

# Vedlegg

## Referanser

Baekdal, T. (2011). *The Shift - From print to digital... and beyond*. Baekdal Media.

Poppendieck, M., & Poppendieck, T. (2003). *Lean Software Development: An Agile Toolkit*. Addison Wesley.

## Forprosjektrapport

**FORSIDE ER UTELATT VEDLEGGET**

## Presentasjon

Vi er en gruppe på tre dataingeniørstudenter som har tatt på oss oppgaven med å utvikle en konseptuell plugin til WordPress som kan importere nyhetselementer som følger standarden NewsML-G2.

### **Aptoma AS**

Vår oppdragsgiver er Aptoma AS, som er et programvareselskap som fokuserer hovedsakelig på nyere media.

*Geir Berset*  
CEO  
geir@aptoma.com

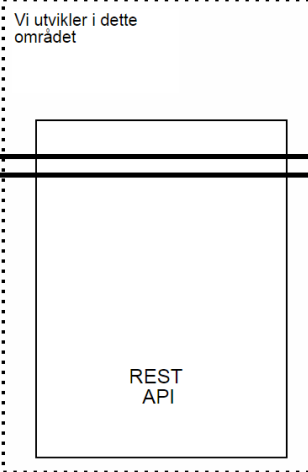
*Stefan Grunert*  
Veileder hos bedrift  
stefan@aptoma.com

## Sammendrag

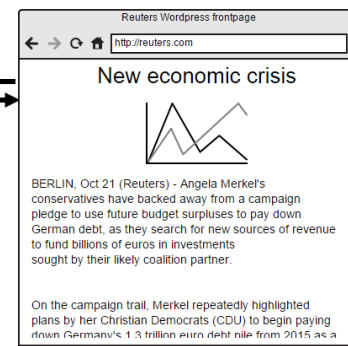
Vår konseptuelle plugin skal gjøre det mulig å bruke et API for importering som kan tas i bruk med Representational State Transfer (REST) over HTTP, for deretter å lagre (i databasen til WordPress) og publisere nyhetselementene når de mottas.

## Abstrakt

Program som eksporterer NewsML-G2



WordPress-basert nettsted



Nyhetselementene følger standarden NewsML-G2 satt av International Press and Telecommunications Council (IPTC) og brukes i dag av blant annet Associated Press (AP), Norsk Telegrambyrå (NTB) og Reuters. NewsML-G2 formatet er implementert som XML.

Omfattende standard medfører planleggingsutfordringer og vi går derfor for Extreme Programming med modulær iterativ utvikling.

## Dagens situasjon

Mediehus som blant annet NRK, Aftenposten, VG og Dagbladet møter utfordringer med at kundene deres i økende grad mottar nyheter digitalt. Kundene forventer at nyheter på nett skal være gratis, og de møter konkurranse fra bloggere og nettbaserte bedrifter. Dette stiller krav til en effektiv produksjon og distribuering av nyheter. Avisene jobber derfor med å forenkle prosessen og har også i stor grad gått over til å motta nyheter via nyhetsbyråer. Umiddelbar distribusjon av nyhetssaker har blitt et krav, samtidig som man fortsatt har strenge krav til nyhetene. De må være nøyaktige, lovlige og ikke grovt krenke personer eller grupper.

Aptoma bidrar til å løse utfordringer mediehus møter ved å levere produkter for forvaltning og produksjon av innhold til web. De fokuserer på intuitive og effektive verktøy, og deres produkter er mye brukt av større og mindre mediehus.

Aptoma er interessert i å fremme standarden NewsML-G2 for å formidle nyheter. Dersom alle følger én standard kan tekst, bilder, video og metadata lett tas ut og behandles etter behov. Deres verktøy kan da tas i bruk av flere. Standarden er allerede tatt i bruk av noen av de største mediehusene og byråene i verden, blant annet Reuters, Associated Press, NTB, VG og Eurovision.

Vår rolle er her å lage en plugin til WordPress som direkte importerer nyhetslementer som følger standarden. WordPress er et content management system (CMS) og bloggverktøy. Av de 10 million mest besøkte sidene i verden, brukes WordPress av hele 23%. Slik kan brukere verden over (inkludert store mediebedrifter) publisere nyheter rett fra kilde til nett med presis og korrekt metadata i høyt tempo.

## Mål og rammebetingelser

### Oppgave

Utvikle en plugin til WordPress som kan importere data fra nyhetsstandarden NewsML-G2 satt av International Press and Telecommunications Council (IPTC).

### Krav

Grunnet omfanget av oppgaven og standarden er det en fungerende prototype / konseptuell plugin som skal utvikles. Det er ikke krav om å ha fungerende import av bilder, lyd og video, men en beskrivelse av hvordan dette kunne ha vært gjort i praksis er ønsket.

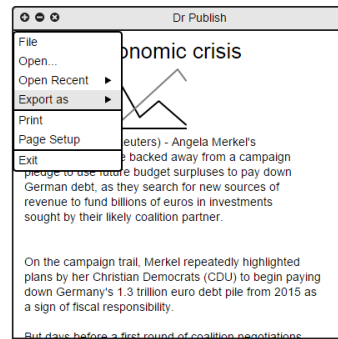
Koden skal legges ut, med en lisens for åpen kildekode.

