

# Linked data - et nettverk!

[Innledning](#)

[Om prosjektet](#)

[Bakgrunn](#)

[Mål med prosjektet](#)

[Oppbygning av prosjektet](#)

[Workshops og møter](#)

[Første samling](#)

[Andre samling](#)

[Tredje samling](#)

[Fjerde samling](#)

[Videre samlinger](#)

[Videre arbeid](#)

## Innledning

Denne rapporten er en intern rapport skrevet for ansatte på NTNU Universitetsbiblioteket. Hensikten er å informere de ansatte om Linked data - et nettverk! og rapporten er en samlet oversikt over prosjektet og arbeidet som er gjennomført så langt. Den er bevisst ikke skrevet på et veldig detaljert nivå fordi man ønsker at alle skal kunne forstå innholdet i rapporten.

Materialet det siktes til i rapporten er ulike kilder informasjon som finnes i våre spesialsamlinger. I prosjektets tilfelle kan være alt fra brev mellom kjente personer, manuskripter, bilder, kart og noter som på noen måte har tilknytning til Sangerfestene i Bergen og Trondheim. Dette materialet digitaliseres slik at et ellers beskyttet og skjørt materiale lettere kan fremstilles og benyttes av publikum.

## Om prosjektet

Linked data - et nettverk! er et ettårig samarbeidsprosjekt mellom NTNU Universitetsbiblioteket (NTNU UB), Universitetsbiblioteket i Bergen (UBB) og Bergen offentlige bibliotek (BoB). Prosjektet har fått midler (500 000kr) fra Nasjonalbiblioteket og skal resultere i en demo som viser hvordan samlet data fra de tre institusjonene ved hjelp av Linked data vil gi publikum en ny opplevelse av tilgjengelighet og fleksibilitet.

## Bakgrunn

Linked data - et nettverk! er et resultat av en prosjektidé som oppstod under en workshop mellom NTNU UB og UB Bergen. Denne workshopen omhandlet også Linked data og var i forbindelse med arbeidet med Digitalmanuskripter. På denne samlingen ble det tydelig at det finnes mye materiale som omhandler samme hendelser/personer spredt på de ulike bibliotekene, og at alle jobber for å formidle dette digitalt til publikum.

Ideen ble derfor at en samlet fremstilling ville kunne gi publikum en mer helhetlig opplevelse av materialet noe som forhåpentligvis ville føre til økt bruk og mer kjennskap til informasjonen vi sitter på.

Basert på dette og det faktum at formidlingsønsket har fått flere institusjoner i Norge (og ellers i verden) til å begynne å snakke om og å jobbe med Linked data, ble man enig om å søke midler til et samarbeidsprosjekt. [Søknaden](#) ble innvilget og tildelingsbrevet fra Nasjonalbiblioteket satte krav til at vi skulle involvere flere prosjekt nemlig Realfagstermer i Oslo og Bergens musikkhistorie samt inkludere Nasjonalbiblioteket i sentrale diskusjoner.

Fra før har alle de tre institusjonene andre prosjekter som omfatter Linked data og fått tro på at dette kan være veien å gå. Ved å benytte de samme standardene for metadataarkitektur vil vi kunne få et materiale som kan knyttes sammen uten at vi trenger å jobbe likt eller ha nøyaktig samme mening om like typer materiale. De ulike samlingene med data kobles sammen ved at man finner likheter i datasettene som blir koblet sammen slik at informasjon fra ulike kilder tilsammen gir et helhetlig inntrykk.

I tillegg ønsket alle å lære mer om Linked data og å bli kjent med andre som jobber med dette.

### Hva er Linked data

Linked data beskriver en måte å strukturere, publisere og koble sammen data på. Linked data i seg selv er altså bare et begrep som omhandler ulike standarder man kan benytte for at data enkelt kan struktureres, deles og kobles sammen.

Enkelte standarder er mer benyttet enn andre. Eksempler på disse er [RDF](#) (Resource Description Framework) for representasjonen/oppbygging av dataene og spørrespråket [SPARQL](#) for å hente informasjonen ut av dataene.

All metadata struktureres fra bunn av i såkalte tripler bestående av små utvetydige deler som presenteres som nettlener. Disse lenkene kan igjen lenkes til andre informasjonsskilder med informasjon om tilsvarende materiale. Fordi hver datadel er en selvforklarende lenke, får man en stor mulighet for utvidelse og fleksibilitet. Dette forenkler arbeidet med å gjøre helhetsopplevelsen bedre og forståelsen for materialet større for publikum.

Fordi Linked data sier noe om selve dataene er det ikke nødvendigvis slik at man kan se på en nettside om Linked data er benyttet eller ikke. Utseende på en nettsiden kan være likt uavhengig av det som ligger bak. Likevel vil man kunne oppleve en forskjell. Dette fordi man får sammenkoblet informasjon fra ulike kilder, nye sider ved materialet og en helhet som kan oppstå uten å være et resultat av tett samarbeid mellom institusjoner. Det setter ingen begrensninger til geografisk lokasjon, kan koble sammen ulike språk, gjøre søk med perfekt presisjon på en enkel måte og vise utfyllende informasjon om det man er interessert i.

