

Lösningförslag för e-arkiv baserat på Archivematica, AtoM och utveckling av inleveransportal

Digisam och Demokratijubileet 2018-2021

2019-05-14

Bakgrund

Digisam utreder möjligheterna att etablera en gemensam lösning för att säkra lagring och åtkomst av vårt digitala kulturarv. Inom tidsramarna för Demokratijubileet föreslås en pilotlösning för 30-100 aktörer för lagring av 1-100 TB, där denna lösning succesivt vidareutvecklas för att stödja behoven.

Archivemata och AtoM är byggda på öppen källkod (open source) och kan tillsammans eller var för sig användas för e-arkiv och arkivredovisning. Archivemata stödjer pre-ingest, ingest, arkivering och tillgängliggörande, enligt OAIS-modellen och möter Riksarkivets krav på e-arkiv. Arkivredovisningssystemet AtoM stöder arkivredovisning enligt allmänna arkivskemat eller verksamhetsbaserad arkivredovisning.

I Sverige ökar intresset för open source-lösningar. Flera organisationer, bland annat Sveriges geologiska undersökning (SGU) och Sveriges Lantbruksuniversitet (SLU) har valt Archivemata som e-arkivlösning inom sina verksamheter.

Kungliga biblioteket (KB) använder bl.a. AtoM för att redovisa författares arkiv.

ArkivIT har erfarenhet av att i samarbete med Safespring etablera drift av e-arkiv baserade på Archivemata och AtoM, erbjuda support, och även av att utveckla webbaserade lösningar för att förenkla uppladdning och registrering av digitalt material för e-arkivering.

Inom *Enskilda arkivet* (enskildaarkivet.se) samarbetar Arkiv Sörmland, Arkiv Gävleborg och Folkrorelsearkivet för Uppsala län kring en lösning med Archivemata och Atom, där ArkivIT och Safespring tillhandahåller driftsplattformen för lösningen. Även dessa tre föreningar analyserar nu möjligheter och förutsättningar kring utveckling av en webbaserad lösning för att förenkla uppladdning till Archivemata.

I detta dokument ges förslag på vilka funktioner som kan utvecklas och hur ett utvecklingsprojekt kan delas upp i olika faser för ett stegvis införande.

Förutsättningar

Följande lösningförslag utgår från att open source-systemen Archivematica och AtoM används som e-arkiv. Den lösning som utvecklas (av ArkivIT) ska vara webbaserad, integrera med Archivematica och AtoM, och drifas inom en gemensam plattform (av Safespring).

En annan viktig aspekt är att andra accessmöjligheter ska vara möjliga i denna lösning, bland annat för att kunna göra informationen i e-arkivet tillgänglig till externa arkiv- och söksystem och även kunna tillhandahålla materialet enligt standarder för Öppna data. Detta kan lösas med öppna gränssnitt (API) för kommunikation mellan system, både med de gränssnitt som redan ingår i Archivematica och AtoM, men även med nyutvecklade gränssnitt om så behövs.

Inom utvecklingsprojektet ska varje delmål (delprojekt) utgå från en avgränsad mängd funktioner som till största del ska vara fullt användbara vid leverans utan krav på ytterligare vidareutveckling. Detta för att lösningen ska kunna finansieras och utvecklas stegvis enligt en agil och iterativ modell. Hela projektet kan omfatta ett flertal års arbete, men utvecklingen bör ske i avgränsade delprojekt om ca 2-6 månader.

Samarbetsformer och kravställning

De kunder som väljer denna gemensamma plattform ska kunna köpa e-arkivtjänst, support och drift baserat på volym och funktion, med priser som varierar och är relevanta i förhållande till kundens organisations storlek och behov. Utveckling av tjänsten kommer dels finansieras av intäkter från tjänsten, men dels också som sponsrade utvecklingsprojekt med extern finansiering, exempelvis från nya eller befintliga kunder. De kunder som vill avbryta tjänsten ska kunna exportera informationen till andra lösningar eller teknikneutrala format.

ArkivIT och Safespring kan ansvara för förvaltning, drift, support och vidareutveckling av denna e-arkivplattform. Den applikation som utvecklas kan också införlivas i användarföreningen för Archivematica och AtoM, AOAIS (<https://www.aoais.se>) för gemensam koordinering av utvecklingsplan för denna produkt.

Utvecklingsalternativ

Nedan redovisas olika alternativa vägar för utveckling.

