

# Komplexa digitala objekt i öppna arkiv inom det konstnärliga forskningsområdet

## Rapport

2011-12-03

OpenAccess.se-projekt, diarienumr: 51-480-2009

Stefan Andersson

Uppsala universitetsbibliotek, Enheten för digital publicering

Per Larsson

Katharina Nordling

Högskolan i Borås, Bibliotek & läranderesurser

## Innehållsförteckning

Beskrivning av projektet .....	3
Denna rapport.....	3
Vad är komplexa digitala objekt?.....	4
Bakgrund: Vad har gjorts i Sverige? .....	5
DiVA.....	5
Programvaror för öppna arkiv .....	6
Beskrivande metadata i de öppna arkiven .....	6
Platta och XML-baserade metadataformat .....	7
Metadatastandarder för att komplettera Dublin Core.....	8
Komplexa digitala objekt och långtidsbevarande.....	9
Specifika frågor för konstnärliga publikationer .....	10
Publikationstypen ”Konstnärligt forsknings- och utvecklingsarbete” .....	10
Fysiska objekt i relation till digitala avbildningar.....	11
Särskilda rättighetsfrågor.....	11
Publikationstyper för icke-textbaserade objekt i de lokala systemen.....	11
Samlingsposter.....	11
Gemensamma element för alla publikationstyper .....	11
Specifika nya element för olika publikationstyper .....	12
Musiktryck, noter.....	12
Ljudinspelning-musik eller – ej musik .....	12
Bild .....	13
Rörlig bild, video .....	13
Föremål .....	13
Programvara/multimedia .....	13
Blandat innehåll .....	13
Kompatibilitet med SwePub .....	13
Publikationstyper .....	14
Befintliga publikationstyper i SwePub .....	14
Andra publikationstyper (ej textbaserade).....	14
Exempel.....	15
Samlingsposter.....	15
Upphovsman.....	15
Språk .....	15

Fysisk beskrivning.....	15
Prototyp .....	16

## Beskrivning av projektet

I arbetet med SwePub-projektet, liksom i det tidigare SVEP-projektet samt andra mindre utvecklingsprojekt, har det tydligt framkommit att de konstnärliga forsknings- och utvecklingsarbeten som produceras vid ett antal svenska lärosäten är svåra att passa in i de publiceringssystem som används för vetenskapliga publikationer och examensarbeten. Dessa system är ursprungligen utvecklade för att ta hand om textbaserade publikationer vilket innebär att metadatamodeller liksom filuppladdnings- och arkiveringsfunktioner är inte anpassade till icke-textbaserade publikationer.

Detta projekt har kartlagt vilka typer av publikationer som produceras inom den konstnärliga utbildningen och forskningen för att få en grund till en så heltäckande lösning som möjligt vad gäller metadatabeskrivning och använda filformat. Kontakter har också tagits med de konstnärliga lärosätena vars synpunkter har bakats in i våra rekommendationer.

I detta perspektiv är det mycket viktigt att försöka hitta modeller som bygger på interoperabilitet för att kunna införliva dessa publikationstyper i nationella system som SwePub och Uppsök som framgångsrikt har skapat nationella metadataformat för textbaserade dokument. Eftersom dessa nationella system tekniskt bygger på Libris, det nationella bibliotekssystemet, skall en föreslagen modell också anpassas till detta.

En annan aspekt som måste utredas är förhållandet mellan de fysiska objekten (målningar, textilier, installationer etc.) och digitala, sekundära, versioner av dessa och hur förhållandet mellan dessa skall beskrivas. Vi har även konstaterat att det finns särskilda behov för dessa publikationer vad gäller upphovsrättsfrågor och att publiceringssystem och metadatamodeller måste kunna stödja uppmärkning av restriktioner på filnivå för att göra det möjligt att skilja open access-material från det som inte är det och därmed möjliggöra registrering och arkivering av båda typerna.

Slutligen avser vi ta fram en systemoberoende prototyp för användarvänlig registrering av komplexa objekt samt bra visuell presentation som stödjer de krav som framkommit i projektets tidigare delar - samt i den föregående pilotstudien vid HB - både vad gäller registreringsdelen och det publika sökgränssnittet. Prototypen skall också utvärderas och anpassas i samverkan med forskare.

## Denna rapport

Rapporten omfattar alla delar i projektet utom själva prototypen för visuell presentation.

Sammanfattning av resultaten:

- De öppna arkiven (DiVA/Fedora samt BADA/DSpace) kan hantera komplexa objekt
- SwePub:s formatspecifikation täcker redan de behov som finns för icke-textpublikationer
- Icke-textpublikationer kan behöva beskrivas med egen beskrivande metadata i systemen
- Det skall vara möjligt att ange hierarkiska relationer i metadata
- Begreppet *konstnärlig forskning* bör definieras i förhållande till publikationsregistrering

- Några särskilda element för icke-textpublikationer måste läggas till (objekttyp, material och teknik samt fysisk beskrivning)
- Den visuella presentationen av icke-textpublikationer är eftersatt i systemen

## Vad är komplexa digitala objekt?

Med komplexa digitala objekt i detta projekt avses "objects that aggregate data, metadata and sometimes services into a single, logical digital entity" såsom de definieras av Michael L. Nelson och Herbert Van de Sompel i introduktionen till *IJDL special issue on complex digital objects* (International Journal on Digital Libraries (2006) 6(2): 113–114)<sup>1</sup>. Poängen är alltså att flera digitala filer utgör en logisk enhet i motsats till exempelvis en digital avhandling som består av en enda enskild PDF-fil.

