

# Protokoller

Z39.50, SRU og OAI-PMH

Tor Arne Dahl, Høgskolen i Oslo

Biblioteklaboratoriets workshop

Bergen 12.-13. november 2007

# Hva er en protokoll?

Regler som beskriver hvordan to datamaskiner kan opprette forbindelse, kommunisere og utveksle data ved hjelp av meldinger. Meldingene har en syntaks som kan tolkes av begge datamaskinene

# Eksempler

- Z39.50
- SRU (Search/Retrieve via URL)
- OAI-PMH (Open Archive Initiative - Protocol for Metadata Harvesting)

# Teknisk utvikling etter Z39.50

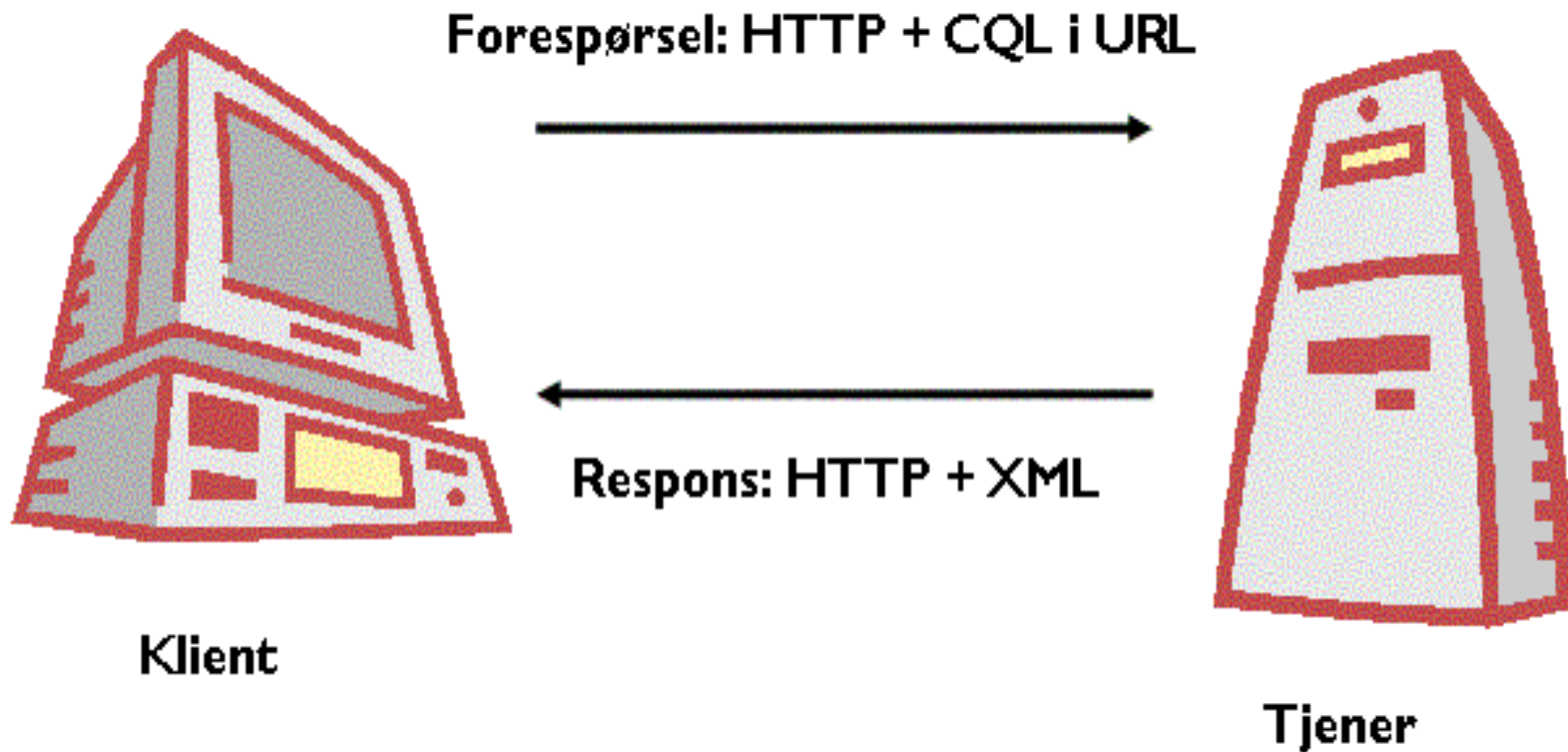
- World Wide Web
- XML
- Web services
  - REST-orientert (REpresentation State Transfer)
  - SOAP-baserte (Service Oreiented Architecture Protocol eller Simple Object Access Protocol)

