

# Pode

Statusrapport, juli 2009

# Innhold

1.	Innledning.....	3
2.	Mål .....	3
	Resultatmål.....	3
	Retningslinjer .....	3
3.	Aktiviteter.....	4
	Gjennomførte aktiviteter fase 1 .....	4
	Samarbeid med Biblioteklaboratoriet.....	4
	Workshop #1 Oslo 19.-20. juni 2008 .....	4
	Workshop #2 Oslo 30.-31. mars 2009.....	4
	Samarbeid med Høgskolen i Halden.....	5
	Engasjering av programmerer .....	5
	Planlagte aktiviteter fase 2 .....	5
4.	Foreløpige resultater.....	5
	Problemstillinger .....	6
	Tilgang til egne katalogdata .....	6
	Eksport av Deichmans katalog.....	6
	Katalogisering – formater og praksis .....	7
	Kompetanse .....	7

# 1. Innledning

Podeprosjektet (heretter: Pode) startet opp i mars 2008. Prosjektet var opprinnelig ett-årig, men etter søknad om prosjektmidler til fase 2 ble Pode tildelt midler for ytterligere et år.

Etter reorganisering av prosjektet i januar 2009 og påfølgende behandling i styringsgruppen (26/2-09), ble det besluttet at fase 1 og fase 2 kunne ses samlet i forhold til fremdriftsplan og sluttrapportering.

Denne statusrapporten markerer at vi beveger oss over mot arbeidsoppgaver i fase 2. Prosjektet vil fortsatt jobbe med aktiviteter knyttet til fase 1 utover høsten.

Statusrapporten oppsummerer prosjektets overordnede mål og resultatmål, gjennomførte aktiviteter i fase 1, planlagte aktiviteter i fase 2, foreløpige resultater og problemstillinger. I tillegg rapporteres det på fremdrift, usikkerhet i prosjektet og hovedposter i regnskapet per mai 2009.

## 2. Mål

### Pode skal

- Legge til rette en plattform for utvikling av mashupapplikasjoner.
- Utvikle to eller tre mashupapplikasjoner
- Utvikle et nettsted som illustrer bruken av disse applikasjonene
- Prøve ut relevant funksjonalitet i det frie biblioteksystemet Koha

### Resultatmål

- [www.bibpode.no](http://www.bibpode.no) → Podes nettsted. Demo av utviklede mashups. Utvikles som et frittstående topplag over Koha/SRU-protokollen. Nettstedet skal også inneholde et repository for applikasjoner/mashups og dokumentasjon av disse.
- [www.bibpode.no/deichman](http://www.bibpode.no/deichman) → Bibliofil + mashup. Demo av utviklede mashups mot Bibliofils katalog ved bruk av z39.50-protokollen.
- [dev.bibpode.no](http://dev.bibpode.no) → Podes Kohainstallasjon. Demo av utviklede mashups integrert i Koha.
- [www.bibpode.no/blogg](http://www.bibpode.no/blogg) → Fortløpende dokumentasjon av prosjektets fremdrift
- Dokumentasjon av API-ene vi knytter oss mot.
  - Egen bibliotek katalog (z39.50 (RSS))
  - Det frie biblioteksystemet Koha (SRU)
  - Eksterne innholdsleverandører
- Sluttrapport

### Retningslinjer

Alle mashups lages mot både z39.50 og SRU-protokollen

All kildekode kommenteres. Kildekoden publiseres og gjøres tilgjengelig for nedlasting.

Det skal fremlegges entydig dokumentasjon av API-ene vi knytter oss mot.

### 3. Aktiviteter

#### **Gjennomførte aktiviteter fase 1**

Behov for å avgrense prosjektet i forhold til opprinnelig idé førte til at Musikkonline og Filmarkivet ble valgt bort som eksterne kilder. For å løsrive prosjektet fra Deichman valgte prosjektgruppa å benytte et åpent biblioteksystem som eksempel-katalog. Dette var også nødvendig da Bibliofil ikke har en SRU-protokoll. Pode vil benytte Deichmans katalog som eksempel for mashups mot z39.50-protokollen.