Komplexa objekt kan följaktligen bestå av en mängd olika sorters data, metadata, filtyper etc. men just det utgör egentligen inte något problem i detta sammanhang. Svårigheten med att beskriva och lagra komplexa objekt i "digitala bibliotek" ligger istället i att på ett effektivt sätt redogöra för relationerna mellan de olika ingående komponenterna i denna sammanhörande enhet. En fullt utvecklad modell måste också dels kunna innehålla beskrivande metadata av olika typ dels tekniska uppgifter om de olika ingående filerna samt slutligen även själva filerna i sig. Hur man anger bibliografiska metadata för varje enskild ingående enhet är sedan en fråga för sig. Det är också just den specifika delen med inriktning på konstnärliga publikationer som är huvudsaken i detta projekt och *inte* det övergripande formatet för hur komplexa digitala objekt skall behandlas eftersom det senare, som det kommer att framgå, redan finns på plats i publiceringssystemen. Huvudproblemet är alltså inte att objekten är komplexa utan att det saknas "publikationstyper" för att beskriva icke-textbaserade publikationer och därmed också en del specifika element för detta.

Nelson och Van de Sompel beskriver hur, på samma sätt som olika metadataformat har uppstått inom olika domäner, olika format för komplexa objekt har utvecklats för olika syften. Inom biblioteksområdet finns i huvudsak två format: METS och MPEG-21 DIDL. METS (Metadata Encoding and Transmission Standard)<sup>2</sup> underhålls numera av Library of Congress. Ursprungligen utvecklades formatet inom projektet *The Making of America II (MOA2)*<sup>3</sup>. MPEG-21 DIDL, en officiell ISO-standard<sup>4</sup>, är en del av MPEG-21:s standardiseringsmodell som främst drivs av den digitala underhållningsindustrin.

Artikelförfattaren Van de Sompel har själv i olika sammanhang förespråkat användandet av MPEG-21 DIDL för digitala bibliotek. Se exempelvis *Using MPEG-21 DIDL to Represent Complex Digital Objects in the Los Alamos National Laboratory Digital Library* (D-Lib Magazine (November 2003) 9 (11))<sup>5</sup> som även innehåller en övergripande beskrivning av formatet eller utvärderingsrapporten av det KB-finansierade projektet *Långtidsbevarande - från prototyp till fullfärdiga verktyg*<sup>6</sup>.

<sup>1</sup> DOI 10.1007/s00799-005-0127-y

<sup>2</sup> Se: <http://www.loc.gov/standards/mets/>

<sup>3</sup> Se: <http://sunsite.berkeley.edu/MOA2/>

<sup>4</sup> XML schema: [http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-21\\_schema\\_files/did/didl.xsd](http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-21_schema_files/did/didl.xsd)

<sup>5</sup> Se: <http://www.dlib.org/dlib/november03/bekaert/11bekaert.html>

<sup>6</sup> Se: [http://www.kb.se/dokument/Om/projekt/langtidsbevarande\\_utvarderingsrapport.pdf](http://www.kb.se/dokument/Om/projekt/langtidsbevarande_utvarderingsrapport.pdf)

Relaterat till dessa två format kan även nämnas Open Archives Initiative Object Reuse and Exchange (OAI-ORE)<sup>7</sup> vars syfte är att exponera komplexa objekt ("aggregations") på webben från den akademiska världen för "web 2.0" och så kallade sociala medier. För detta ändamål har en generell modell baserad på antingen Resource Description Framework (RDF)<sup>8</sup> eller Atom<sup>9</sup>, som Web 2.0-baserade tjänster förväntas kunna kommunicera med, tagits fram. RDF är ett generellt språk för XML-uppmärkning av information på webben medan Atom ursprungligen är ett XML-format för syndikering av "feeds" på samma sätt som RSS. Det skall alltså noteras att OAI-ORE, precis som namnet anger, syftar till att återanvända och utbyta information i andra sammanhang än de lokala arkiven snarare än att primärt beskriva eller lagra den som i METS och DIDL. Genomslaget av detta initiativ förefaller, ännu så länge, begränsat.

## Bakgrund: Vad har gjorts i Sverige?

En översikt över projekt vid svenska lärosäten med anknytning till komplexa digitala objekt som bedrivits under 2000-talet.

### DiVA

Utvecklingen av ovan nämnda format hänger i första hand samman med den snabba ökningen och användningen av XML-baserade applikationer under 2000-talet. I Sverige startade under hösten år 2000 projektet DiVA (Digitala vetenskapliga arkivet)<sup>10</sup> vid den nystartade Enheten för digital publicering vid Uppsala universitetsbibliotek vars mål var att skapa ett publiceringssystem för universitetets avhandlingar och andra publikationer i digitalt format. Redan från början var förutsättningen att systemet skulle utgå från XML och komplexa objekt men något "färdigt" format fanns inte tillhanda när det började utvecklas. Därför togs ett nytt eget sådant, DiVA Document Format, fram som redan då stödde beskrivning av komplexa objekt på samma sätt som METS och DIDL nu gör, dvs. att bibliografisk och administrativ metadata kunde kopplas till data om själva filerna och att även hela dokumentets text i XML-format ingick som en del i det samlade digitala objektet. Eftersom denna funktionalitet redan fanns på plats i bakgrunden kunde också DiVA leverera komplexa objekt i exempelvis METS-formatet när detta hade etablerats (2003).

