Avsnittet inleds med argumentation i frågan om det behövs en nationell aggregator och fortsätter med att lista fem möjliga kompletterande roller som K-samsök kan ha. Avslutningsvis presenteras några vägval på kort och lång sikt.

### 3.1 Behövs en nationell aggregator?

I dagsläget är K-samsök den enda möjligheten de flesta museer att tillhandahålla data kring sina samlingar. Samtidigt är det resurskrävande att ta fram portar från museisystem till K-samsök (om en sådan inte redan tillhandahålls av museisystemet). Ett synsätt från några intervjuer är att det inte är mer kostsamt att tillhandahålla sina data som LOD på egen hand än att koppla dem till K-samsök.

Det finns ett antal fördelar med att museerna själva publicerar information om sina samlingar som LOD:

- Ägandet av informationen tydliggörs, vilket kan ha fördelar bl a för datakvalitet och förädling av data.
- De enskilda institutionerna kan tillhandahålla data på det sätt som passar de egna samlingarna bäst (samtidigt som de även tillhandahåller den data som en nationell aggregator kan ha nytta av). Den informationsmängd som finns lokalt är i många fall större än den som kan fångas i en nationell aggregator, åtminstone som K-samsök är uppbyggd idag.
- Andra aggregatorer kan tillkomma, som tillhandahåller andra perspektiv (baserat på, till exempel, mer specifika domäner).

Om alla museer tillhandahåller LOD, behöver vi då en aggregator på nationell nivå? Skulle inte Europeana kunna hämta direkt från innehållsleverantörerna och utvecklare plocka hem data från de museer de är intresserade av? Det finns ett antal argument för att behålla en nationell aggregator för kulturarvsdata:

- Värdet i att ha en gemensam ingång för svenska samlingar.
- Även om inte alla behöver använda samma vokabulär, behöver det finnas någon typ av samordning för att underlätta återanvändandet.
- Intressanta sammanlänknings underlättas av att auktoritetslistor tas fram och sprids.
- Det behövs sannolikt ändå en tjänst för att utvecklare ska hitta datakällorna och förstå hur de kan användas.
- En gemensam ingång kan vara mer stabil och tidseffektiv för utvecklare. Samtidigt kan man välja att hämta data från specifika museer för att t ex få en mer finkorning beskrivning.
- I ett kortsiktigt perspektiv kommer Europeana att föredra att skörda från en nationell aggregator snarare än enskilda museer. I ett längre perspektiv är rollen som kunskapspridare och kommunikatör viktigare.

- Europeana är inte moget att ersätta det den nationella aggregatorn: den skördar inte data inkrementellt.
- Aggregatorn blir en naturlig mötesplats för nationellt samarbete, den kan agera motor för att sprida kunskap kring länkade data, licenser, och annat som rör hantering och återanvändning av information kring kulturarvet.

Sammantaget finns det både ett stort värde i att institutionerna själva publicerar LOD och i att ha en nationell aggregator. Genom uppdateringar i protokoll kan man minska problemet att mycket information försvinner när data aggregeras i två nivåer (där den låga granulariteten kan sedan leda till att det är svårt att använda tillhandahållen data).

### 3.2 Roll 1: Aggregator

Denna roll är baserad på den primära rollen som K-samsök har idag, men i ny tappning, där länkade öppna data har en mer central roll. En fördel med en lösning baserad på länkade data är att ägandeskapet tydligare kan placeras hos innehållsleverantören. Det ökar även möjligheten att inte utgå från minsta gemensamma nämnare vad gäller protokoll.

