



# Teknisk utvärdering av elektroniska publiceringsplattformar

Av Peter Berkesand

## Bakgrund

Linköpings universitet började redan 1996 med elektronisk förlagsverksamhet vid Linköping University Electronic Press (E-press). Detta uppmärksammades internationellt och verksamheten har presenterats vid ett flertal internationella konferenser. E-press har åtagit sig att garantera säkerhet och tillgänglighet på en fast URL-adress under en längre tid och sedan starten är URL-adressen <http://www.ep.liu.se/>.

Förlagets publicering är bred:

- Monografier och artiklar i sju olika serier.
- Fyra elektroniska tidskrifter.
- Doktors- och licentiatavhandlingar.
- Rapporter.
- Examensarbeten/opsatser från 6 olika institutioner. Ytterligare fyra institutioner börjar publicera hösten 2003.

I dag (den 12 september 2003) använder sig förlaget av ett automatiserat publiceringssystem där studenterna själva publicerar sina uppsatser. Studenten registrerar uppgifterna om examensarbetet på en webbsida (<http://www.ep.liu.se/exjobb/>) och skickar sedan arbetet som pdf-fil till en server. Publiceringen sker sedan automatiskt när institutionens administratör, via en webbsida, vidimerat att uppsatsen är godkänd. Det finns således ingen möjlighet att publicera någon icke godkänd uppsatt. Alla examensarbeten är sökbara via söksidan <http://www.ep.liu.se/exjobb/search.html>.

Avhandlingar och artiklar i övrigt publiceras för närvarande manuellt. Dessa är **inte** sökbara då de inte hanteras i någon databas.

Det nuvarande automatiserade systemet för elektronisk publicering klarar således endast examensarbeten. Volymen av avhandlingar och artiklar ökar från termin till termin, vilket får till följd mängden icke sökbara avhandlingar och artiklar ökar varje månad. Förlagets publiceringssystem är föråldrat och måste konstrueras om för att anpassas till dagens behov och nya krav från forskarna. En utveckling av det befintliga publiceringssystemet blir både arbetsintensivt och kostnadskrävande. Ett helt nytt publiceringssystem är därför nödvändigt.

I samband med universitetets utredning [Universitetets informationsförsörjning](#) tillsattes resursgrupper. En av dessa grupper var *tekniska gruppen* som fick i uppgift att se över vilka publiceringssystem som finns. I gruppen ingår:

- Peter Berkesand, redaktör, E-press
- Kent Engström, säkerhetsansvarig, UNIT
- Anders Fåk, 1 fo.ing., universitetsbiblioteket
- Jens Larsson, 1 fo.ing, ISY
- Ursula Nielsen, bibliotekarie, universitetsbiblioteket



Gruppen har undersökt vilka publiceringsplattformar tillgängliga för elektronisk publicering av vetenskapliga publikationer som kan vara intressanta för E-press. Inom gruppen har diskussioner förts kring de olika plattformarna.

## Publiceringsplattformar

Tre publiceringsplattformar har undersökts och har bedömts som intressanta för E-press. Dessa är EPrints (<http://www.eprints.org/>), Digitala vetenskapliga arkivet, DiVA utvecklat av Uppsala universitet (<http://www.diva-portal.se/index.xsql?lang=sv>) samt DSpace (<http://www.dspace.org/>).

Den 13 februari genomfördes ett studiebesök vid Sveriges Lantbruksuniversitet och Uppsala universitet i syfte att titta närmare på publiceringsplattformarna EPrints och DiVA. Beträffande DSpace har gruppen tyvärr inte hunnit vara lika noggrann när det gäller tester av systemet.

Flera krav har ställts för att programmen ska vara intressanta. De viktigaste var låg kostnad både vad gäller inköp och drift; hög säkerhet för obehörigt intrång; driftsäkerhet, användarvänlighet, redigeringsmöjligheter av programvara, goda sökfunktioner, enkla rutiner vid upplägg av vetenskapliga arbeten, statistikmodul samt inbyggda funktioner för OAI-protokoll (Metadata) samt möjlighet att generera fil för Libriskatalogen (Marc 21) och till den lokala bibliotekskatalogen.