## **Mål med prosjektet**

Som navnet tilsier er det et mål for prosjektet å etablere og utvikle et nettverk innen Linked data slik at vi blir kjent med andre som jobber med dette og kan dele erfaringer, kunnskap og ideer. Dette innebærer å skape en felles forståelse av prosesser, begrep og metoder vi benytter oss av. Alt dette vil gi oss en kompetanseutvikling eller en kompetanseheving og muligheten for omstilling til annen teknologi og/eller arbeidsmetoder.

Dataene blir åpne og tilgjengelige og det blir eventuelt spesifisert og designet moduler som må til for integrering med andre systemer dersom dette er ønskelig.

Det kanskje mest synlige resultatet vil vær en demo som samlet viser materiale de tre institusjonene har for Sangerfestene i Bergen (i 1863) og Trondheim (i 1883). Denne demoen vil vise hvordan Linked data kan gi en bedre fleksibilitet og tilgjengelighet for vårt materiale.

Med en demo mener man her en nettside hvor informasjon om Sangerfestene vil være søkbar og bli presentert på en informerende, selvforklarende, spennende og kanskje uventet måte.

## **Oppbygning av prosjektet**

Linked data - et nettverk! består av en prosjektgruppe med fire personer fra NTNU UB, fire personer fra UB Bergen og tre personer fra Bergen offentlige bibliotek. Prosjekteier og prosjektleder er ved NTNU UB av to årsaker, nemlig at midlene fra Nasjonalbiblioteket forvaltes herfra og at NTNU UB har mest erfaring basert på tidligere Linked data prosjektet. Videre er det en koordinator på hver av de andre institusjonene med ansvar for å styre arbeidet der. På denne måten kan vi ha felles mål og milepæler underveis uten at vi trenger å jobbe på samme måte eller i samme tempo.

I tiden før sommeren har vi samlet materiale, økt kunnskapen og strukturert en plan for gjennomføring av prosjektet. Alt materiale er i denne stund i ferd med å bli ferdig scannet og registrert slik at det er klart til konvertering og implementering fra høsten av.

Det jobbes parallellt med metadata og demo, men en ferdig demo er likevel det siste som

kommer til syne i prosjektet. Dette betyr at det for folk flest vil være lettere å se styrken og fleksibiliteten til det vi jobber med nå mot slutten av prosjektet.

## Workshops og møter

Til nå i prosjektet har det vært fire samlinger av ulike typer og tilholdssted har hovedsakelig vært i Bergen. I tabellen vises en oversikt over hva som har vært avholdt når:

Dato	Workshop	Møter	Annet
24.01.2012		Prosjektmøte/ oppstartsmøte	
15.02.2012-17.02.2012	Med Magnus Stuhr (to dager)	Prosjektmøte	
23.04.2012-24.04.2012			Koordinatorer fra hver institusjon møttes
03.05-2012-04.05.201 2	Kort med Robert Engels	Prosjektmøte, Styringsgruppemøte	

### Første samling

Første møte handlet mest om å få en oversikt over prosjektet, møte de som skulle involveres og ta stilling til de nye kravene som fulgte med tildelingsbrevet fra Nasjonalbiblioteket. Vi så på hva slags kunnskap de ulike institusjonene satt med og hvordan vi kunne hjelpe hverandre. I tillegg ble det klart at UB Bergen i forbindelse med et annet Linked data prosjekt hadde begynt å planlegge en workshop i midten av februar som viste seg å være av aktuelle interesse og nytte også for Linked data - et nettverk! Det ble derfor bestemt at denne workshopen skulle vinkles mot og inkluderes i dette prosjektet.

### Andre samling

På andre samling fikk vi avklart samarbeidet med de nye aktørene fra tildelingsbrevet. Vi ble enige om å inkludere Realfagstermer, Bergens musikkhistorie og Nasjonalbiblioteket i våre workshops, noe de også så som en god løsning. Dette innebærer at vi deler kunnskap med hverandre.

I prosjektmøte ble prosjektet mer formelt organisert, vi valgte prosjektleder og ble enige om hvordan vi kunne arbeidet mest effektivt. Vi diskuterte en samarbeidsavtale mellom de tre

institusjonene og ble enige om at vi trenger å ha en slik avtale. Vi diskuterte en fremdriftsplan og felles tema for materialet vårt. Temaet som ble valgt var Gullalderen (Nasjonalromantikk i Norge) og vi ble enige om å til neste gang ha materialet klart fra hver institusjon samt en påbegynt strukturering.