### Våre mål

- Plugin skal ha et REST API.
- Plugin skal bruke WordPress sitt API for utvikling og tåle oppdatering av kjernen til WordPress.
- Koden skal lett kunne vedlikeholdes av andre.
- God feilhåndtering som gir meningsfulle tilbakemeldinger.
- Vi ønsker at det skal være enkelt å ta i bruk.
- Tilfredsstillende god sikkerhet på APIet
- Plugin skal ha mulighet for manuell opplastning.

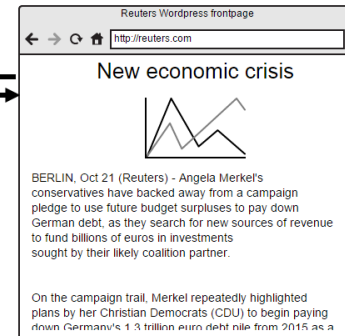
## Abstrakt

Program som eksporterer NewsML-G2



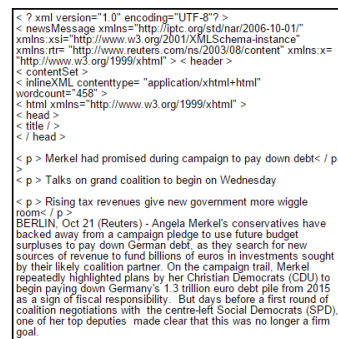
Vi utvikler i dette området

WordPress-basert nettsted



## Teknisk

NewsML-G2 fil



REST API

WordPress-basert nettsted



## Våre rammer

- Minimum arbeidstid
  - 09:00 – 16:00
  - Mandag – Torsdag.
  - Hos Aptoma

## Teknologier og avhengigheter

- PHP samt XPath
- Markeringsspråk (XML, HTML)
- CSS
- JavaScript (jQuery, AJAX)
- MySQL
- Apache

## Standarder og rammeverk

- NewsML-G2
- WordPress
- HTTP protokollen

**Miljø og verktøy**

Arbeidsplass i kontorlandskap hos Aptoma som inkluderer kontorrekvisita, lunsj og tilgang på møterom. Vi bruker tekstredigeringsprogrammet Notepad++ og Sublime Text.

## Løsninger / alternativer

**Løsning på oppgaven**

Standarden NewsML-G2 er svært omfattende med flere hundre tagger med data og beskrivelser av data. Dette medfører store utfordringer med å både forstå og implementere standarden. Som følge av dette har vi valgt å holde langsiktig planlegging til et minimum, vi har også fått dette anbefalt av oppdragsgiver. Derfor har vi ofte møter og forsøker heller å bruke tiden på iterativ og modulær utvikling, med løse koblinger slik at modulene enkelt kan erstattes hvis en bedre løsning skulle dukke opp. Dette gir oss muligheten til å jobbe parallelt med lite dødtid. Extreme Programming er derfor utviklingsmetodikken vi legger oss etter, da det også er svært motiverende å se resultater kontinuerlig.

Vi har valgt å bruke PHP, da rammeverket WordPress er skrevet i PHP og dette derfor var nødvendig.

**Arbeidsmetoder**

- Daglige møter
- Extreme Programming
- Alle skal alltid ha en arbeidsoppgave
- Ingen krav til tidsbruk på arbeidsoppgave
- Fleksibel håndtering av arbeidsoppgaver

**Kunnskapsrelaterte utfordringer**

- Lite eller ingen tidligere erfaring med PHP
- Ingen erfaring med NewsML-G2 / XML
- Lite eller ingen erfaring med WordPress.

## Analyse av virkninger

Aptoma har gitt mye informasjon som forteller hvordan NewsML-G2 kan implementeres. Og etter en del samtaler ser det ut som begge parter har sett for seg samme løsning på problemet. Siden vår plugin lages for WordPress falt valget på PHP siden det er språket WordPress bygger på. Vi startet med å lage en plugin med manuell opplasting av filer.

For å gjøre det enklest mulig for brukerne valgte vi at vår plugin skal være RESTful. Dette medfører at man kan sende nyheter rett fra ulike publiseringsverktøy og direkte med kommandolinje på Linux eller

Windows rett inn i WordPress databasen. Vi ønsker også at personer som bare har tilgang til selve NewsML-G2 filene vil kunne laste opp nyheter til WordPress.

Aptoma selv kommer sannsynligvis ikke til å bruke vår plugin direkte, men det vil kunne brukes av alle med interesse for NewsML-G2 og WordPress.

Den største fordelen vil komme i form av at en stor plattform som WordPress vil fungere sømløst med standarden.

Vi tror at løsningen vi har valgt og målene vi har satt er lar seg gjennomføre. Dersom vi ikke skulle klare å gjennomføre prosjektet, vil det ikke påvirke bedriften direkte negativt.

### Scenarier

1. Produktet er så mangelfullt at det ikke kan brukes til noe.
2. Produktet oppfylder delvis kravet til oppgaven.
3. Produktet oppfylder kravet til oppgaven eller mer.
  - Vi beviser at standarden kan brukes i WordPress
  - Brukere kan bruke vilkårlige verktøy som eksporterer NewsMLG-2 for å lage innlegg i WordPress.
  - Standarden kan bli tatt i bruk av langt flere.

### Milepælsplan

Det er vanskelig for oss å sette en konkret milepælsplan da det er vanskelig å planlegge noe særlig i dette prosjektet. Vi jobber parallelt med både produktet og dokumentasjon til fristen under.