Efterhand som DiVA-systemet sattes i drift och började användas även av andra lärosäten än Uppsala universitet visade det sig dock att DiVA-formatet inte var tillräckligt generellt när det gällde den bibliografiska beskrivningsdelen som initialt enbart var framtagen för avhandlingar och andra examensarbeten och då inte stödde andra typer som exempelvis artiklar eller konferensbidrag. Detta gjorde att formatet inte blev tillräckligt flexibelt och det blev uppenbart att det var en nackdel att specificera det bibliografiska metadataformatet i samma schema som den komplexa strukturen. Om man jämför med METS och DIDL gör man inte heller så där utan det finns endast "behållare" för bibliografiska metadata i strukturen där sedan valfritt metadataformat kan användas vilket gör att formatet blir helt generellt.

Detta gör att METS och DIDL kan fungera på vilka komplexa digitala objekt som helst men ur interoperabilitetssynpunkt måste man då också komma överens om en gemensam bibliografisk

---

<sup>7</sup> Se: <http://www.openarchives.org/ore/>

<sup>8</sup> Se: <http://www.w3.org/TR/rdf-syntax-grammar/>

<sup>9</sup> Se: <http://www.ietf.org/rfc/rfc4287>

<sup>10</sup> Se: <http://www.dlib.org/dlib/november03/muller/11muller.html>

metadastandard om även metadata från olika datakällor skall kunna samordnas. Man kan jämföra med The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)<sup>11</sup> som definierar hur metadataposter skall kunna hämtas men inte hur metadata i sig skall utformas utan i elementet *metadata* i OAI-PMH-formatet kan olika format plockas in.

## Programvaror för öppna arkiv

Som Nelson och Van de Sompel också noterar går utvecklingen av system för lokala arkiv hand i hand med hanteringen av komplexa objekt och nämner Fedora<sup>12</sup> och DSpace<sup>13</sup> som två publiceringssystem med stor användarbas. Fedora listar (2010) 172 kända projekt och DSpace 876 kända installationer på respektive webbplats.

Dessa två ligger numera också till grund för de publiceringssystem som används av en stor majoritet av de svenska lärosätena: DiVA<sup>14</sup> (används av 27 lärosäten) som helt bygger på open source-programvaror med Fedora som bas samt BADA<sup>15</sup> vid Högskolan i Borås och MUEP<sup>16</sup> vid Malmö Högskola som båda använder DSpace.

Skillnaden mellan dessa system ligger främst i att, som Nelson/ Van de Sompel beskriver den, Fedora fungerar som "*middle ware*" som kräver integration med en lokal miljö medan DSpace är ett "*ready-to-deploy*" system.

## Beskrivande metadata i de öppna arkiven

Utvecklingen av det nya DiVA-systemet (2008) tog naturligtvis med sig de erfarenheter som gjorts i "gamla" DiVA och de problem med ett strikt metadataformat specificerat enligt ett särskilt schema som nämnts ovan har nu lösts på det sättet att ett internt "enkelt" metadataformat i XML används i Fedora som i huvudsak håller reda på hierarkiska relationer samt interna ID:n men utgör även basen för sökning av posterna. Detta format visas sedan aldrig externt utan där används olika varianter av mer standardiserade utbytesformat som t.ex. MODS för export till SwePub eller MARC-XML för export till Libris liksom andra enklare format för informationsspridning som Dublin Core, RSS och Atom. Detta innebär att man inte behöver anpassa primärdata till färdigspecifierade scheman vilket gör det mycket enklare att lägga till nya element eller relationer och man behöver inte kompromissa med vilka element som skall tas med eller inte (eftersom alla befintliga scheman i praktiken saknar något element som behövs ändå). Utgångspunkten är alltså att man skall kunna registrera de data som behövs istället för att utgå från vilka data som är tillåtna enligt ett visst schema. Detta medför visserligen att mycket arbete måste läggas ner på att utveckla gränssnitten till systemet men i det långa loppet blir det avsevärt enklare att bygga på och lägga till saker. I synnerhet som det hela tiden ställs nya krav på olika sorters funktionalitet och uppgifter i systemen är detta en stor fördel och den flexibilitet som tidigare saknades har nu uppnåtts.

Vad gäller DSpace är såväl gränssnitt som metadata-schema som redan nämnts "färdigförpackade" i systemet men kan självfallet modifieras efter lokala behov. Detta betyder att systemet rimligtvis är

---

<sup>11</sup> Se: <http://www.openarchives.org/OAI/openarchivesprotocol.html>

<sup>12</sup> Se: <http://www.fedora-commons.org/>

<sup>13</sup> Se: <http://www.dspace.org/>

<sup>14</sup> Se: <http://www.diva-portal.org/>

<sup>15</sup> Se: <http://bada.hb.se/>

<sup>16</sup> Se: <http://dspace.mah.se/>

enkelt att sätta upp men att flexibiliteten i gränssnitten och metadataformatet trots allt kan vara mer begränsad om man skall utnyttja denna fördel fullt ut.

Det finns generellt sett tre olika sorters metadata<sup>17</sup> där systemet, enligt dokumentationen, för beskrivande metadata stödjer ”*multiple flat metadata schemas for describing an item*”. Ett qualified Dublin Core metadata schema baserat på the Library Application Profile<sup>18</sup> erbjuds som default.

