

## DIGITALA VETENSKAPLIGA ARKIVET

Några fakta, nuläge, teknisk beskrivning, problemområden och framtidsplaner

### Enheten för digital publicering

*Några fakta om oss:*

Som resultat av *Utredning om universitetets vetenskapliga publicering på www*<sup>1</sup> finns det vid UUB sedan september 2000 en enhet för digital publicering. Sedan november 2000 arbetar vi i fullbemannning (3 tjänster) med utvecklingen av nya tekniska lösningar för elektronisk publicering av i första hand sammanläggningsavhandlingar och elektronisk spikning. Samtidigt har vi i drift en enkel äldre lösning som vi anpassade för spikning och elektronisk publicering av fulltextfiler i pdf-format.

Uppdraget, enligt utredningen, är följande:

#### Uppdrag:

- ⇒ skapa tekniska lösningar och ett fungerande arbetsflöde för elektronisk spikning samt fulltextpublicering av avhandlingar, examensarbeten, arbetsrapporter och andra typer av vetenskapliga publikationer
- ⇒ bevaka utvecklingen på områden som har anknytning till digital publicering

Under en begränsad period finansieras enheten via centralt tilldelade medel, därefter kommer publiceringen troligen att avgiftsbeläggas för att täcka driftskostnader samt vidareutveckling av systemet.<sup>2</sup>

Detta krav, att på en relativt kort utvecklingsperiod och en snabb övergång till produktion, som kan bli självfinansierad, ställde oss inför ett par strategiska val som speglas både i utformningen av hela konceptet, men också i valet av konkreta tekniska lösningar.

Därför bestämde vi oss att redan från början satsa på en bredare tolkning av vårt uppdrag och skapa ett system som kan hantera vilket fulltextdokument som helst.

Så tillkom konceptet av digitala vetenskapliga arkivet.

---

<sup>1</sup> se: <http://w3.ub.uu.se/diverse/publ.pdf>; för mera bakgrundsinformation se:

<http://publications.uu.se/epcentre/background.xsql?lang=sv>

<sup>2</sup> När det gäller vidareutveckling av systemet nämner man dock i utredningen på s. 26 att “*I uppbyggnadsfasen bör tillräckliga utvecklingsresurser tillföras via de föreslagna utvecklingsprojekten. I uppbyggnadsskedet bör även extra resurser finnas för externt konsultstöd och investeringar. Medel måste också reserveras för den framtida driften av servrar etc för publiceringen.*”

Utredning om universitetets vetenskapliga publicering på www.

URL:<http://w3.ub.uu.se/diverse/publ.pdf> (2001-10-04), s. 26

När det gäller val av teknik satsar vi på en säker lagring i databasmiljö, som ger oss frihet att redan från börjankunna använda alla nya teknologier och standarder, även de som är under utveckling. På detta sätt kan vi säkra att en väsentlig del av den publicerade informationen utan större problem blir kompatibel även med nya kommande teknologier och standarder. När det gäller hela fulltextfiler kommer de initialt att lagras, spridas och arkiveras i pdf. Men vi är beredda att övergå till andra lagringsformat (XML) så fort stabila standarder tas fram.

Sammanfattningsvis kan man säga att våra lösningar är väldigt pragmatiska och att vi försöker skapa balans mellan användning av nya teknologier och säkring av driften ur ett långtidsperspektiv. Ambitionen är att utforma lösningar som kan möta både dagens och framtida behov och standarder.

### **Digitala vetenskapliga arkivet (DiVA)**

#### **Målet:**

- ⇒ Skapa ett sökbart arkiv över alla vid Uppsala universitet publicerade digitala dokument som är samsökningsbart i strukturerad sökning med bibliotekskatalogen och andra relevanta söktjänster samt tillgängliggöra metadata för informationstjänster som är relevanta för spridning av information om Uppsala-forskning
- ⇒ I samarbete med Kungliga biblioteket garantera långtidslagring och tillgänglighet i långtidsperspektiv
- ⇒ Fungera som kompetensskälla på områden inom digital publicering, digitala bibliotek

Arbetsflödet för inflöde samt utprodukter är dock anpassat till vårt huvuduppdrag – elektronisk spikning och fulltextpublicering av sammanfattningsdelen av sammanläggningsavhandlingar.

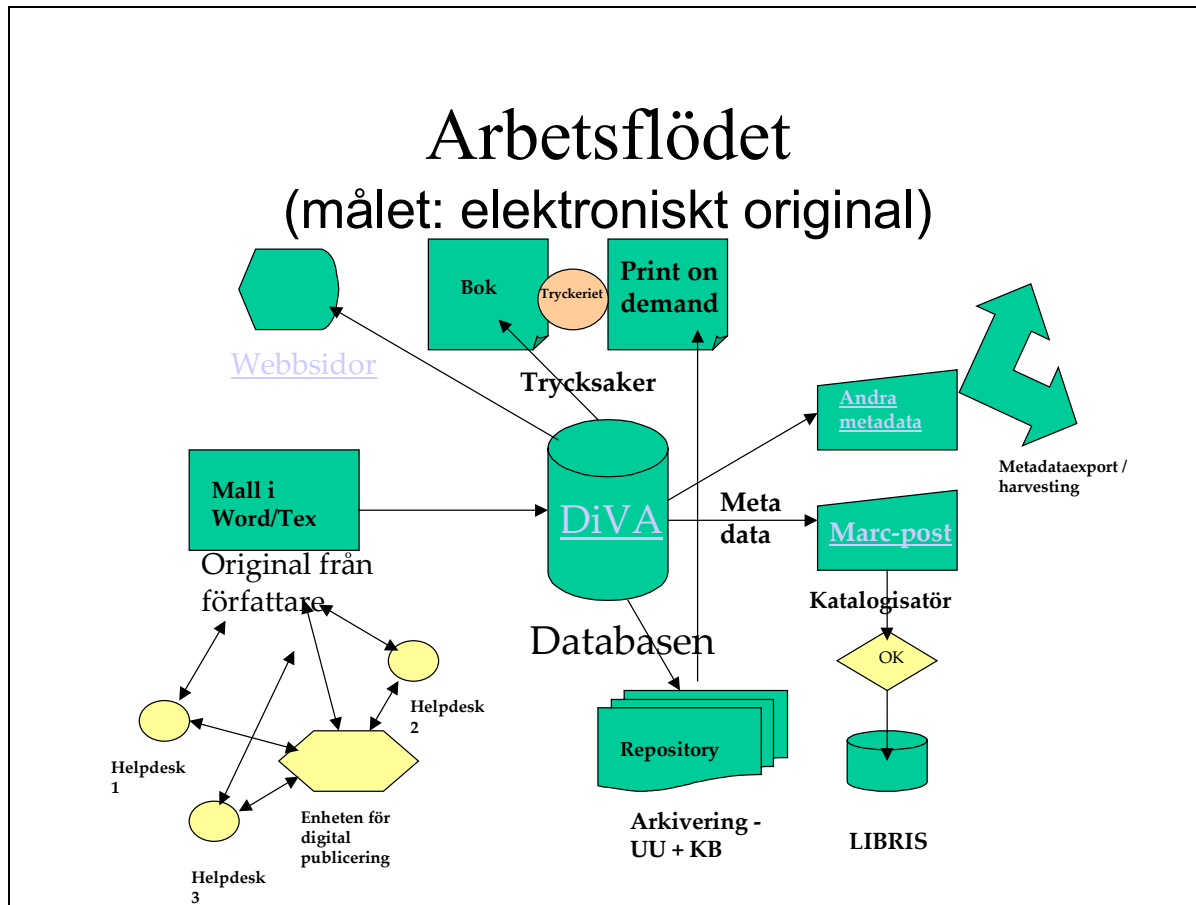
Genom utformningen av våra tekniska lösningar är vi förberedda att publicera även andra typer av publikationer. Framtida utvecklingsinsatser kommer då att koncentreras kring arbetsflöden som kan variera beroende på publikationstyp

#### **Vision:**

Genom att implementera ett bra arbetsflöde (dvs. mellan upphovsmän - biblioteket - tryckeriet - slutanvändaren) skall våra tekniska lösningar vara användarvänliga för alla inblandade parter samtidigt som de kommer att kräva minimalt underhåll.

Lösningar skall utformas så att man kan möta både dagens och framtida behov och standarder

## Arbetsflödet



## Nuläge

**Ett bassystem** är utvecklat och kommer att lanseras den 22 oktober

(<http://publications.uu.se/theses/>) .

Den bygger på en Oracle databas som lagrar strukturerad information i Unicode som tas in med hjälp av XML. Även för skapande av dynamiska webbsidor används nya teknologier.

Ett nytt sökgränssnitt som tar hänsyn till sökning på flera språk har utvecklats.

Fulltexter lagras i pdf.

(se detaljerad beskrivning av systemet s.5).

**En ny spikningsmall** introduceras nu. (Word mall som taggar upp information i bakgrunden för konvertering till XML). Fulltextmallen (Word, Tex) är under utveckling.

**Ett nytt arbetsflöde** mellan författare – Enheten för digital publicering – Grafiska redaktionen/ Tryckeriet som fokuserar på elektroniskt original är under diskussion och kommer förhoppningsvis att införas snart. För att kunna genomföra ändringar i arbetsflödet krävs det dock en del formella beslut som vi otåligt väntar på.

**Ett nytt arbetsflöde** inom UUB när det gäller rutiner för spikningen har införts. Spikningskvitto genereras automatiskt från databasen. Vi arbetar även på att generera automatiskt underlag för UMI agreement form.

**Ett nytt arbetsflöde** mellan UUB och KB införs från och med oktober. Det gäller katalogisering samt långtidslagring av fulltextpublicerade dokument.

En katalogpost enligt MARC format genereras automatiskt och överförs till den nationella katalogen LIBRIS.

KB kommer även ta emot en arkivkopia på alla fulltextpublicerade avhandlingar (och eller andra dokument som kommer att fulltextpubliceras via DiVA).

I samarbetet kommer praktiskt att testas användning av URN och överföring av information till KB:s ”URN resolution service”.

**En helpdesk** för elektronisk publicering är under utredning. Frågan är hur komplex helpdesken skall vara? Om komplexiteten är stor påverkas omfattningen av arbetet på andra enheter inom universitet... (Grafiska Redaktionen, Informationsavdelningen etc.)

**En dokumentserver** som möjliggör publicering av käll- och annat material som författarna vill hänvisa till i sina referenslistor kommer att sättas upp. I detta fall kommer endast bibliografiska uppgifter att lagras i databasen.

