

Konferenspapper: **Solidarisk E-publicering**

Denna konferens arrangerades av Världsbiblioteket som ett led i dess arbete för en mer öppen och solidarisk syn på elektronisk och digital publicering. Föreläsningarna spelades in och denna text är en utskrift av ljudupptagningen. Tidigare konferenser har behandlat alternativa synsätt på upphovsrätt – bland annat licensavtalet Creative Commons – samt försökt sätta in denna, för vår tid så viktiga, fråga, i ett större perspektiv.

Se även länkar till kompletterande PPT-presentationer från konferensens föreläsare. Mer information om Världsbibliotekets verksamhet hittar du på Solidaritetshusets hemsida.



Världsbiblioteket
Solidaritetsrörelsens Hus
Tegelviksgatan 40
116 41 Stockholm
08-4620787
vbib@solidaritetshuset.nu
www.solidaritetshuset.nu



Detta verk är licensierat under Creative Commons Erkännande-Ickekommersiell-Dela Lika 2.5 Sverige licens. För att se en kopia av denna licens, besök <http://creativecommons.org/licenses/by-nc-sa/2.5/se/> eller skicka ett brev till Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

Inledning av Arwid Lund, Världsbiblioteket

Välkommen till Världsbibliotekets seminarium på temat Solidarisk E-publicering. Kul att se att så många har kunnat komma. Vi var lite rädda när vi satte ihop programmet att det inte skulle komma så mycket folk. Människor brukar ju vara lite rädda för datorer och Internet.

Vad menar vi då med Solidarisk E-publicering? Som vi ser det på Världsbiblioteket så går det att ringa in det solidariska med några enkla begrepp. Det första är *Open Access*, öppen åtkomst, som står för ett ideal om fritt tillgänglig information. Det är ett begrepp och en trend som i hög grad kommer från den akademiska världen, där man har haft problem med att offentligt finansierad forskning har privatiserats genom att den publiceras i kommersiella vetenskapliga tidskrifter. Men tankegångarna bakom begreppet går tillbaka på en äldre tradition som betonar att utbytet mellan forskarna i den akademiska världen ska vara fritt och icke-kommersiellt.

Detta tema kommer *Jan Hagerlid* från Kungliga Biblioteket (KB) att utveckla i sin inledning till denna konferens.

Ett annat begrepp som ringar in det solidariska är *Open Source*, Öppen Källkod. För att publicera resurser elektroniskt krävs programvara och denna programvara kan antingen ha öppen eller stängd källkod, antingen är det ett gemensamt eller ett kommersiellt och stängt projekt - en vara.

Aina Svensson från Sveriges Lantbruksuniversitet kommer att presentera ett fritt publiceringsverktyg här idag.

Vi på Världsbiblioteket tycker även att det solidariska kan vara att underlätta för andra att ta del av de elektroniska resurser som man själv erbjuder via nätet. Eller att tillgängliggöra och synliggöra andra förenings resurser.

Peter Hansson från Chalmers kommer att prata om hur man kan erbjuda beständiga länkar till sina resurser och öka interoperabiliteten mellan olika digitala förråd. Medan *Stefan Andersson* från Uppsala Universitetsbibliotek kommer att ta upp en del informationstjänster som man kan konstruera med ny informationsteknologi.

Men det verkligen fina i kråksången med det solidariska är att det som är solidariskt också är det som är effektivt när det kommer till informationsarbete.

Vi som anordnar konferensen heter *Världsbiblioteket* och *Globalportalen*. Världsbiblioteket är ett litet bibliotek med 7.000 media. En av anledningarna till att vi satsar på det digitala arbetet är att vi vill nå ut till nya grupper inom ett större geografiskt område. Vår uppgift är annars främst att serva det 30-tal föreningar som finns i Solidaritetshuset i Stockholm och se till att deras informationsbehov är tillgodosedda. Vår hemsida hittar ni på www.solidaritetshuset.nu.

Dessutom är vi med i nätverket Globalportalen som Johanna Lindell från LO/TCO:s biståndsnämnd och Globalportalens styrgrupp nu kommer att presentera närmare.

Presentation av Globalportalen av Johanna Lindell

Globalportalen består av ett 30-tal olika organisationer som presenterar olika nyhetsnotiser om vad som händer i världen. Om någon organisation som finns representerad här idag skulle vilja bli del av Globalportalen så går det bra att prata med mig i pausen. Vill ni veta mer om oss kan ni gå in på www.globalportalen.org.

Föreläsning om Open Access av Jan Hagerlid, Kungliga Biblioteket

[Utskrift av ljudupptagning. PPT: www.solidaritetshuset.nu/usrd/uns666.ppt]

Jag jobbar på Kungliga Biblioteket där jag har ansvar för att stödja våra insatser för att stödja elektronisk publicering vid svenska universitet och högskolor. Jag ska prata om vad det här med Open Access egentligen är för något, ge en bakgrund, historik och titta lite på vad man lägger i begreppet och vad det betyder idag. Men samtidigt under vägen ska jag peka på en del källor som ni kan använda er av för att hitta fritt tillgängligt material som ni förhoppningsvis kan komma att ha användning av i ert arbete. Så jag kommer att gå igenom Open Access-tidskrifter och Eprint-arkiv och bibliotekens roll i sammanhanget med kostnadsspiralen kring vetenskapliga tidskrifter som gör oss och många forskare så frustrerade. Jag kommer också att ta upp lite av de politiska nivåerna i denna fråga.

Så det blir en blandning av konkreta tips och en liten historik över området eller frågan. När man tittar tillbaka så blir man förvånade över hur länge processen har pågått. Vi tänker ibland i sådana här webbår som är ungefär två månader eller så där. De första e-tidskrifterna fanns före webben! De kom redan runt 1990. Bland annat *Psycoloquy*. Så det här har funnits rätt länge. Och det var fritt tillgängliga tidskrifter på Internet med en primitiv teknik. Men det var en första start och man kan säga att hela den här utvecklingen drivs av två krafter. Den ena är att forskare har velat använda sig av Internet för att sprida, kommunicera med varandra och den andra är just bibliotekens missnöje med hur marknaden fungerar.

Redan i början 1990-talet kom det första fritt tillgängliga arkivet för elektroniska artiklar. Och det var väldigt typiskt att det uppkom inom fysikområdet där man har behov av att snabbt kommunicera. Det händer mycket inom området. Har någon gjort en upptäckt så måste den snabbt visa det för kollegerna så att man blir först om den. Det fanns en tradition att forskarna skickade så kallade pre-print, det vill säga artiklar som de ville ge ut som vetenskapliga artiklar. Dessa skickade man med post till varandra. Nu tänkte man att man kunde lägga in de här pre-prints i en databas och göra dem tillgängliga för alla. Och det gjordes redan från 1991 av en pionjär som heter *Paul Ginsparg*.

Efter hand kom det sedan in post-print, och det betyder alltså en artikel som är accepterad för publicering, den behöver inte vara tryckt och ligga på disken, men den är accepterad för publicering. Hela begreppsparet pre-print och post-print är lite konstiga eftersom vi pratar om elektronisk publicering, men det är de begrepp som vi använder i sammanhanget.

Så den här databasen i fysik innehåller idag mycket post-print, det vill säga man har en modell där en forskare lägger ut en pre-print, det vill säga en artikel som en tidskrift ännu inte har tagit in, som bara har skickats in till en tidskrift, för att sedan när den blivit accepterad för publicering kompletteras av en post-print. Forskaren knyter ihop de två versionerna och sedan ges artikeln ut i en fysiktidskrift som begränsas av prenumerationer eller licensavtal även om den är i tryckt eller elektronisk form. På detta sätt når man omedelbart alla forskare inom fysik som är intresserade. Det är ett väldigt snabbt, billigt och effektivt sätt. Men man får komma ihåg att man inte gör något av det som en tidskrift gör, man gör ingen kvalitetsgranskning. Det finns en basal form av granskning men ingen egentlig kvalitetsgranskning.

Inom vissa områden inom fysiken så är den här databasen det första stället man tittar på om man vill veta vad som händer. Den har en ganska stor omfattning och den växer hela tiden. Idag finns ca en 400.000 artiklar här. Jag vet inte hur många av er som är intresserade av fysik men det här är så att säga modellen. Det här är eprint-arkivens urmoder skulle man kunna säga.

Sedan växte det efter hand upp liknande arkiv inom andra ämnesområden som hade detta arkiv som förebild. Ett som är väldigt stort heter *RePec*, Researchpaper in Economics, och det är ett nätverk av över 50-talet institutioner som bidrar med material till detta. Det är både rapporter och artiklar. Har finns det en stor mängd litteratur. Jag nämner även ett annat som heter *CogPrints* inom kommunikationsvetenskap. Det finns ytterligare ett antal sådana här ämnesbaserade eprintarkiv där forskare lägger in antingen pre- eller post-printmaterial eller bådadera.

Då hoppar vi snabbt. När det hade skapats ett antal sådana här arkiv, så var man angelägen om att knyta samman dem. Man ville att det samtidigt skulle gå att söka i dessa arkiv. Och då skapades en teknisk lösning för att medge detta. Och det är något som kallas för Open Archives Initiative (OAI) som är ett protokoll för metadata-harvesting. Alltså hösta in metadata. Metadata är här som kataloguppgifter som står på ett katalogkort. Man höstar in datum och publikation och på basis av den gör man en databas som man söker, och när man har hittat en publikation så går man via en länk då till fulltexten och får fram den. Och det här protokollet har spelat en väldigt stor roll för hela den här rörelsen. Det har gjort det möjligt att sammanknyta en mängd sådana här fria eller öppna arkiv. Först var det framför allt sådana som var baserade på ämne, och sedan kom det i allt högre grad sådana som var förankrade i en institution. Det vill säga ett universitet, en akademi eller ett forskningsinstitut lade ut sina skrifter och gjorde dem tillgängliga genom detta OAI-protokoll. Och det har då växt väldigt mycket.

Det finns inga säkra uppgifter om hur många sådana här öppna arkiv som det finns idag: 400 till 600 det finns olika uppgifter. Det är väldigt många universitet inte bara i främst Västeuropa och USA ska jag säga, utan även i Brasilien, Kina och Indien finns det. Det är inte enbart i väst. Sedan uppkom det ganska snart efter det här med OAI-protokollet sådana här programvaror som skulle göra det möjligt för till exempel ett universitet att skapa ett publiceringsverktyg som är fritt tillgängligt. Aina Svensson kommer att prata om det största, det som har flest användare, EPrints som utvecklats i Southhampton. Det finns ett annat som är väldigt använt som heter DSpace som är utvecklat vid MIT i samarbete med dataföretaget Hewlett Packard. Så den här utvecklingen eskalerade. Från de här enstaka open access-tidskrifterna och de här enstaka e-printarkiven som bands samman med Open Archives Initiative. En våg sattes igång där en mängd universitet satte igång att skapa öppna arkiv. På engelska talar man om Digital Repositories, Academic Repositories och Institutional Repositories. Det är svårt att hitta bra ord på svenska men jag föredrar termen Öppna Arkiv. Repositorium låter lite som en begravningsplats.

Det finns en tjänst som heter *OAIster*. Och det är en liten ordvits då på OAI och en mussla, eller förlåt ostron, och detta är väl den största söktjänsten som finns för att söka av öppna arkiv som finns i världen. De säger att de har över 10 miljoner poster från 701 institutioner. Tjänsten drivs vid University of Michigan.

