

Erik Sandewall

Delrapport för OpenAccess.se-projektet
Domänmodellering av rättigheter och bivillkor
vid parallellpublicering av vetenskapliga artiklar

This series of memos reports administrative and organizational issues
in the Division of Publication Infrastructure

The present memo can persistently be accessed as follows:

Memo persistent URL: <http://piex.publ.kth.se/reports/rapp/007/>

Date of manuscript: 2009-01-16

Related information can be obtained via the following www sites:

The PLEX website: <http://piex.publ.kth.se/>

The author: <http://www.ida.liu.se/~erisa/>

Följande är en övergripande och kort delrapport för delprojektet ”Domänmodellering av rättigheter och bivillkor vid parallellpublicering av vetenskapliga artiklar”. Den är avsedd både som formell rapport till anslagsgivaren och som information till samarbetspartner och andra intresserade. Det ursprungliga projektförslaget refereras som bilaga 1 nedan. Bilaga 2 och 3 är tekniska delrapporter i nuläget. Rapporteringen avser projektläget 2009-01-15.

Kort beskrivning av projektet

Den här rapporterade aktiviteten utgör ett delprojekt inom projektet *OA-publicerade domänmodeller avseende vetenskaplig publicering och gruppstruktur* som beskrivs i det ursprungliga projektförslaget (bil. 1). Projektet avser att samla och strukturera information inom givna områden och göra den fritt tillgänglig i ett så kallat Common Knowledge Library i en form som gör att den kan inkorporeras i andra programvarusystem.

Det här rapporterade delprojektet har rubriken *Domänmodellering av rättigheter och bivillkor vid parallellpublicering av vetenskapliga artiklar*. Dess mål är dels att samla in, strukturera och komplettera information som avser rättigheter och villkor för parallellpublicering av vetenskapliga artiklar, dels att göra denna information operativt tillgänglig så att den kan användas i flera typer av programvarusystem, främst:

- System som administrerar institutionella arkiv för open access ska erbjudas en tjänst som identifierar huruvida publicering i arkivet av en given, redan tidskriftspublicerad (eller tidskriftsaccepterad) artikel är förenlig med det normala publiceringskontraktet mellan respektive tidskriftsförlag och författare.
- System som är avsedda för forskare och som assisterar med hanteringen av artikelmanus och artiklar ska på motsvarande sätt kunna ge kontraktsinformation till sina användare.

Information av detta slag finns i den mycket användbara, webbaserade tjänsten SHERPA/Romeo, men denna är endast organiserad för att man ska kunna slå upp information manuellt från fall till fall. För att man ska kunna implementera de nämnda nya tjänsterna behövs ytterligare arbete på följande aspekter:

- Den verbalt uttryckta informationen på SHERPA/Romeos webbplats behöver omvandlas till en strukturerad form.
- Komplettering med information från ytterligare förlag behöver göras.
- Eftersom förlagen ibland ändrar sina regler, och eftersom uppköp och sammanslagning av förlag kan leda till ändring av reglerna för en given tidskrift, behövs ett system som arkiverar kopior av de kontrakt som gäller vid en viss tidpunkt och som automatiskt samlar och arkiverar nya standardkontrakt när sådana affischeras på respektive förlags webbplats.
- Programvara som interpreterar och tillämpar nämnda information för en given artikel och dess bibliografiska metadata behöver implementeras.

Detta är i korthet målsättningen för det nu aktuella delprojektet, vilket stöds av utvecklingsprogrammet OpenAccess.se som leds av Kungl. Biblioteket (KB). I projektet ska också samarbete etableras med SHERPA/Romeo och med användare av data, liksom med det likaledes OpenAccess.se-finansierade projektet *Parallellpublicering av vetenskapliga artiklar* (PAVA).

Som bas för projektet finns en tidigare genomförd omvandling av SHERPA/-Romeo-informationen till strukturerad form, och ett författarstödssystem (MADMAN) som kan vara lämpligt att använda för test med denna ansats. Befintliga system för administration av institutionella arkiv, t.ex. det svenska DiVA-systemet utgör också naturliga resurser för projektet.

Resultat av det inledande arbetet

Projektet organiseras med följande deltagare:

- Erik Sandewall (KTH och Linköpings universitet): projektledare, samt ansvarig för vidareutvecklingen av ett system för författarstöd (MADMAN) i vilket de nämnda IPR-tjänsterna ska inarbetas.
- Gunnar Eriksson (SICS): Implementering av den server som ger tillgång till informationen i IPR-regel-databasen.
- Preben Hansen (SICS): Utveckling av lämplig struktur för insamling och arkivering av publiceringskontrakt; insamling av sådana kontrakt från en ”testbatch” av förlag och extraktion av IPR-informationen i dem som komplettering av den redan tillgängliga informationen från SHERPA/Romeo.