## EPrints

EPrints' funktioner studerades närmare vid studiebesöket på [Lantbruksuniversitetet i Uppsala](#) och deras publiceringssystem [Epsilon](#). EPrints är gratis och kan laddas ned från Internet (<http://www.eprints.org>). Information om programvaran finns på <http://software.eprints.org/#ep2>. De kostnadskrävande delarna för att börja använda systemet är installation och anpassning av programvaran, hårdvara samt löpande uppgraderingar av både hård- och mjukvara.

## Operativsystem och programvaror

EPrints är skrivet i [Perl](#) och är utvecklat under operativsystemet [Linux](#). EPrints använder sig av webservern [Apache](#) och databasen [MySQL](#) samt flera fria programvaror.

I sitt grundutförande består EPrints av färdiga, inbyggda funktioner med sida för registrering, söksida med presentation, sidor för redigering m.m. Det finns även möjligheter att ändra i scripten för att anpassa funktioner, utseende och händelser allt efter egna behov och krav.

Programmets konfigurationsfiler styr

- Den grundläggande webbdesignen.
- Vilka metadatafält ett arkiv ska omfatta, vilka av dem som är obligatoriska och vilka som vid inmatningen endast är synlig för särskilt behöriga (t.ex. katalogiserande bibliotekarie).
- Hjälptexter för respektive metadatafält.
- Vilka fält som ska tillhandahållas för OAI-harvesting.
- Vilka format på uppladdade filer som tillåts eller är obligatoriska.
- Vilka fält som ska var sökbara/ingå i bladdringslist/ingå i visningsformat och hur de ska sorteras.



Ändringar i själv programkoden krävs för

- Förekomst av vissa element på sidor som genererats av olika funktioner i systemet som t.ex. knappar, ta bort eller förändra text. Man kan alltså styra **hur** saker och ting visas men inte alltid **vad** som visas t.ex. fält kan inte tas bort.
- Ändringar av namn på knappar

Ändringar på databasnivå (MySQL)

- Ändringar av fältnamn i efterhand.
- Lägga till eller ta bort fält.

## Publicering

Publikationen är skapad efter i förväg gjorda mallar för examensarbeten, licentiat- och doktorsavhandlingar. Mallarna kan laddas ned från Internet ([http://epsilon.slu.se/0210\\_styles.html](http://epsilon.slu.se/0210_styles.html)). Forskare, doktorander och studenter lägger själva in alla nödvändiga uppgifter om sin publikation. (På <http://epsilon.slu.se/Submit-form.pdf> framgår hur en sida för inmatning i Epsilon ser ut.) Därefter laddas själva filen med publikationen upp till särskild plats på servern. Biblioteket på SLU får sedan information om att en ny publikation ska läggas upp. Någon ansvarig person går då in och kontrollerar att alla uppgifter är korrekta, rättar ev. felaktigheter och godkänner sedan arbetet för uppläggning på Internet. Hela systemet är helt webbaserat. Man använder sig av färdiga mallar för t.ex. det vetenskapliga arbetet och titelblad.

Avtal finns mellan SLU och författaren. Författaren läser ett avtal publicerat på SLU:s hemsida (<http://epsilon.slu.se/diss/> och <http://epsilon.slu.se/vfak/>). I samband med att författaren fyller i alla uppgifter om arbetet kryssar författaren i ruta och godkänner därmed den i förväg lästa överenskommelsen. Inga underskrifter förekommer (avtalen finns på [http://epsilon.slu.se/0210\\_avtalstext.html](http://epsilon.slu.se/0210_avtalstext.html) och [http://epsilon.slu.se/vfak/avtal\\_grundmall\\_ex.html](http://epsilon.slu.se/vfak/avtal_grundmall_ex.html)).

Epsilons statistikmodul finns på sidan [http://utb32.bibul.slu.se/analog/epsilon/epsilon\\_access\\_0301.html](http://utb32.bibul.slu.se/analog/epsilon/epsilon_access_0301.html).

I Sverige används EPrints av Sveriges Lantbruksuniversitet, Lunds universitet och Handelshögskolan i Göteborg. Ytterligare information om systemet finns på [http://epsilon.slu.se/0210\\_omepsilon.html](http://epsilon.slu.se/0210_omepsilon.html).