Ja det här var då utvecklingen från forskarhåll. Sedan formerade de amerikanska forskningsbiblioteken 1998 en Scholarly Publishing and Academic Resources Coalition, förkortat *SPARC*, som då låter väldigt, betyder gnista, visar på någon slags kampvilja. De ville åstadkomma rimligare förhållanden på tidskriftsmarknaden. Man ville främja konkurrens och lägre priser. Så bland annat gav man stöd till nya vetenskapliga tidskrifter. Tidskrifter som var billigare och som bevisade att man kunde publicera väsentligen billigare tidskrifter. Det hände bland annat att hela redaktioner hoppade av från etablerade vetenskapliga tidskrifter och startade en ny med samma inriktning och annat namn. Det förekom en del

spektakulära avhopp vid den här tiden. Man kan säga att nu är SPARC helt inne på att stödja Open Access-arbete. Det ska vara fritt tillgängliga vetenskapliga tidskrifter och det stöder man genom information, lobbyarbete och stöd till dem som vill starta nya tidskrifter. I SPARCs regi finns också den bästa nyhetskällan om man vill följa utvecklingen på detta område. Det är något som kallas för *Open Access News* och är också en blogg som drivs av en oerhört energisk man som heter *Peter Suber*, som är någon slags filosof. Men det är en jättebra källa, den följer inte bara Open Access utan även Open Source, copyright-frågor, det är en lite bredare källa alltså. Och har ofta mycket intelligenta och tänkvärda kommentarer till utvecklingen.

Varför var då biblioteken så bekymrade, förargade eller upprörda? En snabb summering över det man kan kalla det vetenskapliga tidskriftsproblemet ger vid handen att vi har cirka 24.000 kvalitetsgranskade vetenskapliga tidskrifter idag och dom producerar 2,5 miljoner artiklar per år. Det här ökar med 3,5 procent per år. Det är en väldig massa, väldig volym av den vetenskapliga publiceringen. Priserna ökar högt över inflationen sedan lång tid tillbaka. Man kan säga att det mattats lite de senaste åren då biblioteken har hittat en form av lösning genom att bilda konsortier. I Sverige har man till exempel ett konsortium i Bibsams regi som gemensamt handlar upp tidskrifter. Man kan säga att man får fler tidskrifter för pengarna och har bromsat prisökningarna. Men de är fortfarande över inflationen. Och hur man än behandlar det här så kan man fastslå att inget bibliotek har medel att köpa in allt som efterfrågas. Och det gäller även de rikaste universiteten i de rikaste länderna. Som ni förstår så gäller det i än högre grad för fattigare länder och fattigare universitet. I hela tredje världen är tillgången till vetenskapliga tidskrifter ett jätteproblem. Det finns en del program där man erbjuder prenumerationstidskrifter till billigare kostnader, reducerade kostnader, men det är fortfarande ett gigantiskt problem.

Kan ni se det här? Det är en klassisk graf när det gäller tidskrifternas prisutveckling. Det här visar tidskrifts- och monografikostnader vid amerikanska forskningsbibliotek 1986 till 2002. Ni kan se att den här stapeln är tidskriftskostnaden och den rakar konsekvent i höjden jämfört med andra kostnader. Och sen har vi standardprisutvecklingen. Det är den blå kurvan. Under den här perioden har den varit 68 procent, den allmänna inflationen, medan enhetskostnaden för en tidskrift har varit 215 procent under den här perioden, vilket är väldigt mycket över inflationsgenomsnittet. Monografierna har inte alls ökat så mycket. Och det man har råd att köpa är väsentligen mindre. De här kurvorna här nere är de tidskrifter man har haft råd att köpa. Det här är, kan man säga, en illustration över bibliotekens situation. Utgifterna rakar i höjden och man är tvungen att skära bort en hel del tidskrifter, man kan inte köpa lika mycket monografier. Den här kostnadsökningen är sammansatt då det finns kommersiella förlag som har en ännu starkare, högre relativ prisutveckling. Och sedan finns det icke-kommersiella vetenskapliga sällskap som har en måttligare utveckling. Så jämför man prisutveckling mellan dessa grupper så visar det sig att de kommersiella tidskrifterna har en ännu kraftigare, ännu högre relativ prisnivå.

Man kan säga att det här är en process. Ungefär fram till andra världskriget så var internkommunikationen inom vetenskapssamhället helt och hållet en fråga som sköttes av akademien. Det fanns inga pengar, inga vinster att hämta och prisnivåerna var rätt obetydliga. Sedan kom det en snabb expansion av forskningen efter andra världskriget och de gamla modellerna klarade inte riktigt av att hantera flödet av artiklar. Kommersiella förlag såg att det här var en möjlighet, en affärsmöjlighet. Så då fick man gradvis en utveckling där den här publiceringen togs om hand av kommersiella bolag. Men man hade ändå kvar den här

traditionen att det rörde sig om internkommunikation; det vill säga författare i vetenskapliga tidskrifter, redaktörer, de som utvärderar – ingen får betalt. Så det här har varit en modell för internkommunikation. Och det lever kvar, de tar inte betalt. Priserna sätts efter betalningsvilja egentligen, vad som går att ta ut.

De här defacto-monopolen som ett antal förlag då har gör att man kan ta ut oerhört höga vinstmarginaler. Så de största kommersiella förlagen ligger då bland de allra mest vinstgivande företagen av alla företag över huvud taget. Fortfarande gör de det om man tittar på sådana här listor. Universiteten får då betala löner för forskare, författare, redaktörer, utvärderare, och i andra änden också betala dyra prenumerationskostnader. Så universiteten får ändå inte ut det som är målet för forskarna såväl som universiteten – det vill säga en maximal spridning på de forskningsresultat man har åstadkommit.

Fråga från publiken: Har du någon aning om hur stor del av forskningen som blir till artiklar i dagspress och liknande?

Jan Hagerlid: Nej, jag har ingen siffra på det, på hur många som kommer ut i dagspress? Som blir nyheter? Det är en försvinnande liten del, försvinnande liten andel, måste det ju vara som blir riktigt nyhetsstoff. Man ser ju ofta sådana här nyheter och då är det ofta artiklar som har publicerats i Science eller Nature. Men det är ju en promille, skulle jag tro. Man får ju komma ihåg att det mesta av det här är ju vetenskapliga specialområdets internkommunikation. Men vissa delar har ju ett bredare intresse. Det beror lite på ämnesområdet. Jag tror att få kan förstå avancerad matematik och fysik, medan klinisk medicin finns det en mängd läkare och sjuksköterskor utanför universiteten som kan ta del av. Men det som kommer ut i massmedia är en väldigt liten del vill jag bestämt påstå.

Det var alltså forskare då, om jag fortsätter med biomedicin, som var oroadе ansåg att det här inte var ett bra läge. Man ville uppnå en situation där hela den vetenskaplig publicering inom ett område var fritt tillgänglig, man fritt kunde söka igenom det, men det fick man inte för materialet var så att säga inlåst i olika tidskrifters avtal. En nobelpristagare som heter *Harold Varmus*, han var då chef för National Institute of Health, som är världens största finansierare av biomedicinsk forskning – även svensk forskning får pengar från National Institute of Health faktiskt – han lanserade då idén om PubMedCentral. Det skulle vara ett fritt tillgängligt arkiv av tidskriftsartiklar. Han ville hitta en form för att samarbeta med förlagen och säga att ”visst, vi vill bara att ni deponerar era artiklar, helst 6 månader men högst 12 månader efter publicering”. Och sedan därefter skulle de vara fritt tillgängliga. Men det var alltså ingen stor anslutning till förslaget från förlagen. Idag finns det då 185 tidskrifter i det här arkivet. Man bildade då en lobbygrupp, *Public Library of Science* och hotade med en bojkott, man gick ut med ett öppet brev på nätet och sade att om förlagen inte går med på det här kravet att deponera artiklarna inom 12 månader efter publicering, så tänker vi inte längre skriva i era tidskrifter och vi tänker inte längre redigera era tidskrifter och vi tänker inte bistå er med utvärderingar. Och det blev en väldigt debatt. Det här satte frågan i fokus i forskarsamhället. Nu var det 26.000 forskare från i stort sett de flesta länder på jorden som skrev under det här öppna brevet. Så jag tror det var väldigt viktigt för att väcka intresset för den här frågan. Sen tror jag inte alla uppfyllde det de hade sagt i det här brevet, jag vet inte hur många som faktiskt hoppade av, som kanske fortsatte att skriva ändå.

Den här gruppen *Public Library of Science*, som bestod av mycket erkända medicinska forskare, gick vidare och skapade fritt tillgängliga tidskrifter som hade avsikten att ligga på

den högsta statusnivån. Så att de skulle vara i klass med Nature och Science. Och de har lyckats ganska bra. *Public Library of Science - Biology* och *Public Library of Science - Medicine* är väldigt högt rankade tidskrifter idag. Det finns ytterligare några.

Ett annat viktigt opinionsmässigt steg var något som hette *Berlindeklarationen* som kom 2003 och det var initierat av tunga forsknings och akademiska institutioner i framförallt Tyskland och Frankrike. Det finns något som heter Max Planck-institutet, som ni kanske har hört talas om, som är en jättelik organisation i Tyskland med en massa forskningscentra, som står bakom detta. Det europeiska atomforskningsorganet *CERN* i Schweiz står bakom detta och forskningsinstitutet *CNRS* samt *INSERM* i Frankrike står bakom detta. De enades om en deklaration som är en slags avsiktsförklaring om att arbeta för Open Access. Och den här deklarationen skrevs under av generaldirektörer och motsvarande för de här organisationerna. I deklarationen sägs: "Our mission of disseminating knowledge is only half complete if the information is not made widely and readily available to the society", och man vill uppmuntra forskarna och de som mottar forskningsbidrag att publicera sitt arbete enligt principerna i ett open access-paradigm.

Det här fick väldigt stort genomslag och efter hand anslöt sig universitetsrektorerna och forskningsråd från ett stort antal länder vars institutioner, viktiga organisationer i forskarsamhället, idag har skrivit på den här deklarationen, jag tror att det rör sig om en 200 eller så där.

Den här deklarationen var ganska vag egentligen och det finns en uppföljning vid ett möte i Southampton som liksom lyfter nyckelfrågorna egentligen, och man menar där att man ska ha en policy som kräver att forskare deponerar en kopia av alla publicerade artiklar i ett öppet arkiv, det är det första kravet, det andra är att man ska uppmuntra forskare att publicera sina artiklar i fritt tillgängliga tidskrifter om det finns en lämplig sådan. Man måste göra detta möjligt. Det är alltså lättare att ha obligatorium om att publicera din artikel i den tidskrift där du finner det lämpligast, men deponera den i ett öppet tillgängligt arkiv. Forskare väljer väldigt mycket var de vill publicera sig, om tidskriften är rätt ämnesmässigt, om den har god status i forskarsamhället – och det är inte säkert att det finns en fritt tillgänglig tidskrift som svarar mot de kraven. Därför pratar man bara om att uppmuntra när det gäller att publicera i Open Access-tidskrifter.

I Sverige så har då SUHF:s, som är Sveriges universitets- och högskoleförbund – alltså en intresseorganisation för universitet och högskolor, skrivit under det här, Vetenskapsrådet har gjort det, Svensk biblioteksförening, Kungliga biblioteket och nu senast Stockholms universitet. Och SUHF har också anslutit sig till de här två punkterna som jag nämnde alldeles nyss: att man ska rekommendera forskare att deponera en kopia av varje publicerad artikel i ett öppet arkiv och uppmuntra att publicera i tillgängliga tidskrifter. Så det finns idag en ganska stor enighet bland ledande aktörer inom forsknings- och biblioteksvärlden att stödja detta. Sen är det ändå en ganska lång väg att gå i praktiken för att få det att slå igenom i stor skala. Vetenskapsrådet har till exempel ännu inte ändrat de krav man har i sina kontrakt, men man överväger att göra det, och skulle man göra det så får det väldigt stor betydelse eftersom de delar ut så mycket pengar till forskning. Andra forskningsråd, i Storbritannien till exempel har ställt sådana här mer obligatoriska krav: Får ni pengar av oss ska ni också se till att er artikel är fritt tillgänglig på något vis. Jag återkommer till praktiken, men jag tror inte att detta är så långt borta.