#### **Samarbeid med Biblioteklaboratoriet**

Pode har arrangert to arbeidsseminarer i samarbeid med Biblioteklaboratoriet, samt deltatt på et arbeidsmøte med Summa og Öppna bibliotek 20. november 2008. Workshop #1 ble arrangert av Biblioteklaboratoriet. Invitasjon ble sendt ut på e-postlisten Bibliotek-Norge. Workshop #2 ble arrangert av prosjektgruppa og var for spesielt inviterte; en representant per biblioteksystem pluss studenter med utviklerkompetanse, prosjektgruppa og Biblioteklaboratoriet. Alle workshops og møter ble avholdt ved Høgskolen i Oslo.

#### **Workshop #1 Oslo 19.-20. juni 2008**

Formålet var å analysere og dokumentere systemer og datakilder til bruk i en norsk (katalog)kontekst generelt og Pode spesielt, samt å kartlegge og dokumentere svakheter ved eksisterende systemer og gjennom dette liste opp ønsker for fremtidige OPACer.

#### **Oppgaver**

1. Installasjon og dokumentasjon av metasøkesystemet [VuFind](http://www.vufind.org/)<sup>1</sup>
2. Installasjon og dokumentasjon av metasøkesystemet [LibraryFind](http://libraryfind.org/)<sup>2</sup>
3. Undersøkelse av [Summa](http://www.statsbiblioteket.dk/summa)<sup>3</sup> (her hadde vi bare tilgang på god dokumentasjon).
4. Undersøkelse og dokumentasjon av eksterne datakilder, som Bokkilden og Wikipedias dbpedia.org.
5. Kartlegge ønsker for Podes grensesnitt og samtidig (advare mot) svakheter ved eksisterende systemer: Diskutere og finne momenter til en Bad usability calender med et OPAC-perspektiv.

#### **Workshop #2 Oslo 30.-31. mars 2009**

Oppgaver og program ligger i kapittel 8 Vedlegg.

Detaljer og resultater for hver enkelt oppgave er dokumentert på Biblioteklaboratoriets wiki: <http://wiki.biblioteklaboratoriet.no/index.php/Kategori:Pode-workshop>

---

<sup>1</sup> <http://www.vufind.org/>

<sup>2</sup> <http://libraryfind.org/>

<sup>3</sup> <http://www.statsbiblioteket.dk/summa>

## **Samarbeid med Høgskolen i Halden**

I januar fikk vi forespørsel om å være oppdragsgiver for et hovedprosjekt ved avd. for Informasjonsfag, Høgskolen i Østfold. Pøde formulerte en oppgave som i hovedtrekk gikk ut på å teste hvor ”mashupvennlig” protokollene som gir tilgang til katalogdata er, illustrert ved en ”reiseplanlegger”. Erik Eiesland, Lars Emil Knudsen og Øystein Wormstrand valgte vår oppgave og oppstartsmøte var 10. februar. Samarbeidet har fungert bra – deler av prosjektgruppa har reist til Halden for oppfølging og studentene har deltatt på workshop i Oslo. Samarbeidet varte i perioden 10. februar – 28. mai 2009. Sluttrapporten er vedlagt.

## **Engasjering av programmerer**

For å løsrive prosjektet fra Deichman og Bibliofil valgte prosjektgruppa tidlig å installere Koha siden vi får bedre kontroll med en egen katalog. Vi får også testet den moderne protokollen SRU i mashupsammenheng. I den forbindelse var det nødvendig å styrke Koha- og programmeringskompetansen i gruppa. Magnus Enger er engasjert som programmer for en 3-månedersperiode i 50% stilling (juni – september, inkl sommerferie).

## **Planlagte aktiviteter fase 2**

- Verkspresentasjon i trefflister (FRBRisering)
- Undersøke muligheter for relasjoner mellom poster
- Undersøke FRBRisering på postnivå i Koha
- Vi ønsker FRBR på trefflistenivå. Er det mulig i Koha eller må vi teste ut annen programvare?

Prosjektet ønsker å arrangere et seminar som er åpent for alle i samarbeid med Biblioteklaboratoriet i starten av september. Tema for seminaret: FRBR og FRBRisering

Del 1) Teoretisk innføring i FRBR (verktøy, status i Norge/verden, regler for FRBR-registrering osv)

Del 2) Praktiske eksempler.