## Platta och XML-baserade metadataformat

Att använda ett *platt* format som Dublin Core innebär vissa problem med att leverera metadataposter enligt SwePub-MODS-formatet<sup>19</sup> eller andra tjänster där relationer mellan olika element är viktiga. Det kan ses som aningen anmärkningsvärt att arkitekterna bakom format eller applikationer som OAI-PMH, DSpace, DIDL osv. har negligerat denna aspekt när det gäller beskrivande metadata när den är helt central för själva protokollen eller applikationerna i sig. Anledningen är sannolikt att utvecklingen har varit mer teknikdriven än innehållsdriven samtidigt som Herbert Van de Sompel har pekat på problemen att få in metadata av hög kvalitet<sup>20</sup>. Detta hänger troligen också samman med att applikationer som DSpace i första hand används för självregistrering av metadata och publikationer av forskare/författare snarare än professionell katalogisering av bibliotekarier eller andra specialister och att därför en lägre kvalitetsnivå förväntas redan från början. Utgångspunkten har också varit att det skall vara ”enkelt för forskaren” att registrera och tonvikten i de öppna arkiven har, i alla fall tidigare, mer legat på att få en stor kvantitativ mängd publikationer snarare än korrekta bibliografiska referenser. I vilket fall har detta har bidragit till att Dublin Core blivit en de facto-standard för exempelvis både OAI-PMH och DSpace.

Normalt sett har detta inte heller så stor betydelse för sökbarheten eller identifieringen av ett visst digitalt objekt men i fallet publikationsdatabaser är den bibliografiska kvaliteten avgörande och det är en nödvändig funktion att kunna koppla exempelvis en viss författare till en viss institution som i exemplet nedan. I den platta strukturen kan endast de olika grupperna räknas upp var för sig men det är inte möjligt att koppla vilken författare som hör till vilken institution.

Just koppling mellan författare och institution utpekas också som en särskilt viktig relation i rapporten *Kvalitet och publikationsdatabaser*<sup>21</sup>. Enligt rapporten bör denna uppgift göras obligatorisk i publikationsdatabaserna om posterna skall kunna användas för bibliometriska undersökningar.

### Hierarkisk struktur

- Författare nummer 1
  - Institution nummer 1
  - Institution nummer 3
- Författare nummer 2
  - Institution nummer 1
- Författare nummer 3
  - Institution nummer 2

<sup>17</sup> Se: [http://www.dspace.org/1\\_6\\_1Documentation/ch02.html#N101F1](http://www.dspace.org/1_6_1Documentation/ch02.html#N101F1)

<sup>18</sup> Se: <http://www.dublincore.org/documents/library-application-profile/>

<sup>19</sup> Se: [http://www.kb.se/dokument/Libris/teknisk\\_information/swepub\\_format.pdf](http://www.kb.se/dokument/Libris/teknisk_information/swepub_format.pdf)

<sup>20</sup> Se: [http://www.kb.se/dokument/Om/projekt/langtidsbevarande\\_utvarderingsrapport.pdf](http://www.kb.se/dokument/Om/projekt/langtidsbevarande_utvarderingsrapport.pdf)

<sup>21</sup> Se: [http://gupea.ub.gu.se/bitstream/2077/22420/4/gupea\\_2077\\_22420\\_4.pdf](http://gupea.ub.gu.se/bitstream/2077/22420/4/gupea_2077_22420_4.pdf) s. 25

### Platt struktur

- Författare nummer 1
- Författare nummer 2
- Författare nummer 3
- Institution nummer 1
- Institution nummer 2
- Institution nummer 3

## Metadatastandarder för att komplettera Dublin Core

I projektet *Modell för att hantera komplexa digitala objekt i öppna arkiv – med utgångspunkt i konst och design*, som utfördes vid Högskolan i Borås och som är en direkt föregångare till detta projekt, undersöktes om det fanns lämpliga befintliga metadatastandarder att komplettera Dublin Core med för ämnesområdet konst/design. Dublin Core är, som redan sagts, det metadataformat som används i Borås lokala arkiv BADA (DSpace).

Via ett utvalt testobjekt<sup>22</sup> konstaterade man att de element som finns i Dublin Core inte räcker till för att beskriva detta. Resultatet, när objektet analyserats tillsammans med upphovsmännen, blev att fyra ytterligare element som saknades i den befintliga DC-modellen lades till.

Denna situation är analog med de problem som tidigare nämnts i samband med "gamla" DiVA-systemet: om man från början bestämt sig för ett standardiserat format kommer man förr eller senare att upptäcka att något element saknas i publiceringssystemen. Eftersom dessa system utvecklas så att ständigt nya typer av objekt skall integreras eller införlivas i strukturen är det viktigt att det finns en beredskap för detta.

Inom projektet gjordes också ett studiebesök hos Kultur Project<sup>23</sup> i England (som avslutades 2009) vars övergripande mål var att skapa en modell för ett öppet arkiv för forskningsresultat inom konst och konsthantverk. Intressant att notera är att man inom det projektet också försökte utgå från "forskarna och vad de ville beskriva"<sup>24</sup> snarare än en färdig metadatastandard och sedan försökte "mappa" detta mot Dublin Core (eftersom man utgår från systemet ePrints som också bygger på DC) som metadataformat.

I Borås-projektet tittade man på andra befintliga metadatastandarder för att hitta de element som saknades i DC och hittade två format som skulle vara aktuella utifrån de riktlinjer som definierats: *CDWA (Categories for the Description for the Works of Art)*<sup>25</sup> samt *VRA Core (Visual Arts Association Core)*<sup>26</sup>.