**En referensgrupp** för elektronisk publicering vid Uu har bildats och består av forskare och experter på områden som har anknytning till elektronisk publicering och digitala bibliotek.<sup>3</sup>

**Ett samarbete** mellan de svenska universitetsbiblioteken är under diskussion.

**Internationellt samarbete** har inletts med NDLTD (<http://www.theses.org>) och PhysNet (<http://physnet.uni-oldenburg.de/PhysNet/>). Enheten är även i kontakt med andra projekt från Holland, USA och Tyskland.

---

<sup>3</sup> För mera information se bilagan.

## TEKNISKA LÖSNINGAR

### Serverar

För digital publicering används två serverar. En databasserver och en webbserver.

#### Databasserver

På servern har bara kommersiell programvara installerats. Både databashanteraren och backupprogrammet körs på den.

#### Programvara

- Windows 2000 Advanced Server  
<http://www.microsoft.com/windows2000/advancedserver/default.asp>
- Oracle 8i RDBMS (relational databas management system)  
<http://technet.oracle.com/products/oracle8i/content.html>
- Veritas Backup Exec 8.5  
[http://www.veritas.com/products/category/ProductDetail.jhtml?productId=bent2000&\\_requestid=51298](http://www.veritas.com/products/category/ProductDetail.jhtml?productId=bent2000&_requestid=51298)

#### Tekniska data

Servers viktigaste komponenter:

- Intel dualprocessor moderkort
- 2 Intel Pentium 3 CPU med 900MHz
- 1 GB internt minne
- 2 nätaggat som kan bytas under driften (hotswap)
- [Adaptec Raidcontroller 3200S](#) med 64 MB cache
- 5 18GB SCSI 160 hårddiskar i hotswap ramen som bildar en RAID5 med en hårddisk som hotspare.

#### Backup

Som backupmedier används DLT IV medier och en Tandberg DLT1 (<http://www.tandberg.com/>) bandspelare. Med DLT kan man göra snabba backuper på stora datamängder. En DLT IV sparas upp till 80 GB data.

Backupen sker dagligen (inte på helger) efter grandfather-father-son (GFS) principen med roterande band. Måndag till torsdag görs det differentiella backuper som sparas en vecka. På fredagar görs det en fullständig backup som sparas 4 veckor. Sista fredagen i månaden görs det en fullständig backup som sparas ett år.

### Webbserver

På servern har bara open source programvara installerats. Den sköter kommunikationen mellan användare och databasen och framställer våra webbsidor dynamiskt.

### Programvara

- Redhat Linux 7.0 med brandvägg <http://www.redhat.com/>
- Apache webbserver <http://httpd.apache.org/>
- Tomcat <http://jakarta.apache.org/tomcat/index.html>
- FOP <http://xml.apache.org/fop/index.html>
- Oracle XDK [http://technet.oracle.com/tech/xml/xdk\\_java/content.html](http://technet.oracle.com/tech/xml/xdk_java/content.html)

### Tekniska data

Servers viktigaste komponenter:

- Gigabyte moderkort
- Intel Pentium 3 CPU 500MHz
- 512MB internt minne
- [Adaptec Raidcontroller 2100S](#)
- 2 36GB SCSI 160 hårddiskar som speglas

## Databasen

Databasmanagementsystem (RDBMS)

Som databasmanagementsystem används Oracle 8i Standard Edition i sin aktuella version 8.1.7. Att använda Oracles databashanterare har den stora fördelen att man kan välja mellan olika operativsystem. Oracle 8i har en bra skalbarhet och det finns flera inbyggda funktioner som gör utvecklingen av ett system för digital publicering enklare. Funktionerna är:

- Unicode (UTF-8 och UTF-16) och national language support (NLS)
- Webbserver
- Oracle Text
- XML development kit (XDK)

### *Unicode*

Oracle 8i kan spara alla tecken i unicode. Det är en viktig förutsättning för att kunna spara texter som innehåller olika språk och dess tecken. För mer information titta på <http://www.unicode.org/>.

### *National language support (NLS)*

NLS är en förutsättning för att kunna ta emot texter i olika språk och att kunna hantera texterna på rätt sätt. Det handlar inte bara om tecken. Det handlar också om:

- sorteringen i olika språk
- datumformatet
- tidsformatet
- nummerhanteringen

Beroende på språket kan man automatiskt framställa olika format. Detta används för spikbladsframställning på svenska, engelska, franska och tyska.

### *Oracle Text*

Med Oracle Text är det möjligt att göra databasen sökbar. Där finns det stöd för olika språk och deras egenheter. Det är möjligt att göra indexering beroende på språk. Man kan söka efter diakritiska tecken (t ex ö, å, é, ç) utan att man måste skriva in dem. Det räcker att man söker efter deras bastecken (o, a, e, c). Dessutom kan man skapa flera thesauri. Beroende på dessa kan databasen indexeras. Dessutom är indexering och sökning av XML-taggar möjlig. För mer information titta på <http://technet.oracle.com/products/text/content.html>.

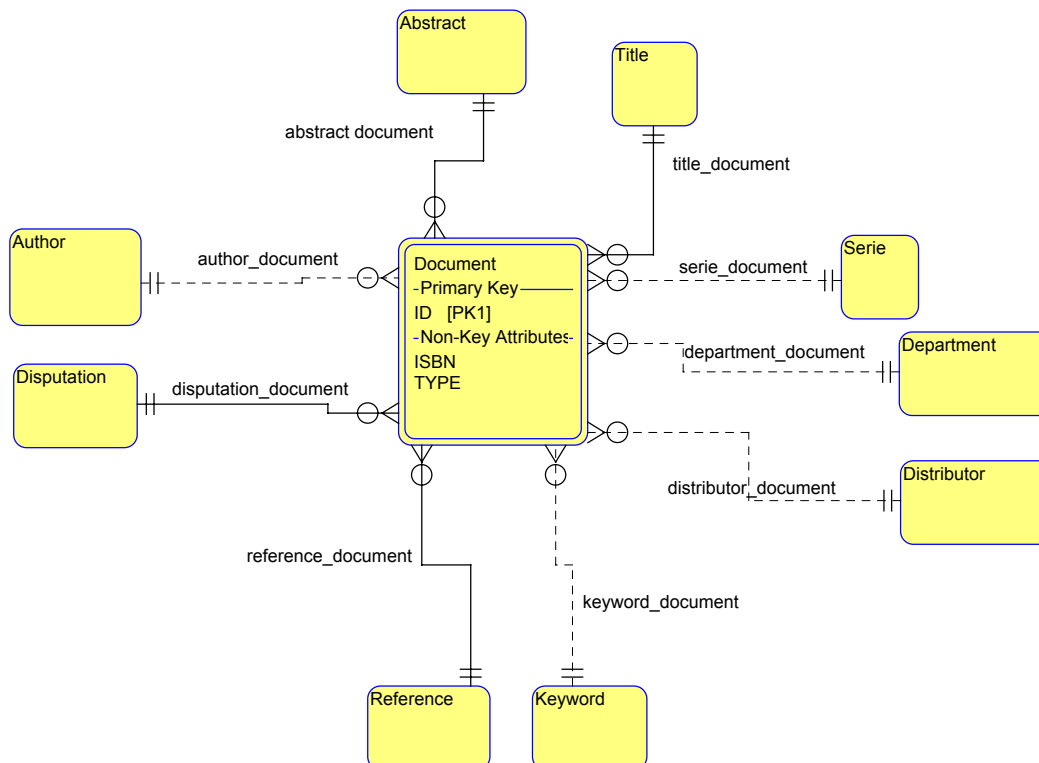
### *XML development kit (XDK)*

XDK gör det möjligt att hantera hela databasen via XML. Man kan importera texter från XML-filer och exportera texter till XML-filer. Användningen av XSL gör det möjligt att framställa

olika format för både visning och tryckning (HTML, PDF, XML) från en och samma källa. Alla våra webbsidor skapas med den tekniken. Webbservern får XML sidor från databasen och med hjälp av XSL framställs sidorna. För mer information titta på [http://technet.oracle.com/tech/xml/xdk\\_java/content.html](http://technet.oracle.com/tech/xml/xdk_java/content.html)

### Databasens struktur

Efter en ingående undersökning av Uppsala universitets spikblad skapades en första prototyp av databasen. Databasdesignen följer normaliseringsreglerna (tredje normalform). Databasen förbereddes för olika typer av publikationer. Den är inte enbart avsedd för avhandlingar. Alla typer av publikationer har likadana element som t.ex. rubrik, referenser, författare och text. I databasen kan sparas bland annat examensarbeten, rapporter och även enstaka artiklar som sedan kan sammanställas till elektroniska tidskrifter. För dessa publikationer måste dock eget arbetsflöde skapas och därmed krävs kanske ytterligare en eller flera tabeller.



Spara texter i databasen

För att kunna få in publikationer och deras metadata i databasen behövs en uppmärkning av hela publikationen. Detta krävs för att man automatiskt ska kunna identifiera alla metadata. Här står man inför ett problem:

Hur får man in publikationer utan att det krävs mer arbete av författarna?

Ett sätt att lösa problemet är att man skapa mallar för det program som författarna använder. I de mallarna sker markeringen genom att författarna använder definierade format för t ex rubriker, listor, tabeller och författare. Med hjälp av dessa skapas sedan XML-filer som innehåller både texten och metadata.

De flesta ( mer än 90%) som skriver en avhandling skriver den med Microsoft Word (PC och Mac). Därför skapades först en mall för Word. Den första mallen skapades för spikbladet eftersom där finns all nödvändig information för att skapa en post i databasen. Med hjälp av spikbladet testades informationsflödet som är följande:

1. Doktoranden fyller i mallen
2. Filen skickas till Enheten för digital publicering
3. En makro sorterar informationen
4. Med programmet R2Net (<http://www.logictran.com/>) skapas en XML-fil
5. XML-filen skickas via en http-post till databasen
6. XSQL omvandlar XML strukturen i filen med hjälp av XSL till en struktur som krävs för databasen
7. Ett PL/SQL program sparar data i databasen

## Sökningen

I en databas kan man söka på olika sätt. Man kan söka med vanliga SQL frågor som är snabba. Men för fritext och strukturerad sökning krävs det mer än de vanliga indexen i en databas. En viktig fråga är hur man hanterar ord med diakritiska tecken som å, é, ç. Vi har implementerat tre sorters sökning. Sökningen handlar alltid om hur man indexerar. Ett bra index gör sökningen enkel och snabb.

