

Een Fedora-depot voor architectuurarchieven

Rapportering resultaten 1^e fase

Auteur:

Annelies Nevejans
i.s.m. Jeroen Bekaert

Antwerpen, 14 juni 2010

Met steun van de
Vlaamse overheid



Centrum Vlaamse Architectuurarchieven

Inhoud

1. Inleiding

1.1 Het CVAa

Het Centrum Vlaamse Architectuurarchieven (CVAa) werd in november 2003 opgericht onder de koepel van het Vlaams Architectuurinstituut (VAi). Het CVAa is een cultureel thema-archief voor architectuurarchieven.

Het CVAa wil kennis, internationale expertise en best practices over archieven van de gebouwde omgeving en architectuurarchieven in het bijzonder verzamelen, ontwikkelen en verspreiden in Vlaanderen. Onze doelgroep zijn in de eerste plaats de bewaarinstellingen die verspreid over heel Vlaanderen kleine of grote collecties architectuurarchieven beheren. Daarnaast staan we ter beschikking van archiefvormers, –gebruikers en diverse overheden die met het geheugen van de ruimte in aanraking komen.

Het CVAa functioneert als een expertise- en coördinatiecentrum en bouwt geen eigen collectie architectuurarchieven op.

1.2 Nood aan expertise over digitale architectuurarchieven

Sinds de late jaren 1980 produceert de gemiddelde praktijk van een architect heel wat digitale objecten. Deze digitalisering van het architectuurontwerp is de afgelopen jaren nog toegenomen. Jonge bureaus ontwerpen vandaag soms nagenoeg uitsluitend met behulp van gespecialiseerde software.

Een archiefbeleid met toekomstvisie voor het brede thema van de architectuur en stedenbouw kan niet zonder kennis over en aangepaste instrumenten voor een duurzaam beleid voor digitale archieven over de gebouwde omgeving.

Omdat de nood hoog is wil het CVAa kennis ontwikkelen over de archivering van digitale architectuurarchieven (met zowel 'digital born' als gedigitaliseerde data). Belangrijke vragen hierbij zijn:

- Wat zijn de grootste moeilijkheden bij het ontwikkelen van een digitaal depot?
- Welke aspecten aan het ontwikkelen van een digitaal depot zijn specifiek voor architectuurarchieven? Welke aspecten zijn generiek voor alle sectoren?
- Welke rol moet het CVAa op landelijk en internationaal vlak spelen in de archivering van digitale architectuurarchieven?

Om antwoorden op bovenstaande vragen te kunnen formuleren en daarbij niet te vervallen in een eindeloze en intensieve studie van andere voorbeelden en onderzoeken – ervaring is de beste leerschool – werd beslist een beperkte testomgeving te ontwikkelen. Hierbij was het in de eerste plaats de bedoeling om antwoorden te vinden op de vraag wat de moeilijkheden zijn bij het ontwikkelen van een digitaal depot.

De testomgeving werd zo gekozen dat ze met een beperkt budget en binnen een gelimiteerd tijdsbestek kon worden ontworpen. In dit kader werd geopteerd om als use case een relatief eenvoudige en duidelijk afgebakende digitale collectie te nemen, met name de digitale collectie van de eigen vereniging (VAi). De eenvoud van deze collectie biedt het voordeel dat de complexiteit van een aantal vraagstukken (zoals de definitie van metadata standaarden en content modellen, en de ondersteuning en preservatie van bestandsformaten (zie verder)) tot een minimum wordt herleid. Het is met dit project immers niet de bedoeling om een volledige wetenschappelijke discussie te voeren omtrent elk aspect eigen aan een digitaal depot; maar veeleer een poging om een aantal vraagstukken die zich stellen bij de ontwikkeling digitale depots in kaart te brengen. Door middel van verdere studies en projecten kan vervolgens elk van deze vraagstukken meer uitgediept worden.

2. Selectie van een softwaresysteem

2.1 Voorwaarden testcase

Met een beperkt budget (10.000 € voor fase I) en met een beperkte timing (6 maanden) moest een testomgeving worden ontwikkeld waarbinnen kon worden getest en geëxperimenteerd, om zodoende verdere ervaring op te doen in het opzetten en beheren van digitale architectuurarchieven.

Het e-depot moest bij voorkeur aan het OAIS-model beantwoorden.¹ Dit gestandaardiseerde model biedt een theoretisch referentiekader voor het ontwikkelen van archiefbeheersystemen en legt vijf sleutelprocessen vast die voorzien moeten zijn in een dergelijk systeem. Daarnaast moesten ook andere internationale interoperabiliteitsstandaarden geïntegreerd worden met het oog op de uniforme uitwisseling van metadata en digitale objecten met andere digitale depots en archieven. We denken hierbij onder meer aan OAI-PMH² voor het delen van metadata, of OAI ORE³ voor het uitwisselen en hergebruiken van digitale objecten. Deze interoperabiliteitsstandaarden zijn cruciaal om ons digital depot, in de toekomst, te integreren in zogenaamde depot- of archief federaties.