Av dessa valdes CDWA Lite, en förenklad version av CDWA. Här hittade man fält för *Object/Work Type, Materials & techniques* samt *Current location*. Dock återfanns inget relevant fält för *Physical description* så därför skapade man ett eget lokalt metadata-schema med detta element. Dessa har sedan lagts till befintliga DC-element. Som framgår i posten<sup>27</sup> finns nu alltså tre olika namnrymder i

<sup>22</sup> Se: <http://hdl.handle.net/2320/3925>

<sup>23</sup> Se: <http://kultur.eprints.org/index.htm>

<sup>24</sup> Se: [http://bada.hb.se/bitstream/2320/5255/1/Suld\\_%20Arvidsson\\_slutrapport.pdf](http://bada.hb.se/bitstream/2320/5255/1/Suld_%20Arvidsson_slutrapport.pdf)

<sup>25</sup> Se: [http://www.getty.edu/research/conducting\\_research/standards/cdwa/cdwalite.html](http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite.html)

<sup>26</sup> Se: <http://www.vraweb.org/projects/vracore4/index.html>

<sup>27</sup> Se: [http://bada.hb.se/handle/2320/3925?mode=full&submit\\_simple=Visa+fullst%C3%A4ndig+post](http://bada.hb.se/handle/2320/3925?mode=full&submit_simple=Visa+fullst%C3%A4ndig+post)



metadatatypen: dc, cdwalite samt bada. Detta är en ”tekniskt” helt korrekt men i praktiken lite otymplig lösning som troligen kan innebära vissa interoperabilitetsproblem om posten exporteras på detta sätt. Hursomhelst är det en utmärkt illustration till problemet med ett på förhand strikt specificerat metadataformat. Som också konstateras i rapporten skulle säkerligen ytterligare saknade element identifierats om man tittat på fler poster än en.

## Komplexa digitala objekt och långtidsbevarande

I vår ursprungliga projektansökan ingick också att undersöka den aktuella utvecklingen inom området långtidsbevarande, med specifik hänsyn till konstnärligt icke-textbaserat material, men denna del skars bort av Styrgruppen för KB:s program för Open Access i Sverige<sup>28</sup> med hänvisning till att Kungl. biblioteket kommer att lösa den.

Det skall påpekas att ett antal KB-finansierade projekt mellan åren 2003 och 2007 som bedrevs vid Enheten för digital publicering vid Uppsala universitetsbibliotek då grundligt utredde denna fråga och inom uppföljningsprojektet *Långtidsbevarande – från prototyp till fullfärdiga verktyg*<sup>29</sup> dessutom levererade flera konkreta tekniska lösningar och applikationer till KB för att framställa komplexa objekt och omvandla dem till leveranspaket (*Submission Information Packages, SIPs*) samt applikationer för att ta emot/kontrollera sådana digitala leveranser enligt OAIS-modellen<sup>30</sup>. Även dessa applikationer använde METS respektive DIDL för att sätta samman de komplexa objekten.

Detta arbete var avsett som en grund till en infrastruktur för digitala leveranser av komplexa objekt från de lokala publiceringssystemen till Kungl. biblioteket (eller andra arkiv) men av en mängd, i huvudsak organisatoriska och kommunikationsmässiga skäl (som till viss del även framgår av projektutvärderingen)<sup>31</sup>, kom aldrig dessa lösningar att testas eller vidareutvecklas på KB.

Eftersom digital arkivering nu håller på att införas i större skala vid flera lärosäten, bland annat Uppsala universitet, och publiceringssystemen då måste kunna leverera både data och metadata i arkivpaket för exempelvis avhandlingar och examensarbeten till arkiven har fokus i denna fråga, åtminstone för DiVA:s del, flyttats till samverkan med universitetsarkiven istället.

För de specifikt konstnärliga utbildningarna är dessutom behovet av förbättrade arkivlösningar än mer angeläget med tanke på de lagringsmedier som i dagsläget används. Publikationer levereras och sparas i dagsläget huvudsakligen på fysiska medier som CD eller DVD-skivor eller till och med disketter.

Man kan även tänka sig att man har olika varianter av filerna för lagring och visning där lagringsvarianten är den ursprungliga. En annan tanke som framförts är att lagra filerna men bara visa smakprov eller trailers (av upphovsrättsskäl eller kvalitetsskäl). Hög ljudkvalitet eller bildkvalitet kan i högsta grad vara del av det konstnärliga uttrycket och det är inte alltid möjligt att åstadkomma via webben. En del filer kan vara väldigt stora och inte så lämpliga att ladda upp via ett

---

<sup>28</sup> Se: <http://www.kb.se/bibliotek/referensgrupper/Styrgruppen-Open-Access-/>

<sup>29</sup> Se: [http://www.kb.se/dokument/Om/projekt/open\\_access/langtidsbevarande\\_stefan\\_andersson.pdf](http://www.kb.se/dokument/Om/projekt/open_access/langtidsbevarande_stefan_andersson.pdf).

Beklagligtvis fungerar inte de interna länkarna i rapporten längre men all källkod och alla applikationer har överförts till Kungl. biblioteket.

<sup>30</sup> OAIS (Open Archival Information System) är ett ramverk för elektroniska arkiv som hanterar digital information ursprungligen utvecklad av NASA och en de facto standard för e-arkiv.

<sup>31</sup> Se: [http://www.kb.se/dokument/Om/projekt/langtidsbevarande\\_utvarderingsrapport.pdf](http://www.kb.se/dokument/Om/projekt/langtidsbevarande_utvarderingsrapport.pdf)

webbformulär utan andra lösningar kanske måste sättas upp för sådana överföringar. Likaså är sannolikt streaming det bästa sättet att tillgängliggöra många filer på webben.