Den här frågan om fri tillgång har varit uppe på politisk nivå och i det brittiska parlamentet var det en underhuskommitté för vetenskap och teknik som gjorde en utmärkt utredning för

några år sedan där man frågade alla som ingår i branschen och hade ganska skarpa förslag och rekommendationer. I USA:s representanthus har frågan också varit uppe och där ville man då ställa ett hårt krav på nyss nämnda National Institute of Health om att alla som fick pengar för biomedicinsk forskning också skulle se till att resultaten var fritt tillgängliga. Och nu finns det flera förslag som väntar på beslut i kongressen. Men man har alltså inte riktigt tippat över balansen, samtidigt är det en väldig lobbyverksamhet från de största förlagen mot den här utvecklingen. Framför allt i USA och Storbritannien. Hos oss märker vi inte så mycket av det. EU-kommissionen gjorde i våras en utmärkt rapport om den vetenskapliga publiceringsmarknaden, som har varit ute på remiss nu. Man kan säga att den överlag är kritisk till att det är för mycket monopol på den här marknaden och man förordar att man ska arbeta för fri tillgänglighet till offentligt finansierad forskning.

Om man då skulle sammanfatta vad de grundläggande motiven för Open Access är, så är det väl att man vill uppnå maximal spridning av forskningsresultat. För forskaren och universiteten så är det att uppnå en maximal synlighet. Det är synlighet som är forskarens belöning i det här sammanhanget. Och uppmärksamhet. Och ju vidare du kan sprida din forskning ju större chans är det att någon citerar den. Och det visar sig också att om man har sina artiklar fritt på nätet så får man fler citeringar idag. Sen är det väl också så att på något sätt så finns det ju, det måste ju betalas för vissa saker i det här sammanhanget, man måste på något sätt betala för en organisation för att organisera kvalitetskontroll, men mycket talar för att Open Access är en effektivare modell i en digital miljö. Därför att om man nu lägger betalningen, även om det skulle kosta lika mycket att ge ut Open Access-tidskrifter som *"toll access-tidskrifter"* så skulle man ändå för samma pengar nå en maximal spridning. Man skulle nå alla tänkbara läsare. Sen tror jag också att om man vänder på det och lägger betalning där man producerar kunskap så finns det också större förutsättningar för att få någon form av effektiv konkurrens. Då kan man konkurrera om att vi, vår tidskrift är lika bra som den där tidskriften och vi kan faktiskt ge ut det här med mindre kostnader.

Ett viktigt motiv för forskare är principen om öppen tillgång till forskningsresultat. Det ingår i forskarvärldens grundvärden. Sen tror jag också att vi har en situation idag där användare över huvud taget förväntar sig att hitta det mesta fritt på nätet. Det finns så väldigt mycket tjänster som baserar sig på att de ligger fritt på nätet och sedan tar man in underlag eller pengar på något annat sätt. Jag tar Google som ett exempel.

En annan viktig sak är att vetenskaplig information behövs i allt högre grad utanför högskolan. Ett exempel är då läkare, men det finns många andra, det finns akademiska yrkesgrupper som även sen de lämnat universitetet behöver tillgång till vetenskaplig information. Och det finns alla möjliga grupper och ni är väl en av de, om man är specialist på ett område, en solidaritetsorganisation, det kan vara patientföreningar, det finns en mängd föreningar där man faktiskt behöver tillgång till vetenskaplig källinformation. Sen varierar det väldigt mycket beroende på ämne i vilken mån man kan ta till sig det. För forskare ser man också en möjlighet att utnyttja webben för att länka samman forskning. Man talar om *datamining*. Det ges ut så fruktansvärt mycket inom medicin till exempel så att man för att över huvud taget kunna tillgodogöra sig det, behöver hjälp av artificiell intelligens för att plöja igenom tusentals och tusentals artiklar, och leta efter kombinationer av vissa ämnen och sådant där för att få fram innehållet på ett annat sätt. Man har behov av att kunna plöja igenom hela den vetenskapliga litteraturen på nätet helt enkelt.

Snabbare spridning, när vi har elektronisk form så sprids informationen snabbare till minskade kostnader, är viktigt för hur behoven ändras. Lite grann tror jag att vi är i en

situation som vi var i på flygområdet för några år sedan med lågprisbolagen och de etablerade bolagen. Och nu har vi sett att det går att erbjuda flyg till en billigare kostnad och man får någon slags reell konkurrens i det här.

Så hur ser läget ut idag då när det gäller Open Access? Man kan säga att av de kvalitetsgranskade vetenskapliga tidskrifterna så är det cirka sju procent som är fritt tillgängliga idag. Och de kan ha olika finansieringssätt, de kan vara kommersiella eller de kan vara ideella, det är väl det vanligaste. I vissa fall tar man ut en publiceringsavgift från författaren eller hans institution eller forskningsfinansiär, i vissa fall tar man inte ut någon avgift alls. Det finns många olika modeller. Ofta drivs sådana är tidskrifter egentligen därför att forskaren bjuder på sin tid som betalas av hans institution.

Sen har vi då exemplen med forskare som lägger ut kopior av forskningsartiklar i öppna arkiv. Det kan vara sådana som finns på högskolorna, inom ämnesområden som jag nämnt tidigare. Det vanligaste är faktiskt att forskare lägger ut artiklar på sin personliga hemsida eller institutionens. Så det flyter omkring väldigt mycket fritt material på nätet. Man tror att omkring 20 procent av den årliga produktionen av vetenskapliga artiklar är fritt tillgängliga på något sätt.

Hur ser det ut i Sverige då? Ja man kan säga att de flesta högskolor och universitet har någon form av akademisk elektronisk publicering. Biblioteken har i regel huvudansvaret för det. Det finns avhandlingar, rapporter och artiklar och volymen ökar hela tiden. Det finns några olika publiceringsplattformar. DiVA-konsortiet som drivs från Uppsala och som är de som omfattar de flesta i Sverige, sedan har vi EPrints och DSpace med egen utveckling. Det finns lite olika lösningar.

Det här är ingångssidan till DiVA. Rätt många universitet där man till exempel kan söka efter avhandlingar, här samsöker man dem.

Under åren 2003-2005 genomfördes ett stort nationellt projekt för att samordna e-publiceringen inom högskolorna. *SVEP-projektet* arbetade med samordning av metadatastandarder, verktyg för långsiktig tillgång och nya söktjänster. Projektet finansierades av KB/BIBSAM.

Byte av band. Avbrott.

Jag kan nämna ett resultat av det här projektet. Och det är att det finns en söktjänst, *Uppsök*, där man kan söka efter uppsatser från svenska universitet och högskolor. Finns drygt 20.000 uppsatser i den här, och det är på olika nivåer, det slog an ganska fort. SVEP-projektet har efterträtts av ett program som heter *OpenAccess.se* och det är KB som finansierar det, men det genomförs i samarbete med SUHF, Vetenskapsrådet och Kungliga Vetenskapsakademien. Och där vill man då på olika sätt arbeta för maximal synlighet för de arbeten som publiceras av lärare, forskare och studenter. Man sysslar i det här programmet med standarder, verktyg, en viktig uppgift är att se till att innehållet i de här arkiven ökar snabbt. Vi försöker stödja nya tjänster för att ge tillgång till sådant material, det är kvalitetsfrågor, det här med långsiktig tillgång är också en viktig fråga. En ny fråga för oss är att vi vill stödja en övergång till Open Access-modeller för svenska vetenskapliga tidskrifter.

Det här är *ScieComs* hemsida, Svenskt resurscentrum för vetenskaplig kommunikation, som KB har stött ett antal år. Det ligger i Lund. Här finns det väldigt mycket information om hela den här processen. Webbplatsen är för närvarande fryst men det finns mycket information att hämta här. www.sciecom.org.

En tjänst som är viktig, tror jag, för er kan vara *Directory of Open Access Journals*, som drivs av Lunds universitetsbibliotek med stöd av KB men även av andra. Här har man samlat ihop alla de fria, vetenskapliga och kvalitetskontrollerade tidskrifter som finns. Man kan söka på olika ämnesområden. Man kan också söka direkt i artiklarna på en tredjedel av de här. Och idag finns det ungefär 2.500 vetenskapliga tidskrifter sökbara i den här tjänsten. Och den är internationellt väldigt välkänd. Så *Directory of Open Access Journals*, *DOAJ* alltså. Man kan komma med i den här tjänsten om man har någon form av kvalitetskontroll. Det behöver inte vara peer-review utan kan vara en redaktionell kvalitetskontroll om det är en vetenskaplig tidskrift man driver.

Slutligen en tjänst som givetvis är användbar när man vill hitta fritt tillgängligt material och det är Google. Och kanske i synnerhet *Google Scholar*, vilket är en ny tjänst som Google har dragit igång där man avgränsar sökningarna till det vetenskapliga materialet. Här kan man både hitta sådant material som är fritt tillgängligt, men också sådant som är begränsat i tillgänglighet. Men i allmänhet kan man alltid få fram abstracts eller något. Den har väldigt stor täckning. Den är naturligtvis väldigt snabb. Och du har också någon slags ranking av materialet så att man sätter det mest intressanta högst, vilket är värdefullt, det finns även citeringsdata, det vill säga hur många har citerat den här artikeln. Man kan gå vidare och följa sådana här citeringslänkar. Så det är väl ändå den mest omfattande tjänsten du kan ha idag för att hitta det här materialet. OAIster som jag nämnde tidigare har också väldigt bred täckning. Och det skapas ständigt nya tjänster. Ny är *BASE*, Bielefeld Academic Search Engine, där kan man använda en modell där man kan tala om att man bara vill ha fritt tillgängligt material, man klickar bort andra alternativ i sökrutan. Så det finns många källor att hämta fritt tillgänglig vetenskaplig litteratur idag. Och jag ser det väl som att det är en politisk process, det är krafter i rullning som gör att vi, jag tror att vi kommer att få en allt större del av den vetenskapliga produktionen fritt tillgänglig i framtiden. Vilket jag tror är samhällslyst nyttigt. Ja, snabbt avslutar jag där. Tack.

Fråga: I USA finns en tradition av att man lägger ut myndigheters material i "Public Domain"? Hur hänger det begreppet ihop med detta?

Jan Hagerlid: Ja det är lite separat egentligen. Vi har inte Public Domain i Sverige eller i Europa. Men är det Public Domain så är det ju i och för sig Open Access också. Men det ligger en del annat i begreppet Public Domain än att det bara är fritt tillgängligt. Jag vet inte hur det är med upphovsmannarätten till exempel. För är det Open Access så har du ju fortfarande som upphovsman en ideell upphovsrätt; det är ditt verk, citeras det måste man tala om att det är du som har gjort det och man får inte förvanska det. Undrar om inte Public Domain är lite mera anonymt i det avseendet. Men jag är inte någon expert på det där. Det är väl i princip material som inte har någon upphovsrätt, men jag är lite osäker på vad man får göra med det.

Fråga: En sådan tradition i USA borde göra deras myndigheter mer benägna att tvinga högskolorna att lägga ut material?

Jan Hagerlid: Jo, det är ju så och därför har det ju kommit så långt som det har gjort i kongressen. Alltså det finns något som heter FRPAA, någon ACT som är uppe för diskussion i kongress. Den har lagts fram av den här senatören Lieberman som höll på att bli bortröstad,

men som faktiskt kom in som oberoende kandidat. Det är ett långtgående förslag. Där kräver man att alla federala forskningsfinansiärer ska ha ett bestämt krav på fri tillgång till det material man ger forskningspengar till. Så det är fullt möjligt att det slår igenom i full skala.