Bijkomende voorwaarde was dat we gezien het beperkte budget en de beperkte timing specifieke (op maat gemaakte) ontwikkeling zoveel mogelijk willen vermijden. Ook het onderhoud nadien moet immers tot een minimum beperkt worden.

Gezien deze beperkingen leek het raadzaam zoveel mogelijk gebruik te maken van kant-en-klare softwaresystemen. Ook het gebruik van open source software leek ons aangewezen. Niet alleen het beperkte budget was hiervoor een drijfveer. Een open source systeem wordt veelal ontwikkeld binnen een zogenaamde 'community', een netwerk van gebruikers. Binnen een dergelijke community van gebruikers kan men steunen op de reeds ontwikkelde expertise en kunnen ook gebruikerservaringen worden gedeeld.

¹ OAIS staat voor Open Archival Information System. Het is een abstract referentiemodel dat als basis kan dienen voor de ontwikkeling van archiefbeheersystemen. <http://www.classic.ccsds.org/documents/pdf/CCSDS-650.0-B-1.pdf>

² OAI-PMH staat voor Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Het is een eenvoudig HTTP en XML gebaseerd protocol voor het delen van metadata tussen digitale 'repositories' op een gestandaardiseerde manier. <http://www.openarchives.org/OAI/openarchivesprotocol.html>

³ OAI-ORE staat voor Open Archives Initiative Object Reuse and Exchange. OAI-ORE definieert standaarden voor de beschrijving en uitwisseling van aggregaties van web-resources. <http://www.openarchives.org/ore>

2.2 Keuze van de software

Tijdens het vooronderzoek werden drie bestaande open source depot-systemen nader bekeken: DSpace, Fedora Commons en EPrints. Het is hierbij belangrijk op te merken dat elk van deze softwaresystemen initieel werd ontwikkeld als een 'institutional repository' systeem voor het ontsluiten van wetenschappelijke publicaties van onderzoeksinstituten.

2.2.1 DSpace

DSpace is een *institutional repository system*, ontwikkeld door MIT Libraries i.s.m. Hewlett Packard.

Op het eerste zicht leek DSpace de eenvoudigste oplossing. Deze repository software kan *out of the box* geïnstalleerd worden op een eenvoudige Windows of Linux server, en kan rekenen op een zéér uitgebreide community van gebruikers. Bovendien beschikt DSpace over een ingebouwde web-interface.

Helaas ondersteunt DSpace geen complexe – hiërarchische gestructureerde – data modellen; en ook de integratie van metadatastandaarden (op Dublin Core na) en versie-ondersteuning blijken problematisch. ⁴

2.2.2 EPrints

Eprints, de voorloper van DSpace, is een *institutional repository system* ontwikkeld aan de University of Southampton. Het is een erg eenvoudig platform, maar het is, net zoals DSpace, in de eerste plaats gericht op het ontsluiten van wetenschappelijke publicaties van onderzoeksinstituten, zonder hierbij de noodzakelijk aandacht te geven aan de kernelementen en kernprocessen die worden vermeld in het OAIS model.⁵

2.2.3 Fedora Commons

Fedora Commons is een *repository system* ontwikkeld door Cornell University en de University of Virginia. Fedora staat voor Flexible Extensible Digital Object Repository Architecture. Fedora Commons richt zich niet enkel tot het toepassingsdomein van institutionele archieven voor wetenschappelijke publicaties, maar wordt ook aangewend als repository systeem voor het bewaren van gedigitaliseerde erfgoed- en museumcollecties.

Fedora Commons vraagt meer installatie- en programmeerwerk dan DSpace, maar laat ook meer flexibiliteit toe om op een vrij eenvoudige manier achteraf nog modules bij te bouwen met nieuwe functionaliteiten. Een gebruiksvriendelijke front end op de repository moet bijkomend worden geprogrammeerd. Ook de datamodellen moeten afzonderlijk ingesteld worden. ⁶

Opmerking: Ondertussen is er een samenwerkingsverband opgestart tussen Fedora Commons en MIT/DSpace: DuraSpace

2.3 Keuze: Fedora

2.3.1 Voordelen

We sommen hier enkele voordelen van de Fedora software op:

⁴ <http://www.dspace.org>

⁵ <http://www.eprints.org>

⁶ <http://www.fedora-commons.org>

- Het is een modulair systeem. Als er bijkomende functionaliteiten nodig zijn in het e-depot, dan kunnen die op een vrij eenvoudige manier toegevoegd worden.
- De repository is opgebouwd uit self-contained objects. Daarmee wordt bedoeld dat alle essentiële eigenschappen van het object in één xml bestand kunnen worden gebundeld.
- Fedora biedt een digitaal objectmodel dat complexe objectstructuren en een verscheidenheid aan datastromen ondersteunt.
- Het datamodel van Fedora laat toe om versies van (deel)objecten en datastromen te beheren. (In het OAIS model noemt dit Provenance Information).
- Elk object in de repository krijgt een unieke identifier.
- Fedora laat toe om zelf "services" en "methodes" te definiëren en te ontwikkelen waarmee je objecten die in Fedora worden bewaard, in allerlei varianten kunt opvragen. Mogelijke voorbeelden zijn een vertaalservice waarbij een document dat wordt bewaard in het digitale depot wordt vertaald in een taal naar keuze; of een service waarbij een gearchiveerde foto (op hoge resolutie) wordt omgezet naar een zgn. 'thumbnail' of preview-foto op lage resolutie; of een preservatie-service die het mogelijk maakt om bepaalde tijdelijke bestandsformaten om te zetten naar meer robuuste archiveringsformaten (bijvoorbeeld JPEG naar JP2000); enzovoort.
- Er zijn API's (Application Programming Interface) voor het beheren en ontsluiten van objecten.⁷
- Fedora ondersteunt de eerder vermelde standaarden OAI-PMH en OAI-ORE.
- Fedora is een modulair systeem dat eenvoudig verder kan worden uitgebreid met andere componenten.

2.3.2 Nadelen

We zetten de belangrijkste nadelen op een rijtje:

- Fedora bevat alleen een *storage layer* en *middleware* (de *repository*). Een *front end* of interface om objecten uit het depot te bekijken en te bewerken is niet voorzien en moet dus bijkomend ontwikkeld en/of geprogrammeerd worden.
- Er zijn geen pasklare instrumenten om het Fedora datamodel toe te passen op een bepaalde dataset. De gebruiker moet zelf regels en relaties definiëren.

⁷ Een API is een Application Programming Interface. Het is een interface tussen twee computerprogramma's die er voor zorgt dat beide met (bepaalde onderdelen van) elkaar kunnen communiceren.

3. Use case: invoer van objecten Jaarboek Architectuur Vlaanderen

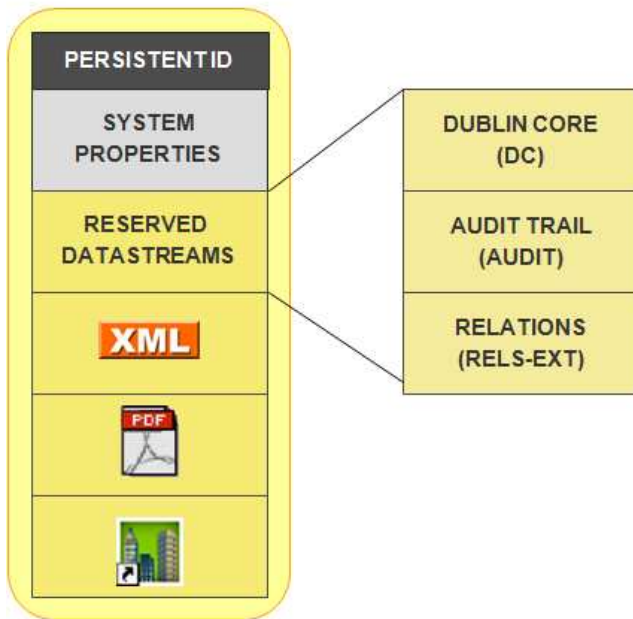
3.1 Inleiding

Als volgende stap in de ontwikkeling van de testcase moesten we uiteraard digitale objecten hebben die we wilden invoeren in het op te zetten systeem. De keuze viel op de digitale objecten die het VAI ontvangt naar aanleiding van het Jaarboek Architectuur Vlaanderen. Dit is een tweejaarlijkse publicatie over hedendaagse architectuur in Vlaanderen. Via een oproep aan alle actieve ontwerpers wordt gevraagd om een portfolio in te sturen van maximum drie projecten. Elke portfolio bestaat uit een reeks digitale bestanden: een inschrijvingsformulier met contactgegevens van de deelnemer, een projectformulier met gegevens over het bouwproject, een concepttekst over het bouwwerk, enkele ontwerptekeningen en enkele foto's van het uitgevoerde project.

De collectie is momenteel klein, maar heeft een hoge culturele waarde aangezien ze een doorsnede bevat van de hedendaagse architectuurproductie in Vlaanderen. Deze collectie wordt dan ook veelvuldig geraadpleegd, zowel intern door medewerkers bij voorbereiding van andere projecten als extern door onderzoekers, de pers,...