Dette vil i praksis si en liste over hva slags typer materiale hver institusjon har om temaet. Med strukturering mener man en ide om hva slags informasjon det er viktig å ha med for hver materialtype og hvordan disse bitene med informasjon kan sees i forhold til hverandre. For eksempel vil et dokument ha en tittel, et årstall og en forfatter (også flere punkter som ikke nevnes her), og forfatteren vil igjen ha et navn, et fødselsår og et dødsår (også flere punkter som ikke nevnes her).

Workshopen som fant sted under denne samlingen var med Magnus Stuhr fra Computas. Han fortalte generelt om Linked data, hva det er og hvordan man kan jobbe med det. Vi fikk god tid til å stille spørsmål og på denne måten øke vår kunnskap om temaet. Magnus Stuhr sa videre at han er tilgjengelig dersom vi ønsker å benytte oss av hans kunnskap og det er også mulig å leie han inn dersom man ønsker lignende foredrag eller workshop for å forklare linked data mer inngående.

### **Tredje samling**

I Trondheim hadde vi et lite møte med koordinatorerne fra hver institusjon. Samtidig fikk vi mulighet til å ha et lite møte med prosjektdeltakere fra Trondheim som ikke hadde mulighet til å bli med på neste prosjektmøte i Bergen.

Her kom det fram at Gullalderen hadde vist seg å være et litt for bredt tema til at det var enkelt å få oversikt over alt materialet og det ble derfor bestemt at vi skulle begrense tema til Sangerfestene i Bergen og Trondheim. En liste over alt materiale innenfor dette temaet skulle derfor være klart til neste samling.

I tillegg fikk vi i denne runden sett nærmere på fremdriftsplanen, satt inn noen flere punkter, sett litt på og kommentert arbeidet vi hadde gjort så langt samt planlegging av neste samling og møter i Bergen.

### **Fjerde samling**

Den siste samlingen før sommeren inkluderte et styringsgruppemøte, et prosjektmøte og en liten intern workshop med Robert Engels fra popsenteret i Oslo. Han fortalte litt om hva han arbeider med og hvordan de har inkludert og benyttet Linked data.

I tillegg bidro han med forslag til hvordan han mente vi kunne fremstille vårt arbeid og hvilke problemer vi eventuelt kunne unngå å støte på. Robert fortalte også om andre prosjekter som på noen måte kunne være av interesse for oss. Samtidig så han på våre modeller (våre struktureringer av metadata) og kom med innspill og kommentarer som hjalp oss videre på

veien. Robert Engels gjorde det også klart at det var mulig for oss å involvere han i vårt arbeid videre, noe vi ser på som høyst aktuelt.

Samme periode i Bergen diskuterte vi hva vi forventer oss av demoen, hvordan vi ser for oss arbeidet med den og hva vi ønsker å få formidlet ut mot publikum. Vi tenker oss at nettsiden må være appelerende både for forskerne og for den interesserte amatør. Fremstillingen må være intuitiv og spennende samtidig som den bør vise hvor materialet kommer fra slik at det blir tydelig at vi har samlet informasjon fra ulike kilder.

## **Videre samlinger**

Videre planlegger vi flere samlinger utover høsten og den første vil være en workshop i månedsskiftet august/september.

## **Videre arbeid**

Innen juni skal alt materialet som skal brukes i prosjektet være registrert og digitalisert. Videre arbeid ut over høsten vil innebære konvertering og lagring av datasett, arbeid med søk mot datasettet, arbeid med demo og fremstilling av materialet på en måte som publikum vil kunne dra nytte av og sette pris på. I praksis betyr dette at det synlige arbeidet for publikum fortsatt gjenstår.

Vi har ikke tenkt å benytte oss av en registreringsklient, det vil si at informasjonen om materialet vi bruker skal struktureres i et excel-ark og hver celle skal derfra konverteres til RDF slik at vi får våre datasett.

Vi har ingen krav til at vi må gjennomføre alt arbeid selv, f.eks design/programmering/ferdigstilling, fordi vi er enige om at vi ønsker å begrense tidsbruken for oss selv. Samtidig vil det å inkludere ekstern hjelp i noen tilfeller kunne gi et bedre og mer troverdig resultat. I tillegg skaper vi da også et nettverk og kunnskapsdeling med andre aktører samtidig som vi sprer litt informasjon om vårt materiale.

Uavhengig av om vi gjør arbeidet selv eller setter det bort, må vi lage klare rammer for hva vi forventer ut av dataene våre og av demoen. Dette innebærer også å sørge for at det som lages er åpent og så enkelt at det kan inkluderes i andre systemer senere dersom dette er ønskelig.

Etter hvert som vi arbeider med prosjektet vil vi også videre holde andre ansatte ved institusjonene informert om det vi gjør. Med resultatene fra prosjektet håper vi å øke interessen og forståelsen for Linked data slik at vi på sikt kan involvere større deler av biblioteket i lignende arbeid.

Neste offisielle rapport er en statusrapport som skal sendes til Nasjonalbiblioteket. Frist for denne rapporten er 15. september.