

# Forskningspublicering: Spredning af meta-data ved hjælp af et 'Static Repository'

Af Mikael K. Elbæk og Lars Nondal

**Hvordan synliggør man forskningen i internationale søgemaskiner? Handelshøjskolens Bibliotek høster meta-data til et arkiv med working papers.**

**Mikael K. Elbæk**  
Fagreferent for Informations- kommunikationsteknologi  
Handelshøjskolens Bibliotek i København  
mke.lib@cbs.dk

**Lars Nondal**  
Projektleder  
Handelshøjskolens Bibliotek i København  
ln.lib@cbs.dk



Mange forskningsbiblioteker har påtaget sig at opbygge 'Institutional Repositories', hvor man samler og arkiverer institutionens forskningspublikationer, eksempelvis working papers, i fuld-tekst. Formålet med denne artikel er at vise, hvordan man relativt let kan få høstet publikationers metadata ved hjælp af OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) – og derved få gjort oplysninger om ens forskningsproduktion søgbar i internationale søgemaskiner. Vores eksempel fra Handelshøjskolens Bibliotek (HBK) er delvist automatiseret, men i praksis kan alle opgaverne i processen udføres manuelt.

Der findes to måder at få metadata høstet på i forhold til OAI-PMH konventionen. Enten som et fuldt OAI-PMH kompatibelt 'Repository' [1] eller som et 'Static Repository' [2].

Vi valgte at etablere et Static Repository af to årsager. For det første er det den metode, der umiddelbart kræver mindst mulig teknisk indsigt og den mindste indsats. For det andet står HBK over for at skulle skifte til nyt bibliotekssystem, der er fuldt OAI-PMH kompatibelt (Book-IT fra Axiell), og derfor ønskede vi ikke at bruge tid på at udvikle noget, vi forventer at få leveret som pakkeløsning om et år.

## Hvad er et Static Repository?

I OAI-sammenhæng er et repository en database, der kan indeholde både metadata og fuldtekstdokumenter, eller blot metadata. Metadataene skal være tilgængelige i form af et XML-dokument på en web-server, der kaldes for en Data provider. Disse data skal så 'høstes' eller overføres til en såkaldt Service provider, som samler metadata

sammen fra flere leverandører og gør dem søgbare i en portal som eksempelvis Oai-ster.org. I et fuldt OAI-PMH kompatibelt repository skal data provider serveren kunne behandle en række simple forespørgsler fra service provideren, så denne f.eks. kun henter nye eller opdaterede poster.

Et static repository er derimod "statisk", da data provider serveren ikke kan behandle forespørgsler, men blot gøre metadata tilgængelige i et XML-dokument. I dette tilfælde skal XML-dokumentet tilknyttes Static Repository Gateway. Gateway'en behandler forespørgsler fra service provideren og sørger for, at data kan høstes.

XML-dokumentet skal være tilgængeligt på en web-server med en fast URL. I Handelshøjskolens tilfælde: <http://ep.lib.cbs.dk/oai.xml>. XML-dokumentet skal som minimum have et simpelt Dublin

## Katalogisering af working papers på HBK

På HBK katalogiserer biblioteket forskernes working papers som en del af højskolens forskningsregistrering.

..... >

Katalogiseringen sker i danMARC2-format i en working paper base udviklet af bibliotekets IT-afdeling. Databasen bygger på en MySQL-database og dokumentserver. De katalogiserede poster fra denne base eksporteres dagligt til bibliotekets katalog. I working paper basens web-baserede katalogiseringsmodul uploades dokumenterne som PDF-filer til en dokumentserver. Hvert dokument tildeles en stabil URL baseret på ISBN-nummer, og der linkes til denne URL fra katalogposterne.

Fra Bibliotekskatalogen sendes dagligt opdateringer vedrørende working papers til et XML-dokument der udgør vores Static Repository. Alle yderligere rettelser, tilføjelser mm. i posterne sker herefter i Bibliotekskatalogen, så der altid er overensstemmelse mellem beskrivelsen af et bestemt dokument i henholdsvis Bibliotekskatalogen og Static Repository.

Core-format, men man kan også anvende andre metadataformater end Dublin Core.