Statusrapport	24.10.2014
Prosjektskisse	05.12.2014
Forprosjekt	23.01.2015
<b>Innlevering av prosjekt og prosjektrapport</b>	<b>22.05.2015</b>
Prosjektrapport	26.05.2015
Presentasjon	08.06.2015

Hendelse	Sannsynlighet	Konsekvens	Forebygge	Tiltak
Sykdom	Middels	Forsinkelse av arbeidsprosessen	Jobbe jevnt for å unngå unødvendig stress. Si i fra ved behov for pause.	Arbeidet til den syke blir fordelt.

Harddisk problematikk	Lav	Vi blir nødt til å gjøre alt arbeidet på nytt.	Versjonshåndtering og sikkerhetskopi i nettsky.	Skifte harddisk og laste ned kode og dokumentasjon på nytt
Motivasjon	Lav	Vi mister arbeidskraft.	Ha pauser ofte, hjelpe hverandre, skape godt arbeidsmiljø, ha faste arbeidsrutiner og helt fri ellers.	
Virus og andre ondsinnede programmer	Middels	Vi blir forsinket i arbeidet	Anti-virus, og sunt nettvett.	Reformatere infiserte maskiner
Arbeidslokalet forsvinner	Lav	Vi må finne oss et nytt sted og sitt når vi skal jobbe		Finne et nytt stede vi kan sitte
Konflikter	Lav	Det blir ikke noe godt arbeidsforhold i gruppa	Sørge for at det ikke er mye kringling innad i gruppa	Finne en løsning slik at man kan fortsette arbeidet
Tyveri	Lav	Forsinkelse i arbeidet	Ha godt holdninger i forhold til fysisk sikring	Skifte passord, laste ned kode, og fortsette utvikling på andre maskiner
Vi blir ikke ferdig med prosjektet innen tiden	Lav	Vi får ikke vist alt vi kan og klarer ikke å lage et tilfredsstillende produkt	Holde oversikt over hva som må gjøres og hvor mye tid som gjenstår	
For kompleks problemstilling	Middels	Vi bruker masse tid på løse et problem som ikke kan løses og ender å måtte gjøre noe helt annet	Holde kontroll på hva vi jobber med. Lese dokumentene angående NewsML-G2 grundig.	Dialog med Aptoma, endre kode, beskrive problemstillingen i dokumentasjon.

## Prosjektskisse

## Gruppemedlemmer

Michael Pande (s188077)

Petter Lundberg Olsen (s188115)

Diego Pasten (s188071)

## Oppdragsgiver

Geir Berset

geir@aptoma.com (CEO, Aptoma AS)

Aptoma AS

Grensen 18

0159 Oslo

Tlf: 404 80 015

## Kontaktperson

Stefan Grunert

stefan@aptoma.com

## NewsML-G2 plugin for WordPress

### Litt om oppdragsgiver:

Aptoma utvikler og tilbyr konsulent tjenester og har siden det ble grunnlagt i 2004 posisjonert seg sterkt innenfor digital publisering av innhold. Kunder inkluderer VG, NRK, Aftenbladet, og E24.

### Prosjektets plass i oppdragsgivers virksomhet:

Aptoma ønsker å fremme standarden NewsML-G2 ved å gjøre en WordPress plugin tilgjengelig som importerer nyhetsartikler i dette formatet til WordPress. Standarden utvikles og vedlikeholdes av International Press Telecommunications Council ([IPTC](#)) som er en sammenslutning av noen av de største nyhetsbyråene i verden og andre interessenter i samme bransje. Ønsket er at vår plugin skal fungere som konsept på standardisering av nyheter på internett slik at man kan effektivisere publisering og spredning av disse. Ved prosjektslutt blir koden lagt ut som Open Source.

### Om prosjektet:

Prosjektet går ut på å utvikle en plugin/modul til WordPress som kan importere en standard for utveksling av nyheter i formatet NewsML-G2 som for øyeblikket er implementert som XML.

[23%](#) av alle nettsteder bruker publiseringsplattformen WordPress. Å utvikle en plugin til publiseringsplattformen medfører derfor at svært mange får mulighet til å benytte seg av den, og dermed får en mer sømløs overføring av data mellom WordPress og NewsML-G2 standarden. NewsML-G2 er del i en serie av standarder. I tillegg til NewsML, finnes også EventsML-G2 og SportsML-G2. Ønsket med standardene er egentlig å kunne lage nyhetsobjekter som inneholder tekst, bilder, grafikk, video osv og som sammen med metadata deklarerer sammenhengen i innholdet.



Eksempler på metadata er revisjonsnummer, statusdetaljer for objektet, og mer administrative detaljer (opphavesrett, forfattere osv).

Internett har gjort at nettaviser og andre nyhetskilder trenger å produsere større mengder nyheter enn før og derfor vil det være nyttig med én standard.

**Mulige krav til maskinplattform, dataverktøy og andre teknologier:**

Databaser og Web (PHP, XML, NewsML-G2, CSS, Javascript, HTML, SQL)

## Prosjektdagbok

### **Dagbok 18/3 Petter**

Har brukt hele dagen på å skrive unit test til newItemParse klassen. Jeg hadde skrevet noe kode fra før, men det viste seg at denne koden var overflødig. Opprinnelig lå testen inne i klassen de skal teste, men jeg har beste meg for å legge alle testene på et separat sted slik at tester for alle klasser kan samles på ett sted.

### **Dagbok 17/3 Petter**

Vi har brukt hele dagen i dag på lage skjellsetet til sluttdokumentasjonen. Vi har begynt med dokumentasjonen nå for å ikke havne i en tidsklemme når vi kommer til Mai.