## Diskussion kring systemets för- och nackdelar

Koden i EPrints kan ändras för att anpassa systemet efter egna behov och önskemål. Om det skett några ändringar eller tillägg i programkoden kan detta ställa till problem när det kommer en ny version av programmet. Alla förändringar i koden måste vara väl dokumenterade eftersom den nya versionen skriver över tidigare ändringar. De modifieringar som är gjorda i programmet måste sedan läggas till på nytt. När ändringarna har lagts till är det inte alltid säkert att programmet fungerar som det är tänkt. Det finns versionshanteringssystem som kan hantera ovanstående problem. Verket bör undersökas närmare för att se om det går att applicera på EPrints.

Det kan vara svårt att anpassa knappar till de egna behoven. Även här är det inte alltid säkert att det som ändras visas som det är tänkt.

Alla program, som ingår i systemet, är gratis. Kostnaden blir för personal, drift och hårdvara. Filformaten som används är html-, pdf-filer samt bildfiler i gif- och jpg-format.



Systemet tillåter även en mängd andra filformat för digital publicering och det går att styra vilka filformat som ska tillåtas.

Den löpande kostnaden för administration av publiceringsplattformen bör kunna hållas på en rimlig nivå eftersom systemets grundkomponenter är gratis. Programvaran används av en mängd olika institutioner världen över. Uppdateringar, ändringar och nyheter finns på [EPrints' hemsida](#).

Det finns en e-postlista för att ställa frågor och få råd i frågor rörande programmets funktion, hjälp med programkodning o.s.v.

Om programmet anpassas efter E-press nuvarande system, där studenter (och i framtiden även forskare) själva sköter större delen av publiceringsarbetet blir även personalkostnaden låg.

OAI-protokollet för Metadataharvesting är inbyggt. EPrints klarar inte av att generera Marc 21-filer. SLU hade löst detta med ett eget script. Däremot klarar systemet av att generera en textfil till den lokala bibliotekskatalogen.

## Digitala Vetenskapliga Arkivet (DiVA)

DiVA är ett digitalt publiceringssystem som Uppsala universitet utvecklat med ekonomiskt stöd från bl.a. BIBSAM. Detta system är mycket omfattande och avancerat. Mer information om systemet finns att läsa på <http://publications.uu.se/>. Till DiVA är, förutom Uppsala universitet, Stockholms universitet, Södertörns högskola, Umeå universitet, Örebro universitet anslutna samt Åhus universitetet i Danmark. Universitetet i Trondheim är intresserade att ansluta sig till DiVA.

## Operativsystem och programvara

DiVA använder databasen [Oracle](#) som körs på [Windows 2000 Advanced Server](#). DiVA självt och webgränssnittet körs på [Linux](#) och använder [Apache](#) Java och [Tomcat](#).

Systemet bygger på färdiga mallar:

- Spikningsmall för att spika elektroniskt, vilket är obligatoriskt.
- Fulltextmall som används för att skriva ramberättelsen till sammanläggningsavhandlingen.
- Separat mall för att skriva listan på de artiklar som sammanläggningsavhandlingen är baserad på (se sidan <http://publications.uu.se/forauthors/index.xsql?lang=sv>).

## Publicering

Mallarna finns tillgängliga via Internet (<http://publications.uu.se/forauthors/templates/>) tillsammans med instruktioner för författare. Mallarna finns för Word och Latex. Mallen för spikbladet består av åtta flikar med fält i olika antal (se bild 1 nedan). Denna mall fylls i av författaren. På sidan med instruktioner för författare finns all övrig nödvändig information om hur publiceringen går till. I dokumentmallen registreras sedan de uppgifter nödvändiga för publicering av det vetenskapliga arbetet. Mallen består av åtta flikar där varje flik innehåller flera fält för registrering. Flera fält är obligatoriska och systemet visar ett fält i rött om författaren glömt att fylla i eller om det är felaktiga uppgifter. Utöver mallar finns utförliga instruktioner för bildhantering.