## Exakt sökning med diakritiska tecken

En relationsdatabas består i stort sett av tabeller, relationer mellan dem och indexerade fält. Med enkla frågor hittar man snabbt olika poster. Söker man efter ett namn så söker man i namnfält, söker man efter ett pris så söker man i prisfältet osv. Men problemet är att man måste ställa ganska exakta frågor för att hitta det man söker. Man måste veta hur namnet man söker stavas. Och har man ett tangentbord som inte kan framställa alla bokstäver man behöver måste man söka med hjälp av wildcards (trunkering).

Denna typ av sökning har implementerats på <http://publications.uu.se/theses/index.xsql> när man söker exakt.

## Sökning utan diakritiska tecken

Sökning utan diakriter betyder att man kan hitta ord med bokstäver som å, é, ç genom att söka efter ordet skrivet med a, e eller c. Det är viktigt för de användare som inte kan framställa alla bokstäver med sitt tangentbord. Men man måste också ta hänsyn till de olika språken. På tyska är det vanligt att man kan skriva ö som oe. Eller det danska æ söker man intuitivt som ae. Därför är det viktigt att man kan göra indexering beroende på språket som vi gjorde med hjälp av Oracle Text (<http://technet.oracle.com/products/text/content.html>). Man kan testa det på våra sidor med att aktivera funktionen *booleskt utan diakriter* på <http://publications.uu.se/theses/index.xsql>.

## Fritextsökning

Fritextsökning är lämplig för bredare sökningar. För fritextsökningen är det viktigt hur man indexerar texten. Man kan skapa följande index:

- Indexera alla ord som finns i texter (även: och, en, ett osv.)
- Indexera alla ord som finns i texter (men inte: och, en ett osv.)
- Indexera med hjälp av en tesaurus

Med Oracle text är allt detta möjligt beroende på språk eller oberoende på språk. I en så kallat stoplist ange man alla ord som får inte indexeras. Man kan ange en tesaurus för olika ämnesområde. För mer information titta på <http://technet.oracle.com/products/text/content.html>.

Vår fritextsökning finns på <http://publications.uu.se/theses/searchfree.xsql>.

## Dynamiskt webbgränssnitt med XML

### Några fördelar med att använda XML

- Dokumentstrukturen skild från layouten
- Samma dokument kan presenteras i olika format och layout
- Välstrukturerad data
- Samma information kan användas i olika sammanhang
- Plattformsberoende

### XHTML

Databasens innehåll, med undantag för fulltextfilerna, presenteras främst via HTML-sidor (<http://www.w3.org/MarkUp/>) avsedda för webbläsare. Den standard vi valt för uppmärkningen är XHTML 1.0 (<http://www.w3.org/TR/xhtml1/>) som är en XML-anpassning av HTML version 4.01. XHTML 1.0 ”kombinerar styrkan hos HTML 4 med kraften hos XML” (fritt översatt från W3C). Olika varianter används för olika gränssnitt:

XHTML 1.0 Transitional används för anpassning av de grafiska sidorna till den grafiska profilen. Vissa layout-justeringar görs i koden för anpassning till äldre grafiska webbläsare.

XHTML 1.0 Strict används för textversionen. Denna variant ger en rent strukturerad märkning, helt fri från layout-element.

Eftersom alla filer vi arbetar med är XML-baserade är det naturligt att använda XHTML (som ju är HTML som XML) som utmatningsformat, även om det är fullt möjligt att använda HTML 4 som alternativ.

Då sidorna hämtas i XML kommer det att kontrolleras att de är well-formed, dvs. att XML-strukturen och teckenanvändningen är helt korrekt. I annat fall visas ett felmeddelande istället för webbsidan. Detta garanterar att HTML-kodningen är korrekt strukturerad.

För att kontrollera att sidorna använder XHTML-taggar korrekt är de validerade med hjälp av <http://validator.w3.org/>.

### XSL (Extensible Stylesheet Language)

XSL är ett språk för att beskriva s.k. style sheets, <http://www.w3.org/Style/XSL/>. Dessa style sheets beskriver hur en XML-fil skall visas i ett visst format.

Sidorna, t.ex. en lista över aktuella avhandlingar, är XML-filer som skapas dynamiskt från databasen med hjälp av en SQL-fråga (i detta fall: hämta alla avhandlingar publicerade under de tre senaste månaderna).

För att visa filen som en webbsida kopplas den till ett antal olika XSL-filer som anger hur den skall se ut. Detta sker i flera steg:

1. I vår lösning börjar man med att kontrollera vilken webbläsare som används för att bestämma vilken teckenuppsättning (encoding) som skall väljas. Om webbläsaren är t.ex. IE5, som stöder UTF-8 Unicode, väljs denna teckenuppsättning för sidan. Är det en äldre läsare som t.ex. Netscape 3 visas sidan med teckenuppsättningen ISO-8859-1. Detta gör att sidorna dels kan utnyttja den utvidgade teckenuppsättningen i Unicode för att visa specialtecken där det är möjligt, dels att sidorna kan visas korrekt i äldre webbläsare, dock utan specialtecken som inte ingår i Latin-1 men med t.ex. alla svenska tecken. Specialtecken som inte kan tolkas ersätts då av ? (frågetecken). Om alla sidor hade visats i UTF-8 hade de svenska bokstäverna Å, Ä och Ö inte heller visats korrekt i äldre läsare.

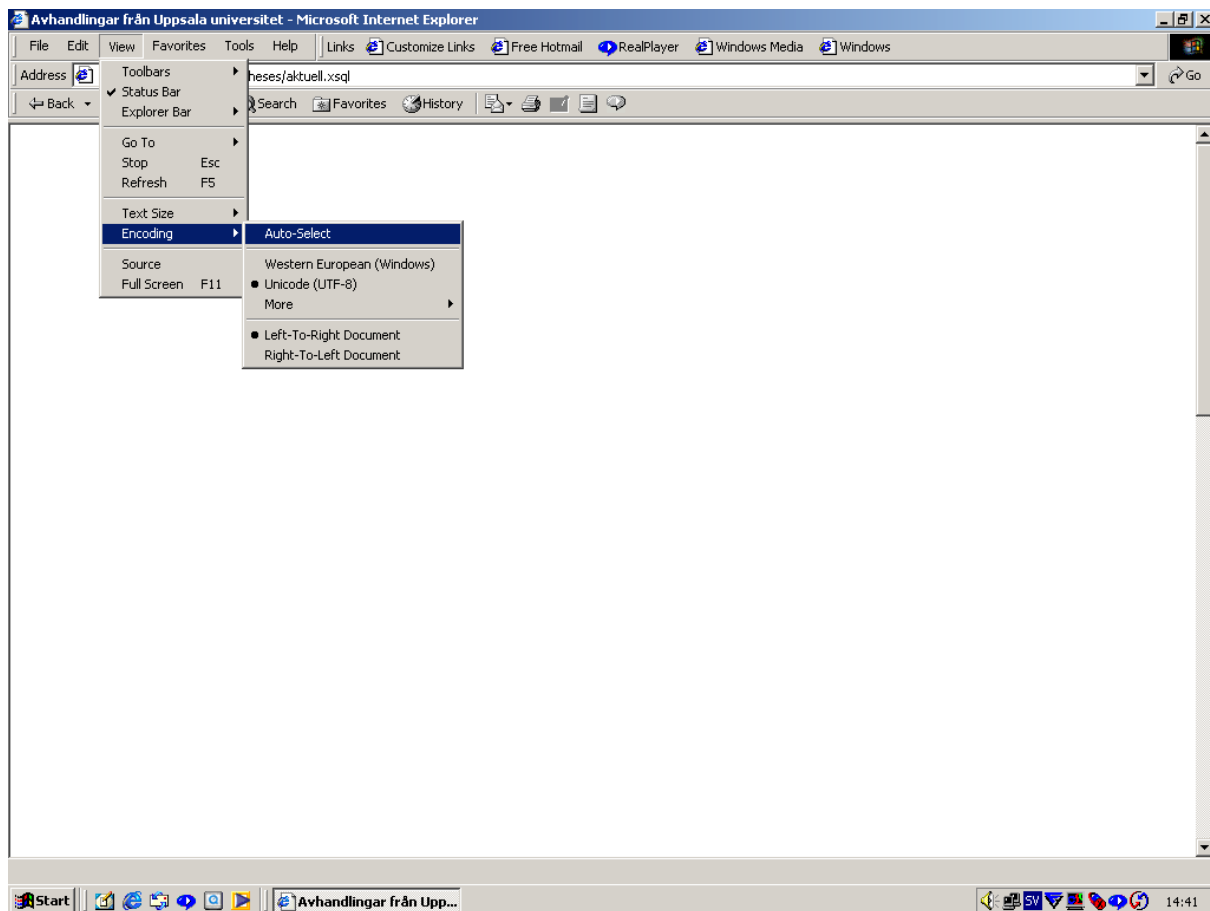


Bild 1. Encoding bestäms automatiskt i webbläsaren beroende på version. OBS: Detta syns ej om man inte väljer "View" eller motsvarande i verktygsfältet.

2. Nästa steg är en XSL-fil som bestämmer i vilket format sidan skall visas. Om det är en HTML-sida väljs i vårt fall inte HTML utan XML eftersom vi använder XHTML istället för HTML 4 för uppmärkning och XHTML är XML (se även ovan).

3. Till detta kopplas sedan en XSL-fil innehållande den övergripande sidlayouten (sidhuvud, sidfot och navigering). Denna fil innehåller t.ex. HTML, HEAD och BODY-taggar. Den innehåller också, i förekommande fall, en länk till ett externt Cascading Style Sheet (CSS, <http://www.w3.org/Style/CSS/>) som beskriver t.ex. teckensnitt och färger.