Andre valg som har blitt tatt mellom denne og forrige dagbok:

Jeg har sette gjennom objektorienter php for å se om dette kan gjøre det lettere når vi skal overføre informasjon for NewsML parseren til RESTApi-et. Fram til nå har vi overført informasjonen i en array som ikke er langt fra å komme ut av kontroll. Håpet var at objektorientering kunne gjøre det lettere, men siden dette innebærer å skrive om både parseren og RESTApi-et valgte vi å ha ting som de er akkurat nå. Skulle vi fått noe ut av objektorientert php måtte vi implementert dette før.

### **Dagbok 4/3 Petter**

I dag her jeg endret arrayen som sender informasjon slik at et bilde ikke ligger som en egen index men i stede, men i stede ligger under den posten de tilhører. Dette er for å gjøre det lettere for Michael å vite hvilket bilde som skal limes inn i hvilken nyhetssak.

Vi fant også ut at DOMXPath ikke tar med HTML tagger når det gir svaret på en spørring og jeg måtte finne en metode på Stack Overflow som gjorde dette for oss.

### Dagbok 04/3 Michael

I dag er jeg tilbake etter å ha vært syk fra 26/2

I dag:

- Koblet Diego sin dokumentasjonsside sammen med demosiden
- Fikset viktig 500 internal server error feilen som gjorde at pluginen ikke fungerte på UNIX baserte servere
- Ordnet en primitiv implementasjon av bilder
- Kommentert på masse med PHPDOC
- Ordnet en mulighet for å overkjøre oppdateringsbegresningen med newsitem's version
- Hatt møte med veileder hos Aptoma (Hvordan vi lå ann, IPTC konferanse og møte, forespørsel etter flere eksempler fra store mediahus. m.m)
- 

### Dagbok 4/3 Diego

Siste uken i februar ble brukt til å få til en overhaling av nettsiden som ble laget i januar og i dag har Michael og jeg slått sammen demosiden og dokumentasjonssiden. Dette tok krevde en del innsats men til gjengjeld er løsningen blitt ekstremt god. Nå kan endringer pushes rett ut på nett med git.

### Dagbok 3/3 Petter

Har startet med dokumentasjonen rundt newsitemParse.php som skal ligge på den offisielle siden til pluginen. Her tar jeg for meg de viktigste tingene man trenger å vite når man i senere tid skal gjøre endringer eller lese koden.

### Dagbok 26/2 Petter

Brukte dagen på å optimalisere den koden jeg har skrevet. Dette er i hovedsak å endre noen if strukturer. Har også lest litt mer på schema.org for å se om vi kan få mer bruk form det.

### Dagbok 25/2 Petter

Gjorde noen endring i forbindelse med publisering status utifra de dataene som finnes i NewsML. Har også gjort endringer i forbindelse med bruker sin jeg siden jeg sist skrev dagbok. Nå er alle brukerne i en liste uavhengig som de er creator eller contributor. Dette er fordi WordPress i utgangspunktet bare har en forfatter, og nå vill personen som er creator bli satt opp her. Listen er laget slik at creatororen ligger først med alle contributores etter.

Det har også blitt gjort en endring i forbindelse med xml namespaces. Siden det står i NewsML dokumentasjonen at NewsMessage er frivillig settes nå det ytterste namespace avhengig av hva som står der, eller det settes ikke i det hele tatt om den ytterste taggen ikke har xmlns.

### **Dagbok 25/02 Michael**

I dag har jeg jobbet med å forbedre workflowen vår, og å få startet ordentlig på demosiden. Dette innebar å deploye fra git repository, og hoste hele wordpress demoen rett fra git repo.

Dette gav meg en del utfordringer:

- Hardkodet URL for site\_url og home i WordPress, måtte bli gjort dynamisk.
- Windows og Linux har forskjellig oppsett av url-adresser.
- Hosting siden må støtte SQLite slik at vi enkelt kan holde den lokale versjonen og den ferdige versjonen lik, uten noe ekstra arbeid.

Workflowen er nå som følger:

Plugin -> Git Repo (Plugin)

Demo med plugin -> Git Repo (Demosiden) -> Deploy

I morgen skal jeg jobbe videre på workflowen til det punktet der vi jobber direkte på demoen, og endringer i pluginen enkelt kan ble reflektert i demonstrasjonen. Dette vil trolig spare oss mye tid.

### **Dagbok 24/02 Michael**

I dag har jeg ferdigstilt forfattere, det ser ut til å fungere som ønsket, men det tok ganske mye av dagen. Årsaken til dette var manglende støtter for flere spørringer i spørresystemet til WordPress.

Jeg har også startet å lage en demonstrasjonsside som er svært portabel, den bruker SQLite istedenfor MySQL som er standard for WordPress. Dette betyr at avhengigheten til MySQL som database er borte. Hosting av en demosome vil derfor være vesentlig enklere.

## Dagbok 23/02 Michael

I dag har jeg jobbet med å få implementert forfattere. Dette har vist seg å være krevende. Årsaken til dette er at det ikke er støtte for flere forfattere på et innlegg i wordpress.