Fråga: Hur går det för användarna av allt det här, det finns så många söktjänster, kommer det i framtiden att finnas en gemensam söktjänst?

Jan Hagerlid: I princip har vi ju det med Google. Dom stora sökmotorerna fungerar så. Men de har ju kvalitetsbrister. Det är kanske klokt att kombinera de här allmänna sökmotorerna med att gå in på specialiserade tjänster som kanske ger en djupare beskrivning eller har en bättre kvalitet på metadatan.

Peter Hansson: Problemet med att använda Google är att man inte får reda på vad de inte indexerar, och man kan inte veta vad de indexerar.

Fråga: Men hur tror ni att det blir i framtiden då?

Jan Hagerlid: Förmodligen går det mycket mot att man ska skapa samlade ingångar. När det gäller universitetsvärlden så finns det ju ett gemensamt projekt att man ska skapa en portal som heter *Samsök*. Men den är tillgänglig bara om man är inskriven vid ett universitet. Idag. Det är möjligt att de fria resurserna i *Samsök* ska ligga ute. Det har i alla fall varit en tanke att de ska göra så. Där söker man så att säga både i fria resurser och i alla specialdatabaser som finns på olika områden. Som då är mera djupa i sin beskrivning. Portaltänkandet är ju starkt och tanken att det ska finnas One-Point-Entry, ett ställe där du kan få allt du behöver. Men jag tror aldrig att vi kommer att få den perfekta lösningen. Jag tror att vi alltid kommer att ha ett antal olika konkurrerande tjänster. Det uppstår ju nya tjänster hela tiden som tillför något nytt.

Fråga: En komplikation med det, om det bara skulle finnas en söktjänst, är väl att man inte kan göra så bra sökningar då? Det blir svårare att hitta det du är intresserad av.

Jan Hagerlid: Du kan inte vara så preciserad. Det går ut över precisionen.

Peter Hansson: Men man ser också att det har blivit fler och fler internationella projekt för att skapa nationella sökportaler. Och det kan ju öppna upp för internationella sådana.

Jan Hagerlid:

Vi strävar ju mot det i KB med *Samsök*, men vi försöker också få in så mycket som möjligt i Libris. Även e-publiceringen vill vi ju ha i Libris. Så när det gäller den svenska utgivningen så finns den i Libris och vi hoppas på att även digitala svenska publikationer ska levereras till KB som pliktleverans, så att vi ska få en fullständig täckning på det också.

Föreläsning om publiceringsverktyget EPrints av Aina Svensson, Sveriges Lantbruksuniversitet

[Utskrift av ljudupptagning. PPT: www.solidaritetshuset.nu/usrd/nnn960.ppt]

Jag heter Aina Svensson och kommer från Ultunabiblioteket. Jag hoppas att ni avbryter om det är något som är krångligt och så får ni gärna ställa frågor.

Det jag tänkte prata om här är först en bakgrund till, sedan en översikt över publiceringsverktyg och sedan lite mer specifikt om EPrints. Och hur vi har tillämpat EPrints vid SLU eller Epsilon som vi kallar vårt e-publiceringssystem.

Publiceringsverktyg har precis som antalet arkiv, öppna arkiv, som Jan pratade om, ökat otroligt mycket de senaste åren. Tekniken har följt med och det finns idag ett stort antal fria publiceringsverktyg. Länken på denna skärmdump (SE PPT på www.solidaritetshuset.nu) visar till den här sidan där man kan hämta en guide till *Institutional Repository Software* och den är från 2004 så den är ju inte helt ny men den tar upp ett stort antal fria programvaror och är man bara medveten om att uppdateringar kan ha skett så finns grunderna för jämförelserna här. Dessa kan även laddas ner och hämtas då i PDF-fil eller via HTML. Det de har gemensamt de här verktygen är det som Jan pratade om: alla stödjer OAI-PMH-protokollet, vilket innebär att använder man dem så blir det lättare att samsöka via OAIster eller andra söktjänster. Och att de är fria. Att man kan gå till hemsidan och ladda ner och hämta de här programvarorna. Och att det är fritt att göra egna anpassningar.

Ett verktyg som har kommit på lite senare tid, efter den här senaste versionen 2004, som jag bara vill nämna är Greenstone på Nya Zeeland. Två stycken av de här fyra programvarorna som Jan nämnde tidigare finns i Sverige och det är DSpace och EPrints.

Fråga: Vad menar du med att de finns i Sverige?

Aina Svensson: Ja, alltså att de används, förlåt jag var lite otydlig där, de används vid svenska universitet och högskolor. Precis.

Det här är en annan översikt som man också mycket väl kan använda sig av. Den är ett resultat av SVEP-projektet som Jan tidigare nämnde. Det där projektet, det svenska projektet som är avslutat nu, men som bland annat ledde fram till att göra denna översikt. Och även om man kan hämta den här från...

Jan Hagerlid: Det går att hämta den på Open Access-programmet, det är nästan det enklaste.

Aina Svensson: Ja, under KB då. Och den är skapad för svenska förhållanden, och då menar jag svenska universitet och högskolor igen. De som var vanligast förekommande vid den tiden, och är det nu också, är EPrints, DiVA och DSpace. Det här är som en tabell där man kan jämföra vilket operativsystem som används för respektive publiceringsverktyg och vilka funktioner som finns. Man ska vara medveten om att den har sidan inte heller har uppdaterats sedan 2004.

Och här är då DSpace och EPrints respektive hemsidor där man kan hämta de fria programvarorna. Jag återkommer lite till EPrints senare, men jag kan ju nämna att DSpace används bland annat av Malmö Högskola, Borås Högskola, Göteborgs Universitet och Halmstads Högskola. DSpace är också väldigt stora internationellt sett. Om man jämför med

EPrints så kan man säga att den skapades efteråt. EPrints var före, DSpace kom till två år efter. Och det var framför allt för att passa stora institutioner, multidisciplinära, mer komplexa institutioner som då inte EPrints klarade av och att man också fokuserade ganska mycket på långtidsbevarande.

EPrints då. Till att börja med skulle jag vilja påpeka att EPrints kan användas i lite olika betydelser. När jag nu pratar om *EPrints* så är det programvaran eller publiceringsverktyget jag menar, men e-prints kan också användas i betydelsen elektroniskt publicerad vetenskaplig publikation eller till exempel i begrepp som e-printsarkiv. De här likartade stavningarna för olika betydelser kan ibland vara lite förvirrande, för även om man pratar om en e-prints-publikation så behöver den inte vara publicerad med publiceringsverktyget EPrints. Det kan vara bra att hålla isär det här.

Lite kort om programvaran. Utvecklat vid universitetet i Southampton. Första versionen kom år 2000 och då var man i princip först. Det fanns nästan inga andra publiceringsverktyg. Så man har hållit på rätt länge. Senaste versionen kom alldeles nu i dagarna och det är EPrints version 3.0 i en Betaversion. Anledningen till att man körde igång var för att vem som helst skulle kunna sätta upp ett öppet arkiv. Och då var det egentligen främst tanken att forskare skulle kunna lägga ut sådana här e-prints, alltså artiklar, och lägga ut dem i arkivet elektroniskt. Sedan har då andra krav kommit till efter hand, men det var det som var tanken till en början. Och idag finns det närmare 200 arkiv runt omkring i världen, en väldigt bred bas. I Sverige är vi väl inte lika många, men vi är några stycken: Handelshögskolan i Göteborg, det är en del av Göteborgs Universitet som annars har DSpace, Lunds Universitet, och det är LU research som Jan visade tidigare, Högskolan i Kristianstad, SLU och SICS Swedish Institute of Computer Science från Kista.

EPrints webbsida då. På www.eprints.org hittar man i stort sett all information man behöver om programvaran. Här finns e-postlistor man kan prenumerera på för att hänga med i den tekniska utvecklingen eller för att ställa frågor till andra Eprints-användare. Det finns också en Wiki där man kan läsa mer och som fungerar som en slags FAQ. Det är också här man hämtar och laddar ner programvaran. Sedan de startade år 2000 har man under de senaste åren satsat på att bygga ut servicen och under 2005 skapades EPrints Community som erbjuder utökad service mot betalning. Det finns på så sätt olika supportnivåer. Det finns exempel på större institutioner i olika länder som har köpt tjänster från EPrints Community. Men grundtanken med EPrints är att programvaran är Open Source. Det vill säga programvaran är fri att använda som man vill men utöver det kan man mot en viss kostnad få hjälp med installation et cetera.

Ok, det här är det mest tekniska som jag kommer att säga i den här presentationen. Jag bara nämner det ifall det finns tekniskt intresserade här och eftersom det finns vissa tekniska skillnader mellan EPrints och Dspace. EPrints programvara har utvecklats för Unix/Linux operativsystem och programmeringsspråket är Perl. Dspace har utvecklats för Unix/MacOSX/Windows och programmeringsspråket Java. EPrints använder sig av databasprogramvaran MySQL.

Vad är då fördelarna med EPrints? Ja, i flera utvärderingar där man har jämfört DSpace, EPrints och *Fedora* så har man kunnat konstatera att EPrints är väldigt basalt och enkelt att sätta upp. Och tar man det rakt av som det ser ut så kan man få upp programvaran inom en till två veckor har det sagts i de här utvärderingarna. Det är också flexibelt att göra egna anpassningar. Och så finns det ju många användare. Man har ju funnits under lång tid så det finns ett starkt stöd och support, vilket gör att det känns tryggt och skönt förstås.

Ok, hur har vi då använt oss av EPrints vid SLU? Det kommer att bli en del skärmdumpar när jag visar hur inmatningen går till. EPrints vid SLU, vi kallar alltså vårt system för Epsilon, och det består av, i dagsläget, av två arkiv. Ett består av examensarbeten och ett separat för avhandlingar, doktorsavhandlingar och licentiatavhandlingar, och då har vi anpassat EPrints så att det ska passa de målgrupperna, de som matar in. Studenter är ju en väldigt homogen grupp så där har vi kunnat ändra språket på sidorna så att det ska bli lättare och mer tillämbart för studenten. Likadant då med avhandlingsbasen eftersom vi vet att det är doktorander som vänder sig hit och matar in sina publikationer. Sen har vi ett tredje arkiv på gång som vi hoppas kommer i skarp drift nu i december. Det är ett öppet arkiv för övriga publikationstyper. Det finns en stor efterfrågan, speciellt efter att lägga ut rapporter eftersom det finns så mycket sådana på SLU. Sen får vi hoppas att vi också får in artiklar då och kanske till och med får SLU:s ledning att rekommendera sina forskare att deponera en artikel i vårt öppna arkiv. Så det är nu något som vi har framför oss att jobba på.

Sen kommer det också en svensk översättning. EPrints finns redan i ett par översättningar: franska, tyska och spanska, och vi ska snart ha en svensk översättning klar anpassad till vårt system. Så att när det öppna arkivet kommer så är det en tvåspråkig version.

Från inmatning till publicering – hur går det då till? Hur gör man? Här är en schematisk bild över hur det går till [se www.solidaritetshuset.nu/usrd/nnn960.ppt]. De här gula fälten är alltså EPrints, det är själva systemet, fast det är olika delar av systemet. Den här första inmatningsdelen kallas för *User Area* i EPrints, sedan är det redigeringsdelen, det är där bibliotekarien går in och redigerar och det kallas för *Editorial Review*, och sen har vi då det publicerade arkivet, eller *Repository* som det kallas. Och här har vi studenten, författaren eller annan ansvarig som matar in sin publikation som en PDF-fil.

Om vi börjar med att titta på den här första delen, hur det går till med själva inmatningen. Vi tar ett exempel med en student som ska mata in sitt examensarbete. Hon gör det i tre delar här. Först måste hon registrera sig i systemet för att kunna logga in, men också för att vi på biblioteket ska kunna ha någon kontroll så att inte vem som helst kan logga in och ladda upp vad som helst. Sedan fyller hon i, i ett webbformulär, bibliografiska data om sitt arbete eller publikation. Och den tredje delen är när hon laddar upp PDF-filen direkt i systemet.