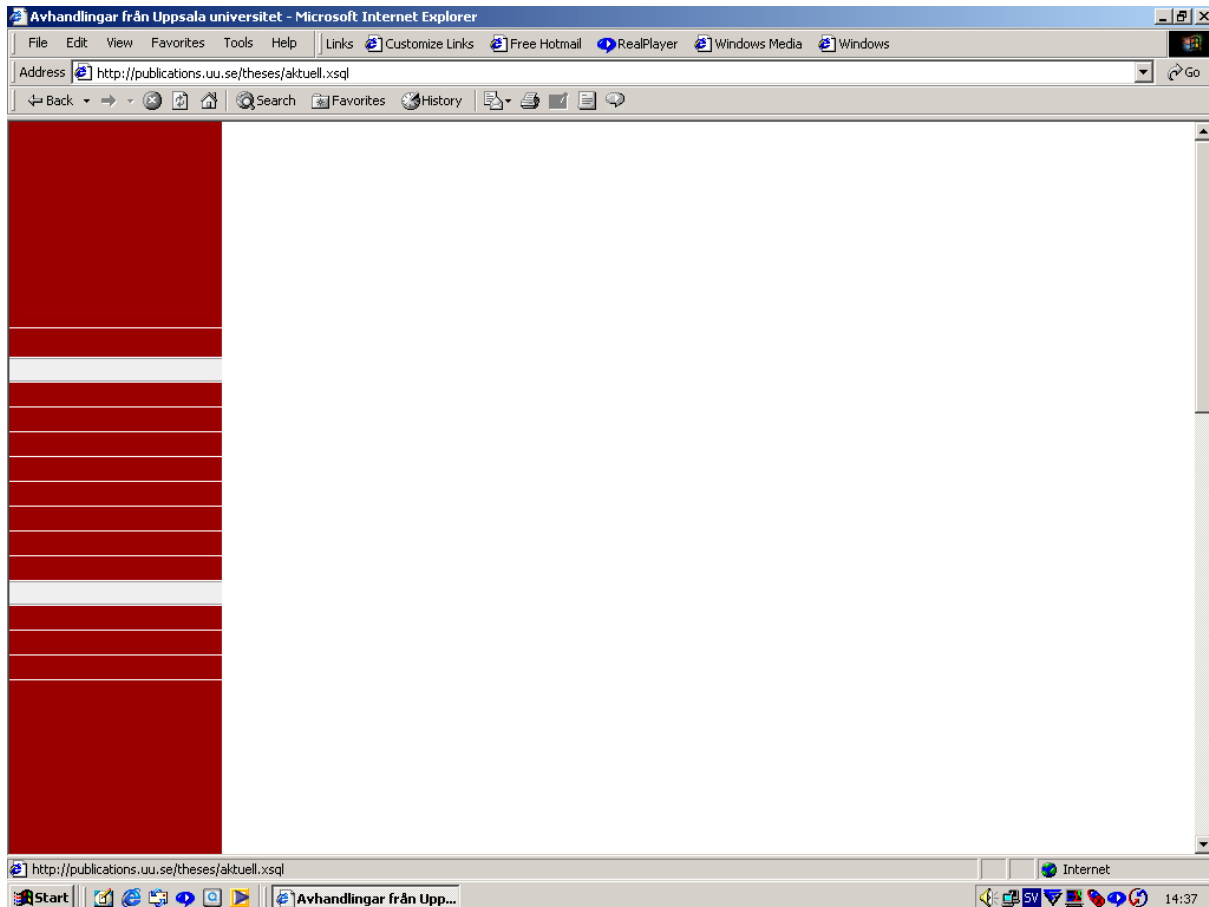


Bild 2. Grundläggande dokumentstruktur och layouten på plats med CSS.

4. Layoutfilen innehåller egentligen ingen egen text utan bara markeringar var texten skall läggas in. Beroende på valt språk hämtas rubriker, länkar och länktexter från en extern XML-fil som innehåller listor över rubriker etc. på olika språk (se även nedan).

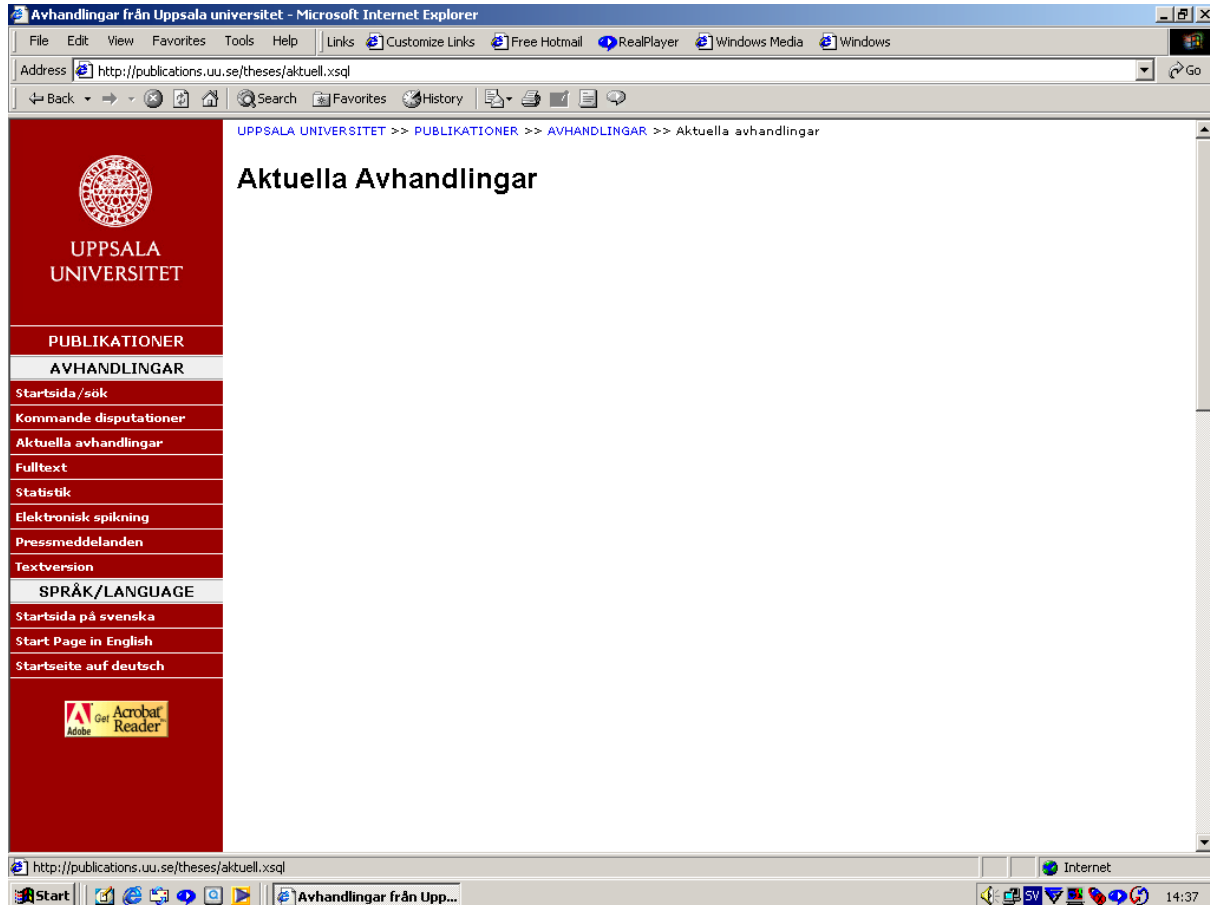


Bild 3: Rubriker, navigering, länktexter och länkar på plats.

5. Nu är det dags att visa de individuella posterna över aktuella avhandlingar. En XSL-fil talar om hur den enskilda posten skall se ut.
6. Delelement i posten (författare, titel, undertitel etc.) formateras på samma sätt i olika sammanhang varför det finns en separat XSL-fil för hur t.ex. författare skall visas (i vårt fall i fet stil följt av kolon). Dessa specifika XSL-filer kopplas ihop och bygger tillsammans upp den bibliografiska posten.
7. De poster som hämtats från databasen och formaterats enligt ovan hämtas och läggs in på rätt plats i layoutfilen.
8. XML-filen visas i webbläsaren som en HTML-sida med den teckenuppsättning webbläsaren stöder och på det språk man valt att visa den på (svenska om inget annat är valt).



Bild 4: Färdig HTML-sida.

## Olika utformat

Med hjälp av XSL kan XML-filer (dynamiskt skapade från databasen eller statiska) presenteras i olika utformat. I vårt fall kan avhandlingarnas abstract-sidor hämtas i HTML-format, PDF eller som "rå" XML-fil.

UPPSALA UNIVERSITET >> PUBLIKATIONER >> AVHANDLINGAR >> 91-554-4906-9

## Avhandlingar från Uppsala universitet

Avhandlingen är publicerad vid [Institutionen för farmaceutisk kemi](#)

**Alterman, Mathias** : Design and synthesis of HIV-1 protease inhibitors. - Uppsala, 2001. - 70p. - (Comprehensive summaries of Uppsala dissertations from the Faculty of Pharmacy, ISSN 0282-7484 ; 245)  
ISBN 91-554-4906-9

**Abstract:** [HTML-format](#) | [HTML-text](#) | [PDF-format](#) (kräver [Acrobat 5](#)) | [XML-format](#) | [Se bibliotekskatalogen](#)  
**Fulltext:** [PDF-format](#) (hämta gratis, kräver [Acrobat 5](#)) | [Köp boken online](#)

REFERENS TILL FULLTEXTDOKUMENTET ☒

Alterman, Mathias (2001). Design and synthesis of HIV-1 protease inhibitors. URL:<http://w3.uu.se/theses/abstract/91-554-4906-9.pdf> (2001-10-10)

### ABSTRACT

Human Immunodeficiency Virus (HIV) is the causative agent of Acquired Immune Deficiency Syndrome (AIDS). The C<sub>2</sub>-symmetric HIV-1 protease is one of the prime targets for chemotherapy in the treatment of the HIV infection. Inhibition of HIV-1 protease leads to immature and non-infectious viral particles. Design and synthesis of a number of C<sub>2</sub>-symmetrical C-terminal duplicated HIV-1 protease inhibitors and subsequent biological evaluation is presented in this thesis.

A versatile three step synthetic route has been developed using a carbohydrate as an inexpensive chiral starting material thus allowing inhibitors with the desired stereochemistry to be obtained. By this efficient method a series of tailor-made P<sub>2</sub>/P<sub>2</sub>' modified inhibitors were synthesized, and these were evaluated on purified HIV-1 protease and in HIV-1 infected cell assays. Highly active HIV-1 protease inhibitors were identified among the tested compounds. Analyses of the X-ray crystal structures of two of the most active compounds, as complexes with the protease, guided the further design of P<sub>1</sub>/P<sub>1</sub>' elongated inhibitors. Substitutions in the para-position of the P<sub>1</sub>/P<sub>1</sub>' benzyl groups were promoted efficiently by microwave-irradiated of palladium-catalyzed reactions. Particular modifications in the P<sub>1</sub>/P<sub>1</sub>' region of the inhibitors resulted in a 40-fold increase of the anti-viral activity on HIV-1 infected cells. Furthermore, a fast, efficient, and general one-pot microwave enhanced synthesis protocol for transformations of organo-bromides to tetrazoles was developed and applied on the inhibitor scaffold. Attachment of linker molecules to the P<sub>1</sub>/P<sub>1</sub>' benzyl groups of one inhibitor was used to develop of sensitivity enhancer tools in surface plasmon resonance biosensor

Bild 5: Abstract i grafiskt HTML-format

Abstractet hämtas från databasen som en XML-fil vilken kopplas till en XSL-fil som omvandlar den till en HTML-sida. Layout och navigering hämtas på samma sätt som tidigare.