Vidare är Oscar Täckström från SICS adjungerad i projektet då han har intresse av en möjlig vidareutveckling av detta i riktning mot automatisk extraktion av villkor ur digitalt tillgängliga kontraktstexter.

Det inledande arbetet på Preben Hansens del visade att den information som finns på förlagens webbplatser eller som vi kan få fram på annat sätt har en betydligt mer komplex struktur än vad vi först räknade med, och att arbetet med insamling av informationen därför kommer att vara tidskrävande och komplext. Vi har därför beslutat att göra en tydlig separation mellan punkt 3 ovan och de två första punkterna, så att systemutvecklingen görs baserad på den redan tillgängliga informationen som är extraherad från Romeo/Sherpa. Arbetet på insamling och tolkning av kontrakt har visat sig så omfattande att det föreliggande projektet måste koncentrera sig på att lägga en systematisk grund för fortsatt arbete, och på att komplettera befintlig databas med ett begränsat antal väl genomarbetade fall.

Arbetet på systemimplementering enligt de första två punkterna fortskrider i stort sett planenligt.

Sammanfattning av arbetsläget

Arbetsläget per 2009-01-15 är som följer.

Preben Hansen har i sitt delprojekt studerat en testbatch av 31 förlag som inte redan ingår i SHERPA/Romeo. Han har identifierat ett antal olika modeller för hur förlagen delger sina IPR-villkor på webbsidor och på annat

sätt. Denna strukturering är viktig för att definiera det ramverk i vilket en större mängd kontrakt o motsv kan laddas ner, analyseras och arkiveras. Jämsides med detta har han studerat det faktiska innehållet i villkoren från de 31 förlagen. Detta arbete rapporteras i bilaga 2 vilken föreligger i preliminär version vid rapporteringsdatum.

Gunnar Eriksson implementerar i sitt delprojekt ett serverprogram som kan ta emot förfrågningar (till att börja med sända som enkla http-requests) om vad som gäller för parallellpublicering av en artikel specificerad med följande parametrar (preliminärt):

- tidskrift (angiven med ISSN)
- artikelns utgivningsår och -månad
- anslagsgivare i arbetet (t ex ERC)
- typ av webbplats där man vill lägga upp artikeln (institutional, private, osv)

Servern svarar med det man behöver veta, speciellt:

- om och när parallellpublicering är OK
- vilken version av artikeln som får användas (t ex last submitted, eller publisher's final)
- information som behöver läggas in i den upplagda artikeln, t ex hänvisning till förlagets IPR, dess webbplats, eller DOI
- information som behöver läggas in i den kontext, t ex webbsida i vilken artikeln ingår.

Arbetet har kommit en bra bit och integration med system enligt nästa punkt avses göras i början av februari.

Erik Sandewall vidareutvecklar MADMAN-systemet för författarstöd där författaren kan hantera sina artiklar, formatera, göra i ordning motsvarande webbsidor, osv. Detta system kopplas nu till föregående, så att man dels kan ge användaren information om huruvida parallellpublicering är OK för en given artikel, dels också att man kan påverka utformningen av artikel och webbsidor så att de krav som förlaget har ställt faktiskt blir uppfyllda och detta med automatik. För närvarande används lokalt lagrad IPR-information; koppling till servern enligt föregående punkt avses göras i början av februari.

Utöver dessa huvudaktiviteter finns en explorativ aktivitet med deltagande av Preben Hansen och Gunnar Eriksson i samråd med Oscar Täckström att göra preliminär analys av texterna i tillgängliga kontraktsvillkor som förstudie till projekt för automatisk analys av texterna.

Fortsatta planer

Vi avser att kunna demonstrera framtaget system och att rapportera arbetet senast den 15 februari 2009, med möjlighet till några mindre kompletteringar under de följande månaderna.

Bilagor

1. Erik Sandewall: OA-publicerade domänmodeller avseende vetenskaplig publicering och gruppstruktur. Organizational Memo number 4, Division of Publication Infrastructure, KTH, Stockholm. Tillgänglig på <http://piex.publ.kth.se/reports/rapp/004/>.
2. Preben Hansen, Gunnar Eriksson och Oscar Täckström: Steps towards automatic acquisition and recognition of IPR conditions for parallel publishing. Kommande projektrapport från SICS; preliminär version föreligger.
3. Erik Sandewall: Support for IPR and Parallel Publishing Management in the MADMAN Research Author Support System. Tillgänglig på <http://www.ida.liu.se/ext/caisor/pm-archive/adept/002/>.