Och då ser det ut så här, för att ta det stegvis, när man registrerar sig får man upp ett sådant här formulär där man fyller i sin e-postadress, väljer ett användarnamn och ett lösenord. När man har gjort det så får man ett automatiskt mejl där man ska klicka på en länk för att bekräfta lösenordet. Detta är en vanlig rutin som förekommer när man ska registrera sig för olika tjänster på nätet. Sedan är man tillbaka i systemet och klickar på den här länken *logga in*. När man har gjort det så kommer man till sin egen *User Area*, home page, eller vad man ska säga. Det är alltså min egen, och ingen annan kan se den här, eller komma åt den här, och då klickar jag på den här länken *Open a new thesis form*. Jag kan säga att vi har instruktioner ute på nätet som de kan använda sig av och följa steg för steg. Då klickar man på den här länken och öppnar ett tomt webbformulär som man då fyller i med sina uppgifter: författare, titel och vad det kan vara. Och här kan jag säga att det här inte är EPrints standard, den kan man ju använda sig av, men vi behövde anpassa och lägga till vissa fält. Vi tog även bort vissa fält. Så vi har gjort en del ändringar här.

Ja, här är det bara några exempel på vad vi har för fält; vilken fakultet man tillhör, institution och så vidare. Längst ner i det här fältet så har vi också lagt till det här fältet: *Accept the agreement*, där man väljer ja eller nej. I fält-texten finns en länk till ett avtal som vi har formulerat, som ska vara ett avtal mellan oss och författaren. Så att vi får tillstånd att

publicera det här examensarbetet elektroniskt på nätet, och också att inte studenten när som helst kan, utan anledning, dra tillbaka sitt arbete. Men självklart så står det ju också där att studenten behåller sin rätt till att använda arbetet. Och det står också något om att vi har rätt att, eller att våra användare har rätt att ladda ner och skriva ut och så vidare. Så det där är ett egensnickrat avtal som vi använder oss av.

Vi har gått från steg ett till två, formuläret med de bibliografiska uppgifterna, uppgifterna om dokumentet. Och nu kommer det sista steget, att lägga till eller ladda upp PDF-filen. Då går man vidare här och väljer filformat. Det behöver ju inte vara i PDF, det finns ju andra filformat också. Men det är det vi har valt. Filen hämtas via den här *browse-knappen* och laddar upp, *Upload file*, så finns då den här filen länkad. Det sista man gör då är att kontrollera uppgifterna, så man får upp en sida här där man har möjlighet att backa tillbaka om det är så att något är fel. Eller så kan man, om man tycker att man är klar, klicka på den länken, *Deposit item now*. Då har man gjort hela den här proceduren som författare.

Och då hamnar mitt arbete i delen som kallas för Editorial review. Där kommer inte den här personen som matade in den att se den, när de går in i sin User area så finns inte arbetet kvar där utan har flyttats över hit. Det som händer nu är att det går ut ett mejl till respektive bibliotekarie i Alnarp, Umeå eller Ultuna beroende på vilken fakultet som studenten hade valt vid inmatningen. Och då loggar den bibliotekarien in i systemet och hämtar dokumentet. Här finns ett arbete som låg inne igår eller om det var i förrgår, med titeln *Ett objekt*, det fanns en undertitel också, som jag inte riktigt minns nu, något med arkitektur. För närvarande ligger det totalt två arbeten inne här och man kan se när det är inmatat på datumet här. Normalt ligger de inte kvar här så länge för de bibliotekarier som jobbar med det här är ganska snabba på att få undan. Jag hade tur i alla fall att det fanns två här när jag gick in för att skapa skärmdumpen. Men det ligger kvar ett från den nionde oktober och det beror på att det var något problematiskt med just det här examensarbetet. Det var något som saknades på titelsidan eller något sådant. Och då väntar man med att publicera det. Då har den studenten kontaktats och så väntar vi på att få in en bättre version. Det är inte så ofta det händer men det är därför det ser ut så.

Det man gör när man ska redigera ett arbete är att klicka på titeln för att få fram arbetet igen. Och när man då loggar in som bibliotekarie, då loggar man in som editor, då har man mera rättigheter och kommer åt att göra mer saker i EPrints. Till exempel får man då möjlighet att se de här knapparna här nere, man har möjlighet att *Bounce item back to submitter*, jag skulle kunna skicka det tillbaka till författaren om jag vill, med en kommentar om att det här var inte fullständigt eller så. Man kan ta bort det också. Men nu ska jag ju titta på det här för att redigera det, och då väljer jag den här knappen *Edit item*. Då får jag upp formuläret som den här författaren har fyllt i, Lea Sjölund, så kommer det några andra fält här, till exempel vilket program studenten har valt. Då kan man ju se om det har blivit fel ifyllt någonstans. Om titeln har fyllts i på felaktigt sätt eller något annat ser konstigt ut, så kan man rätta till det. Och sen överst här står det *Accept the agreement*, alltså det sista fältet som var författarens ansvar att fylla i, och nedanför det, när man är inloggad som editor, så har det tillkommit några fält. Och det är några fält som vi har valt ska ligga på bibliotekarien att fylla i därför att vi tror att det är svårt att fylla i för studenten. Om man vill, är det lätt att flytta ett fält så att det blir synligt alternativt osynligt i studentens inmatningsformulär. Det är inget tekniskt krångligt. Däremot är det mycket mer komplicerat att lägga till nya fält eller ta bort fält helt och hållet.

Sen fyller bibliotekarien i de här sista fälten, och om man tycker att allting ser klart ut, så går man tillbaka till den här sista sidan för att publicera. Och då väljer jag *Move item to repository*. Och det är ju det man gör, man flyttar dokumentet så här, skulle jag till exempel råka trycka fel och klicka på *Remove item* som har hänt mig någon gång när man hamnar i webbläsare som man inte är riktigt van vid och knapparna hoppar upp på lite olika sätt så blir man först alldeles skräckslagen och sedan "oj, nu har jag förstört det här och slängt bort det". Men då är det ju så att det finns en *Deleted buffert*, så man kan söka upp det i den här skräpkorgen kan man säga och hämta tillbaka det och lägga tillbaka det i Editorial buffert och sedan till arkivet.

Ett publicerat arbete hamnar då här i arkivet (repository). När det kommer ut i arkivet så finns möjligheten att via en prenumeration, en nyhetsprenumeration då, få ett mejl att nu har det kommit ett nytt arbete. Och det används både av institutionssekreterare för just examensarbeten och används för en del andra i andra sammanhang också. En del prenumeranter som tidigare fick tryckta avhandlingar, kan nu få ett sådant här mejl. Och ett publicerat examensarbete kan då se ut så här. Då finns det författar-, titeluppgifter, alldeles ovanför den här PDF-ikonen, här finns även länken till fulltext och abstract, alltså en html-sida med översiktlig information, om det ingår i en serie till exempel.

Det kan också se ut så här. Det här är avhandlingsbasen. Och ganska nyligen fick man tillstånd att ha en liten bild på framsidan av avhandlingarna på SLU och det tycker forskarna att det är väldigt trevligt att utnyttja sig av. Och då är det några som vid sidan av sin PDF-länk också laddar upp en bildfil, alltså den bild som de har på framsidan. Och det är ju ganska trevligt. Det här är då en flock renar! Det handlar om någon parasit hos renar.

Sökfunktionen. Man kan bläddra på samtliga fält som finns inlagda i EPrints. Och när jag säger bläddra då menar jag att klickar man på författare i det här fallet så får man en alfabetisk lista på författare. Och för examensarbeten så kan det ju vara ganska bra, att per institution se vilka examensarbeten som finns eller per program. Sen finns det då också ett sökformulär. Ett sökformulär som vi har lagt utanför själva EPrints-systemet, eller EPrints gränssnitt. Och för att tala lite negativt om EPrints så kan man säga att sökfunktionen inte är jättebra. Det går inte, till exempel, att trunkera. Den är inte så avancerad. Det går inte att trunkera, ni vet söka på tecken som gör att man söker på olika ändelser. Det går inte att göra kombinationssökningar och mer precisa sökningar. Så det är därför som vi har valt att lägga den här utanför och välja ut en del fält som vi tror är användbara.

Den här bilden är bara för att visa vilka fält som finns för författaren, alltså webbformuläret med de fält som författaren eller studenten fyller i, och vad man gör som bibliotekarie. Det ser ungefär ut så här. Så vi får ju ganska mycket information från den som matar in som man sedan redigerar. För avhandlingarna är det ännu färre fält som bibliotekarien fyller i, det är bara en institutionskod som vi lägger på plus ämnesord från Agrovoc, som är den thesaurus, ämnesordlista, som finns för SLU:s ämnesområden. Som vi använder.

Men även om EPrints själva inte har ett så bra, ett så väl fungerande sökgränssnitt, så hittar man oss ändå från en rad olika håll. Om vi tittar här då, när arbetet finns i EPrints, i vår databas, så kommer vi också att ta över det i Libris, alltså den nationella bibliotekskatalogen. Vi har ju tidigare alltid katalogiserat examensarbeten i Libris, och nu kan vi använda de bibliografiska posterna i EPrints och importera till Libris och vår lokala bibliotekskatalog Lukas. Arbetena kommer alltså att vara sökbara i Libris/Lukas med länkar till fulltexter i Epsilon. De kommer också att presenteras på SLU:s institutioners webbsidor med länkar till Epsilon. Är det då ett examensarbete så kommer de också att skördas av Uppsök, Uppsök

ligger ju under Libris, men den ligger där som en separat bas. Man hittar den i fliken uppe i den övre menyn när man är inne och söker i Libris. Och sist men inte minst så hittar söktjänster på Internet, Google Scholar och även OAlster, men framför allt är det nog Google som hittar oss, och det gör de. Eller vem det nu är som hittar till oss.

Vi tittar på lite statistik här. De tio mest efterfrågade examensarbetena det senaste året då, september 2005 till september 2006, så ligger det ganska högt, det är väldigt högt, och det har faktiskt varit så under senare år, och det är jättestor skillnad mot förut då man spred några få kopior av sitt examensarbete. Men jag tror att dels har SLU ämnesområden som intresserar en stor målgrupp, och en mycket varierad målgrupp. Det kan vara, som Jan nämnde tidigare, att utbildade veterinärer behöver få tillgång till sådant här material, men det kan också vara gymnasieelever, eller det kan vara privatpersoner med en sjuk hund, ja, och många av de här arbetena skrivs också på engelska så det är inte bara i Sverige utan även internationellt som efterfrågan finns. Så det är kul.

För avhandlingar visar vi ännu större siffror. Den första avhandlingen, den som leder, som ligger först på listan, ligger på 55.000 nedladdningar. Och det är en väldigt efterfrågad avhandling. Den handlar om svensk stadsplanering från 1500-1600-talet. Avhandlingen består av en bilaga med enbart kartor över främst svenska städer från denna period och en omfattande textdel.

Avbrott. Byte av band.

Lite kort slutligen om våra egna erfarenheter av EPrints. Det är enkelt att använda och vi tycker att vi har kunnat göra våra egna anpassningar. Det har inte varit några problem egentligen. Men så finns det vissa funktioner som vi har saknat. Statistikverktyg, stöd för import och export mellan olika databaser, och det har vi utlovats. Det mesta som har efterfrågats har utlovats till kommande version.

Fråga: Men inte förbättring av sökfunktionen?

Aina Svensson: Nej vi satt och tittade på den där listan om vad som hade kommit, jag tror att det finns med, inte i den här första betaversionen som har släppts, men som ett senare steg. För det har påpekats i flera olika sammanhang.