På HTML-sidan är det också möjligt att skapa dynamiska länkar, t.ex. till Uppsala universitets katalog och universitetsbibliotekets online-katalog. Genom att klicka på institutionsnamnet kopplas man vidare till universitetskatalogen och får fullständiga adressuppgifter m.m.

Om avhandlingen inte finns tillgänglig i fulltext är det möjligt att via bibliotekskatalogen direkt se var boken finns och om den är tillgänglig.

De avhandlingar som ingår i Acta Universitatis Upsaliensis kan också beställas i tryckt form direkt online via ett separat beställningssystem.

HTML-sidan kan också visas i en enklare helt textbaserad layout utformad enligt Web Accessibility Initiative (se nedan) men också lämplig för t.ex. utskrift.

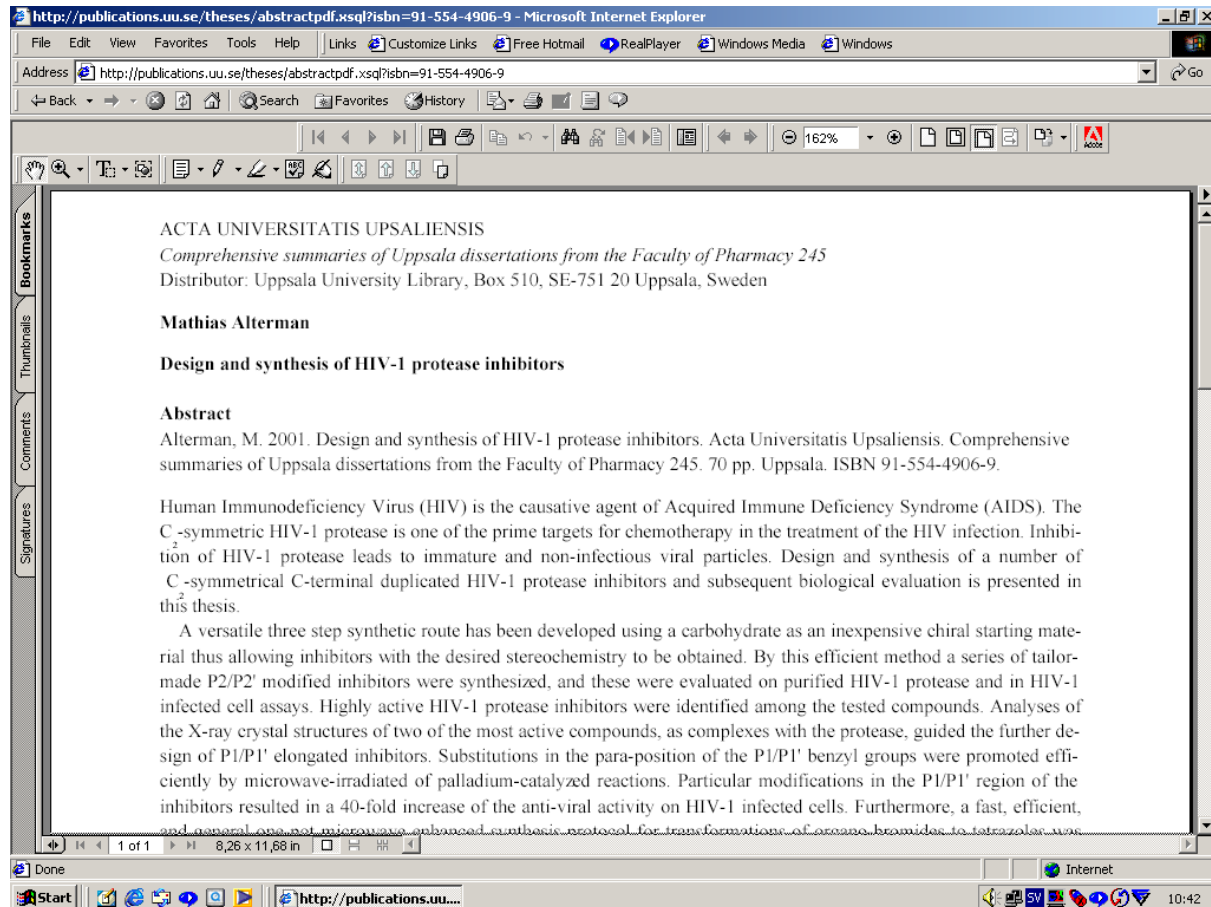


Bild 6: Abstract i PDF-format (= spikblad färdigt för tryckning).

Med hjälp av XSL och XSL-FO, <http://www.w3.org/Style/XSL/XSL-FO/>, skapas en PDF-version av abstractet. Denna PDF-fil är alltså inte förhandslagrad som en separat fil utan skapas på samma sätt som HTML-sidan ovan dynamiskt från databasen genom att lägga ett antal XSL style sheets på XML-filen, som hämtas via en fråga till databasen.

Denna metod kommer att användas för att framställa ett enhetligt färdigformat tryckunderlag till det obligatoriska spikbladet.

Även omslag, titelsida och abstractsida i fulltextavhandlingar kan genereras på detta sätt utifrån de uppgifter författaren registrerat i inmatningsformuläret.

Acrobat version 5 rekommenderas för att få se abstractet i PDF-format.

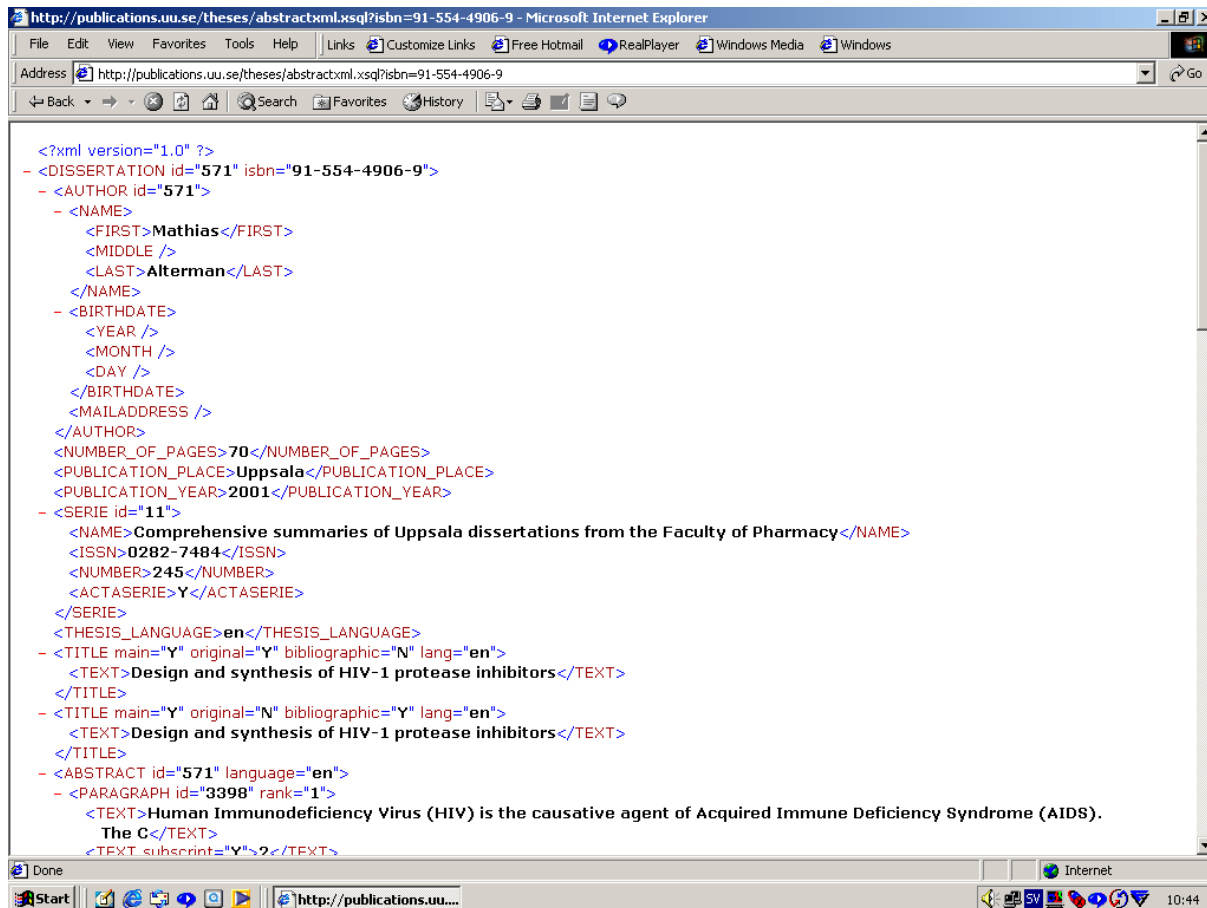


Bild 7: Abstract i XML-format.

Abstractet kan också ses som en ren XML-fil. Detta kräver IE 5 eller annan läsare som kan visa filer i XML-format. Observera att abstractet *inte* ligger lagrat som en textbaserad XML-fil utan även detta format skapas dynamiskt från databasen enligt samma metod som ovan.

## Flerspråkigt gränssnitt

Istället för att göra olika webbsidor för olika språk har vi valt en lösning där man hämtar texten från en separat XML-fil och infogar denna i layout-filen eller i löpande text på sidan. Detta gör att man inte behöver uppdatera flera sidor parallellt utan endast en fil även om man visar sidorna på flera olika språk. Olika språk markeras i URI:n med hjälp av motsvarande språkattribut.

### Exempel

The screenshot shows a Microsoft Internet Explorer browser window displaying the Uppsala University Theses website. The address bar shows the URL <http://publications.uu.se/theses/index.xsql?lang=sv>. The page title is "Avhandlingar från Uppsala universitet". The main content area features a search form titled "Sök efter avhandlingar" with the instruction "Fyll i ett eller flera fält (automatiskt OCH mellan fälten)". The search form includes the following fields and options:

- Författare: [text input] (efternamn) [text input] (förmamn)
- Titel: [text input] (eller ett ord från titeln)
- Serie: [text input]
- Utgivningsår: Välj år: [dropdown menu]
- ISBN: [text input] (inkludera bindstreck)
- Institution: Välj institution: [dropdown menu]
- \* = Institution som upphört, delats eller slagits samman
- Visa: 20 [dropdown menu] avhandlingar per sida
- Sök:  exakt med diakriter  booleskt utan diakriter

Below the search form, there are links for "Gå till: [Fritextsökning](#) | [Söktips](#)". The page also contains a sidebar with navigation links such as "PUBLIKATIONER", "AVHANDLINGAR", "Startsida/sök", "Kommande disputationer", "Aktuella avhandlingar", "Fulltext", "Statistik", "Elektronisk spikning", "Pressmeddelanden", "Textversion", "SPRÅK/LANGUAGE", "Startsida på svenska", "Start Page in English", and "Startseite auf deutsch". The footer includes copyright information: "© 2000-2001 Uppsala universitetsbibliotek | Enheten för digital publicering | Kontakta".