Ressurspersoner/-miljøer kan være: Bibits med Micromarc3, Bibsys med Ask2, Trond Aalberg, Knut Hegna, Mikael Preminger med Griegsamlingen, Reindex. Vi kan også kanskje løfte problemstillingen opp på et nasjonalt nivå og høre hva NB/Bibsys tenker om muligheten for et åpent ISBN-register. Kanskje noen kan si noe om RDA og FRBR, for eksempel Tore Brattli på Dokumentutdanningen i Tromsø som har introdusert RDA i undervisningen.

## **4. Foreløpige resultater**

- Eget domene der kildekoden og demonstrasjon av mashups skal publiseres; [www.bibpode.no](http://www.bibpode.no)
- En egen blogg hvor prosjektet dokumenteres fortløpende. utfordringer og resultater deles med alle; [www.bibpode.no/blogg](http://www.bibpode.no/blogg)
- Installasjon av Koha: import av data fra Deichmans katalog
- Studentenes reiseplanlegger: <http://torfeus.deich.folkebibl.no/~erike/mashup/reiseplanlegger.php>  
Overordnet oppgave: å teste ”mashupvennligheten” til de ulike måtene å hente ut katalogdata på - RSS, z39.50 og SRU.  
Prosjektgruppa vil jobbe videre med denne mashupen høsten 2009.

- Resultater fra workshop # 2
  1. Anbefalingslister: <http://torfeus.deich.folkebibl.no/~gruppe1/skjema.php>
  2. Norgeshistorielinje: gjenstår å koble bøker fra samlingen til periodene på linja. [http://torfeus.deich.folkebibl.no/~gruppe2/tidslinje/simile\\_test.html](http://torfeus.deich.folkebibl.no/~gruppe2/tidslinje/simile_test.html)
  3. Eventyrmashup: katalogdataene var for lite detaljerte til å få noe ut av det og Dok.pro hadde ingen åpen API.
  4. Musikk mashup: utkast <http://torfeus.deich.folkebibl.no/~gruppe4/>. Prosjektgruppa vil jobbe videre med denne mashupen høsten 2009.

## **Problemstillinger**

### **Tilgang til egne katalogdata**

Lukkede biblioteksystemer og manglende tilgang til egne data har vist seg å ikke være så vanskelig som vi først antok. Selv om Bibliofil per i dag ikke tilbyr et åpent API eller SRU-protokollen er det mulig å lage mashups basert på bibliotekataloger via z39.50 protokollen. Studentene har bevist dette via "reiseplanleggeren" hvor de har sammenlignet RSS, z39.50 og SRU med hverandre i forhold mashupvennlighet og hastighet mot Deichmans katalog.

- RSS slik den er implementert i Bibliofil er lite egnet til å lage mashups med. Treg, upresis og ufullstendig. Se 4.1.1 i studentenes sluttrapport.
- z39.50, sammen med PHP/YAZ, tilbyr mye fleksibilitet fordi man kan hente ut katalogdata i flere formater og syntakser: "Med små endringer kan reiseplanleggeren eller andre mashups som bruker Z39.50 hente data fra andre, eller flere, bibliotek." Se rapportens punkt 4.2.
- SRU var desidert raskest og enklest å jobbe med. Protokollen fremstår som både veldokumentert og fremtidsrettet. Se rapportens punkt 4.3.

Tidsmessig kom SRU best ut på gjennomsnitt, z39.50 var ikke langt unna, mens RSS var tregest. z39.50 har vist seg ikke å være så vanskelig å bruke som først antatt og har den fordel at den er relativt godt utbredt i bibliotekverden. Det varierer fra leverandør til leverandør om z39.50 følger med installasjon eller ei, og dermed kan det forekomme bibliotek i Norge som ikke har z39.50-server. Norzig er ansvarlig for norske profiler av både z39.50 og SRU, om biblioteksystemleverandører velger å følge denne standarden varier. Aleph – som er et internasjonalt system – følger ikke den norske versjonen. For detaljer se rapporten "Biblioteksøk – Sluttrapport fase 1 Vedlegg 4: Kartlegging av funksjonalitet i biblioteksystemene"<sup>4</sup> som gir et bilde av situasjonen.