Då tackar jag för mig om det inte finns några frågor.

Fråga: De här spridningsvägarna? Om du backar några bilder. Andra som är intresserade av eller använder sig av er Epsilon-bas. Det är bara akademiker, eller?

Aina Svensson: Google är ju den allra största om vi tittar här på var de som söker kommer ifrån, det är väldigt mycket mer därifrån än de som kommer via bibliotekskataloger. Och det tror jag att man har sett vid Uppsala universitet också, eller? Att det är väldigt få som söker via bibliotekskataloger. De flesta kommer via söktjänster som Google och då vet man inte, de behöver inte vara akademiker. Men jag kan tänka mig att, vi har mycket forskning som berör tredje världen, vi har ett stort utbyte med gästforskare sådär, så det kommer nog folk från hela världen. Nu kommer jag inte ihåg riktigt hur fördelningen mellan enskilda länder ser ut, men Sverige utgör förstås en stor del och internationellt kommer många sökningar från USA och Kanada.

Fråga: Men det jag tänker på är att vår organisation och det kanske gäller många av föreningarna här, producerar ju inte i första hand akademiska texter och då undrar jag om det här är någonting som främst används inom den akademiska världen?

Aina Svensson: Du menar publiceringsverktyget?

Frågeställaren: Ja, precis. Databaserna, om de är någonting som, hur ska jag säga, finner sin bästa tillämpning inom akademien?

Aina Svensson: Det är väl där det har börjat. Eftersom hela den här utvecklingen har startats inom den akademiska världen, och det är också så för oss på biblioteken, först hade man de tryckta samlingarna att försöka få ordning i och få överblick över, och sen så har ju allt det elektroniska kommit till, och det måste man också försöka att få någon slags kontroll över. Och då har ju det här med e-publicering tillkommit.

Fråga: Finns det några exempel på arkiv som inte är akademiska, kan våra föreningar söka samarbete med den här typen av verktyg?

Peter Hansson: Om man till exempel går till Greenstone som programvara så finns ju, det är FN som har ... utvecklingsländerna... där finns det ju andra slags institutioner... jag tror att det fanns en ganska bra lista på deras hemsida... slå Greenstone.org eller något.

Aina Svensson: Vi är själva lite hemmablinda också i högskolevärlden...

Jan Hagerlid: Där var det en NGO... Det verkar väl vara en del annat, eller hur? Den här programvaran kanske är mera lämpad för er.

Aina Svensson: Ja, jag vet inte så mycket, jag vet att vår tekniker har en artikel om Greenstone och EPrints, där Greenstone får bra betyg.

Stefan Andersson: Jag skulle vilja säga att EPrints och olika system, att de mest är till för att organisera sina samlingar. När det gäller det där med att söka då är det vanliga att man hittar informationen via Google eller så, de här arkiven och systemen är ju mer till för att hjälpa en institution, organisation eller förening att organisera sina egna publikationer.

Fråga: Innebär det att man måste komma upp i en ganska stor volym innan det blir en fördel att använda sig av de här verktygen? Är det en skillnad i hur många sökträffar man får om man lägger något på en hemsida än om man har dem i ett öppet arkiv?

Stefan Andersson: Nej det behöver du inte. När det gäller antal dokument så vet jag inte var gränsen skulle gå, men jag tror att den egentligen är ganska låg. Några hundra dokument alltså.

Jan Hagerlid: Jag tror att man kan säga att det som de här systemen producerar, jag menar att de är tillgängliga enligt det här OAI-protokollet, men för att komma in på Google behöver

du inte OAI-protokollet, för Google använder inte det, de indexerar ju sidorna. Jag vet inte exakt vad som krävs för att bli inplockad i Google.

Aina Svensson: Ja, visst, det beror ju på vad man har för hemsida. Vi är ju då indexerade av Googles sökrobotar, och då borde vi ju ligga högre upp än andra hemsidor. Det skulle jag kunna tänka mig, faktiskt.

Peter Hansson: Det pågår också försök att indexera *the invisible web*. Alltså det man inte kan nå direkt via html-sidor. För att nå det använder man bland annat OAI-protokollet för att nå den informationen.

Arwid Lund: Ja, Google når ju inte åt att indexera många databaser.

Jan Hagerlid: För att nå ut på Google är det väl bra om varje publikation är en egen webbsida?

Stefan Andersson: Ja, det är ju det bästa.

Jan Hagerlid: Men har man allt i en enda lång svit så fungerar det inte så bra.

Peter Hansson: Man kan ju länka till webbsidan.

Stefan Andersson: Man kan ju använda den som någon form av index över alla dokument som man har. Så att Google kan hitta dem. Till exempel författare.

Aina Svensson: Sedan kan man ju göra det olika mycket krångligt. Vi lägger ju ganska mycket tid på att redigera och det behöver man ju kanske inte göra heller. Det finns olika exempel där också. Det går ju att använda systemet, menar jag, på en enklare nivå.

Jan Hagerlid: Fördelen här är väl att du kan dela upp ansvaret mellan författaren och utgivaren. Författaren kan bidra med så mycket som möjligt och uppgiftsgranskaren, den som är utgivare så att säga, kan lägga till ytterligare fält. Det kan gälla för universitet. Man behöver ha en kontrollfunktion, kvalitetsfunktion, för att det ska fungera bra. Och då är ju sådana här system riggade för att klara det här arbetsflödet.

Aina Svensson: Om man tittar på det från vårt håll så har vi redan slutat att ta emot tryckta examensarbeten från en fakultet, och det kommer väl på sikt att bli så att rapporter kommer vi inte att ta emot. Det blir ett elektroniskt bibliotek där vi har kontroll över SLU:s publicering.

Fråga: En annan sak som är bra är ju det här med abstract. Med sammanfattningar, att man kan ha sådana.

Föreläsning om beständiga identifikatorer och uppslagstjänster av Peter Hansson, Chalmers.

[Utskrift av ljudupptagning. PPT: www.solidaritetshuset.nu/usrd/nse137.ppt]

Jag heter Peter Hansson och jag jobbar på Chalmers bibliotek för tillfället, jag jobbar mycket som systemutvecklare. Det här kommer att bli lite tekniskt. Hoppas att det går bra. Jag ska försöka att lyfta upp det lite och göra det lite lättare. Rubriken för det här föredraget är beständiga identifikatorer med tillhörande uppslagstjänster. Och så är det ett litet tillägg, på slutet kommer jag att prata lite om OAI-PMH som vi nyss har diskuterat.

Man kan fråga sig; Varför ska man prata om beständiga identifikatorer? Eller som det kallas i vissa sammanhang: *persistenta identifikatorer*. Det beror helt enkelt på att man vill kunna nå information utan att veta var informationen finns för tillfället. Om man tänker sig att man har en referenslista med vissa referenser, så är det ganska intressant att skriva dit, om det refererar till böcker, att skriva dit ett ISBN som gör att man kan gå vidare till olika katalogtjänster och slå upp och hämta boken. Men när det handlar om elektroniska resurser så blir det ju lite annorlunda och i början av 1990-talet fick man höra talas om URLer, och visst, de kanske fungerade, en dag, en vecka, en månad eller ett år, men sen så fann man inte materialet eftersom URLen pekade mot något som inte längre fanns.

Så en beständig identifikator är ingenting konstigt. Det finns omkring oss hela tiden. I olika sammanhang som exempelvis personnummer. Sedan är det också en ny slags identifikator, eller den tillkom i slutet av 1990-talet, som kallas *URN:NBN*. Som jag kommer att prata lite mera om. På punkt två här så står det att symbolen ska vara konstant. Det innebär att den inte får förändras under tiden och att den helst ska finnas för evigt för att den ska vara beständig. Men frågan är, det kommer de nog inte att göra, men de ska åtminstone finnas under lång tid.

Här har vi ett exempel på beständiga identifikatorer som används. Och, i DSpace, som Aina berättade om tidigare, som är ett sådant Open Source Repository System, använder man något som kallas för *Handle* och det har helt enkelt sådana här nummerserier som anger varje resurs. Jag kommer inte att gå in på det närmare, detaljerat. Sen finns det också ett kommersiellt system som heter *DOI*. Digital Object Identifier, som många förlag använder sig av för att tilldela de här identifikatorerna till olika artiklar exempelvis. Och båda de här, både Handle och DOI bygger på samma system, *the handle system*. URN:NBN kan se ut så här, som hos Växjö Universitet. Och sedan finns det också, jag tror det var OCLC som tog fram en hybrid.

Jan Hagerlid: Du får nog förklara vad OCLC är?

Peter Hansson: OCLC är en organisation som utvecklar, ligger i framkant av utvecklingen...

Jan Hagerlid: Det är ett bibliotekskooperativ i någon mening, ett kooperativt företag kan man säga.

Peter Hansson: Och de konstruerade något som de kallar för *Persistent URL* som i praktiken är en adress då, hela det här är en adress, men det som är unikt är då ändelserna. Som i sig då kan ses som en identifikator.

Men varför ska man inte använda URL? En URL är föränderlig, så därför är det ganska olyckligt att använda en URL som en beständig identifikator. Några exempel: ett företag

köper ett nytt managementsystem, och det tvingar företaget att flytta alla datorerna till en annan server med nya URLer. Eller ett företag köps upp av ett annat företag och måste ändra sin domän. Eller företaget går i konkurs, URLen försvinner, men finns till salu för andra att köpa. Och de lär ju inte använda den för de gamla dokumenten.

Om man kollar in webbens fader, *Tim Berners-Lee* så skriver han att en URL kan ändras, kan peka mot ett annat objekt vid senare tillfälle. Och sedan läste jag också på en mejlinglista för bara en vecka sedan att det är idioti att använda URLer som identifikatorer eftersom det finns bättre identifikatorer, till exempel URN:NBN.

URN, det som står före kolontecknet, är persistent, en adressoberoende identifikator som kan leda till olika slags resurser. Och man ser här också hur en URN är uppbyggd. Att den för samman en prefix, URN, med kolon, följt av något som kallas *Namespace Identifier*, följt av kolon, och till slut följt av något som kallas *Namespace Specific String* [se www.solidaritetshuset.nu/usrd/nse137.ppt].

Och det här ordet, *Namespace Identifier*, kan vara NBN, det kan också vara andra saker. Men det vi ska prata om här är NBN, och det står för *National Bibliographic Number*. Och det är alltså en namnrymd under URN, som kan användas som symbol för beständiga identifikatorer.

För att vi ska bli lite mer konkreta. Här så ser man hur de olika URN:NBN ser ut [se www.solidaritetshuset.nu/usrd/nse137.ppt]. Det finns tre olika varianter. Om ni kollar första varianten, så ser ni att man har "URN:" och sedan har man ordet, alltså NBN, och sedan kolon, följt av en landskod, som definieras i en annan standard. Och landskoden för Sverige är *se*, Norge har till exempel *no*, följt av bindestreck och sedan en så kallad *NBN-sträng*. Och om man kollar exemplet så ser man att det består av både landskoden och den här NBN-strängen, den här NBN-strängen kan man definiera som vad som helst bara man använder sig av de giltiga, håller sig inom de tecken som får användas. Men det är alltså Kungliga Biblioteket som styr och ställer när det gäller URN:NBN-tilldelningen i Sverige.

Om vi kollar på nummer två, så ser man att det har smugit sig in något som kallas för *undernamnrymd*. En undernamnrymd är något som man måste registrera hos Kungliga Biblioteket, oftast är det en organisation eller företag som gör detta. Sedan får denna organisation tilldela egna NBN-strängar under den undernamnrymden, man delegerar den uppgiften åt företaget eller organisationen – att skapa egna URN:NBN.