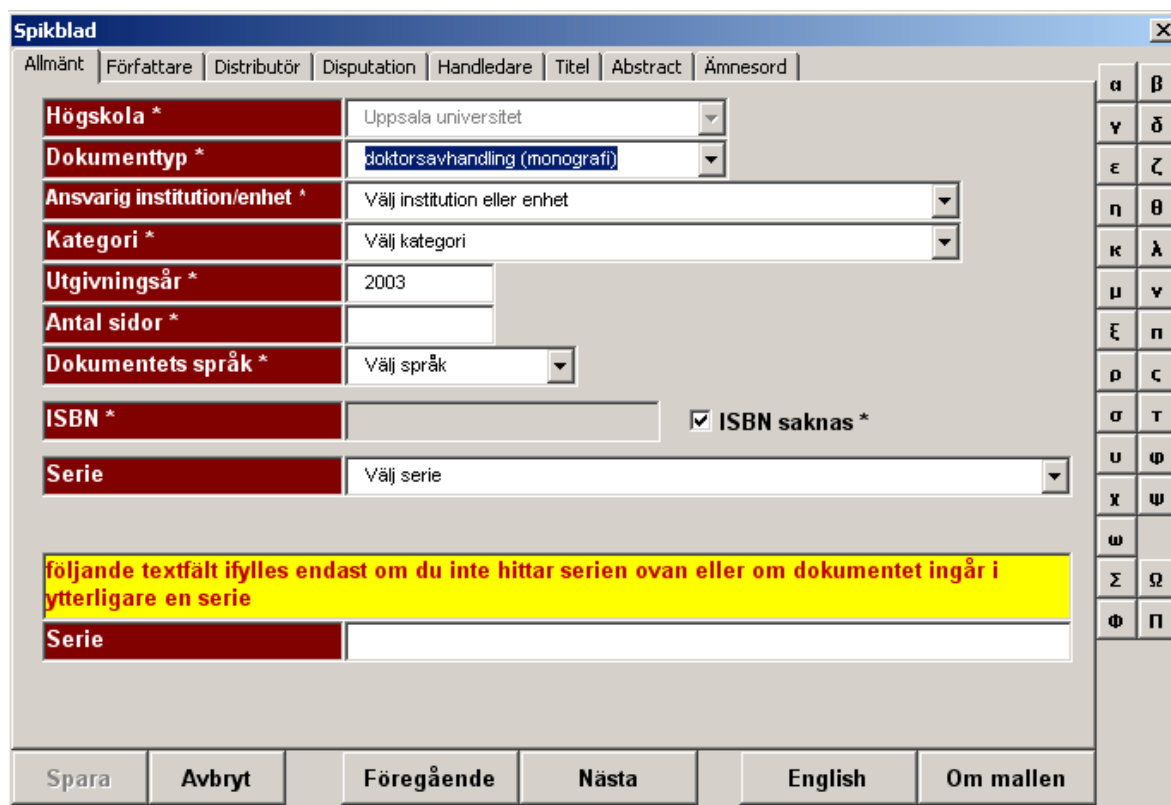


Bild 1.

När fälten är ifyllda skapas ett Word-dokument som tillsammans med pdf-filen skickas med e-post till Enheten för digital publicering vid Uppsala universitetsbibliotek.

Ansvarig person importerar Wordfilen i Diva Manager som laddas ned via webben (se bild 2). Omvägen via Open Office är numera ”osynlig”, då konverteringen sker automatisk på servern, Open Office är installerat.

När filen är importerad kontrolleras att alla uppgifter är korrekta. Därefter laddas filen upp för publicering. Vid publicering genereras en checksumma för varje publikation. Denna visas ännu inte på sidorna. Statistiken är bra, enkel och mycket överskådlig. (<http://publications.uu.se/theses/statistics.xsql>). Den finns på sex språk, där svenska är standard.