Det är också önskvärt att bättre kunna kontrollera hur filerna öppnas från webbläsaren och skapa enhetligare lösningar för det där även "spelaren" ingår i gränssnittet.

## Specifika frågor för konstnärliga publikationer

Några punkter som dykt upp under publikationsinventeringen:

### Publikationstypen "Konstnärligt forsknings- och utvecklingsarbete"

När specifikationen för metadataformat till SwePub gick på remiss kom synpunkten från de konstnärliga högskolorna att publikationstypen *Konstnärlig forsknings- och utvecklingsrapport* dels skulle byta namn till *Konstnärligt forsknings- och utvecklingsarbete* dels inte längre vara en egen publikationstyp utan snarare ett attribut till de övriga som artikel, doktorsavhandling etc.

Detta ansågs som ett utmärkt förslag eftersom det då också blir möjligt att jämföra publiceringen mellan olika lärosäten på samma villkor samtidigt som det då går att skilja ut de publikationer som har ett konstnärligt innehåll. Dock är inte begreppet *konstnärlig forskning* helt oproblematiskt. Se t.ex. Henk Borgdorff: *Artistic Research and Academia: an Uneasy Relationship*<sup>32</sup>. Vid förfrågan till de konstnärliga högskolorna har också framkommit att:

*"Fältet konstnärlig forskning är under formering. Vissa skulle svara ja på frågan om konstnärlig forskning är en egen typ av kunskapsproduktion – andra nej".*

Denna diskussion är i sig inte avgörande för hur begreppet skall behandlas när det gäller registrering av publikationer, om man fortfarande utgår från att det är relevant i detta perspektiv, men däremot krävs i så fall en tydligare definition av vad som skall märkas upp som konstnärlig forskning i publiceringssystemen för att göra en sådan kategorisering meningsfull och särskilja den från forskning *om* konst (typ musikvetenskap eller konstvetenskap) eller forskning som utvecklar det rent instrumentella perspektivet (hantverkskunnandet).

En utgångspunkt skulle kunna vara VR:s beskrivning:

*Konstnärlig forskning tar sin utgångspunkt i den konstnärliga processen och verksamheten. Forskningen, som kan beröra alla konstarter, är praktikbaserad och inkluderar en intellektuell reflektion för att utveckla ny kunskap. Resultatet från konstnärlig forskning redovisas vanligen i både gestaltande och skriftlig form*<sup>33</sup>.

eller:

---

<sup>32</sup> Se: [http://www.ahk.nl/fileadmin/download/ahk/Lectoraten/Borgdorff\\_publicaties/an-uneasy-relationship.pdf](http://www.ahk.nl/fileadmin/download/ahk/Lectoraten/Borgdorff_publicaties/an-uneasy-relationship.pdf)

<sup>33</sup> Definition från VR:

<http://www.vr.se/forskningvistodjer/forskningsomraden/konstnarligforskningochutveckling.4.3376a56c12918b8d17b80001186.html>

*Konstnärlig forskning är forskning som bedrivs av konstnärer som utifrån beprövad erfarenhet och på konstnärlig grund forskar i sin konst<sup>34</sup>.*

Den senare varianten är mer specifik men troligen för snäv för att användas eftersom allt konstnärligt arbete vid lärosätena inte utförs av konstnärer.

Det skall också betonas att även om konstnärlig forskning publiceras i annan form än text är det inte något som behöver vara unikt just för detta område.

Efter diskussion med konstnärliga högskolor har vi även kommit fram till att termen *konstnärligt arbete* är att föredra framför *konstnärligt forsknings- och utvecklingsarbete* eller *konstnärlig forskning* vid publikationsregistrering. Även studentarbeten kan omfattas av konstnärligt arbete. Konstnärligt arbete skall i publiceringssystemen inte heller ses som en "publikationstyp" motsvarande bok, artikel etc. utan som en speciell *typ av innehåll* istället.

## **Fysiska objekt i relation till digitala avbildningar**

När det gäller det fysiska resultatet av arbeten vid de konstnärliga utbildningarna har det konstaterats att i normalfallet behåller upphovsmannen detta (liksom i Borås testpost) och den digitala avbildningen är det format som sparas vid lärosätet. Därför behövs normalt ingen särskild uppgift lämnas om detta i metadatabeskrivningen.

## **Särskilda rättighetsfrågor**

Inom musikområdet har särskilda behov relaterade till rättighetsfrågor framkommit. Inom övriga områden framstår rättighetsfrågor inte som ett problem. När det gäller musik gäller det dels att visst material endast är tillgängligt inom lärosätets domän av upphovsrättsliga skäl. I dagsläget finns inget tekniskt stöd för just detta i publiceringssystem och det har heller inte funnits ett sådant behov tidigare. Dock går det ange olika sorters tillgång på filnivå vilket är en förutsättning (så att posten och eventuellt "fria" filer kan visas även om en viss fil är begränsad). För musik kan det också vara viktigt att olika typer av upphovsmän eller andra roller kan anges i metadata.

## **Publikationstyper för icke-textbaserade objekt i de lokala systemen**

### **Samlingsposter**

Resultaten från de konstnärliga utbildningarna oftast består av en textdel och en icke-textdel och dessa då måste kunna kopplas ihop. Precis som en sammanläggningsavhandling eller en antologi kan också exempelvis en musik-CD ha en egen titel men sedan innehålla ett antal separata spår som kan behöva beskrivas var för sig precis som en artikel i en sammanläggningsavhandling beskrivs som en egen post och med en egen "fulltext" kopplad. Ett annat exempel är en utställning som har ett eget namn osv. men som även består av individuella verk som kan kopplas till en "huvudpost".