Bilaga 2 innehåller en beskrivning av möjliga tekniska funktioner som en sådan lösning kan innehålla, framtagen av Metasolutions. Förutom möjliga funktioner förklaras även möjliga kombinationer av funktioner, med deras för och nackdelar. Dessa kombinationer är (för närmre genomgång av funktionerna, se bilagan):

- **Naiv** (skördning + cache + SPARQL endpoint) – Bygger på att alla källor finns som länkade data och att det huvudsakliga syftet är att ge en enhetlig ingång och att SPARQL frågor kan spänna över flera källor på ett effektivt vis.
- **Övertagande** (import + mappning + minting + lagring + SPARQL) – Engångsimport och mappning av data med nya URI:er, sökningsbara visa SPARQL.
- **Icke-erkännande** (import + ny bas-URI + lagring) – Liknar övertagande, förutom att källorna redan är tillgängliga som länkade data. Trots detta skapas nya URI:er, med effekten att källan inte erkänns.
- **Pragmatisk** (skördning + mappning + minting + cache + SPARQL + export) – Hanterar att det finns källor som inte redan finns tillgängliga som länkade data, vilket gör att det utöver skördning även krävs mappning och myntning av URI:er där sådana saknas. En exportfunktion tillhandahåller datadumpar, vilket är användbart bl a för utvecklare.
- **Komplett** (skördning + mappning + minting + koreferenser + berikning + materialisering + cache + look ahead + export + SPARQL + text-sökning) – Utöver det som ingår i den pragmatiska lösningen hanteras även duplikat (när samma sak ingår i flera källor, med olika identifierare) genom en koreferensfunktion. Data kan berikas manuellt eller automatiskt, med hjälp av t ex datalingsvistiska metoder. För att undvika att klienter behöver göra onödiga anrop för att få relevant data, kan en look ahead-funktion användas för att inkludera exempelvis typer och labels för länkade resurser. Eftersom enkla klienter inte kan dra ens grundläggande inferenser kan en materialiseringsfunktion göra implicit information som klass- och egenskapshierarkier explicita. En text-baserad sökning kompletterar SPARQL-sökningar.

Metasolutions förslag är att i alla fall sikta på den pragmatiska kombinationen, vilket utöver det som görs idag skulle innebära att man lägger till möjligheten att skörda från datakällor som tillhandahålls som länkade data, att man har en offentlig SPARQL endpoint samt att man tillhandahåller en exportfunktion för att ladda ner datadumpar. Om man vill driva på

användandet av länkade data inom sektorn kan målen vara mer ambitiösa, med en lösning som ligger närmre den kompletta kombinationen av funktioner, där den väsentligaste funktionen rör stöd för koreferenser, eftersom det stödjer innehållsleverantörer i att producera mer användbara data, vilket i sin tur ger bättre resultat för användare av tjänsten.

Förutom rent tekniska funktioner kopplat till rollen som aggregator finns även funktioner kopplade till kommunikation och rutiner.

### 3.3 Roll 2: Auktoritetslistor – samordnare och/eller ägare

För att länkade data ska funka fullt ut måste man i någon utsträckning använda samma auktoritetslistor (gemensamma resurser för exempelvis plats och tid – ett slags master data) – eller specificera hur olika auktoriteter förhåller sig till varandra. För att auktoritetslistor ska komma till användning måste de upplevas som relevanta och långsiktiga, för vilket det krävs samordning och ägarskap. Oavsett ägarskap handlar det inte om att skapa en enda stor kulturhistorisk ontologi, utan att se till att gemensamma basdata finns på plats samtidigt som andra uppdelningar av koncept är möjliga. Ett sätt att uppnå långsiktighet är att koppla till wikidata<sup>5</sup> (där det eventuellt redan finns en del poster som kan användas).

Hur K-samsök agerar i frågan om auktoritetslistor är en avvägning som både har att göra med ansvar och kostnad, men också att dra fördel av den mångfald länkade data kan erbjuda samtidigt som man faciliterar en samsyn när sådan är möjlig och till hjälp för innehållsleverantörer.