Noen hovedpunkter:

- Måtte opprette/finne brukere igjennom REST APIet
- Satte hovedforfatter (Det som wordpress støtter) til creator
- Visning av flere forfattere på samme sted, med lenker til forfatterens arkiv
- Erstattet arkivfunksjonaliteten på forfattere med egen (Uferdig)
- Måtte overstyre globale variabler i wordpress
- Det skal ikke være mulig å ha duplikatforfattere (må ha unik epost)

Det krever derfor at brukerne som skal benytte seg av pluginen må endre på temaet de bruker til wordpress for å få vist flere forfattere på et innlegg. For å løse dette har jeg valgt å fokusere på å gjøre det enkelt for brukerne å implementere denne funksjonaliteten, helst bare ved å lime inn kode to steder i temaet de bruker til WordPress. Det er dessverre ikke mulig å gjøre dette på en enklere måte. Jeg valgte derfor å overkjøre noen av de globale variablene til WP. De jeg overkjørte gjorde det mulig for meg å bestemme og vise hvilket forfatter arkiv man ser på og ikke bare innholdet, jeg valgte også å overkjøre standardvisningen av innlegg for en bruker, slik at innleggene nå vises basert på en unik spørring som genereres av en måte jeg skrev. Dette ble jeg ikke helt ferdig med, og det finnes i hvertfall to særtilfeller jeg må ta høyde til for å oppnå et tilsynelatende fullt fungerende resultat.

Noe av det som har vist seg å være krevende her er manglende dokumentasjon og eksempler på hvordan lage gode spørringer i WP sitt eget spørringssystem, samt hvordan overkjøre metoder og variabler uten å ødelegge annen funksjonalitet, samtidig som man gjør det enkelt å vise alt innholdet.

Særtilfelle 1: Visning av alle innlegg til første forfatter (Det virker som wordpress her ikke klarer å finne IDen til forfatteren, og viser derfor alle innlegg, siden alle IDer matcher på ingen)

Særtilfelle 2: Hvis artikkelen er opprettet i WP av en bruker, og ikke via REST APIet.

I morgen blir det finpuss, testing, bugfixing og noe videre utvikling av forfatter funksjonaliteten, slik at den i størst mulig grad blir komplett.

## Dagbok 10/2 Michael

I dag har vi planlagt veien fremover, og fått snakket en god del med veilederen vår hos Aptoma. Vi har fått vite at det kan være ønskelig å få på plass mye av basis funksjonaliteten frem til midten av mars. Årsaken til dette er en IPTC konferanse som skal avholdes, der det er et ønske om å vise frem vårt arbeid.

Vi skal derfor prøve å strekke oss etter å få implementert både kategorier og forfattere til den tid, slik at svært mye av den aktuelle funksjonaliteten i NewsML-G2 er implementert i WordPress.

Utover dette har vi oppdaget noen potensielle feil med pluginen vår på tvers av plattformer, dette må vi se nærmere på. Problemet oppstod på en maskin med OS X Yosemite og en svært ny versjon av Apache og PHP. Det er ingen åpenbar grunn til at dette skulle ha feilet. Vi frykter at vi kan møte svært mange feiltilfeller og bugs fremover i utviklingsprosessen, og testing og debugging vil derfor trolig ta en svært stor andel av tiden.

Jeg fikk hjelp fra IPTC angående bruk av IPTC CV i forhold til uppercase av første bokstav i subjectcodes (<name>), da det i den engelske CV har småbokstaver på alt, men på både tysk og fransk har stor forbokstav. Dette ble oppklart, og vi har fått vite at vi kan ta stor forbokstav. Dette ble også implementert i dag.

Vi har også fått utforsket Dr.Publish ytterligere, og laget et eksempel til.

Fra og med i dag tar vi alle fri frem til mandag den 23/2. Vi har forskjellige årsaker til dette, min er eksamenslesning til teknologiledelse.

### **Dagbok 9/2 Petter**

Startet dagen med å kommentere deler av koden. Brukte litt tid etter på fikse subjects pga et de kan ha qcode, men trenger ikke nødvendigvis å ha navn. Måten vi sjekket om navn var satt ga ikke riktig resultat fordi navn er en array som man hentet i en foreach, og siden lengden på navn-arrayen var 0 gikk man aldering inn i denne delen av koden. Det var også noe usikkerhet om vi skal vise subject navn med små eller store bokstaver.

### **Dagbok 09/2 Michael**

Etter kategorier virket til å fungere på onsdag (04/2 – skrev ikke dagbok) har jeg oppdaget mange særtilfeller som feiler. Dette har krevd mye tid, og har derfor tatt en stor del av dagen. Nå ser det ut til at kategorier det fungerer godt.

IPTC sitt kontrollerte vokabulær har i dag ikke vært tilgjengelig og det har vanskeliggjort noe av testingen rundt dette, men vi fant midlertidige løsninger. Vi har også kontaktet en nyhetsgruppe angående IPTC sin CV i forhold til store forbokstaver på Subjects for kategoriene våre. Vi kommer trolig til å sette stor forbokstav på alle kategorier for å holde det pent da vi anså det som rimelig å anta at de fleste vil ønske å ha dette i storforbokstav, dette skal enkelt kunne overkjøres i WordPress med litt kode.

Ut over dette har jeg startet litt med forfattere, og funnet en teoretisk mulig løsning på utfordringen.

### **Dagbok 03/02 Michael**

I dag har jeg fått ordnet en ganske god wrapper klasse rundt en SQLite database, som snart er god nok til bruk. Det viste seg å være svært krevende, da vi må gjøre mange tusen spørringer per import av KnowledgeItems (inneholder masse Qcodes).