Pilotprojekt

ArkivIT har tidigare som första alternativ föreslagit ett pilotprojekt omfattande 1 000 timmar för att iterativt utveckla en helhetslösning som kan börja användas och utvärderas av de aktörer inom Digisam som har material att leverera in.

Detta förslag finns beskrivet i tidigare offert, se [1]. Här beskrivs även drift, support, förvaltning, genomförande av workshops med mera.

Omfattning: 1000 timmar

Etapper för inleveransportal

Ett alternativ är att istället dela upp utvecklingen i ett flertal etapper. Nedan ges förslag på fem (5) nivåer som i stora drag överensstämmer med slutmålet för ovanstående pilotprojekt.

Nivå 1 – Webbformulär för metadatasättning

Webbgränssnitt för att registrera metadata och generera indatafiler till Archivematica (AM). Målet är att kunna utföra en inleverans utan att behöva skapa metadata-fil för hand.

1. Peka ut en mapp med filer på AM-servern (via WinSCP, FileZilla eller liknande)
2. Varna om filer eller mappar med ogiltiga namn
3. Registrera en uppsättning Dublin Core-termer (15 st) i ett webb-formulär för samtliga filer i mappen. Några termer går att repetera flera gånger, som dc.creator, dc.subject, dc.coverage...
4. Starta arkivering från inleveransportalen
5. Skapa *metadata.csv* (indata-fil till AM)
6. Automatisk uppladdning i AtoM – alla filer hamnar på samma nivå i arkivbeskrivningen

Omfattning: ca 250 timmar (varav hälften inkluderar grundläggande applikationsstrukturer)

Nivå 1 berör inte de krav ställda från Digisam, utan är en grund för föreningsarkivens behov.

Nivå 2 – Metadata per fil samt uppladdning från egen dator

Registrera enskilda filer och ladda upp med webbläsaren.

1. Peka ut en mapp med filer på egen dator
2. Döpa om filer med ogiltiga namn
3. Registrera Dublin Core-termer för varje fil (utan lagring i databas).
4. Starta arkivering från inleveransportalen (som ovan)
5. Skapa metadata.csv (som ovan)
6. Automatisk uppladdning i AtoM (som ovan)

Omfattning: ca 200 timmar

Nivå 2 introducerar de krav Digisam har för enklare uppladdning av filer.

Nivå 3 – Uppladdning med hierarkisk filstruktur

Bibehåll befintlig mappstruktur.

1. Peka ut en mapp med filer på egen dator med mappstruktur.
2. Döpa om filer med ogiltiga namn (som ovan)
3. Möjlighet att registrera Dublin Core termer för varje fil (som ovan)
4. Starta arkivering från inleveransportalen (som ovan)
5. Skapa metadata.csv (som ovan)
6. Skapa structmap.xml (indata-fil till AM)
7. Automatisk uppladdning i AtoM – filerna får mapp-struktur i AtoM

Omfattning: ca 200 timmar

Nivå 4 – Värdemängder för metadata – med databas

Stöd för värdemängder, taxonomier.

1. Peka ut en mapp med filer på egen dator med mappstruktur (som ovan)
2. Döpa om filer med ogiltiga namn (som ovan)
3. Möjlighet att registrera Dublin Core-termer för varje fil – med lagring av metadata i databas
4. Möjlighet att välja värderna ur drop-listor, t ex Kommun, Län, Ämnen, peka mot externa taxonomier.

5. Starta arkivering från inleveransportalen (som ovan)
6. Skapa metadata.csv och structmap.xml (som ovan)
7. Automatisk uppladdning i AtoM (som ovan)

Omfattning: ca 250 timmar

Nivå 5 – Behörighetssystem för externa användare

Bibehåll befintlig mappstruktur.

1. Inloggning i inleveransportalen med olika behörighetsroller
2. Språkstöd i GUI – svenska, norska...
3. Peka ut en mapp med filer på egen dator med mappstruktur (som ovan)
4. Döpa om filer med ogiltiga namn (som ovan)
5. Möjlighet att registrera DC-termer för varje fil med lagring i databas (som ovan)
6. Möjlighet att välja värden ur drop-listor (som ovan)
7. Starta arkivering från inleveransportalen (som ovan)
8. Skapa metadata.csv och structmap.xml (som ovan)
9. Automatisk uppladdning i AtoM (som ovan)

Omfattning: ca 200 timmar

Efter nivå 5 täcks huvuddelen av de funktioner som ingår i ovan nämnda pilotprojekt, dock utan arbete med en inledande utredning för anpassad kravspecifikation, samt projektledning, användarhandbok och hjälpfunktion anpassade för slutanvändare utan viss arkivkunskap.