Hvornår er det interessant at lave et Static Repository? Det er vel primært interessant, hvis man har et mindre antal dokumenter, som ønskes eksponeret i forskellige søgemaskiner, eller hvis man har en dokumentsamling, som er statistisk. Hvis arbejdet skal laves manuelt, ligger smertegrænsen vel ved 100-200 titler/poster. I HBK's tilfælde, hvor processen er delvist automatiseret, har vi pt. ca. 750 titler i vores Static Repository.

### Hvordan gør man?

I det følgende eksempel tages udgangspunkt i, at vi ønskede at eksponere Handelshøjskolens Working Papers i Oaister.org [3], men ellers er fremgangsmåden normativ og kan anvendes generelt til at lave et Static Repository.

OAISTER er udviklet af University of Michigan, og projektet er finansieret af

Andrew W. Mellon Foundation og er et forsøg på at opbygge en søgeportal til frit tilgængelige forskningsressourcer ved hjælp af OAI-protokollen. OAISTER opdateres månedligt.

Efter at vi havde besluttet os til at lave et Static Repository, tog vi kontakt til Kat Hagedorn, som er kontaktpersonen på Oaister.org., og hun formidlede en kontakt til Tom G. Habing fra University of Illinois at Urbana-Champaign, som bestyrer en Static Repository Gateway.

I resten af artiklen vil vi fokusere på, hvordan dette XML-dokument udarbejdes, og hvilke Dublin Core felter vi valgte at tage med, med andre ord vores oversættelse fra danMARC2-format til et simpelt DC format.

Tilretning af dokumentet vil i praksis kunne gøres manuelt ved hjælp af en 'notepad' eller en XML-editor ved hjælp af 'cut/paste', navnlig hvis man har et begrænset antal poster.

### XML og metadata

Man kan anvende forskellige former for metadata-formater fra helt simpelt Dublin Core til fuldt udbygget MARC, så længe man kan konvertere det til et OAI-PMH kompatibelt XML-dokument. I HBK's Static Repository's XML skema valgte vi at oversætte vores danMARC2 format til Dublin Core. Anvendelsen af Dublin Core er ikke uproblematisk, men den diskussion vil vi ikke gå ind i her. Men nøjes med at henvise til tabellen nedenfor, der viser oversættelsen af de anvendte 'tags' i vores XML-skema og sammenhængen mellem Dublin Core og bibliotekets udgave af danMARC2. Det er dog værd at nævne, i forbindelsen med oversættelsen af vores danMARC2 poster til DC, at OAISTER ønskede at få vores danske emneord med.

Skemaet kan genanvendes af andre, hvis blot de lokalspecifikke felter ændres.

OAI:Identificer	Tekst/ Konvention		Beskrivelse
<oai:repositoryName> </repositoryName>	Copenhagen Business School Working Papers		Titel på Repository
<oai:baseURL>	http://ep.lib.cbs.dk/oai.xml		Den url som vores XML fil kommer til at ligge på
<oai:adminEmail>	workpap.lib@cbs.dk		Kontaktperson
<b>OAI:header</b>			
<oai:earliestDatestamp>	2004-08-10		Dato for oprettelse
<oai:identifier>	OAI:x656444133		En unik identifikation til hver post, vi valgte at anvende ISBN eller X-nummer
<oai:datestamp>	2004-08-10		Dato for postens oprettelse i XML format
<b>OAI:metadata</b>			
danMARC2 "Hermes"	Delfelter	Dublin Core Element	Beskrivelse
001	*d	<dc:datestamp>	Postens oprettelses dato
008	*l	<dc:language>	Dokumentets hovedsprog
100/ se også 700	*a *h	<dc:creator>	Kan gentages. Skal indeholde *a *h adskilt af , og blanktegn. Ansvarlig for indhold. Både forfattere og redaktører
245	*a *c	<dc:title>	*a og *c adskilles af : [kolon] og blanktegn Titel på ressourcen
260	*c	<dc:date>	Kun årstal medtages i **** format Tidspunkt for ressourcens oprindelse
440	*a	<dc:type >	Serietitel her brugt som publikationstype
440	*v	<dc:relation>	Serienummer
504	*a	<dc:description>	Abstract
630	*a	<dc:subject>	Kan gentages. Emneord – danske emneord skal medtages
631	*a	<dc:subject>	Kan gentages. Ukontrollerede emneord
700	*a *h	<dc:creator>	Se 100
710 440	*a *æ	<dc:publisher>	Enhed ansvarlig for udgivelsen.
856	*u	<dc:identifier>	Ressource identifier = URL Efter denne skabelon: http://ep.lib.cbs.dk/paper/ISBN/x656444133