Det finns flera möjligheter kring vilka funktioner K-samsök kan ha för att driva frågan om auktoritetslistor:

- Att K-samsök aktivt leder arbetet för att ta fram auktoritetslistor för de domäner som är relevanta. Förutom svenska institutioner bör man undersöka om det finns redan framtagna listor som med små medel skulle kunna anpassas till svenska förhållanden.
- Att K-samsök koordinerar arbetet med att respektive institutioner tar fram och tillhandahåller auktoritetslistor inom sin domän, till exempel att Livrustkammaren håller i regentlängden och Riksantikvarieämbetet i lämningstyper. Här har K-samsök sedan en funktion som spridare av vetenskap och användande av listor till andra institutioner. En sådan lösning bygger på att flertalet museer är intresserade av att ta fram och underhålla auktoritetslistor och att det drivs från institutionsnivå och inte som personbundna projekt.
- Att K-samsök tydligt överläter arbetet med auktoritetslistor till innehållsleverantörerna eller till annan myndighet. I detta fall finns ytterligare vägval:
  - Att K-samsök föreslår användning av listor som tagits fram eller nu håller på att tas fram (antingen internationellt eller nationellt).
  - Att K-samsök inte föreslår listor för användning.

Oavsett vilken väg K-samsök tar ovan kan man även välja om man vill att aggregatorn ska cacha de auktoritetslistor som används. På så vis tar en roll som aggregator även för listorna, där man utöver cachning skulle kunna hantera kluster av sameAs-länkar och eventuellt föreslå användning av en viss stabil URI för en auktoritet.

K-samsök kan även ta denna roll utan att vara aggregator, men om arbetet som aggregator ska bli framgångsrikt måste man aktivt förhålla sig till denna fråga. Här krävs samordning, bli innehållsleverantörer, Riksarkivet och Kungliga biblioteket.

---

<sup>5</sup> <https://www.wikidata.org/>



### 3.4 Roll 3: Mötesplats för innehållsleverantörerna

I arbetet med K-samsök har det även funnits behov att arbeta med exempelvis informationshantering allmänt inom sektorn och information om vad olika öppna licenser innebär och har för fördelar och nackdelar. En möjlig roll för K-samsök är att fortsätta och fördjupa detta arbete, där ett ökat fokus sannolikt kommer att läggas på stöd kring länkade data. Funktioner om man har rollen som mötesplats för innehållsleverantörer kan vara:

- Utbildning (exempelvis kring LOD)
- Utvecklingsstöd (exempelvis expertis kring vad som behövs för att utveckla med hjälp av LOD)
- Fysiska möten
- Stöd för arbetsgrupper i specifika frågor

En sak att ta ställning till är för vem mötesplatsen är till för. Denna roll kan tas utan att vara aggregator, men om man tar aggregatorrollen behöver i alla fall några av funktionerna ovan ingå på ett väldefinierat vis.

### 3.5 Roll 4: Stöd för utvecklare

Ett sätt att nå ut till allmänheten med det material som tillhandahålls kan vara genom applikationer framtagna utanför K-samsöks ramar. Det kan till exempel vara ett museum som använder K-samsök för att bygga en app för en utställning, en entusiast med programmeringskunskaper som vill nå ut i sin niche, eller en region som vill sprida information om kulturarvet i turismsyfte. För att K-samsök ska komma till användning måste vetenskap om det vara spritt och dokumentation måste skapa en låg tröskel för användande.

Funktioner för rollen kan vara:

- Ta fram material som informerar om möjligheterna med K-samsök, där fokus sätts mer på innehållet än tekniken.
- Tillhandahålla dokumentation som är lätt att sätta sig in i.
- Stödja kunskapsutbyte mellan utvecklare som använder K-samsök.
- Aktivt ta in feedback från de som använder API:et för att kunna göra förbättringar
- Synas i relevanta sammanhang – hackathons och konferenser, grupper i sociala media, etc.

Är man inte aggregator blir denna roll mindre relevant. Om innehållsleverantörer enskilt tillhandahåller länkade data kommer det finnas ett behov av en samlad ingång för var man hittar respektive dataset och hur man arbetar med dem.

