

Praktisk workshop i (data)modellerering

5. mai 2017

Om modeller

Når en ønsker å studere egenskaper ved en bygning, en båt eller en gjenstand lager en gjerne en fysisk skalamodel som kan testes under kontrollerte forhold. I våre dager er den fysiske modellen ofte erstattet av en digital modell og matematisk simulering. Værvarslingstjenester som yr.no er basert på matematiske modeller av atmosfæriske fenomener kombinert med digitale modeller av terrengformasjoner. Tilsvarende vil en funksjonell og hensiktsmessig databaseløsning kreve en gjennomtenkt modell av den delen av virkeligheten som dataene i den skal beskrive.

Datamodeller finnes på flere nivå. I programvareutvikling er datamodeller formaliserte beskrivelser av hvordan data skal organiseres. I utviklingen av slike datamodeller vil fokus være på datakonsistens, enklest mulig vedlikehold samt på plassbruk og rask prosessering av data. Men i et godt informasjonssystem bør datamodellen også være i samsvar med brukernes mentale modell av de oppgavene systemet skal løse. Utviklerne må prøve å forstå denne mentale modellen, og i det minste bruke den som et utgangspunkt når den mer tekniske datamodellen skal implementeres i systemet.

Grensen mellom tekniske datamodeller og formaliserte begrepsmodeller er uklar. I informatikk er det en lang tradisjon for å låne termer fra psykologi, lingvistikk og filosofi men bruke dem metaforisk. Denne praksisen kan ofte gjøre det vanskelig å vite om en snakker om den virkelige verden, modeller av verdenen eller implementasjoner av slike modeller. Dette er spesielt synlig i felt som kunstig intelligens, 'knowledge engineering' og semantiske teknologier som er noen av de mest innovative, men samtidig mest spekulative feltene i informatikk. Termen 'ontologi' er et godt eksempel. Ontologi betyr læren om væren og er egentlig navnet på et vel etablert felt innenfor filosofi. I dag brukes ordet alt fra emneordslister, datamodeller og formaliserte begrepsmodeller. Det er i den siste betydning vi vil bruke ordet i denne workshopen.

Om workshopen

I workshopen vil vi gå gjennom hvordan en kan definere begrepsmodeller (ontologier). Vi skal ikke altså ikke arbeide med tekniske datamodeller som beskriver hvordan en skal organisere tabeller i en relasjonsdatabase og heller ikke på teknikker for å implementere ontologier og data i grafdatabaser/RDF(S). Men siste del av workshopen er en gjennomgang av et verktøy vi kan bruke til å omforme (mappe) eksport fra en (relasjonsdatabase) til RDF i henhold til en gitt ontologi.

Vi vil bruke begrepsmodellen CIDOC-CRM som eksempel. Denne modellen er utviklet i regi av ICOM-CIDOC og er basert på faktisk dokumentasjonspraksis i museer. Formålet for utviklingen av CIDOC-CRM var å lage en felles modell slik at en kan integrere data fra databaser i kulturminneinstitusjoner. CIDOC-CRM kan uttrykkes i RDF(S) og OWL, men også tjene som en guide for 'best practise' når en skal designe tradisjonelle relasjonsdatabaser for museer.

Vi vil først se på hva vi mener med såkalt ontologisk analyse og deretter gå gjennom designprinsippene bak CRM. CRM har i forskjell fra mange tradisjonelle modeller klasser for hendelser og abstrakter (f.eks. motivet i et bilde) og skiller mellom navnet på en ting og tingen selv.

Deretter vil vi gå gjennom en del praktiske eksempler på bruk av CRM til å modellere vanlige prosesser i museer og arkiv.

Den siste delen av denne gjennomgangen vil kort beskrive familien av CRM utvidelser som for eksempel FRBRoo for bibliotek og åndsverk. I den forbindelse kan det også være interessant å se på metadatamodellering for bildende kunst og utøvende kunst.

Den siste delen av workshopen er viet et system utviklet i Nederland og i Hellas for å transformere data til en CRM kompatible form i RDF. Dette systemet er generisk og en kan godt bruke det til å transformere data til andre modeller slik som EDM.

Det er ikke tatt høyde for rene 'hands on' øvelser. Men det er meningen at det skal være dialog mellom deltakerne og workshopleder. Graden av dette vil være avhengig av antall deltakere.

En praktisk orientert workshop om (data)modellering

Stikkord : ontologisk analyse, begrepsdannelse, hendelsesorientert modellering , CIDOC-CRM, eksempler fra museum, arkiv og bibliotek, skjemamatching og mapping av data til RDF

Del 1

- Introduksjon til ontologisk analyse og hendelsesorientert datamodellering
- Integrasjon av informasjon om kulturarv
- Basisbegreper for å lage en dokumentasjonsstandard
- CIDOC-CRM
- Eksempler

Del 2

- Familien av CIDOC-CRM kompatible ontologier
- Hvordan modellere metadata for åndsverk, kunst og utøvende kunst

Del 3 Mapping – transformasjon av datasett

- Introduksjon
- X3ML - en formalisme for å beskrive datatransformasjon
- verktøyet 3M basert på X3ML, utviklet ved FORTH å transformere datasett til CIDOC-CRM og andre ontologier. Kort gjennomgang og praktisk demonstrasjon