Här ser vi då att SLU har registrerat *slu:epsilon* som undernamnrymd. Och jag ser också här att Växjö som ingår i DiVA-projektet har *vxu:diva* och så vidare. Alla är uppbyggda enligt samma sätt. Det här har organisationen bestämt själv. Jag tror att det är så i Tyskland att sista siffran är en kontrollsiffra av samma sort som finns i ett personnummer, medan i den svenska här, är det ingen kontrollsiffra med.

Och den tredje varianten är om man vill registrera en annan slags profil än den här landskoden då. Jag känner inte till något sådant exempel, men det står i specen i alla fall.

Jan Hagerlid: Men om man väljer tvåan, som DiVA gör till exempel, då använder man inte den här första modellen, man kan inte använda båda samtidigt så att säga? På KB använder man inte undernamnrymden, men varför gör man inte det? Men det kanske inte du vet?

Peter Hansson: Jo, vi kommer till det senare, men, låt oss säga att man bara har en enda publikation som man vill ha en URN:NBN till, då kan KB se till att den får en enstaka URN:NBN och då kommer den automatiskt att hamna hos...

Så ser det ut. Det här är då hur identifikatorn är uppbyggd, men en identifikator är ju egentligen inget värd om man inte kan använda sig av den. Och det gör man med hjälp av en *Resolution Service*, en uppslagstjänst som det också kallas på svenska. Tidigare så har man ju använt sig av bibliotekskataloger när man har letat efter fysiska resurser. När man är i den elektroniska världen så kan man faktiskt nå dokumentet med detsamma. Traditionellt då användaren frågar en uppslagstjänst om var en resurs eller en viss identifikator finns, får han tillbaka olika slags adresser då, eller placeringar. Sen får man helt enkelt gå och hämta dokumenten på hyllan. Det som händer i den elektroniska världen är att allting kan ske på en gång. Användaren skickar in en identifikator i en webbaserad uppslagstjänst, uppslagstjänsten returnerar dokumentet på en gång.

När det gäller långtidsbevarande, och sådana framtidsfrågor, så är det ju så att KB har ett jättestort arkiv, och så småningom så kommer det också, förmodligen, att bli en e-pliktlag skriven i Sverige. Då kan ju KBs adress, om den normala adressen försvinner, vara ett alternativ nummer två. Alltså att det finns ett dokument i KBs digitala arkiv. Så man kan säga att en sådan här *Resolution Service* egentligen inte löser problemet med att URLerna försvinner – då måste det ändå finnas en kopia någon annanstans. En kopia hos en organisation som har det som uppdrag att bevara elektroniska resurser.

Det betyder i alla fall att användaren inte behöver tänka på vilken adress som gäller för en resurs, det kan den här tjänsten göra i stället. Så, här har vi ett konkret exempel: KBs uppslagstjänst, den ser ut så här. Jag klistrar in här, beroende på vilken form av identifikator som avses, i det här fallet är det en URN, kan man välja... och så trycker jag på slå upp. Då kommer jag till denna sida, det här URN:NBN som man matar in här är alltså kopplat till denna webbsida som är en elektronisk resurs. Och...

Fråga: Det känns inte som det är en resurs...

Peter Hansson: Nej, jag förstår. Det står i KBs rekommendationer att om det är ett någorlunda konstant och oförändrat material så får man ge en identifikator. En viss förändring av data får finnas i en sådan här resurs.

Fråga: Men då måste det vara startsidor, eller?

Peter Hansson: Ja, du får höra med KB, vad de tycker där.

Fråga: Om jag skulle vilja ha ett antal dokument där jag jobbar, då kan jag höra av mig till KB och få någon underrymd och samtidigt kan jag då publicera de här dokumenten som befinner sig i en databas? Är det så det fungerar?

Peter Hansson: Du har ju det elektroniska dokumentet själv. Och sedan vill du ha ett URN:NBN till det. Då går du in på KB:s webbsida.

Fråga: Det måste alltså gå genom KB?

Peter Hansson: Ja, för det är KB som har utgivningsrätten i Sverige. Som ska se till att tillhandahålla URN:NBN. De förväntas ha en organisation som kan bevara det.

Stefan Andersson: URN:NBN är detsamma som ett ISBN till en bok.

Peter Hansson: Och det är just ISBN-centralen som har hand om URN:NBN. Så här ser ansökningsformuläret för URN:NBN ut, och jag tror att vi kan göra så här, här ser ni, här kan man fylla i undernamnrymd, som jag tänkte vi skulle göra alldeles strax, och här kan man hämta enstaka URN:NBN. Och så fyller man i det här formuläret och så får man ett enstaka URN:NBN.

Fråga: Vad händer i nästa fas, om man till exempel blir uppköpt?

Peter Hansson: Då får du anmäla den nya adressen till KB.

Fråga: Men om man bara flyttar dokumenten innan man blir uppköpta så håller väl Google koll på var de hamnar? Men med det här så måste man själv, om man jämför de två systemen så, vad tillför det här systemet?

Peter Hansson: Det här måste ju administreras med URN:NBN och ses till att resurserna är kopplade till det här URN:NBNet, att de alltid har rätt adress i databasen. Alternativt att de har en undernamnrymd för då kan man ju administrera det lokalt.

Fråga: Men det är det ju inte så många som har, det är ju universitet och så?

Peter Hansson: Ja, och ni skulle ju också kunna ha det. Världsbiblioteket skulle ju kunna ha en sådan undernamnrymd, som ni sedan får använda er av.

Stefan Andersson: Säg att ni har registrerat en publikation här. Men, som det är i Tyskland till exempel, så bygger det där på att även nationalbiblioteket har dokumentet, inte så i Sverige än, men jag tror att det kommer att bli så, och om det då skulle vara så att ett dokument försvinner så skulle det alltså finnas ett exemplar på Kungliga Biblioteket istället.

Fråga: Ja, då blir det ju beständigt också.

Peter Hansson: Den här uppslagstjänsten som finns här på KB, vi kan ju visa den, det fungerar så att den kontrollerar länken som vidarebefordras till höstning, innan den blivit vidarebefordrad till den, så om det skulle visa sig att det fanns tre stycken olika länkar att välja på så börjar den här tjänsten att titta på länk nummer ett. Fungerar den, ja, då använder man den. Men om det skulle vara så att den inte fungerar så tittar den istället på länk nummer två och kollar om den fungerar. Så den här backup-rutinen den finns redan i tjänsten. Det enda som saknas är själva e-plikten. Att man faktiskt måste deponera resurser.

Fråga: Ska den där idenfikatoren också finnas i klartext på en sida i dokumentet?

Peter Hansson: Alltså om du kollar på avhandlingar i Uppsala, så har man ju URN:NBN med på tryckortsidan, där det står copyright och vilken stad och år – är det väl? – när avhandlingen är tryckt.

Fråga: Det betyder ju att då blir det ett sökord också?

Stefan Andersson: Ja, just det, jag skulle rekommendera att man skriver ut URN:NBN någonstans på sidan.

Fråga: Vad händer om samma dokument har olika URL?

Fråga: Det måste väl vara samma sak som om en bok finns på flera olika bibliotek?

Fråga: Om man vill tala om för folk att vi har den här resursen på de här olika ställena?

Peter Hansson: Det finns stöd för det. Nu jobbar jag inte på KB så jag kan inte säga hur, men det jag vet från när vi implementerade den här resolution servicen är att det finns stöd för att lägga in flera olika länkar till varje resurs. Men jag vet inte om KB erbjuder den tjänsten. Men jag tror att i det här registreringsformatet så finns det bara utrymme för en URL.

Fråga: Skulle KB ta emot digitala resurser om man skickade in dem till dem idag?

Peter Hansson: Som det ser ut idag får man teckna separata avtal med KB när det gäller e-plikt. Som till exempel Uppsala har gjort för DiVAs del. Så allt som läggs i DiVA-systemet skickas över till KB.

Jan Hagerlid: KB har även avtal med myndigheter. Man kan säga att KB bedriver försöksverksamhet kring detta med att utveckla leveranser, så man har avtal med Uppsala och några myndigheter och så. Jag vet ju inte att man är intresserad av att ha fler försökspersoner just nu. Det kan jag inte svara på. Vad jag vet är att vi har blivit tillfrågad om att dra igång ordentligt med att lägga fram en e-plikttag, man har precis sagt att man är positiv till en, men det måste trots allt skrivas en lagtext och den måste prövas och stämma mot alla andra lagar. Så det är ett visst arbete med det. Dom säger att, det de har sagt är det ska läggas fram ett förslag till e-plikttag 2007, de har avsatt personal för att jobba med det och så där. Men är man intresserad av att pröva detta, så kan man alltid fråga de som är ansvariga för de digitala leveranserna. Rätt person att prata med hittar man på vår hemsida.

Fråga från publiken om vilken den praktiska nyttan av URN:NBN är idag?

Peter Hansson: Det kanske är enklare att ta ett exempel, om du har ett gäng identifikatorer istället för ett gäng URLer, när det gäller användningen: Om du har identifikatorerna så vet du att du kan lita på den... men har du ett gäng URLer så kan du aldrig vara riktigt säker på att den där teknikern inte flyttar dem, resurserna, på morgonen. Det är den praktiska nyttan som du har av identifikatorerna.

Fråga: Men samtidigt om den där dumma teknikern flyttar dokumentet utan att skicka flyttkort till KB, så är det ju ingen nytta?

Peter Hansson: Ja, i dagsläget ja, men i framtiden kanske man kan falla tillbaka på pliktexemplaret på KB.

Men jag ska inte fastna så mycket på det här. Som ni ser så kan KB:s uppslagstjänst hitta information om det material som uppslagstjänsten har tillgång till, och kan i det här fallet använda sig av identifikatorer och handle för att hitta resurserna. Och då använder den här tjänsten andra tjänster för att hitta dem.

Avbrott. Byte av band.

Peter Hansson: För att uppslagstjänsten ska kunna hitta länken till en resurs så måste det finnas en mappningsfil, med andra ord en fil som innehåller aktuella mappningar mellan URN:NBN och URL, någonstans. Och då har vi de här två olika alternativen. I första fallet har KB hand om matchningsfilen och URN:NBN kopplas mot aktuella adresser, och i andra fallet så har man en egen undernamnrymd och administrerar matchningsfilen lokalt. Och har man ett bra publiceringssystem så kan man låta publiceringssystemet skapa den här matchningsfilen automatiskt och sedan publicera matchningsfilen utåt så att KB:s tjänst kan nå denna via en URL. Och sedan kommer den här tjänsten, Resolution Service-tjänsten, på KB att automatiskt läsa av den här filen med jämna mellanrum för att uppdatera sin tjänst. Så om man byter URLer på sina publikationer ändras mappningsfilen och uppslagstjänsten på KB tar del av den informationen.

Om man inte har ett publiceringssystem som kan generera en sådan mappningsfil automatiskt, kan man faktiskt skapa filen i en vanlig texteditor. Så länge man håller sig till filformatet så går det alldeles utmärkt. Sen lägger man filen på sin webbserver så att den blir nåbar via en URL och så meddelar man KB URLen.

Nu hade jag tänkt... vi måste hinna prata lite om OAI-protokollet också. [*anm. Peter Hansson fortsätter att prata om URN:NBN resolution service*] Jag tänkte visa hur en källfil ser ut. Det är XML som används, det är ett XML-format som man håller sig till. Och det är ganska lätt att förstå. XML är väldigt strukturerat, och i det här formatet har man också valt ganska så beskrivande namn på taggarna. Här kan man se att varje sådan här *record tag* är en resurs. Och resursen har det här namnet. Och här ser man också vilka destinationer som finns. Här ser man att den här resursen har en adress i dagsläget, här har man bara definierat en adress. Om vi tittar på nästa exempel i denna exempelsamling så ser vi att URN:NBN-identifikatorn finns häruppe och här ser vi att det finns en... två... och tre... olika destinationer. Den första är aktiverad, den andra är inaktiverad och den tredje är inaktiverad. Det innebär att man har talat om för resolvern på KB att ni får endast använda den första, de andra två får man inte använda. Och om vi tittar lite mer noga här så ser man på nummer två att den här sidan aldrig har blivit nedsläckt, men man ändå har valt att inaktivera den eftersom man antagligen vill fasa ut den adressen. Medan alternativ tre, om man tittar lite närmare, slutade användas den 31 september 2002. Så om jag skulle klistra in den där adressen i en webbläsare så skulle det stå Not Found.

