

Bilag 10. Beskrivelse af platform til dataopbevaring og -håndtering

Små.TING – platformen

For at kunne lave meget målrettet søgning og filtrering i de svarsæt, som små.TING-databrøndens søgesystem returnerer til brugerne i "pissegratis.dk", etableres der en række komponenter i et samlet flow, der tilsammen udgør den del af platformen under pissegratis.dk, der understøtter søgning, filtrering og visning/afspilning af content og de metadata, der er i små.TING.

De relevante kilder for pissegratis.dk realiseres i små.TING gennem følgende komponente:

Posthus

Kilder lægges grundlæggende ind på to måder. Gennem dataleverancer direkte fra content-ejeren. Dette foregår i en relativt bred vifte af inputformater, XML-varianter, MARC, PDF,.. Den anden metode er, at små.TING's datahøster selv henter kilderne på de sites eller baser, hvor den primære produktion af data foregår. Data konverteres herefter ind i små.TING-brønden's DataDock.

DataDock

Data opbevares dels i den form, som de afleveres/høstes i, og dels normaliseres og konverteres data ind i et ensartet XML-format og Dublin-Core (DC), der anvendes direkte som grundlag for søgesystemets grundfunktioner.

I DataDocken – der er en procesbehandling - foregår også forskellige former for yderligere berigelser af data. Det drejer sig om kobling til, hvad der normalt betegnes som "semantisk web", hvor hele grundlaget for at etablere bedre semantisk sammenhæng imellem metadata/content, teknisk implementering og brugernes mentale modeller for de pågældende områder. De tekniske modeller er fx RDF (Resource Description Framework) og andre 3WC standarder.

Kobling mellem forskellige former for emneordssystemer er centralt her. Det er også her, at ejerkoder etableres og andre former for administrative metadata kobles til, der gør bedre filtrering overfor målgruppen mulig.

Objekt-Repository

Data ender i et objektrepository, der egentlig er den egentlige databrønd for små.TING. Her ligger data i de forskellige formater skabt i DataDocken. De opbevares i et framework, der muliggør alle de behov for at skabe relationer mellem data, der er behov for, efter den model som betegnes Fedora Commons (FC). I FC opbevares metadatastrukturerne, relationerne, beskrivelserne, hvorimod selve content og de oprindelige metadata ligger i selve den underliggende database. Databasen er valgfri. Små.TING implementeres foreløb i Open Source frameworket Postgres, der er et stærkt alternativ til markedets kommercielle løsninger.

Profit Taking Index

Det direkte grundlag for søgning skal foregå i en række specialiserede indexer. Indexerne består af et fælles defaultregister, målrettede indexer rettet imod diverse filtreringsbehov, præformaterede visregistre, facetter, osv.

Der skabes således en vis intelligens direkte ind i indexerne, og konceptet sikrer også hurtige svartider, så brugerne af pissegratis.dk ikke oplever unødige ventetider. Indexerne omdannes herefter til egentlige søgeregistre på søgeplatformen.

Søgemaskinen

Søgemaskinen tager sit udgangspunkt i en tilpasset og videreudviklet version af to Open Source platforme: Lucene, der er den egentlige søgemaskine og SolR, der er en

implementation ovenpå Lucene, der retter sig imod en række funktioner og services rettet imod bibliotekers og andre kulturinstitutioners behov i relation til søgning, visning, facetstrukturer osv. Det er også i denne del af platformen for små.TING, at det abstrakte grundlag, forretningslogikken, for de overliggende webservices er implementeret.

Webservice-lag

Alle funktioner og services i små.TING understøtter pissegratis.dk gennem webservices.

Der etableres to varianter:

- SOAP/XML – der beskrives i en såkaldt WSDL (Webservice Definition Language)
- REST/JSON – der beskrives i et Explain-dokument (XML).