The screenshot shows the Copenhagen Business School library catalog interface. The top navigation bar includes links for 'UDGIVNELSER', 'FORSKNING & VIDEN', 'CBS INTERNATIONAL', 'CORPORATE RELATIONS', 'CAMPUSLIV', 'NYHEDER', 'ALUMNI', 'BIBLIOTEK', and 'OM CBS'. Below this, there are links for 'Gå til biblioteket', 'Fjernadgang', 'Ny litteratur', 'CBS-studerende', 'CBS-ansatte', 'Andre lånere', and 'Om biblioteket'. The main content area is titled 'Bibliotekskatalogen' and shows a search result for a book online. The details are as follows:

Titel/Forfatter	Bog - online
Emnebegreb	<b>TITEL</b> Are Delegation and Incentives Complementary Instruments ?
Udvidet søgning	<b>FORFATTER</b> Henrik Lando
Fl. søgning	<b>SERIE</b> LEFIC Working paper, nr.2004-02
Find tidsskrifter	<b>PUBLDATA</b> København, 2004 , 21 s.
Find artikler	<b>EMNEORD</b> Belønningssystemer; Incitamentsafkløning; Agentteori; Uddelegering; Beslutningsteori; Abstract
E-Journals Service	<b>ONLINE</b> <a href="http://ep.lib.cbs.dk/paper/ISBN/v645152603">http://ep.lib.cbs.dk/paper/ISBN/v645152603</a>

On the left side, there is a sidebar menu with options like 'Databaser', 'Linksamling', 'Resourcereguider', 'Søgevejledninger', 'Forny udlån', 'Se udlån', 'Se reservationer', 'Bestil materialer', 'Spørg biblioteket', and 'Nyheder'.

To screen dumps af samme Working paper, et fra HBK (ovenfor) bibliotekskatalog og et fra OAISTER (overfor).

Et eksempel på, hvordan XML-dokumentet kan se ud, vises nedenfor. Eksemplet er delt op i tre dele.

### 1. Repository identifier

### 2. Metadataformat

### 3. Records <ListRecords metadataPrefix="oai\_dc">, <oai:record> og <oai:metadata> (de egentlige poster).

XML-dokument afsluttes af med dette tag </Repository>.

#### 1. Repository identifier

Denne del af skemaet fortæller service provideren, hvilken standard XML-skemaet er konstrueret efter, og hvordan den skal læse det. XML-skemaet skal understøtte UTF-8 tegnsættet. Almindeligvis kan man kopiere denne del og blot erstatte de felter, der er lokalbetingede her <oai:repositoryName>, <baseURL> og <oai:adminEmail>.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <Repository xmlns="http://www.openarchives.org/OAI/2.0/static-repository" xmlns:oai="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/static-repository http://www.openarchives.org/OAI/2.0/static-repository.xsd">
- <Identify>
  <oai:repositoryName>Copenhagen Business School Working papers</oai:repositoryName>
  <oai:baseURL>http://ep.lib.cbs.dk/oai.xml</oai:baseURL>
  <oai:protocolVersion>2.0</oai:protocolVersion>
  <oai:adminEmail>workpap.lib@cbs.dk</oai:adminEmail>
  <oai:earliestDatestamp>2004-09-21</oai:earliestDatestamp>
  <oai:deletedRecord>no</oai:deletedRecord>
  <oai:granularity>YYYY-MM-DD</oai:granularity>
</Identify>
```

#### 2. Metadataformat

Her fortæller XML-skemaet service provideren, hvilket format metadataene er præsenteret i, og hvilket skema der skal anvendes til at læse det. I fald at man vil anvende Dublin Core, kan denne del kopieres. Andre formater er ikke taget med her, men MARC formatet kan bl.a. også anvendes.