# SRU

- Vedlikeholdes av Library of Congress
  - URL: <http://www.loc.gov/standards/sru/>
- Brokete navnehistorie: ZING og SRW
- I dag: én tjeneste, to transportmetoder:
  - SRU via GET (REST-orientert og enklest)
  - SRU via SOAP (SOAP-basert)
- Eget søkespråk: CQL (Contextual Query Language)



# Arkitektur



# Operasjoner

## 1. searchRetrieve

- Klienten sender CQL-spørring
- Tjeneren returnerer poster

## 2. scan

- Klienten sender et CQL-uttrykk mot en indeks
- Tjeneren returnerer termer fra indeksen

## 3. explain

- Klienten bestiller tjenerens explain-dokument
- Tjeneren returnerer detaljer om tjenestene som tilbys

# Eksempel

Søk på "*Hamsun, Knut*" i Dublin Core-indeksen *creator* hos Library of Congress med 10 poster levert i postskjemaet Simple Dublin Core:

```
http://z3950.loc.gov:7090/voyager?  
operation=searchRetrieve&version=1.1  
&query=dc.creator=%22Hamsun,%20Knut%22  
&recordSchema=dc&startRecord=1  
&maximumRecords=10
```

# OAI-PMH

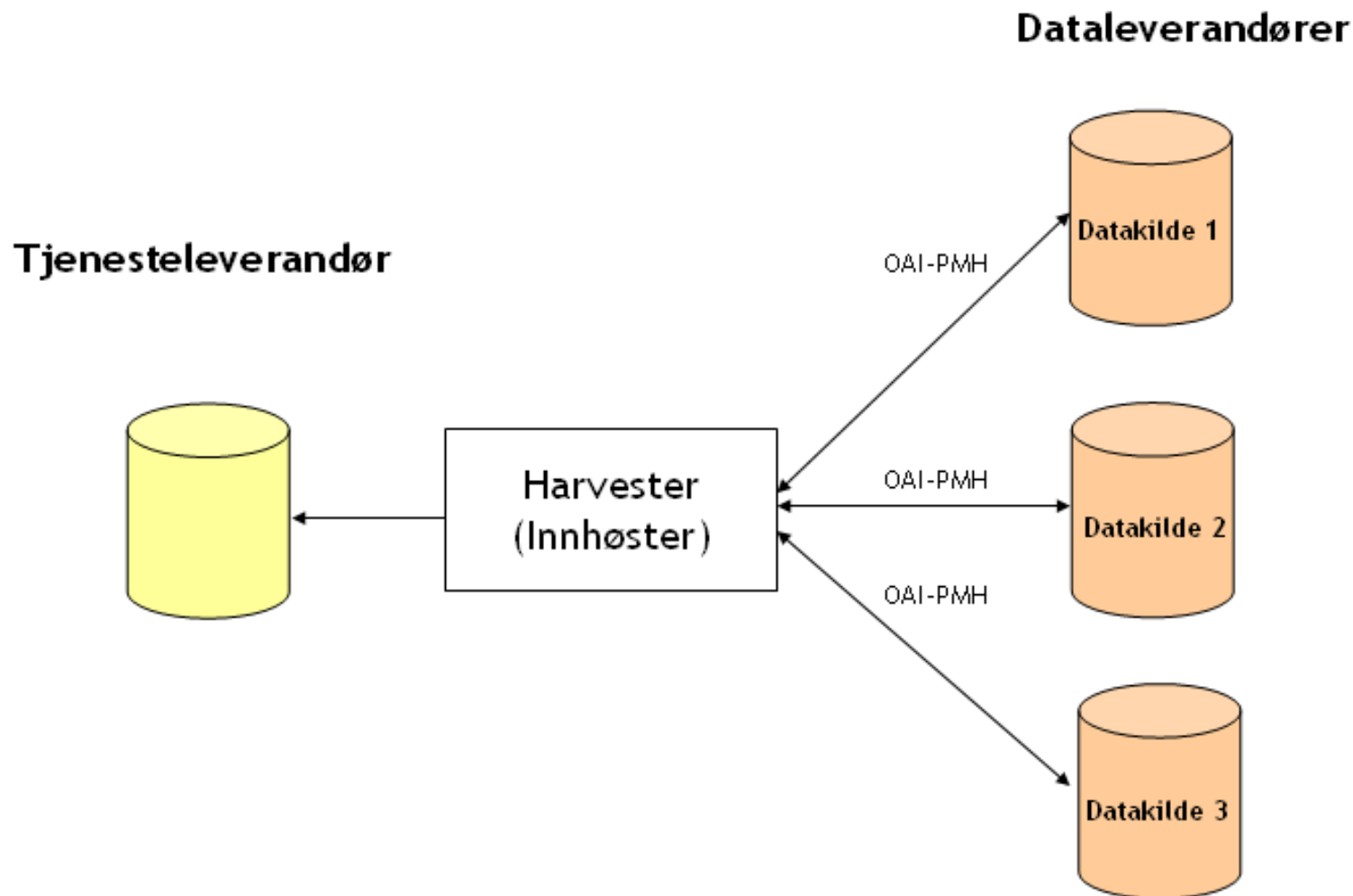
- Vedlikeholdes av Open Archives Initiative
  - URL: <http://www.openarchives.org/>
  - Mål: Spre Open access-innhold mest mulig effektivt
- Opprinnelse: Open archives/Open access-bevegelsen
  - Digitale vitenskapelige dokumenter som gjøres fritt tilgjengelige
  - Oversikt over Open Access-tidsskrifter: DOAJ (Directory of Open Access Journals)