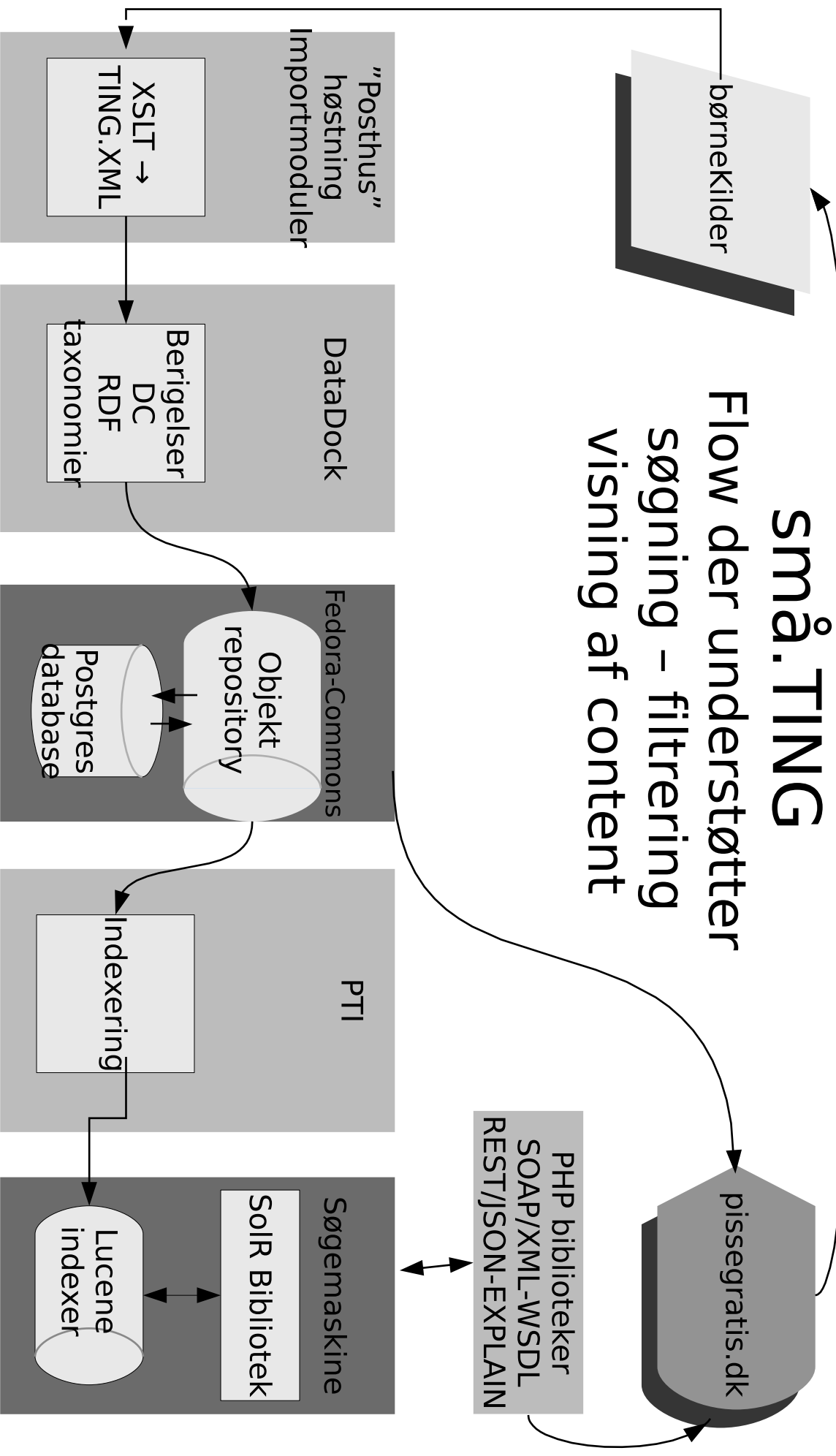
REST = Representational State Transfer

JSON = JavaScript Object Notation

Webservice-laget vil også understøtte adgang til at få udleveret fulde datagrundlag direkte fra Objektrepository´et.

små.TING

Flow der understøtter
søgning - filtrering
visning af content



Supplerende Services

Små.TING integrerer også til en række andre supplerende webservices.

De supplerende services understøtter dels den inspiration, der skal fremlægges for børnene og dels forskellige administrativt rettede funktioner.

Mange af de data, der etableres i disse services er life-data og repræsenteres ikke direkte i databrønden med henblik på søgning og fremvisning sammen med content. I stedet repræsenteres de gennem de fremhævede services, der også i et vist omfang stiller funktioner til rådighed. De supplerende services er:

- Brugerspor, fx ADHL – Andre Der Har Lånt,...
- Bruger Skabte DATA (BSD)
- Bibliotekssystemer
- Linkresolving

En udvalgt del af de data, der hidrører disse services vil blive konsolideret og gjort søgbare, så de kan understøtte pissegratis.dk direkte overfor brugerne i en given sessions søgning og filtrering.

I tilknytning til små.TING-plattformen og pissegratis.dk i det hele taget er der en håndtering af brugerne. Brugerautentifikation, og -autorisation af adgang til betalingsbelagte kilder, som bibliotekerne finansierer og formidler løses gennem forskellige komponenter, dels en Proxy-løsning og dels en autentifikations- og autorisationsmekanisme.

små.TING

Integration til andre webservices