### 3.6 Roll 5: Skapare av kontaktytor mot allmänheten

Syftet med att museer delar med sig av data om sina samlingar är att det ska komma allmänheten till gagn. Till denna roll finns ett antal möjliga funktioner:

- Utveckla generella ingångar till hela aggregatet, som Kringla.
- Utveckla verktyg som gör att man kan bygga nischade applikationer på K-samsök utan större programmeringskunskaper.
- Utveckla verktyg som på ett aktivt sätt samlar in användargenererat innehåll (typ *crowdsourcing* eller *citizen science*), t ex rättningar, kommentarer, bilder, länkar mellan K-samsöksobjekt och externa resurser.
- Sätta upp samarbeten kring hur materialet i Kringla kan användas i undervisning.
- Arbeta aktivt för att nischade applikationer tas fram. Det kan handla om kontakter och aktiviteter med intresseföreningar.

Denna roll är den som i sina aktiviteter är mest perifer till aggregatorrollen, men samtidigt ligger mycket nära syftet med att tillhandahålla data från början. En möjlighet är att dessa

funktioner uppnås genom andra aktörer, där K-samsök indirekt stödjer genom sina andra roller.

### 3.7 Vägval och förslag på kort och lång sikt

Ur ett nationellt perspektiv är det baserat på argumenten i Avsnitt 3.1 relevant att behålla en nationell aggregator, där LOD är ett rimligt grundantagande. K-samsök behåller således sin roll som aggregator, men utvidgade funktioner i enlighet med Metasolutions förslag som beskrivs i Avsnitt 3.2. Utöver det förtydligar man rutiner och kommunikation mot innehållsleverantörer.

K-samsök behöver ta ställning till i vilken mån man vill ta övriga roller som beskrivs, kring auktoritetslistor och i förhållande till innehållsleverantörer, utvecklare och allmänhet. Utifrån utgångspunkten om rollen som aggregator ovan, ges nedan ett antal saker att ta i beaktande på kort och lång sikt.

#### 3.7.1 Förslag på kort sikt

I det korta perspektivet (1–2 år) kan ett fåtal innehållsleverantörer börja tillhandahålla data kring sina samlingar som LOD. Detta bör K-samsök ta tillvara för att uppdatera sina system för att kunna skörda länkade data. Ett vägval här är i vilken mån K-samsök ska engagera sig i vilket vokabulär (protokoll) som används, eller om detta ska överlåtas åt innehållsleverantören för att sedan mappas in vid skördning. Europeana Data Model (EDM) skulle kunna vara ett möjligt val, om man med hjälp av länkade data kan berika EDM-modellen så att man fångar en rikare informationsmängd (EDM saknar en del information som ingår i det nuvarande protokollet). Om man har en överenskommen vokabulär måste man även bestämma sig om man vid skördning endast ska ta in det som ingår i basprotokollet, eller om man ska fånga upp även sådant som inte ingår i basprotokollet. Skördningsformatet skulle alltså kunna tillåta en stor mängd olika RDF-uttryck men uppmuntra att man använder vissa fält, eller beskriva hur man mappar till annat.

Under denna period behöver man även se över URI-tilldelningen. Ska man ha en gemensam bas-URI för hela K-samsök, eller ska innehållsleverantörer som tillhandahåller egna länkade data använda egna domäner? För de som i det korta perspektivet skördas genom portar, ska man se över URI-strukturen (där man i dagsläget har identifierat några brister)?

Man bör även dra lärdom av de erfarenheter som de museer som börjar tillhandahålla LOD gör. Beroende på hur man förhåller sig till Roll 3 kan man även erbjuda en kunskapsutbytesplattform så att informationen kommer andra institutioner till gagn. En risk för K-samsök är att museerna som gör detta arbete inte ser behovet av den nationella aggregatorn och om man vill behålla den rollen behöver man visa på värden som K-samsök kan tillföra samt visa uppskattning och intresse för det arbete som görs.