Vidareutveckling

Lösningen kan sedan succesivt vidareutvecklas med nya funktioner och moduler och implementeras i etapper på motsvarande sätt som för nivåerna ovan, där omfattningen varierar från 50 till 200-300 timmar eller delas upp på flera etapper.

Faktureringsunderlag – Sammanställning av nyttjande och kostnadsfördelning mellan lösningens olika kunder och användare.

Fysiska arkiv – Registrera och söka fysiska arkiv och fysiska objekt och platser. Skriva ut etiketter för fysiska lagerplatser.

Teckenformatering – Validera teckenformat och teckenuppsättningar. Konvertera filnamn som innehåller åäö. Ersätta filnamn med åäö och ogiltiga tecken som ingår i webbarkiv med länkade sidor.

Ladda upp filer med *drag 'n' drop*-gränssnitt.

Tolkning – Extrahera metadata ut filnamn, mappstrukturer, filinnehåll, OCR.

E-post – Hantera arkivering av e-post och konvertering av MSG-filer och inkluderade filer.

Sökning – Sökmöjlighet och översikter över material under bearbetning i inleveransportalen.

Generera *tumnaglar* i inleveransportalen.

Generera *EAD*-filer.

Hantera och generera *PREMIS*-rättighetsfiler.

Behörighetshantering – Inloggning med tvåfaktorautenticering. Tidsbegränsade användarkonton. Hantera glömt lösenord. Granska användarlistor.

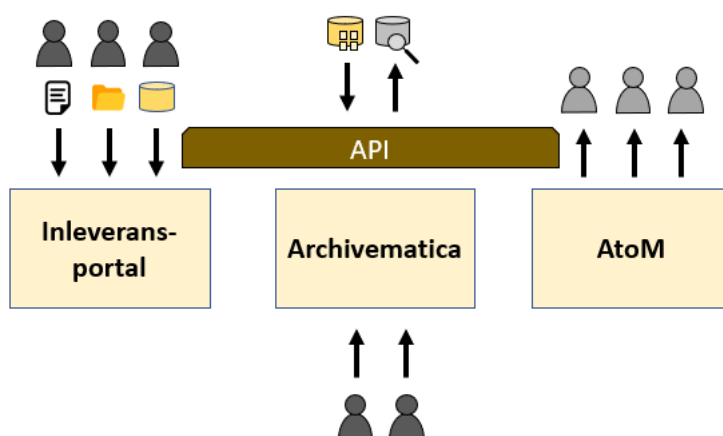
Import – Stöd för att importera metadata från databaser och register. Import av refererade filer.

Export – Exportera till andra system. Export av DIP till annat målsystem än AtoM, Skapa leveranspaket till ALVIN. Skapa EAC-CPF för leverans till Riksarkivet.

Öppna data – Mappning av metadata till DCAT-AP. Export till portaler för öppna data. Stödja API för öppna data.

Lösningssvision

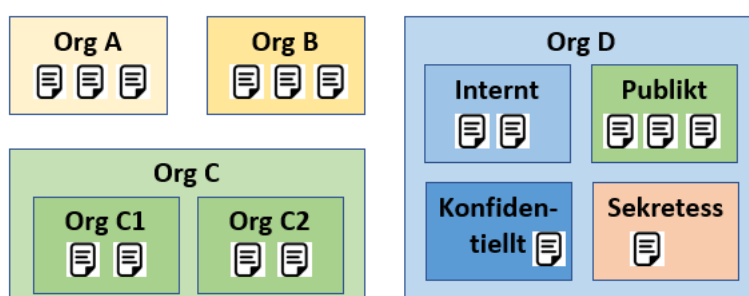
Inleveransportalen kompletterar open source-produkterna Archivematica och AtoM med ett lättanvänt grafiskt gränssnitt för att ladda upp och metadata-sätta dokument. Lösningen nyttjar alla de API:er och metadata-set som ingår i Archivematica och AtoM, bland annat OAI-PMH och PREMIS, samt kommande stöd för datalänkar och RDA, samt att vid behov komplettera dessa med egenutvecklade lösningar. Detta för att stödja extern integration för förädling och vidarenyttjande av den arkiverade informationen.