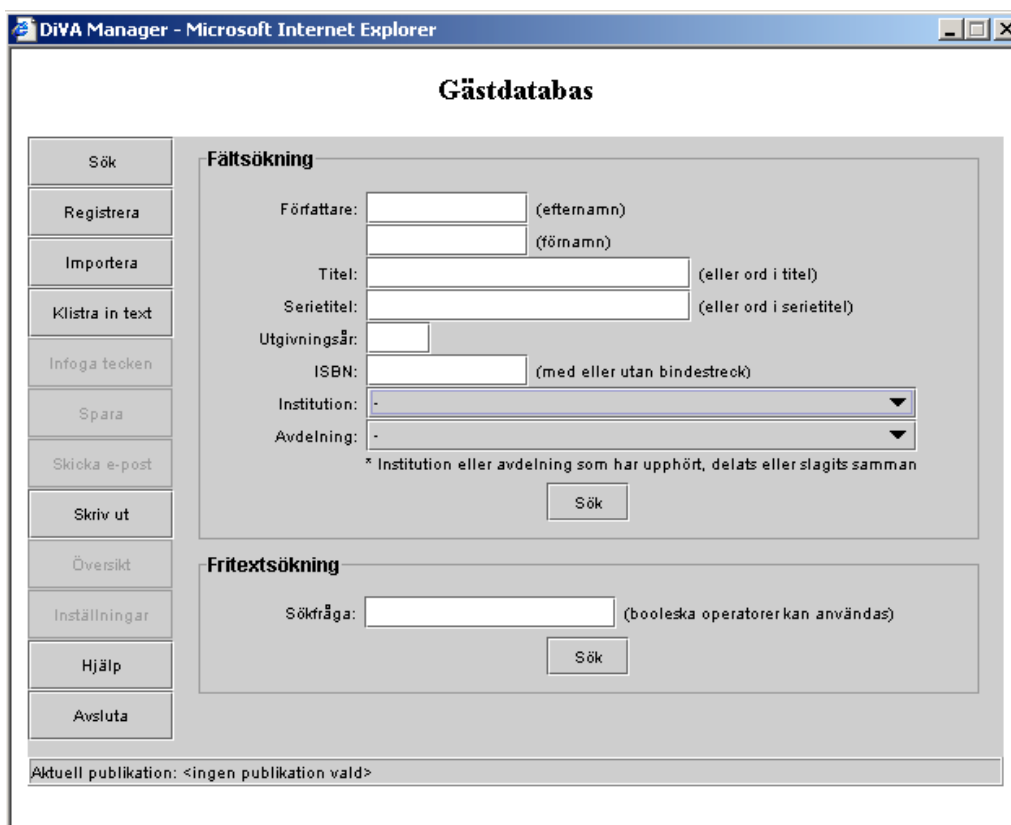
DiVA hanterar unicode (<http://www.unicode.org>) enligt UTF-8. Unicode används i databasen och vid kodningen av html-sidor.

Varje dag levereras metadata om nypublicerade avhandlingar automatiskt till LIBRIS med Marc 21-fil. DiVA hanterar även OAI-metadata.

En prototyp till ett förenklat webbgränssnitt för inmatning av alla typer av publikationer (avhandlingar/examensarbeten/rapporter) har tagits fram under sommaren 2003. I prototypen har antalet metadatafält begränsats för att göra det lättare att lägga in nya poster till DiVA.

DiVA har även en inbyggd URN:NBN Resolver som levererar urn:nbn två gånger per dygn till KB.

Under hösten/våren 2003 kommer fulltexter automatiskt att börja levereras till KB för långtidslagring.



The screenshot shows a web browser window titled "DiVA Manager - Microsoft Internet Explorer". The main content area is titled "Gästdatabas". On the left side, there is a vertical menu with buttons for: Sök, Registrera, Importera, Klistra in text, Infoga tecken, Spara, Skicka e-post, Skriv ut, Översikt, Inställningar, Hjälp, and Avsluta. The main area contains two search sections: "Fältsökning" and "Fritextsökning".

**Fältsökning**

Författare:  (efternamn)  
 (förnamn)  
Titel:  (eller ord i titel)  
Serietitel:  (eller ord i serietitel)  
Utgivningsår:   
ISBN:  (med eller utan bindestreck)  
Institution:   
Avdelning:   
\* Institution eller avdelning som har upphört, delats eller slagits samman

**Fritextsökning**

Sökfråga:  (booleska operatörer kan användas)

Aktuell publikation: <ingen publikation vald>

Bild 2.

DiVA:s webbsidor kan lätt anpassas för nya språk. Under hösten 2003 kommer ett flerspråkigt stöd att implementeras för DiVA Manager. En modul för automatisk tilldelning av ISBN och nummer i serie är under utveckling och beräknas vara i drift under oktober månad 2003.

Samtliga universitet har egna portaler men dessa är placerade under Uppsala universitets adress t.ex. Södertörns söksida <http://publications.uu.se/sh/reports/>.

## Diskussion kring systemets för- och nackdelar

Anslutningsavgiften för DiVA återinvesteras i fortsatt utveckling av systemet. Om flera användare ansluter sig blir kostnaden lägre. Systemet kan anpassas efter önskemål av respektive anslutet universitet. Ett exempel är söksidor med presentation, information, mallar, registrering o.s.v.