Det kan vara så i framtiden att man har en URL till en resurs men att man inte vet vilken URN:NBN resursen har. Men då kan man söka på de URLerna i en historisk databas. Men den tjänsten finns inte nu.

Uppslagstjänsten kontrollerar också så att man verkligen använder rätt undernamnrymd. SLU är förhindrat att använda Uppsala Universitetsbiblioteks undernamnrymd. Det är inbyggt i tjänsten.

Och nu hade jag tänkt att registrera en undernamnrymd för Chalmers faktiskt. För vi har inte registrerat någon sådan. Men istället kommer jag att prata lite mer om det här protokollet.

Open Archives Initiative, OAI, skapades 1999. Och det första protokollet kom 2001. *OAI-PMH 2.0* är den version som används idag och som man rekommenderas att använda, och det finns ingen kompatibilitet bakåt till tidigare versioner. OAI som organisation vill utveckla interoperabilitetsstandarder... ett bra alfapetord faktiskt... Och resultat hittills är alltså protokollet OAI-PMH. Protokollet gäller beskrivningen, metadatan, inte själva digitala objektet i sig utan bara beskrivningen av objektet.

Däremot något som är på gång, som har startats upp som ett nytt projekt, är *OAI-ORE*, där man helt enkelt ska fokusera, istället, på hur man kan sprida de digitala objekten. Ett nytt protokoll som kommer inom några år. Och tanken är då att de här två olika protokollen ska kunna samexistera.

Interoperabilitet, vad är det, jo det är någon form av samordningsförmåga. Om man ska uttrycka det på svenska. Men det finns också lite olika nivåer av interoperabilitet. Och det man fokuserar på i OAI-organisationen är att skapa ett protokoll då. Tyvärr har också OAI bidragit till att dålig metadata har spridits över världen genom att de rekommenderat de öppna arkiven att skriva sin metadata i formatet *OAI_dc*, dc står för Dublin Core. Och det är ett dåligt format. Och det har hämmat utvecklingen av nya häftiga informationstjänster. Men man förstod det inte till en början.

Åter till protokollet. OAI-PMH är ett väldigt enkelt protokoll som ligger ovanpå ett annat protokoll, som ni alla känner till, HTTP. Alltså det protokoll som används när ni surfar på nätet i en vanlig webbläsare. Alla svar man får levereras i XML, och XML är också en sådan här standard. Och som sagt protokollet har i sig flera olika metadataformat, där man alltså rekommenderar att åtminstone *simple Dublin Core* används. Och dessvärre är det många depositories utomlands som bara gör det.

Och observera att detta inte är något sökprotokoll, utan att det är ett protokoll för att hämta metadata. Så det finns ingen sökfunktion över huvud taget. Så söktjänsten, som Stefan kommer att prata mera om sedan, får man bygga själv.

I specifikationen för det här protokollet nämns *Data Provider*, en passiv innehållsleverantör som bara har informationen, och därefter *Service Provider*, som använder sig av Data Providers innehåll. Alltså, Service Provider hämtar innehåll från Data Providers innehåll. Det system Data Provider använder är OAI-PMH. Service Provider som bygger tjänsterna använder en Harvester. Ni ska snart få se hur det ser ut... det här molnet ska föreställa "nätet"... här ser vi aktörerna som gula, leende gubbar, och de gröna boxarna är system då [se www.solidaritetshuset.nu/usrd/nse137.ppt]. Och observera att i definitionen av OAI-PMH står att repositories måste vara nätanslutet, de måste vara anslutet till Internet. Och det är för att OAI-PMH är ett protokoll – gemensamma regler för att föra ut data. Så det som händer är att Harvester gör en request/förfrågan till Repository och Repository svarar.

Här i Repository finns metadata till dokument, när Harvester hämtar metadata, i det här fallet, över OAI-PMH, så läggs metadatan i en egen databas och där kan ju den här Service Providern berika metadatan så att den blir bättre.

Och sen så måste man också veta vad ett *Sets*, en term som används inom OAI-PMH, är. Och det är helt enkelt, det översätts *mängder*, grupperingar av objekt, metadata-objekt, som formerar olika slags nivåer: avhandlingar, avhandlingar från Chalmers, examensarbeten,

inom datavetenskap. Man kan ordna efter ämnessets, men man kan också ha det organisatoriskt som mängder.

Här är sex stycken frågor som definieras av protokollet. De här tänker jag gå igenom lite snabbt här. På vänster sida ser vi tjänsteleverantören, jag tror jag har bytt plats på dem här, anledningen till att gjort det är att vi läser från vänster till höger. Då ska vi se vad som händer när man gör en *Identify*. Och *Identify* kan sägas vara en fråga som ställs till Repository: Identifiera dig!? Alltså Harvestern frågar Vem är du? Och Repositoryet, om det är OAI-PMH-kompatibelt så svarar det. Och det måste svara på ett speciellt sätt. Den måste först svara "Jag är den här ...", "Jag använder den här OAI-informationen?", kanske är Repositoryet bara för avhandlingar eller bara för handböcker. Då kan man skriva det.

Och bara för att man ska få se hur det kan se ut då: Det här är själva svaret. Då är det inte så konstigt egentligen, utan i XML-koden så finns det olika slags element definierade, och om man bara lär sig vad de elementen står för, så kan man läsa av... Här står till exempel att det här repositoryet har kontaktadress på Chalmers.

Och sedan finns det som sagt fem andra anrop som man kan göra. Man kan fråga Repositoryet vilka metadataformat som finns representerade, och Repositoryet svarar med följande, i det här fallet finns bara stöd för OAI_dc på Chalmers.

Jan Hagerlid: Än så länge.

Peter Hansson: Än så länge. När man frågar vilka Sets som finns det definierade i det här Repositoryet, så ser man här vilka Sets som finns. Och vi på Chalmers har valt att använda oss av SVEPs ämnesord, som man kan se som någon form av defacto-standard för Sverige.

Och så fortsätter vi: *List Records*, ge mig de här posterna som har publicerats mellan första januari 2006 och första juni 2006, och så får man alla de posterna. Här ser man då varje Record. Först ser man lite från OAI-protokollet, sedan kommer här själva posten i OAI_dc-format. Och så vidare. Och här gör jag ingen sökning, utan jag gör ett urval. Så det är inte så att OAI-PMH är ett sökprotokoll.

Vi fortsätter. Vilka identifikatorer finns i systemet? Kan man fråga också. Och då kommer man att få en lista över identifikatorer och så vidare. Och här finns det, vad ska man säga, ännu en identifikator. Jag tror att det här kommer att försvinna, jag tror man kommer att sätta en URN:NBN där istället. För det finns ingen mening med att ha en ny identifikator i OAI-protokollet.

Så det är ett ganska simpelt protokoll.

Föreläsning om sökportaler byggda på fria resursers metadata av Stefan Andersson (Uppsala Universitetsbibliotek)

[PPT: www.solidaritetshuset.nu/usrd/vbl555.ppt]

Tyvärr var kvaliteten på ljudupptagningen för dålig för att föredraget skulle kunna transkriberas. Följande anteckningar kommer från mina anteckningar under föredraget. Texten är granskad i efterhand av Stefan Andersson så att inga direkta felaktigheter smugit sig in.

Jag arbetar på enheten för digital publicering vid Uppsala universitetsbibliotek. Vår programvara DiVA gör samma sak som EPrints men är inget Open Source-program. Vi är flera universitet som samarbetar kring DiVA.

Det jag tänkte prata med er om är ett annat projekt som heter *Testsök*, som är ett försök att testa SVEPs formatrekommendationer, som bygger på protokollet OAI-PMH, som alltså är bra på att överföra information men inte på att beskriva information.

Det finns en massa olika publiceringssystem som används: DSpace, EPrints, DiVA. De här systemen stödjer OAI-PMH. Däremot är det viktigt hur metadatan är strukturerad, det finns ingen standard, OAI_dc (dc står för Dublin Core) är dåligt strukturerat. Därför jobbade vi för att publiceringssystemen skulle få en modell för ett gemensamt format. Om metadata ser annorlunda ut i olika system kan man inte göra samsökningssystem där man söker i flera olika källor samtidigt. *Testsök* skapades för att undersöka hur väl olika repositories eller arkiv motsvarade denna standard. Det omfattade alla typer av vetenskapliga publikationer vid 15 stycken universitet.

Testsök har tre komponenter: 1) Harvesting (med OAI-PMH), 2) Indexering och sökning, för de som vill bli serviceprovider, och 3) webbgränssnitt, för att göra tjänsten tillgänglig för allmänheten.

Systemet fungerar som så att man först harvostar metadata från olika öppna arkiv till sin egen databas. Det går till som så att en OAI-harvester regelbundet kontaktar de olika publiceringssystemen, som svarar med att skicka XML-filer innehållande metadataposter. Metadataformatet *MODS* (Metadata Object Description Schema) används i *Testsök* som det gemensamma utbytesformatet och är bättre än Dublin Core.

Databasens indexerare har sedan två huvudfunktioner, dels att söka igenom den skapade katalogen över harvestade metadatafiler efter nya filer som tillkommit och gamla som fallit bort eller ändrats. Detta för att undvika att man hämtar poster som redan finns i databasen. Dels att indexera metadatafilerna/posterna och möjliggöra sökning i dem. Indexeraren gör ett index av dem, som en telefonkatalog kan man likna det vid, så att rätt poster går att hitta. Vid en sökning skickas en fråga till indexeraren som svarar med att skicka tillbaka resultatet i XML. För sökning och indexering använder *Testsök* *Apache Lucene*. Ett tips är även att använda sig av en utveckling av Lucene, *Solr*, som också är Open Source och ger bra sökningar.

Slutligen görs sökfunktionen och databasen tillgänglig över webben. HTML-sidor skapas för sökning, bläddring och sökresultat. Bläddring kan ske när det gäller författare, titel, institution, kategori och datum.

Träfflistan i gränssnittet visar upp nummerföljden på träffarna, författare, titel, institution, publikationstyp och datum (för publicering). Titeln är länkad till en *Fullständig beskrivning* av metadataposten (i form av HTML-sida) där författarfältet är länkat till en träfflista för alla verk knutna till det författarnamnet. Likaså erbjuder fälten institution och kategori länkar till en

träfflista för resurser som ligger under den specifika institutionen eller kategorin. Under fältet kategori finns även ämnesord som länkar till alla resurser som faller in under dem.

Men framförallt finns överst på sidan en länk till fulltextversionen i det öppna arkiv som man hämtat metadatan ifrån. Kopplingen mellan metadata och själva dokumentet är finessen med söktjänster att detta slag. I detta fall är URN:NBN mycket bättre än URL i metadatan eftersom den är mer beständig. I Testsök länkar man bland annat till universitet och högskolor som är med i DiVA-projektet. När man söker på dem i Testsök, klickar på att hämta fulltexten, så kommer man till en HTML-sida hos publiceringsverktyget DiVA. Detta eftersom det kan finnas flera versioner av en resurs: som HTML eller PDF. Dessutom kommer inte formatet PDF att finnas för alltid. Har man då den översiktssida som i DiVA så kan man lägga in framtida format som ersätter PDF.

Följande konferenspapper har nedtecknats efter ljudupptagning på band av Arwid Lund på Världsbiblioteket. Viss redigering för läsbarhetens skull har gjorts.