Den information som lagras i e-arkivet ska vara i maskinläsbar form och ha persistenta identifierare (GUID). Metadata ska stödja ägande/tillhörighet, licensiering/tillgänglighet samt de standarder som ger stöd för öppna data och länkat data. Informationen i e-arkivet ska kunna tillåtas att bearbetas (förbättras och kompletteras) genom interaktioner med externa system och datakällor.

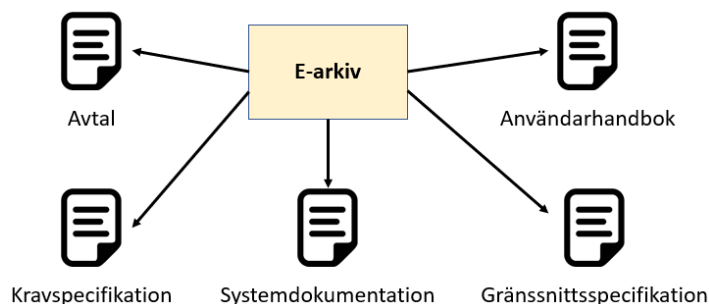
De användare som levererar in information ska ej behöva besitta arkivarietkompetens, och även om gränssnittet ska kunna konfigureras så att det ställer vissa minimikrav på angivet metadata, så ska det också kunna användas som en bearbetningsyta där metadata kompletteras efter hand. Viktigt är att informationen är unikt identifierad och har en initial koppling till ägande/inlevererande organisation. Administratörer kan hantera den arkiverade informationen i samtliga tre delsystem. Nyttjare, externa och interna, kan nå informationen från AtoM eller via något av de tillgängliga API:erna.

Lösningen kan komma att delas av ett flertal statliga och regionala organisationer och även under dem nyttjande organisationer och användargrupper. Därför behöver lösningen en behörighetshantering som garanterar de olika organisationernas åtkomstregler och klassificering för internt och konfidentiellt material relativt publikt material. Vidare ska även denna behörighetshantering och lagringsnyttjande ge stöd för fördelning av kostnader genom underlag för fakturering.

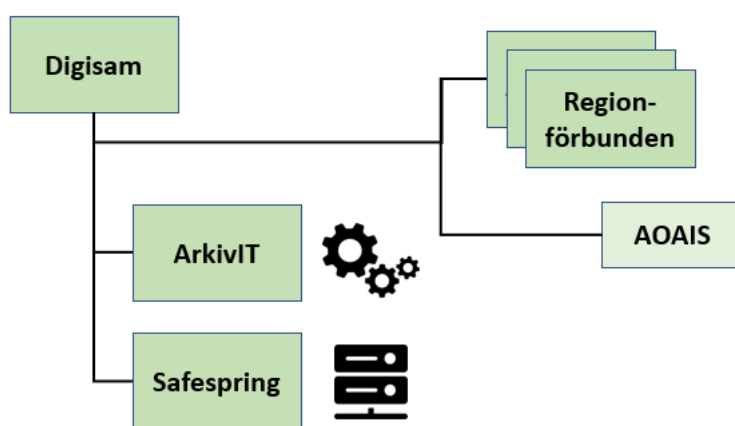


Vid projektstart tas en grundläggande gemensam kravspecifikation fram vilken täcker in lösningens syfte, mål och grundförutsättningar, en eller några inledande iterationer över funktioner som behövs för en första användbar version, samt översiktliga långsiktiga funktionella krav som kommer att behöva stödjas i kommande versioner.

Dokumentation över systemets funktionalitet, gränssnitt, drift och förvaltning hålls ständigt uppdaterade. Användaravtal tecknas med alla nyttjande organisationer. Lösning och drift behöver vara förberedd för en framtida exit vid eventuell migrering till annan arkivlösning.



Förvaltningsorganisation kan utgå från kravställande/finansierande huvudman eller vid behov läggas på ArkivIT som privat aktör samt att lösningen licensieras som open source. ArkivIT och Safespring kan ta ett aktivt ansvar för såväl drift, support, förvaltning, systemutveckling och omvärldsbevakning enligt den omfattning som önskas. Användarföreningen AOAIS kan utgöra stödjande gruppering för spridning av lösningens användande och koordinering av utvecklingskrav/önskemål.



Referenser

- [1] Offert, *Utveckling, installation och drift av e-arkivlösning, Archivematica och AtoM*, 2018-10-22