Noen hovedpunkter:

- Måtte skrive ganske omfattende kode til wrapperen da vi må gjøre mange tusen spørringer på kjappest mulig tid.
- Det er en transaksjonell grense i SQLite på rundt 500 inserts
- PHP støtter ikke overloading
- Testdreven utvikling har tatt en del tid da omfattende testing er svært viktig, da dette skal være en av flere kjernekomponenter i pluginstrukturen. Det måtte testes på diverse språk og deres tegn (opp mot UTF8), og ulovlige tegn.
- Måtte lage en caching mekanisme der en person som bruker wrapperen og skal gjøre mange spørringer kan kombinere de til en. Dette var ganske krevende og viste seg å ta mye tid, og medføre en rekke problemer.
- Ikke alle tegn er lovlig å lagre i en SQLite database, og jeg har måtte passe nøye på for injections m.m

### Dagbok 03/2 Diego

Dokumentasjonssiden som skal legges sammen med plugin ved lansering er klar til bruk.

Om den skal fylles opp nå eller litt senere gjenstår å se. Men den er vertfall klar.

### Dagbok 02/2 Petter

Bruke store deler av dagen i dag på å parse subject og den tilhørende data-en fra et NewsML document. Det ble til slutt en ganske omfattende array struktur, men vi må bare leve med det siden vi har valgt å løse oppgaven i php. Har også fortsatt å gjøre om de gamle kommentarene til å være på phpdoc formatet samt lagt til noen nye.

### Dagbok 02/2 Michael

I dag klarte jeg å hente ut alle qcodes og tilhørende verdier med en ny parser, som parser NewsML-g2 Knowledge Items. Det var første steg i prosessen med å tatt i bruk større mengder data fra NewsML-G2. Jeg kom også på at vi kan bruke SQLite til å ha vår egen lokale database kun for pluginen vår, da den ikke stiller krav til en databaseserver. Idéen ligger i det å lage en enkel wrapper klasse som gjør transaksjoner med databasen svært enkel. Det vil si `$db->set($key, $value)` og `$db->get($key)`. Denne klassen skal da lagre dataen sendt inn i databasen, som da vil fungere som en avansert assosiativ array. Dette gav meg videre idéen om å lagre serialiserte objekter med qcode og språk som nøkkel, noe som gir oss alle muligheten til å lagre både parametere og qcodes på en svært effektiv måte i databasen, og gir muligheten til ganske mye kul manipulering av data.

Noen fordeler med å ha lokal database:

- Vi kan lagre og hente data som ikke er selve innlegget via pluginen på en svært enkel måte
- Vi kan lagre alle typer objekter med en nøkkel eller to nøkler i databasen, og raskt finne frem.
- Vi kan om tiden tillater det implementere ALLE elementer fra NewsML-G2 som er avhengig av QCodes! (Dette kan potensielt sett være svært gunstig)

Dessverre krever det ganske mye tid å lage på en god måte.

## Dagbok 29/1 Michael

I dag har jeg startet for fullt å se på NewsCodes. Det viser seg å være en ganske stor utfordring.

NewsCodes er en form for metadata som beskriver dataen. I dette tilfellet skal vi bruke det til å finne kategorier. Rettere sagt subjects og mediatopics som standarden kaller det. Disse newscodes er satt i q-codes som dette

```
<subject qcode=»subj:14000000">
```

Den første utfordringen ligger da i å klare å hente ut den tilhørende tekstverdien for den qcode'n fra Controlled Vocabulary (APIet) til IPTC. Dette i seg selv er ikke så ille, men dessverre må vi cache verdiene grunnet spørringsbegrensning hos IPTC. Dette er første utfordring. Andre utfordring ligger da i hvordan vi skal lagre det vi har cachet.

Valget stod mellom å lagre den mottatte XMLen som en fil, og lese fra den hver gang vi trenger å slå opp en qcode. Alternativt sett kan vi lagre dataene i en ny tabell i WordPress databasen. Her gikk vi for å lagre dataene i wordpress databasen. Grunnen til det er at det vil være langt mer effektivt, og det sørger for at vi kan oppdatere og hente informasjon fra databasen enklere dersom ikke all dataene skulle være komplett i det innsendte dokumentet, og vi kan sånn sett å økt opplevd verdi. Det gir oss også muligheten for å støtte selvlagde qcodes, noe som det er flere nyhetsbyråer som bruker, om ikke bare internt.

En annen utfordring er situasjoner med

```
<subject qcode=»INTERN QCODE FOR NYHETSBYRÅ»>
```

```
<sameAs qcode=»medtop:20001065">
```

```
<name xml:lang=»de»>Fußball</name>
```

```
</sameAs>
```

```
</subject>
```

SameAS må ikke alltid være inni subject taggen, og det vil være svært krevende å på en pålitelig måte få tak i all dataen vi trenger for dette formålet. For ikke å snakke om at <name> ikke garantert er den samme, dersom medtop qcoden skulle ha vært endret i IPTC CV. Da har vi valgt å prioritere IPTC CVen sin tekst da det er deres standard som skal følges.

Det bør også bemerkes at det var oppgitt språk i <name> noe som gjør at vi må ta høyde for at det kan være flere språk. Noe som ytterligere øker kompleksiteten.

Dette er sålangt vel og bra, men antall potensielle problemer som følge av dette er vanskelig å forutse.

Så i dag gikk mye av tiden til planlegging av hvordan vi skal håndtere denne problemstillingen, og ikke minst hvordan vi skal utvikle fremover for at det skal være skalerbart og utvidbart for oss ettersom vi får økt kompleksitet i koden vår. En av de største utfordringene vi møter fremover nå er å lage en god skalerbar programvare arkitektur, som kan ta høyde for den store mengden av situasjoner som kan dukke opp.