### **Gemensamma element för alla publikationstyper**

De flesta element är gemensamma för alla publikationstyper. Följande element kan tillämpas på alla publikationstyper och fungerar även för icke-textobjekt:

- Typ av innehåll (vetenskapligt/övrigt) + konstnärligt arbete ja/nej

---

<sup>34</sup> Se: Eeva Lilja, september 2007, <http://www.efvalilja.se/pdf/EnKoPalsen07.pdf>

- Upphovsman (fler roller än författare/redaktör skall då kunna anges)
- Titel
  - Huvudtitel
  - Undertitel
- Språk
- Utgivningsår
- Identifikator (intern, URN)
- URL (extern länk)
- Ämneskategori (nationell)
- Ämneskategori (lokal)
- Nyckelord
- Abstract/beskrivning
- Anmärkning

### Specifika nya element för olika publikationstyper

I både DiVA och BADA är det fullt möjligt att ladda upp fil-/dokumenttyper som bilder, video eller ljud som separata filer till en post men det som inte är möjligt är att beskriva exempelvis en bild eller en video på samma nivå som ett textdokument, dvs. med egna beskrivande metadata. I kartläggningen av konstnärliga publikationer, liksom i Borås pilotprojekt, har det tydligt framkommit att det finns ett behov av att göra det. De element som då behövs läggs till är, som redan framgick i Borås pilotstudie, de som gäller *objekttyp*, *material/teknik* samt *fysisk beskrivning* som är annorlunda i förhållande till textdokument.

Nedan följer en lista över förslag till specifika element för olika publikationstyper relaterade till de konstnärliga utbildningarna:

\* = underelement som redan finns för andra publikationstyper.

**Fet stil** = nytt element

#### Musiktryck, noter

- Upplaga\*
- Utgivningsort\*
- Förlag\*
- **Typ (partitur, stämma etc.)** Lokal lista
- **Fysisk beskrivning (antal sidor eller antal enheter)**
- **Genre (ex. jazz, klassisk, pop)** Lokal lista
- Serie\*
- ISBN\*

#### Ljudinspelning-musik eller -ej musik

- Utgivningsort\*
- Förlag\*
- **Fysisk beskrivning (om fysisk enhet, ex. CD)**
- **Längd (tid): hh:mm:ss**
- **Genre (ex. jazz, klassisk, pop)** Lokal lista

## Bild

- **Typ (ex. fotografi, grafiskt blad)** Lokal lista
- **Material (ex. papper, trä, duk)** Lokal lista
- **Teknik (ex. akvarell, olja)** Lokal lista
- **Mått (höjd x bredd x tjocklek)**
- **Övrig fysisk beskrivning (färg, form etc.)**

## Rörlig bild, video

- Utgivningsort\*
- Förlag\*
- **Fysisk beskrivning (om fysisk enhet, ex. DVD, videokassett)**
- **Längd (tid): hh:mm:ss**

## Föremål

- **Typ (ex. skulptur, kläder, väggbonad)** Lokal lista
- **Material (ex. tyg, plywood, marmor)** Lokal lista
- **Teknik (ex. akvarell, olja)** Lokal lista
- **Mått (höjd x bredd x tjocklek)**
- **Övrig fysisk beskrivning (färg, form etc.)**

## Programvara/multimedia

- **Typ (data/programkod/webbplats/multimedia)** Förslag till gemensam lista

## Blandat innehåll

Denna resurstyp används för publikationer som består av flera olika typer av material (text, ljud, bild etc.) om hela publikationen beskrivs i en post eller en gemensam huvudpost.

- **Fysisk beskrivning (t.ex. olika resurstyper som ingår i publikationen)**

I kartläggningen fick vi in ett par exempelpublikationer utformade som webbplatser med länkar till ett antal olika webbsidor eller andra objekt. Rent tekniskt skulle dessa filer kunna laddas upp, beskrivas var för sig och kopplas ihop i systemen enligt ovanstående modell men vi tror inte att det är realistiskt att studenter eller forskare har möjlighet att göra detta för objekt som innehåller ett stort antal filer. Däremot är det önskvärt att en webbplats eller webbsidor också kan visas externt på samma sätt som en "fulltext" i PDF. Vår tanke är därför att upphovsmannen skall kunna ladda upp en webbplats som ett "arkiv" (alla filerna i ett paket) och att den sedan packas upp och publiceras på en särskild extern webbserver som kopplas till publiceringssystemen.

## Kompatibilitet med SwePub

Allmänt skall sägas att SwePub-formatet är ett *minimiformat för interoperabilitet* mellan olika lärosätets poster och att de lokala systemen normalt innehåller betydligt mer information än den som hämtas till SwePub. MODS-formatet som används i SwePub är, precis som MARC-formatet som används i Libris/Voyager, ett generellt format som kan beskriva alla typer av objekt. Detta gör att inga tillägg egentligen behövs göras i SwePub-formatet för att hantera icke-textpublikationer. Däremot bör några förtydligande exempel läggas till i formatspecifikationen.

## Publikationstyper

Indelningen av publikationstyper i SwePub-formatet täcker primärt textbaserade publikationer. Annat hamnar under publikationstypen *Övrigt*.