```
- <ListMetadataFormats>
- <oai:metadataFormat>
  <oai:metadataPrefix>oai_dc</oai:metadataPrefix>
  <oai:schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</oai:schema> <oai:metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc</oai:metadataNamespace>
</oai:metadataFormat>
</ListMetadataFormats>
```

OAISTER  
...find the pearls

Home Search Browse Institutions Help

Your search was in the **Title** field for **"delegation"** and in the **Author/Creator** field for **"lando"**.

You found **1 record**.

- \* Revise your search to retrieve fewer records.
- \* View your results, starting with **records 1 of 1**.

Results by Institution

Copenhagen Business School (CBS) Working Papers  
1 record

Sort by title

**Record 1 of 1**

Title	Are Delegation and Incentives Complementary Instruments ?
Author/Creator	Lando, Henrik
Publisher	Copenhagen Business School. CBS
Year	2004
Resource Type	text
Language	eng
Note	It is natural to suppose that delegation and incentives are complementary both in the sense that when to a lower level of an organizational hierarchy, more use should be made of incentives at that level, and incentives should be accompanied by more delegation. This issue is analyzed within a Principal-Agent framework where two decisions to be made: an effort decision which can only be made by the Agent, and some other decision either the Principal (i.e. be centralized) or by the Agent (i.e. be delegated). Within this framework it is shown that incentives are not necessarily complementary instruments; some decisions should be centralized when
Subject	Belønningssystemer
Subject	Incitamentsafledning
Subject	Agentteori
Subject	Uddelegering
Subject	Beslutningsteori

### 3. Records

Denne del af XML dokumentet viser de enkelte poster. I hver post findes først en unik identifikation i "oai:header" feltet. Derefter kommer selve dokumentets metadata i form af Dublin Core metadata-felter. Se skemaet ovenfor for at se, hvilke Dublin Core felter der korresponderer med danMARC felter.

Denne del af XML dokumentet viser de enkelte poster. Først vises, hvilket metadata format posterne er præsenteret i, hvis vi havde valgt både at tage DC og MARC med, ville posterne være samlet i hver deres klynge. I hver post findes først en unik identifikation i "oai:header" feltet. Derefter kommer selve dokumentets metadata i form af Dublin Core metadata-felter. Se skemaet ovenfor for at se, hvilke Dublin Core felter der korresponderer med danMARC2-felter.

```
- <oai:record>
- <oai:header>
<oai:identifier>OAI:x656444141</oai:identifier>
<oai:timestamp>2004-09-15</oai:timestamp>
</oai:header>
- <oai:metadata>
- <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>Higher-Order Finite Element Solutions of Option Prices</dc:title>
<dc:creator>Raahauge, Peter</dc:creator>
<dc:publisher>Copenhagen Business School. CBS</dc:publisher>
<dc:date>2004</dc:date>
<dc:type>Working paper</dc:type>
<dc:relation>2004-005</dc:relation>
<dc:hasPart>FI</dc:hasPart>
<dc:language>eng</dc:language>
<dc:description>Kinks and jumps in the payoff function of option contracts prevent an effective implementation of higher-order numerical approximation methods. Moreover, the derivatives (the greeks) are not easily determined around such singularities, even with standard lower-order methods. This ... </dc:description>
<dc:subject>kep</dc:subject>
<dc:identifier>http://ep.lib.cbs.dk/paper/ISBN/x656444141</dc:identifier>
</oai_dc:dc>
</oai:metadata>
</oai:record>
```

På <http://ep.lib.cbs.dk/oai.xml> kan dokumentet ses i sin helhed, og det giver måske et bedre overblik

Og hvad får man ud af alt dette arbejde?

På Handelshøjskolens Bibliotek har vi med en forholdsvis begrænset indsats fået spredt metadata inklusive links til egen dokumentserver om Handelshøjskolens Working Papers til i første omgang en enkelt søgemaskine, OAISTER. OAISTER kan vel ikke konkurrere med Google endnu, men vil utvivlsomt blive et relevant søgeinstrument for alle, der søger efter forskningsbaseret litteratur. Pt. indeholder OAISTER 3,5 mio. dokumenter fra ca. 350 institutioner.

Og vores metadata er tilgængelige, så de umiddelbart kan høstes af andre kommende 'service providers', der ønsker at opbygge andre søgeportaler, der måske er fagligt afgrænsede, i modsætning til OAISTER, der medtager alle forskningsressourcer uanset fagområder.

Working papers med adresse på HBK Working Paper-server er i perioden 1.1.2004 – 31.6.2004 blevet downloadet 4.495 gange (eksklusiv besøg af robotter og søgemaskiner). Og det på trods af, at der endnu ikke findes nogen særskilt grænseflade til disse working papers. HBK Working papers har været søgbare i OAISTER siden 2/9 2004, så det bliver spændende at se benyttelsesstatistikken for 2. halvår 2004.