### **Dagbok 29/1 Petter**

Startet dagen med å endre mange av kommentarene jeg skrev i går til å være på forme gitt av phpdoc slik at personer som bruker en IDE i framtiden kan utnytte Intelligent code completion fullt ut. Vi startet senere å diskutere subject taggen i NewsML. Dette viste seg å være en større jobb en det vi hadde forutsett siden IPTC har et ferdig rammeverk for kategorier som vi må bruke. Vi lurte før på om vi måtte lagre det på fil for å lese fra den, men siden det var vanskelig å oppdatere en fil valgte vi å lagre dataen i WordPress databasen i en separat tabell.

### **Dagbok 28/1 Petter**

Brukte hele dagen på å skrive kommentarer på alle funksjonene i Parse.php siden den ikke var godt nok kommentert. Dette gjør at det blir enklere å bruke senere.

### **Dagbok 28/1 Michael**

I dag lagde jeg en DateParser, som mottar en string i formatet som NewsML-G2 bruker og gir tilbake en streng som kan lagres i MySQL databasen til WordPress. Da det å jobbe med datoer ofte resulterer i rare bugs og feil valgte jeg her å bruke testdreven utvikling. Klassen har nå også en mulighet for å innføre nye tester som skal gjøre eventuell debugging vesentlig enklere. Jeg fikk også implementert bruk av DateParseren i REST APIet og såvidt startet prosessen med å implementere NewsML-G2 Newscodes, noe som virker til å være en ganske omfattende oppgave.

### **Dagbok 27/1 Michael**

I dag fikk vi en innføring av bruk av Aptomas verktøy Dr. Publish for å lage artikler av Stefan. Vi fikk også drøftet en del om videre utvikling av struktur og hvordan det kan gjøres.



I dag fikk vi også noen tips om hvordan vi burde håndtere prosjektdokumentasjonen av en ansatt hos Aptoma, vi valgte derfor å ta kontakt med veileder hos HIOA for å få svar på noen spørsmål vi satt igjen med.

Under møte med Geir Skjevling (Veilederen vår på HIOA) fikk vi vite mer om hvordan vi skal føre dokumentasjon, og valgte på grunnlag av den informasjonen å fortsette som før, da det ikke virket så viktig å fokusere mye på fremtidig planlegging og milepæler, men heller god dokumentasjon av arbeidet som blir gjort, og prosessen rundt det.

Jeg har i dag fått tilpasset koden min til å reflektere endringene som Petter gjorde i går og i dag. Jeg fikk også oppdatert debugging til å fungere godt, slik at koden ble langt bedre strukturert for meg selv og andre utviklere. Konvertering av dato i ISO8601 formatet til WordPress sin database er nå påbegynt, utfordringen med dette er at det skal være lagret på to forskjellige måter, en av de er GMT.

### **Dagbok 27/1 Petter**

I dag har jeg lagt til statuskode for overføring i forbindelse med manglende data, satt publisering status i forbindelse med embargo dato og lagt til tager fra en nyhetssak. Siden alle metodene vi har i parse fila er statiske tok det litt tid å finne ut hvordan jeg skulle sette satskodene, men jeg klarte det til slutt ved å sende meta og post arrayene slik at de kunne gjennomgås og endres før de ble satt inn i den overordnede strukturen.

### **Dagbok 27/1 Diego**

Har startet på websiden som skal inneholde veiledning

for implementering av vår plugin og dokumentasjon for de som måtte ha lyst til å videreutvikle prosjektet.

Er ferdig med å sette opp en ryddig meny i AngularJS.

Det viktigste som mangler som jeg ønsker å bygge inn før

det blir skrevet innhold til siden, er en undermeny som

fyller seg selv med nye menyvalg ettersom man legger til innhold i dokumentasjonen.

### **Dagbok 26/1 Petter**

Brukte dagen på å skrive om parse.php slik at den kan ta i mot en newsMessage som inneholder feler newsItem. Den nyeste versjonen inneholder ikke creator og contributor pga vi ikke er sikre på hvordan vi skal lagre det i wordpress databasen

### **Dagbok 26/1 Michael**

I dag har jeg fikset mye på HTTP responsene som REST APIet gir tilbake, og fikset en debug parameter som lar brukeren ved et API kall bruke våre debugmeldinger i tillegg til standard HTTP responser.

Frontend biten på WordPress admin er også forenklet ytterligere.

### **Dagbok 22/1 Michael**

I dag gikk dagen til å ferdigstille forprosjektrapporten.

Forprosjektrapport

### **Dagbok 21/1 Petter**

Har gjort ferdig metadaten i forbindelse med creator og contributor. Måtte oppdatere php for å få alt til å fungere. Har fått en bedre forståelse av xml namespaces i forbindelse med xpath

### **Dagbok 21/1 Michael**

I dag møtte vi et problem med PHP versjoner, der Petter brukte PHP 5.3.x brukte både jeg og Diego PHP 5.5.x og i PHP versjon 5.4 begynte de med array dereferencing, noe som gav oss ytterligere en faktor å ta høyde for med utvikling av plugin som skal brukes av mange.

Jeg har i dag sørget for at API nøkkelen lages av på en kryptografisk sikker måte og lagres i wordpress databasen i motsetning til å være hardkodet som den var i starten. Det er også mulig å generere en ny API nøkkel i grensesnittet til pluginen.