För samtliga anslutna universitet finns också en gemensam sökportal för sökning av samtliga publikationer (<http://publications.uu.se/portal/index.xsql?lang=sv>).

E-press kan låta all digital publicering gå över Uppsala universitet och arkivera all digital publicering på deras server. Uppsala universitet ansvarar för administration, drift och underhåll av servrar, program och backup av databaserna. E-Press' URL-adress, [www.ep.liu.se](http://www.ep.liu.se), kan behållas.

DiVA hanterar Marc 21 med automatisk transaktion direkt till Kungliga Biblioteket som per automatik tar in posterna i Libris-katalogen. OAI-protokoll för harvesting av metadata och Uniform Resource Identifier National Bibliographic Number (URI:NB) finns inbyggt.

Anslutningskostnaden för E-press blir cirka 300 000 (september 2003). Summan är baserad på den totala utvecklingskostnaden av DiVA. I anslutningsavgiften ingår 50 timmar för anpassning av systemet till egna förhållanden. Kostnaden därutöver är 660 kronor per



timme inklusive högskoleadministrativa avgifter. Kostnaden för administrativ samordning av systemet fördelas jämt för alla användare.

För löpande administrativ samordning förväntas inte kostnaden överstiga 100 000 per deltagare och år.

Eftersom DiVA hela tiden utvecklas finns för närvarande ingen riktig heltäckande beskrivning av det kompletta systemet. En dokumentation håller dock på att skrivas som ska innehålla nuvarande funktioner och uppdateras när ytterligare nya tillkommer.

DiVA kan skapa filer i en rad olika format såsom pdf, html och vanligt text.

Det bör påpekas att en viss sårbarhet finns med tanke på att Uppsala universitet har full kontroll av all information som Linköpings universitet lagrar på deras servrar. Hela systemet bygger på kontinuitet och att det alltid finns kompetent personal tillgänglig. Detta är mycket viktigt för framtiden.

## DSpace

DSpace (<http://www.dspace.org>) är ett fritt tillgänglig webbaserat publiceringssystem som utvecklats av Massachusetts Institute of Technology (MIT) Libraries (<http://libraries.mit.edu/>) och Hewlett-Packard Labs (<http://www.hpl.hp.com/>). Utförlig beskrivning finns också i artikeln på sidan <http://www.dlib.org/dlib/january03/smith/01smith.html>.

## Operativsystem och programvaror

Dspace behöver ett UNIX-liknande operativsystem som Linux eller Solaris, webservern Apache och Java.

DSpace är skrivet i Java och hanterar OAI-protokoll, Dublin Core metadata samt de flesta kända filformat för lagring av text och bilder. DSpace stöder inte Marc 21 men programmet håller på att modifieras för att stödja flera system för metadata.

Systemet består av moduler där användaren kan gå in och anpassa det till sina egna behov. Buggar kan rapporteras till MIT för åtgärd, som dock kan ta någon tid. För ytterligare information se <http://dspace.or/what/faq.html>.

De digitala format som DSpace använder sig av är text-, bild-, ljud- och videofiler. Vidare klarar programmet av att hantera och lagra

- Artiklar, preprints, working papers, tekniska rapporter och konferenspapers.
- Böcker.
- Avhandlingar.
- Data sets.
- Dataprogram.
- Visuella simulationer och modeller.

## Diskussion kring systemets för- och nackdelar

I Sverige är det för närvarande ingen som använder DSpace (juni 2003). Det är således helt oprövat. Det nyttjas vid en rad andra universitet bl.a. i U.S.A. och universitetet i Roskilde börjar använda DSpace hösten 2003.

DSpace har inte testats tillräckligt för att det ska gå att ge en rättvis bedömning av systemet. Det bör därför utprovas mer utförligt.



## Sammanfattning

Vid studiebesöket och egna tester framstår för närvarande DiVA som det mest genomarbetade systemet för elektronisk publicering. Vi får ett helt igenom färdigt publiceringssystem som går att använda efter några justeringar för att anpassa det till våra behov.

Både EPrints och DSpace bör utvärderas mer. I synnerhet DSpace som inte ännu används i Sverige. Eventuellt kommer både DSpace och EPrints att installeras och utvärderas närmare under hösten 2003.