# Data- og tjenesteleverandører

- En *dataleverandør* har et repository («åpent institusjonelt arkiv») med metadata og eventuelt fulltekstdokumenter, og gjør disse metadataene tilgjengelig via OAI-PMH-protokollen. Eksempel: [arXiv.org](https://arxiv.org) og [DUO](https://duo.uio.no) (UiO)
- *Tjenesteleverandøren* setter opp et dataprogram som «høster inn» metadata fra én eller flere dataleverandører og lagrer dem i én database. Metadataene hentes inn ved faste tidsintervaller og ved hjelp av OAI-PMH-forespørsler. Tjenesteleverandører kan utvikle og tilby tjenester basert på de innhøstede metadataene. Eksempel: [NORA](https://nora.uio.no) og [OAlster](https://oalster.uio.no)



# Arkitektur



# Begreper

- Ressurs: Et objekt eller «ting» som metadataene beskriver. Ressursene er ikke en del av protokollen
- Objekt: Representerer metadata om en ressurs. På grunnlag av objektet kan det lages poster i forskjellige metadataformater
- Identifikator: En identifikator peker unikt til et objekt i repository
- Post: En manifestasjon av metadata uttrykt i et eller annet metadataformat. Poster er alltid kodet som XML og sendes over som respons til OAI-PMH-forespørsler

# Seks verb

- Identify
- ListMetadataFormats
- ListSets
- ListIdentifiers
- ListRecords
- GetRecord

# Eksempler fra arXiv.org

- Beskriv tjenesten (svarer til explain-operasjonen i SRU)

- `http://arxiv.org/oai2?verb=Identify`

- Hent posten med identifikator  
***oai:arXiv.org:adap-org/9311003***

- `http://arxiv.org/oai2?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:arXiv.org:adap-org/9311003`