Bild 8: <http://publications.uu.se/theses/coming.xsql?lang=sv> visar sidan på svenska.

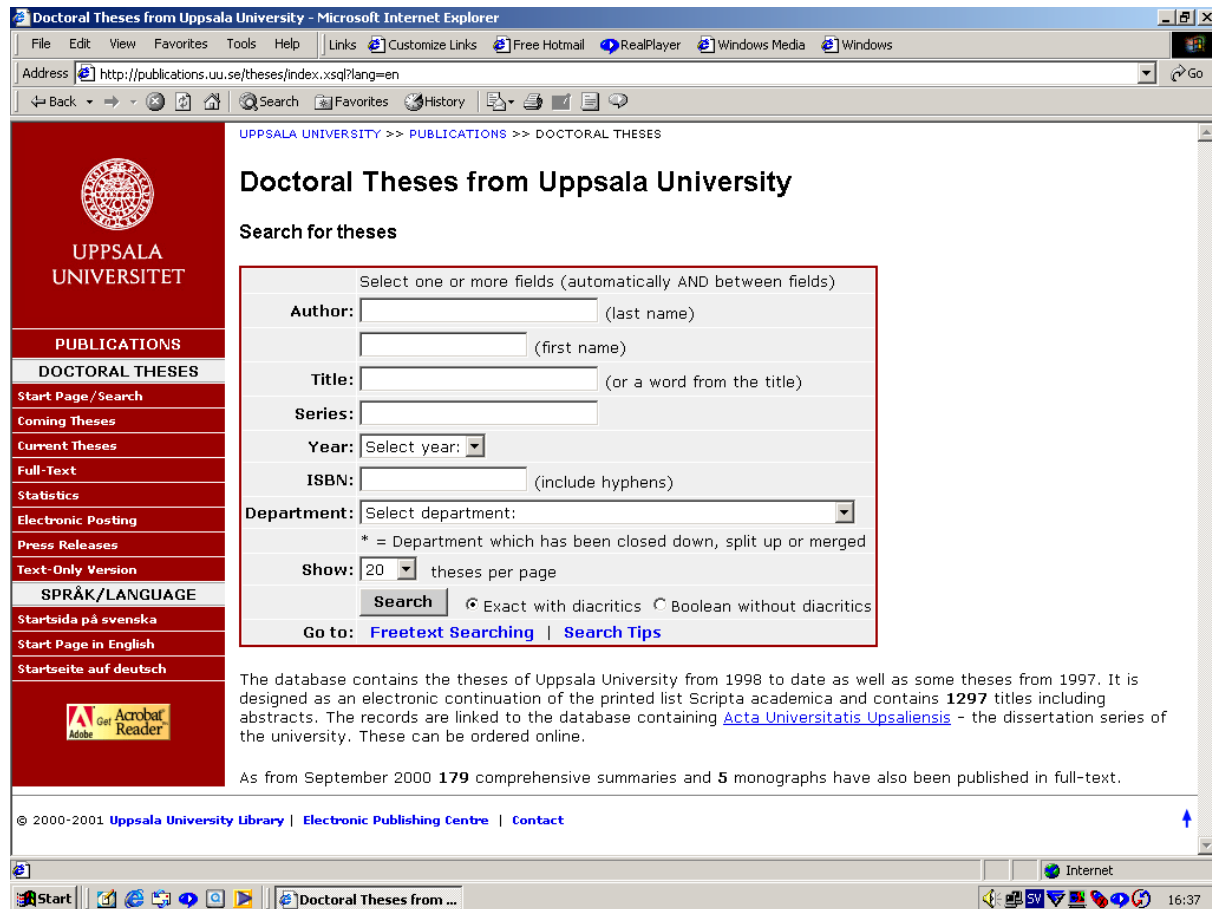


Bild 9: <http://publications.uu.se/theses/coming.xsql?lang=en> visar samma sida på engelska.

När det gäller navigering, rubriker och länkar, korta textsträngar som upprepas många gånger i olika sammanhang, ligger den informationen lagrad i en XML-fil som knyter texten till ett visst värde och ett visst språk. Vid upprepad information kan man med en enda ändring samtidigt uppdatera alla sidor där just den texten förekommer, t.ex. en extern länk som ändrats.

### Exempel

```
<heading id="1" >  
  <text lang="sv">Avhandlingar</text>  
  <text lang="en">Doctoral Theses</text>  
  <text lang="de">Dissertationen</text>  
</heading>
```

För att hämta denna rubrik till webbsidan anger man alltså i XSL-filen att rubrik (heading) nummer 1 skall hämtas med motsvarande språk som valts för sidan.

Vid längre löpande texter, t.ex. hjälpsidor, hämtas istället en färdigformaterad fil innehållande även HTML-taggar beroende på vilket språk som valts. Detta för att få större överskådlighet i sammanhängande texter vid redigering.

## Accessibility

För att göra webbsidorna så tillgängliga som möjligt för alla typer av användare har en speciell anpassning gjorts enligt Web Accessibility Initiative (WAI) <http://www.w3.org/WAI/> och W3C:s Web Content Accessibility Guidelines 1.0, <http://www.w3.org/TR/WAI-WEBCONTENT/>.

Avsikten är att kunna uppnå den högsta nivån, Level Triple-A. Därför har en ren textversion gjorts. I guidelines rekommenderas inte detta som en primär metod, men det främsta skälet till det är att parallella sidor normalt kräver dubbelt underhåll vilket leder till ojämn uppdatering av olika versioner av sidor. I vårt fall gäller inte detta eftersom innehållet hämtas från samma källa. Endast formen, som hämtas från en annan layout-fil, påverkas.

I och för sig har även de grafiska sidorna en hög nivå när det gäller accessibility men på vissa avgörande punkter är de svåra att utforma enligt WAI. Det största problemet är användningen av tabeller (<TABLE>) för att skapa sidlayout. Många grafiska profiler förlitar sig helt på tabeller för att skapa t.ex. fasta navigeringskolumner bredvid själva textinnehållet på en sida. Denna metod är också att föredra framför t.ex. frames (ramar) som utgör ett betydligt större problem när det gäller webbsidors tillgänglighet. Emellertid skapar tabeller svårigheter för andra typer av läsare än grafiska. Rekommendationen är att enbart använda tabeller för ”tabular information”, t.ex. en strukturerad lista över databasposter.

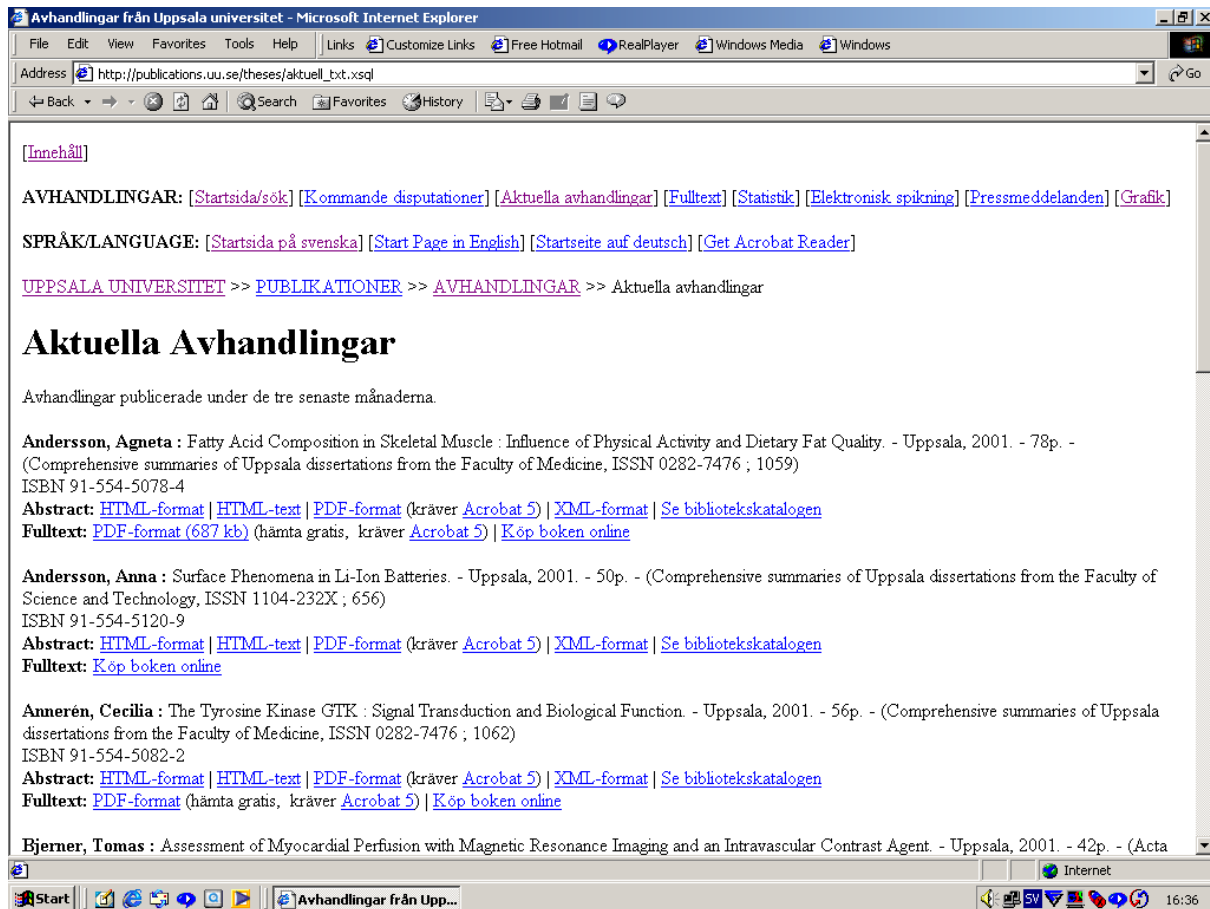


Bild 10: Samma sida som bild 4 med annan layout-mall.

Specifika punkter som i textversionen iakttagits för respektive guideline:

Guideline 1. Provide equivalent alternatives to auditory and visual content.

Sidorna är helt textbaserade vilket innebär att inga speciella anpassningar för ljud eller bild måste göras.

Guideline 2. Don't rely on color alone.