K-samsök behöver i det korta perspektivet tydliggöra sin inställning i frågan om auktoritetslistor, Roll 2, genom något av förslagen som tas upp i Avsnitt 2.3. En möjlighet på kort sikt är att de museer som tillhandahåller LOD också uppmanas att tillhandahålla auktoritetslistor inom något eller några avgränsade områden som berör deras samlingar. Detta bygger kompetens både hos institutionerna och K-samsök, något som sedan kan delas med andra aktörer.

Samtidigt som de första museerna som börjar tillhandahålla LOD på egen hand behöver ett utbildningsarbete ske gentemot andra aktörer (något som sannolikt redan påbörjats), som är en del av Roll 3. Här är det viktigt att se bortom tekniken, att arbeta för en förändrad

inställning, där föremål behöver ha en representation, förutom att ställas ut och förvaltas. För att uppnå detta är det viktigt att man inte bara tar med systemutvecklare och databasadministratörer i arbetet.

I dagsläget är LOD inte allmänkunskap hos systemutvecklare. Det är ett nytt tankesätt på flera sätt, samt teknik som fortfarande är under utveckling. Även om verktyg har genomgått en teknisk mognad de senaste åren, återspeglas detta inte alltid i dokumentation och pedagogik kring lösningen. Domänkännedom är alltid relevant vid systemutveckling, men när det handlar om att ta fram representationer av föremål i samlingar blir domänkunskapen väsentlig. Om man vill satsa långsiktigt på LOD, behöver man komma ifrån att relevant kunskap (hos aggregatorn och innehållsleverantören) är personberoende. Sektorn behöver därför kommunicera behovet av personer med LOD-kunskap, allra helst i kombination med domänkunskap, bl a till högskoleutbildningar (både museiutbildningar och systemutvecklarutbildningar).

För att de datakällorna som tillhandahålls som LOD ska nå sin fulla potential behöver de länkas samman. Idag finns väldigt få länkar mellan objekt i K-samsök, och ägandeskapet för sådana länkar är otydligt. En möjlighet är att ha användargenererade länkar, vilket skulle vara mer experimentellt. En annan startpunkt skulle kunna vara att ha ett ”länkathon med tre till fyra institutioner för att visa på möjligheter med länkande.

### 3.7.2 Analys och förslag på lång sikt

I ett längre perspektiv (här 3-5 år) har flera museer valt att publicera sina egna länkade öppna data, antingen på egen hand eller med hjälp av sitt museisystem. I detta tidsperspektiv levererar flera museer fortfarande data till K-samsök enligt det tidigare formatet, men utfasning sker.

På längre sikt behöver funktioner inom Roll 3 sannolikt få mer utrymme. Eftersom det inte är realistiskt att mindre aktörer läser på och blir bra kravställare för LOD behövs stöd kring vilka krav som behöver ställas vid upphandlingar, och en möjlighet är att det kommer från aggregatorn. När fler vill tillhandahålla LOD (när fler dataset är tillgängliga ser man nytta tydligare) är det viktigt att ha med perspektivet att LOD är en filosofi snarare än en plattform, man behöver se bortom att det är en tjänst som ska köpas in.

Det behöver hanteras att både interna och externa lösningar i dagsläget bygger på K-samsök. Flera museer har byggt egna webbapplikationer på K-samsök, vid en förändring av K-samsök behöver detta hanteras. K-samsök är integrerat i ämbetets egna system och behov.

## 4 K-samsök i förhållande till Europeana

Vad gäller Europeana behöver K-samsök förhålla sig till bl a:

- Hur Europeana skördar data och deras planer vad gäller detta
- Vad Europeana förväntar sig av nationella aggregatorer
- Vilket stöd K-samsök kan få från Europeana
- Europeanas planer vad gäller länkade data
- Europeanas strategi 2015-2020
- Förändringar i Europeanas finansiering

I dagsläget bygger Europeana på att data skördas från nationella (alternativt domänbaserade) aggregatorer. Deras nuvarande organisatoriska och tekniska upplägg håller inte för individuell

skördning från alla europeiska kulturarvsinstitutioner. Den nationella aggregatorn spelar även en roll som nationell kontakt mot institutionerna och för kunskapsutbyte för tekniska plattformar. För K-samsök har Europeana varit till stöd bl a vad gäller kunskap och evenemang. En nackdel med de nationella aggregatorerna är att Europeana inte har direktkontakt med institutionerna. Här ser Europeana att de nationella aggregatorerna ett ansvar att förmedla meddelanden både från institutionerna upp till Europeana och vice versa: problem (exempelvis med specifika dataleveranser eller mer allmänt kring licenser), vad som fungerar väl, men även strategi.