### Befintliga publikationstyper i SwePub

- Artikel i tidskrift
- Artikel, forskningsöversikt
- Artikel, recension
- Bok
- Doktorsavhandling, monografi
- Doktorsavhandling, sammanläggning
- Kapitel i bok, del av antologi
- Konferensbidrag
- Konstnärligt forsknings- och utvecklingsarbete (kombineras med någon av de övriga)
- Licentiatavhandling, monografi
- Licentiatavhandling, sammanläggning
- Manuskript (preprint)
- Patent
- Proceedings (redaktörskap)
- Rapport
- Samlingsverk (redaktörskap)
- Övrigt

### Förslag till ändringar

1. Termen *konstnärligt arbete* ersätter *konstnärligt forsknings- och utvecklingsarbete* i publiceringssystemen
2. Konstnärligt arbete klassas som en särskild *typ av innehåll* som kan sättas till ja/nej (yes/no) istället för publikationstyp
3. En definition med utgångspunkt i ovanstående beskrivningar fastställs på nationell nivå och dokumenteras i SwePubs formatspecifikation som även justeras vad gäller punkt 1 och 2

### Andra publikationstyper (ej textbaserade)

I SwePub MODS finns redan definierat hur andra resurstyper än text skall anges i elementet `<typeOfResource>`. Följande värden är giltiga där och dessa täcker rimligen de varianter som är tänkbara. Genom att ange dessa går det alltså att göra dessa olika typer sökbara i SwePub (i dagsläget finns dock inte denna sökmöjlighet i formuläret).

- cartographic (kartmaterial)
- notated music (musiktryck, noter)
- sound recording-musical (ljudinspelning-musik)
- sound recording-nonmusical (ljudinspelning-ej musik)
- still image (bild)
- moving image (rörlig bild, video)
- three dimensional object (föremål)
- software, multimedia
- mixed material (samling med blandat innehåll)

Frågan är sedan om och, i så fall, hur dessa övergripande typer skall underindelas på motsvarande sätt som textdokument delas in i artiklar, böcker etc. Det behövs sannolikt inte heller på nationell nivå och indelningen av textdokument är i första hand gjord för bibliometriska/bibliografiska syften. I de lokala systemen går det däremot att underhålla egna listor över mer specificerade objekttyper. I testposten från Borås finns exempelvis termerna *Costumes* och *Wall-hanging* som exempel på objekttyper som skulle kunna ingå i en lokal lista. Dessa typer kan också levereras till SwePub i <genre>-elementet om så önskas för sökning.

Den befintliga modellen för publikationstyper och medietyper i SwePub-formatet med tillägg av en särskild innehållstyp för konstnärligt arbete bör således vara tillräcklig för att skilja ut dels konstnärliga arbeten och dels icke-textbaserade publikationer.

## Exempel

### Nuvarande SwePub-format

```
<genre authority="svap" type="publicationType">kfa</genre>
<genre authority="svap" type="publicationType">rap</genre>
<genre authority="svap" type="contentType">vet</genre>
<typeOfResource>sound recording-musical</typeOfResource>
```

### Föreslaget SwePub-format

```
<genre authority="svap" type="artistic">yes</genre>
<genre authority="svap" type="publicationType">rap</genre>
<genre authority="svap" type="contentType">vet</genre>
<typeOfResource>sound recording-musical</typeOfResource>
```

## Samlingsposter

En samling eller "huvudpost" skulle i SwePub-formatet, för tydlighetens skull, kunna märkas upp med attributet collection="yes" i <typeOfResource>-elementet men det är inte helt nödvändigt. I övrigt finns redan <relatedItem>-elementet i SwePub-formatet för att hierarkiskt koppla delar till en samlingspost. Ett problem som kan uppstå här är dock att alla delposter kanske inte uppfyller SwePubs valideringskrav varför informationen i relatedItem i SwePub-poster troligen för säkerhets skull bör begränsas till externa länkar direkt till de lokala systemens poster.

## Upphovsman

I fältet *Upphovsman* bör förtydligas att andra roller (t.ex. kompositör, konstnär, musiker eller producent) än författare och redaktör också bör anges enligt *MARC relators*<sup>35</sup>.

## Språk

Fältet *språk* kan vara ett problem för icke-text publikationer eftersom det är ett obligatoriskt element i SwePub och exempelvis en målning inte har ett "språk". Detta kan lösas på olika sätt: antingen anges det språk som titeln (eller benämningen på objektet) har eller så anges språkkoden "zxx" för "icke-språkligt material". En tredje variant är att ange koden "und" för odefinierat språk.

## Fysisk beskrivning

Icke-textpublikationer kräver en annan fysisk beskrivning än text. När det gäller mått eller tidsuppgifter kan dessa läggas i samma element som antal sidor för böcker: <extent> som är

<sup>35</sup> Se: <http://www.loc.gov/standards/sourcelist/relator-role.html>

underelement till <physicalDescription>. Annan fysisk beskrivning läggs i elementet <note> som underelement till <physicalDescription>. Material får attributet *medium* och teknik får attributet *technique* enligt MODS note types<sup>36</sup>. Exempel bör läggas in i formatspecifikationen.

## Prototyp

Förslag till förbättrad visuell presentation och registrering finns tillgängligt på:

<http://complex.blr.hb.se/>.

Notera att dessa sidor innehåller element i HTML5 och kan därför inte tolkas av vissa webbläsare, bl.a. Internet Explorer. (Dock kan IE 9 som finns för nerladdning i betaversion förhoppningsvis visa sidorna korrekt.) Firefox 3.6+, Safari 5.0+ och Google Chrome 8.0+ är exempel på webbläsare som ska fungera bra.

De olika exemplen visar hur några av ovanstående rekommendationer skulle kunna presenteras konkret på webben.

---

<sup>36</sup> Se: <http://www.loc.gov/standards/mods/mods-notes.html>