Sidorna innehåller inga fasta färginställningar utan anpassas helt efter användarens egna inställningar vad gäller färg, textstorlek och teckensnitt.

Guideline 3. Use markup and style sheets and do so properly.

Sidorna har en strukturerad användning av huvudrubrik, underrubriker, stycken eller datatabeller.

Guideline 4. Clarify natural language usage.

Sidans språk anges dynamiskt genom lang-attribut. Vissa attribut är fasta, t.ex. abstract som alltid är på engelska.

Guideline 5. Create tables that transform gracefully.

Tabeller används endast för strukturerad data, ej layout. Dessa tabeller innehåller summary-attributet som beskriver tabellens innehåll (f.n. endast engelsk version).

Guideline 6. Ensure that pages featuring new technologies transform gracefully.  
Sidorna är inte kopplade till style sheets och innehåller inte skript, applets eller frames.

Guideline 7. Ensure user control of time-sensitive content changes.  
Sidorna innehåller inga blinkande eller rörliga objekt.

Guideline 8. Ensure direct accessibility of embedded user interfaces.  
Sidorna innehåller inga andra programelement.

Guideline 9. Design for device-independence.  
Form-element innehåller tabindex-attributet för korrekt tab-ordning. Användning av accesskey-attributet är under utarbetande.

Guideline 10. Use interim solutions.  
Inga nya fönster öppnas från sidorna. Label-elementet används i forms för att knyta samman rubriken med forms-kontrollen. Tabeller används enligt guideline 5. Näraliggande länkar i navigeringen skiljs av ] [ med ett extra mellanslag emellan.

Guideline 11. Use W3C technologies and guidelines.  
Sidorna är validerade enligt xhtml-strict dtd som är den striktast kontrollerade versionen av XHTML. Alla abstract är tillgängliga i xhtml-strict. Ett undantag från denna guideline är fulltext-filerna som i nuvarande lösning endast är tillgängliga i PDF (eftersom de inte är XML-baserade). Adobes automatiska HTML-konverterare kan inte heller användas då filerna är lösenordsskyddade mot ändringar. När systemet är fullt utbyggt kan dock detta avhjälpas.

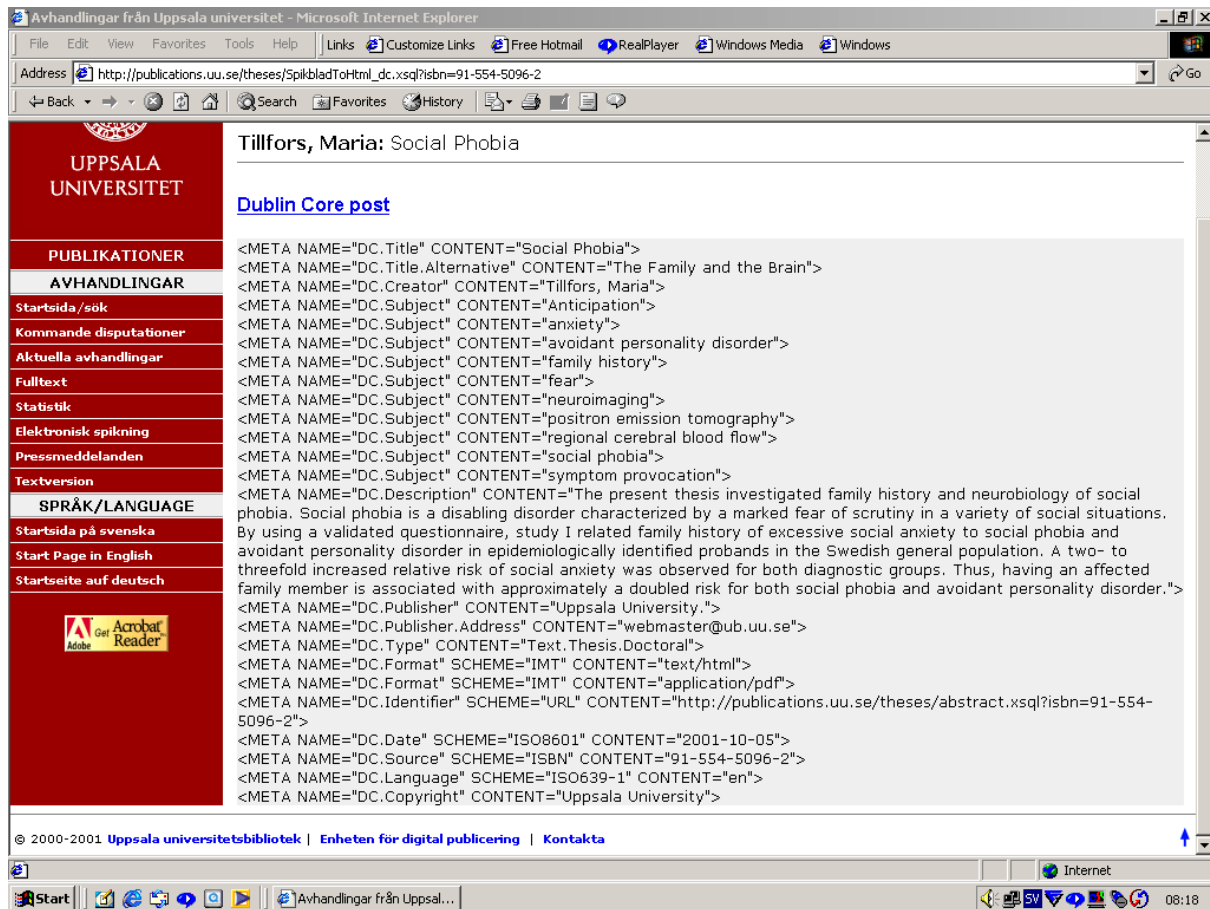
Guideline 12. Provide context and orientation information.  
Sidorna använder inte frames. Label-elementet används i forms. Bibliografiska poster presenteras en och en i datatabell-format.

Guideline 13. Provide clear navigation mechanisms.  
Sidorna har entydiga länktexter. Rel- och rev-attributen används för att relatera olika länkar till varandra. Metadata anges i meta-elementet. Navigeringen är genomgående och ordnad i sammanhängande grupper. Möjlighet ges också till att hoppa över navigeringen och gå direkt till innehållet. Olika typer av sökfunktioner finns för olika nivåer.

Guideline 14. Ensure that documents are clear and simple.  
Sidlayouten är genomgående konsekvent.

## Metadata

Metadata kan hämtas från databasen i olika format:



The screenshot shows a Microsoft Internet Explorer browser window displaying a page from the Uppsala University library website. The page title is "Tillfors, Maria: Social Phobia". The main content area displays Dublin Core metadata in XML format. The left sidebar contains navigation links for "PUBLIKATIONER" and "AVHANDLINGAR", with sub-links for "Startsida/sök", "Kommande disputationer", "Aktuella avhandlingar", "Fulltext", "Statistik", "Elektronisk spikning", "Pressmeddelanden", and "Textversion". Below these are language options: "SPRÅK/LANGUAGE", "Startsida på svenska", "Start Page in English", and "Startseite auf deutsch". An Adobe Acrobat Reader logo is visible at the bottom of the sidebar. The metadata includes fields for Title, Creator, Subject, Description, Publisher, Format, Identifier, Date, Source, Language, and Copyright.

UPPSALA  
UNIVERSITET

PUBLIKATIONER

AVHANDLINGAR

Startsida/sök

Kommande disputationer

Aktuella avhandlingar

Fulltext

Statistik

Elektronisk spikning

Pressmeddelanden

Textversion

SPRÅK/LANGUAGE

Startsida på svenska

Start Page in English

Startseite auf deutsch

Acrobat  
Reader

Tillfors, Maria: Social Phobia

[Dublin Core post](#)

```
<META NAME="DC.Title" CONTENT="Social Phobia">
<META NAME="DC.Title.Alternative" CONTENT="The Family and the Brain">
<META NAME="DC.Creator" CONTENT="Tillfors, Maria">
<META NAME="DC.Subject" CONTENT="Anticipation">
<META NAME="DC.Subject" CONTENT="anxiety">
<META NAME="DC.Subject" CONTENT="avoidant personality disorder">
<META NAME="DC.Subject" CONTENT="family history">
<META NAME="DC.Subject" CONTENT="fear">
<META NAME="DC.Subject" CONTENT="neuroimaging">
<META NAME="DC.Subject" CONTENT="positron emission tomography">
<META NAME="DC.Subject" CONTENT="regional cerebral blood flow">
<META NAME="DC.Subject" CONTENT="social phobia">
<META NAME="DC.Subject" CONTENT="symptom provocation">
<META NAME="DC.Description" CONTENT="The present thesis investigated family history and neurobiology of social phobia. Social phobia is a disabling disorder characterized by a marked fear of scrutiny in a variety of social situations. By using a validated questionnaire, study I related family history of excessive social anxiety to social phobia and avoidant personality disorder in epidemiologically identified probands in the Swedish general population. A two- to threefold increased relative risk of social anxiety was observed for both diagnostic groups. Thus, having an affected family member is associated with approximately a doubled risk for both social phobia and avoidant personality disorder.">
<META NAME="DC.Publisher" CONTENT="Uppsala University.">
<META NAME="DC.Publisher.Address" CONTENT="webmaster@ub.uu.se">
<META NAME="DC.Type" CONTENT="Text.Thesis.Doctoral">
<META NAME="DC.Format" SCHEME="IMT" CONTENT="text/html">
<META NAME="DC.Format" SCHEME="IMT" CONTENT="application/pdf">
<META NAME="DC.Identifier" SCHEME="URL" CONTENT="http://publications.uu.se/theses/abstract.xsql?isbn=91-554-5096-2">
<META NAME="DC.Date" SCHEME="ISO8601" CONTENT="2001-10-05">
<META NAME="DC.Source" SCHEME="ISBN" CONTENT="91-554-5096-2">
<META NAME="DC.Language" SCHEME="ISO639-1" CONTENT="en">
<META NAME="DC.Copyright" CONTENT="Uppsala University">
```

© 2000-2001 Uppsala universitetsbibliotek | [Enheten för digital publicering](#) | [Kontakta](#)

Bild 11: Metadata i Dublin Core formatet

UPPSALA UNIVERSITET >> PUBLIKATIONER >> AVHANDLINGAR >> Abstract >> MARC 21 post

## Avhandlingar från Uppsala universitet

Tillfors, Maria: Social Phobia

UPPSALA UNIVERSITET  
[MARC 21 post](#)