Europeana arbetar sedan tre år med länkade data på olika vis. En drivkraft i detta har varit den nya datamodellen, Europeana Data Model (EDM), som bygger på tankar och tekniker från länkade data. Under de senaste åren har Europeanas arbetsprocesser och verktyg uppdaterats för att använda EDM. EDM har även använts för utvecklandet av en länkade data-pilot.

I Europeanas strategi för 2015-2020<sup>6</sup> är ett av målen att gå från att vara en portal till en plattform, med innehåll som kan återanvändas och byggas vidare på. Denna plattform är tänkt att innehålla tre nivåer:

- En kärna bestående av data, innehåll och teknik.
- En användnings-nivå, med API:er, LOD och villkor för användande.
- En service-nivå, där upplevelsen anpassas för Europeanas huvudsakliga kundgrupper: professionella (innehållsleverantörer), slutanvändare och kreativa näringar.

Kopplat till detta finns ett antal prioriteringar:

- Förbättrad datakvalitet – högre kvalitet på metadata, mer material med öppna licenser.
- Öppna licenser – stödja sina partners att tillhandahålla så mycket material som möjligt för återanvändning.
- Skapa värde för Europeanas partners, exempelvis genom att minnesinstitutioner får uppmärksamhet, eller att entreprenörer skapar nya företag.

Strategidokumentet berör även Europeanas finansiering. Europeana får från 2015 ekonomiskt stöd genom Connecting Europe Facility (CEF), där man måste visa på hållbarhet och där finansieringen från EU kommer att minska. Som en del i att säkra långsiktig finansiering kan partners komma att bjudas in att investera och bli delägare i Europeana, där dessa partnerskap kommer att göras på medlemsstatsnivå. Tanken är att ett delägarskap ger tillgång till ”premium services” i Europeanas moln, som gör att de snabbare kan övergå till nya system, dela data på sätt som förbättrar räckvidd, datakvalitet och effektivitet.

#### 4.1 På kort sikt

Ur det operativa perspektivet sker inga förändringar från Europeana i det korta perspektivet. Från innehållsleverantörer finns önskemål om tydligare rutiner för kommunikation kopplat till Europeana.

Som en del av prioriteringen kring förbättrad datakvalitet, ser Europeana över sin aggregationsprocess, där de vill gå från en centraliserad lösning till en distribuerad arkitektur. Kopplat till detta har Europeana inlett ett arbete där de koordinerar med aggregatorer (nationella och domänbaserade) för att diskutera hur de tycker att aggregeringen ska se ut om några år. Detta skulle kunna handla om att innehållsleverantörerna delar sina data med Europeana Cloud, där både nationella aggregatorer och Europeana kan hämta data. En del i detta är projektet Europeana Cloud<sup>7</sup> (2013-2015), som syftar till att ta fram ett moln-baserat

<sup>6</sup> <http://pro.europeana.eu/documents/858566/640ac847-0dfc-4b01-9f36-d98ca1212ec9>

<sup>7</sup> <http://pro.europeana.eu/web/europeana-cloud/home>

system för Europeana och aggregatorerna. Jag har inte lyckats hitta information om i vilken mån som Europeana Cloud ska bygga på länkade data. Givet att EDM är en naturlig utgångspunkt för Europeana Cloud, och att EDM i sig bygger på filosofin kring länkade data, så förefaller det sannolikt att länkade data spelar en roll i Europeana Cloud.