I dag har jeg også sett litt mer på hvordan man lager et godt REST (Representational State Transfer) API, og hvordan man returnerer korrekte og gode HTTP headers ved bruk av APIet.

Mye av tiden har gått med i å jobbe sammen med Diego på forprosjektrapporten.

### **Dagbok 21/1 Diego**

Versjonshåndtering er satt opp.

Forprosjektrapport er snart klar for levering. En god uke, etter en noe rolig andreuke.

### **Dagbok 20/1 Michael**

I dag føler jeg at vi har fått gjort mye. REST APIet har nå mulighet til å oppdatere en eksisterende nyhetsartikkel ved å bruke to parametere fra NewsML-G2 standarden, version og

GUID. Vi har fått startet å lagre vår egen metadata i WordPress databasen via rammeverket til wordpress, dette gjør at vi får implementert langt mer data om hver artikkel til wordpress. I tillegg har vi fått satt opp et versjonshåndteringsverktøy (git) i dag.

### **Dagbok 20/1 Petter**

I dag har jeg laget en array for metadata som skal sendes samme med post-arrayen og har også begynt på user-arrayen. Vi har i tillegg satt opp Git for prosjektet slik at det vi ikke trenger å redigere på de samme filene.

### **Dagbok 19/1 Petter**

I dag har jeg fått flere eksempler på NewsML, som har ført til at jeg har oppdatert metodene for å hente ut informasjon ettersom at det er flere måter å formatere informasjonen på. Det har også blitt mer klarhet i at vi skal ikke ha så mye ekstern informasjon slik som forfatter og liknende data.

### **Dagbok 19/1 Michael**

I dag fikk vi snakket med veilederen vår Stefan om oppgaven og fått økt klarhet i hva som skal gjøres. Vi fant ut at det var mer aktuelt å sette opp pluginen til å brukes via et REST API. Det ble også klart at vi skulle få mulighet til å bruke et av Aptoma sine redskap Dr. Publish til å sende inn artikler til vårt API. Derfor tok jeg i dag å endret grunnstrukturen rundt kommunikasjon inn og ut av wordpress databasen til å reflektere dette, og på slutten av dagen var det mulig å sende en HTTP POST request til APIet med et lignende utfall som AJAX metoden jeg satte opp på et tidligere tidspunkt.

Det er fortsatt mye som må gjøres på dette tidspunkt med REST Apiet, eksempelvis feilhåndtering ved bruk av HTTP Headers og hvordan vi skal lagre mer av informasjonen i wordpress sin database. Foreløpig lagres bare innholdet og tittel.

### **Dagbok 12/1 Michael**

I dag har vi startet med et grunnleggende rammeverk for videre utvikling av funksjonalitet. Dette rammeverket benytter seg av javascript med biblioteket JQuery. Vi møtte en utfordring i forhold til ajax kall etter fillesning, men har funnet en mulig løsning. Det blir spennende å se hvordan dette implementeres i morgen.

### **Dagbok 9/1 Diego**

Etter første uken var vi kommet godt i gang. Første steg var å kunne trekke ut enkelte deler av en XML og importere denne til WordPress-databasen. Dette har vi fått til. I tillegg har vi startet

på metodene som skal kunne brukes til å trekke ut enkelte deler av XML-filen og for håndtering av ulike feil.

### **Dagbok Petter 8/1**

Den første avgjørelsen som ble tatt i dag var at vi ikke skulle ha database for mellomlagring. Det ble i stede bestemt at vi skulle prøve sette tittel, innhold og slug-line rett inn i WordPress siden det var den informasjonen som er felles for de to formatene. Michael satt opp den grunnleggende applikasjonen som fikk informasjonen fra XML fila til WordPress, Diego jobbet med å lage en feil-logg og hadde ansvaret for å hente informasjonen i XML dokumentet med DOMXPath. Det ble noe komplikasjoner mellom XPath og XML namespace, men vi klarte å løse det og til slutt klarte vi legge inn data i WordPress-databasen slik at det kunne postes som et innlegg.

### **Dagbok Petter 7/1**

I dag startet vi dagen med å lese om NewsML-G2 og WordPress-databasen for å bedre se sammenhengen mellom de to. Hele perioden før lunsj gikk å lesing. Senere på dagen hadde vi en lang diskusjon hvor vi prøvde å komme fram til den beste løsningen for å få informasjonen XML til WordPress. Ideen vi kom opp med var ha en database for mellomlagring der tabellene var XML tagger og radene i tabellene attributter eller innholdet i taggene. Vi innså at denne databasen kom til å bli ganske omfattende, så vi prøvde å finne en måte å konvertere XML til en relasjonsdatabase. Vi fant ingen løsning med en gang og bestemt oss for se mer på det dagen etter.

### **Dagbok Petter 6/1**

Prosjektstart. I starten viste vi ikke helt hva prosjektet gikk ut på så det var litt usikkerhet rundt valg av teknologi. Det ble på et tidspunkt vurdert å gjøre om NewsML dokumentet til objekter (muligens i PHP), og vi lurte også på om vi skulle bruke XSLT siden mye av WordPress databasen var i HTML. Vi kom til slutt fram til å vi skulle bruke PHP for å lese fra XML dokumentet med DOMXPath

### **Prosjektside**

30. november 2014 av Michael

En enkel prosjektside har i dag kommet på plass. Det er nå mulig for alle prosjektdeltagere å publisere innlegg i prosjektdagboka og oppdatere statusrapporten.