### **Eksport av Deichmans katalog**

I forbindelse med eksport/import fra Bibliofil til Koha opplevde vi problemer med ufullstendige poster som skyldtes måten de norske bokstavene var "pakket" i Bibliofil. Bibliotek-Systemers ISO2709-eksport har vært basert på bruk av tegnsettet ISO8859-1 og hjelpeverktøy konverterte automatisk det ferdige resultatet til UTF-8. Dette er nå rettet, slik at vår ISO2709-eksport også fungerer med UTF-8 gjennomgående. Dermed blir lengdene korrekte for alle UTF-8-tegn, inkludert Æ, Ø og Å.

<sup>4</sup> <https://www.abm-utvikling.no/bibliotek/norsk-digitalt-bibliotek-gammel/prosjekter/biblioteksok/fase-1/vedlegg-4-kartlegging-av-biblioteksystemene>

## Katalogisering – formater og praksis

Forskjeller i Marc-formater ved import til Koha og Marc 21: tap av info om dokumenttyper i \*019\$b det alvorligste mangelen. Siden feltet ikke er med i den offisielle Normarc standarden er dette strengt tatt en forskjell mellom BSMarc og Marc 21.

For Poda og folkebiblioteksamlinger er nok mangelen på mulighet til å spesifisere dokumenttyper i tillegg til format et viktig tap av informasjon.

I importen til Koha slik den er per juni 2009 har vi mistet all informasjon om dokumenttyper og formater fordi \*007 er lite brukt på Deichman siden man har mulighet til å registrere dette på en enklere måte i \*019\$b

BSMARC har et eget delfelt for koding av dokumenttyper - 019 \$b *Utvidet informasjonsfelt : dokumenttyper*. Feltet brukes til å presisere type dokument utover det \*007 tillater av formatbeskrivelser. \*007 begrenser seg til å angi feks CD, men ikke videre angivelse om objektet er musikkopptak eller lydbok. \*019 \$bdc,di \$r l : dc = CD og di = lydbok l=roman. \*019 \$b er ikke med i den offisielle [Normarc](#)<sup>5</sup>-standarden.

En utfordring er oversettelsen av delfelt \$b i felt for \*019 *Utvidet informasjonsfelt* til Marc 21. Marc 21 har angivelse av format i 007, men det er ingen mulighet til å skille mellom musikkopptak og lydbok på CD i dette feltet. Lydbok og språkkurs havner i samme kategori.

Amerikanske folkebibliotek (Nelsonville Public Library og New York Public Library) løser dette ved å legge inn *Audio books* i \*650. Dette er ikke en helt heldig løsning - i det [avanserte søket](#)<sup>6</sup> får man for mange uklare. En annen konsekvens av valget er at dokumenttyper som *Audio books* og *Feature films* dukker opp som fasetter under overskriften *Topics*. Dessuten mister man muligheten til å lage versjoner av katalogen på flere språk.

Ikke alle norske folkebibliotek har tatt i bruk \*019\$b - Tynset bibliotek bruker kombinasjonen \*007sc og \*655\$a *Lydbøker*. Dette kan være en mulig løsning med problemet med å uttrykke dokumenttyper i Marc 21 og Koha.

## Kompetanse

I prosjektgruppa og ved Deichman generelt er kompetansen på det bibliotekfaglige og bibliotektekniske høy, mens vi mangler spesialisert programmeringskompetanse. Mens vi for fase 1 har fått tilgang på en freelance programmerer, er dette mer usikkert for fase 2. Se punkt 6 – Usikkerhet i prosjektet.

---

<sup>5</sup> <http://www.nb.no/normarc/010-049.5.php#informasjonskodefelt>

<sup>6</sup> [http://acpl.kohalibrary.com/cgi-bin/koha/opac-search.pl?&expand=su-to#su-to\\_id](http://acpl.kohalibrary.com/cgi-bin/koha/opac-search.pl?&expand=su-to#su-to_id)