Det innebär också att de nationella aggregatorerna har en del i att driva denna utveckling, att dela med sig av sina rekommendationer till Europeana, vilket i sin tur bygger på goda relationer med innehållsleverantörerna.

#### 4.2 På längre sikt

Med en modell lik Europeana Cloud skulle de nationella aggregatorernas roll bli mer inriktad mot kommunikation, kunskapsdelning och utbildning, än mot praktisk datahantering. Det är däremot svårare att bedöma i vilken mån något som liknar Europeana Cloud faktiskt kommer att lanseras.

Sverige och K-samsök behöver förhålla sig till Europeanas förändrade finansiering (kanske redan på kort sikt). Ur båda tidsperspektiven kan K-samsök engagera sig i varierande mån i Europeanas strategi, både på egen hand och med innehållsleverantörerna. De prioriteringar som tas upp i Europeanas strategi förefaller ligga i linje med K-samsöks.

### 5 Summering av vägval

Processen att öppna upp och tillgängliggöra data, bildmaterial och liknande resurser är inte i huvudsak ett tekniskt arbete, även om det finns tekniska problem som behöver lösas. Det handlar lika mycket om organisationers och dess medarbetares förhållningssätt till denna öppenhet – dess möjligheter och risker – men också digital kompetens, licenser, upphovsrätt, samt utåtriktat arbete mot exempelvis andra offentliga verksamheter, kreativa näringar, applikationsutvecklare och de medborgare för vilka kulturarvsresurserna förvaltas. Utbildning krävs på olika nivåer, men även samordning (bl a för att möjliggöra den länkning av resurser som är målet med arbetet – till exempel genom val av vokabulär och arbete med auktoritetsposter). Den digitala infrastrukturen måste vara på plats, men utan ett samspel med icke-tekniska faktorer, bl a de som tas upp ovan, nås inte målet.

För K-samsöks del måste man ta ställning i vilken mån man vill prioritera efter Europeanas mål och förändringsarbete och i vilken mån man vill arbeta utifrån det nationella perspektivet. Här finns det två huvudlinjer, där det idag är något oklart hur väl de överlappar:

- Engagera sig i Europeanas kommande satsningar kring nytt aggregationsupplägg för att det ska passa för svenska intressen. Fördelar med detta är att man blir en del av en större teknisk plattform, med delade kostnader och teknikansvar. Det finns däremot risker kopplade till detta vägval: arbetet kan dra ut på tiden, inte passa Sverige eller kanske inte bli något alls. Om man väljer detta alternativ bör man den närmsta tiden engagera sig i Europeanas utveckling, undersöka i vilken mån man kan gå mot EDM, experimentera med LOD med intresserade institutioner för att vidarebefordra dessa erfarenheter till Europeana, samt förtydliga de roller som inte rör att vara en nationell aggregator.
- Bygga ut den nationella aggregatorn för att tydligare hantera länkade data, där Metasolutions förslag är en startpunkt. På kort sikt experimenterar man med LOD tillsammans med intresserade institutioner, förtydligar sin roll vad gäller auktoritetslistor, samt förtydligar sin roll vad gäller övriga funktioner.

## Bilaga 1: Intervjupersoner

Henrik Summanen (Riksantikvarieämbetet)  
Lars Lundqvist (Riksantikvarieämbetet)  
Rickard Larsson (Riksantikvarieämbetet)  
Johan Carlström (Riksantikvarieämbetet)  
Kerstin Jonsson (Riksantikvarieämbetet)  
Peter Krantz (Kungliga biblioteket)  
Niklas Lindström (Kungliga biblioteket)  
Henning Scholz o Valentine Charles (Europeana)  
Hannes Ebner och Matthias Palmér (Metasolutions)  
Kristina Berg (Arkitektur- och designcentrum)  
Bengt Wittgren (Murberget)  
Magnus Johansson (Världskulturmuseerna)  
Karin Nilsson (Livrustkammaren)