<b>PUBLIKATIONER</b>	008 011002s2001 sw m eng
<b>AVHANDLINGAR</b>	020 \$a 91-554-5096-2
<b>Startsida/sök</b>	100 1 \$a Tillfors, Maria
<b>Kommande disputationer</b>	245 \$a Social Phobia : \$b The Family and the Brain
<b>Aktuella avhandlingar</b>	260 \$a Uppsala : \$b Acta Universitatis Upsaliensis : \$b Univ.-bibl., \$c 2001
<b>Fulltext</b>	300 \$a 72 s.
<b>Statistik</b>	440 \$a Comprehensive summaries of Uppsala dissertations from the Faculty of Social Sciences, \$x 0282-7492 ; \$v 104
<b>Elektronisk spikning</b>	502 \$a Diss. Uppsala, Univ., 2001
<b>Pressmeddelanden</b>	653 \$a Anticipation\$a anxiety\$a avoidant personality disorder\$a family history\$a fear\$a neuroimaging\$a positron emission tomography\$a regional cerebral blood flow\$a social phobia\$a symptom provocation
<b>Textversion</b>	856 4 \$u <a href="http://publications.uu.se/theses/abstract.xsql?isbn=91-554-5096-2">http://publications.uu.se/theses/abstract.xsql?isbn=91-554-5096-2</a>

**SPRÅK/LANGUAGE**

Startsida på svenska  
Start Page in English  
Startseite auf deutsch

© 2000-2001 Uppsala universitetsbibliotek | Enheten för digital publicering | Kontakta

http://www.uu.se/

Bild 12: Metadata i MARC21-formatet

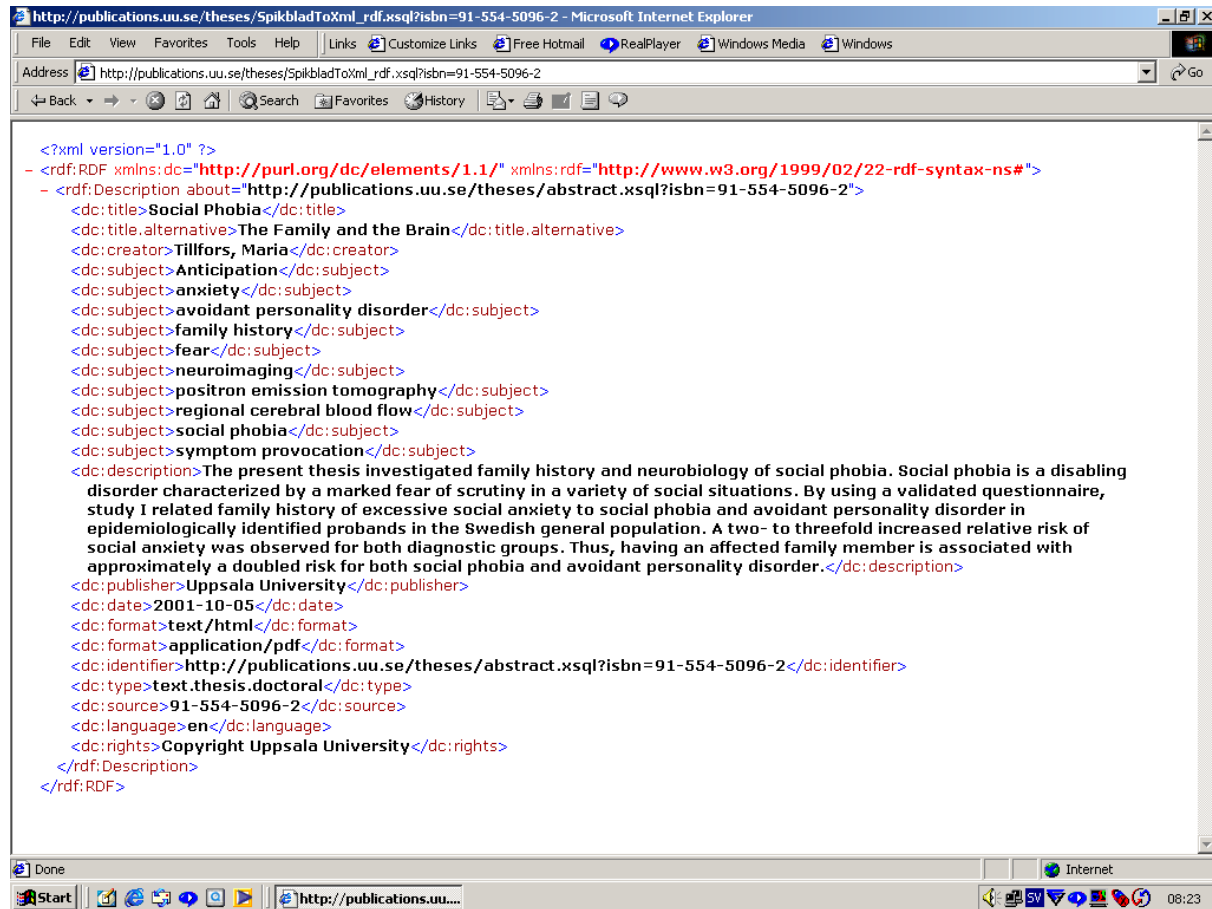


Bild 13: Metadata i RDF (Dublin Core format i XML)

Olika typer av metadata kan sedan exporteras till andra system.

## Katalogiseringsfunktion och överföring till Libris och Disa

I systemet finns en lösning för att överföra posten i avhandlingsdatabasen (som är den första bibliografiska beskrivningen av avhandlingen) till det nationella bibliotekssystemet Libris och vidare till universitetsbibliotekets lokala bibliotekssystem, Disa.

### 1. Posten hämtas via ett MARC-baserat katalogiseringsformulär.

Fields	Ind1	Ind2	Subfields
001			91-554-5088-1
003			U
008			011002 s2001 se w I
041			\$a ger \$b leng
081			\$c [highlighted] \$c [highlighted] \$c [highlighted]
096			\$a [highlighted] \$b [highlighted] \$c [highlighted] \$s g
100	e		\$a Neunsinger, Silke \$c 1970
245			\$a Die Arbeit der Frauen - die Krise der Männer \$b die Erwerbstätigkeit verheirateter Frauen in Deutschland und Schweden 1919-1939
260			\$a Uppsala \$b Acta Universitatis Upsaliensis \$b Univ.-bibl. \$c 2001
300			\$a 294 s.
440			\$a Studia historica Upsaliensia \$n 0081-6531 \$v 198
500			\$a [highlighted]
502			\$a Diss. Uppsala, Univ., 2001
505			\$a [highlighted]
506			\$a Med sammanfattning på engelska
856	4		\$u <a href="http://publications.uu.se/theses/91-554-5088-1/">http://publications.uu.se/theses/91-554-5088-1/</a>

Bild 14: Katalogiseringsformulär för Libris. I detta fall skulle fält 081 och 096 fyllts i. Övriga uppgifter finns redan klara från databasen.



## PROBLEMOMRÅDEN, FRAMTIDSPLANER

### *Problemområden*

#### Interoperabilitet

För att säkra interoperabiliteten har vi valt att lagra information i en databas. Vidare kommer vi att exportera metadata enligt OAI protokoll ( se <http://www.openarchives.org/>) och anpassa till Z39.50 protokollet som används för samsökning med bibliotekskataloger.

#### Semantisk interoperabilitet

Användning av tesaurusar säkrar semantisk interoperabilitet. Därför kommer en utredning om vilka kontrollerade lexikon (tesaurus) är mest användbara i vårt system att göras.

I dagsläget används kontrollerade lexikon som tex. svenska ämnesord

(<http://www.annesord.kb.se/start.html>) samt olika klassningsschema som används vid bibliotek och även i andra sammanhang.

Vidare kommer att användas

- ⇒ Auktoritetsnamnregister
  - ⇒ Personnamn
  - ⇒ Institutionsnamn

#### Flerspråkigt sökgränssnitt

Ett sökgränssnitt som tar hänsyn till olika språk utvecklades i sin första form och kommer att utvecklas vidare.

#### Unicode

Det är enkelt att framställa tecken som sparas i unicode. Det krävs dock att man har typsnitt som kan framställa alla tecken. De är inte många idag. Dessutom behövs det programvara som kan hantera dessa typsnitt. Bara de senaste webbläsarna stöder unicode. Samma sak gäller för pdf-formatet. Man måste importera typsnitt som stöder unicode och bara den senaste Acrobat Reader stöder Unicode.

Det finns också problem med unicode vid sökningen. Med ett vanligt tangentbord kan man inte mata in alla tecken. Dessutom är hanteringen av unicode tecken inte enhetlig i alla program.

#### Matematiska och kemiska formler

För formler finns det idag MathML (<http://www.w3.org/Math/>). Men i dagsläget stöder webbläsarna inte detta format. För att idag visa formler via ett webbgränssnitt måste man hitta andra alternativa lösningar.

## Bilder

Det finns många bildformat idag men ingen vet vilka som kommer att finnas kvar i framtiden. Nu finns det också ett XML-format för grafik (SVG <http://www.w3.org/Graphics/SVG/Overview.htm8>). Men det är inte tillräckligt för högupplösta bilder. Ett sätt är att arkivera bilder i tiff-format och använda ett annat format för visningen på webben.

## Multimedia

Ingen vet hur starkt multimedia kommer att påverka framtida publikationer. För multimedia finns det idag också ett XML-format som heter SMIL. För mer information titta på <http://www.w3.org/AudioVideo/>.

## Säkerheten

En annan fråga är hur man skyddar statiska filer (t ex fulltextfiler) mot obehöriga förändringar. Det finns flera möjligheter. Där finns det koncept med sigelering och checksumma för enstaka filer. Eller så använder man en security server. I alla fall måste man alltid ha en kopia (eller flera) av originalfilen sparad någonstans. I våra lösningar har vi en arkivkopia på UUB och en arkivkopia levereras till Kungliga biblioteket.

## Problem av organisatorisk karaktär

Vid övergång till elektronisk publicering är det viktigt att man kan garantera att samma version är publicerad både på nätet och i tryck form. Därför behöver man ändra på arbetsflödet och skapa både en nätversion men även ett tryckningsunderlag från en och samma källa.

Detta är ett exempel på förändring av arbetsflödet som kommer att påverka arbetssättet på enheter inom UU.

För att genomföra strukturella förändringar i organisationen krävs att dessa frågor utreds och det tar tid.

## Framtidsplaner

- ⇒ Börja publicera licentiatavhandlingar och monografiavhandlingar
- ⇒ Utveckla arbetsflöden för publicering av forskningsrapporter
- ⇒ Utveckla arbetsflöden för publicering av examensarbeten
- ⇒ Utveckla arbetsflöden för publicering av tidskriftsartiklar och framställning av elektroniska tidskrifter