3.2 Objectstructuur

Het Content Model in Fedora legt vast hoe de objecten zijn samengesteld. Alle data moeten immers op een gestructureerde manier worden weergegeven. Op die manier kunnen de data nadien op uniforme wijze worden beheerd en ontsloten. Ons Fedora depot hebben we opgebouwd uit Collecties en Objecten. Beiden zijn gelinkt aan elkaar door middel van relaties (zie verder). Elk object is op zijn beurt samengesteld uit datastromen die ofwel content (data) bevatten ofwel metadata hierover. Schematisch ziet dit er als volgt uit:



Men kan het digitale object vergelijken met een dossier in een analogo archief. Net zoals het object verschillende datastromen bevat, zo omvat een analogo dossier ook verschillende stukken. Uiteraard kan een object ook slechts één stuk bevatten en wordt het stuk/datastroom individueel beschreven.

Een object heeft naast een unieke Digital Object Identifier (unieke identificatiecode) een onbeperkt aantal datastromen waarvan er een aantal gereserveerd zijn voor het vastleggen van systeem-eigenschappen, metadata (Dublin Core), "audit-trail" (of versie-gerelateerde eigenschappen) en relaties.

- Metadata kan in Fedora in om het even welk XML formaat worden vastgelegd. Fedora stelt wél het gebruik van Dublin Core als minimale vereiste (Dublin Core wordt immers vereist door OAI-PMH). Andere metadatastandaarden kunnen echter toegevoegd worden. Een object kan dus door meerdere metadata sets beschreven worden.
- De gereserveerde audit-trail datastroom registreert automatisch elke wijziging die aan het object wordt aangebracht (bijv. toevoegen, wijzigen of verwijderen van datastromen). Een laatste gereserveerde datastroom is die waarin relaties met andere objecten wordt vastgelegd.
- Naast deze gereserveerde datastromen, kunnen nog een onbeperkt aantal datastromen worden toegevoegd. Het kan hierbij zowel gaan om bijkomende metadata sets als om data.

De collectie van het Jaarboek Architectuur Vlaanderen is relatief beperkt en beheersbaar. Bij elke oproep worden zo'n 350 projecten ingezonden. We beschouwen in dit geval een ingezonden project als een uniek Object. Het VAI vraagt de deelnemende ontwerpers foto's, bouwtekeningen en een projecttekst in te sturen. Elk Object bestaat dus uit een bundel van meervoudige datastromen die een logisch inhoudelijk geheel vormen. Naast de Reserved Datastreams met metadata worden dus de foto's, bouwtekeningen en projecttekst als datastromen aan het Object toegevoegd.

De variatie in bestandsformaten van deze datastromen is beperkt. De bestandsformaten voor deze documenten zijn bepaald op JPEG voor foto's en PDF voor bouwtekeningen en projectteksten. In het productiestadium worden bijkomende documenten opgevraagd, zoals foto's in hoge resolutie in TIFF, bouwtekeningen in EPS en projectbeschrijvingen in DOC.

In deze testcase werd de collectie ontsloten tot op objectniveau. De beschrijvende metadata werden toegekend op het niveau van het object (cf. Dossier in een analogo archief) en niet op het niveau van de individuele datastromen (cf. Stukken in een analogo archief). Het is echter wel mogelijk om ook op het niveau van de individuele datastromen metadata toe te voegen.

In het totaal bestaat de collectie uit zo'n 750 objecten die samen zo'n 11.800 datastromen bevatten, met een totale omvang van ca. 50 GB.

Voorbeeld van een Object uit het VAI-eDepot:

MetaData – Dublin Core

| | |
|--------------------|--------------------------------|
| title | Appartementsgebouw Hemelsoet |
| creator | Architettura |
| subject | woning |
| description | Bredestraat 34, Oostakker-Gent |
| publisher | Vlaams Architectuurinstituut |
| date | 2002 |
| identificer | vcol:2006-48 |
| identificer | 227 |
| language | nl |

Datastromen

| Pid | Title | Created |
|------------------------|------------------------------------------------------------------------------------------|--------------------------|
| <i>vbin:2006-48-1</i> | architettura_ bert callens_1.jpg | 2009-09-05T12:41:42.001Z |
| <i>vbin:2006-48-2</i> | architettura_ bert callens_2.jpg | 2009-09-05T12:41:42.001Z |
| <i>vbin:2006-48-3</i> | architettura_ bert callens_3.jpg | 2009-09-05T12:41:42.001Z |
| <i>vbin:2006-48-6</i> | architettura_plannen_appartementsgebouw hemelsoet_01_inplantingsplan.pdf | 2009-09-05T12:41:42.001Z |
| <i>vbin:2006-48-7</i> | architettura_plannen_appartementsgebouw hemelsoet_02_0 en +1.pdf | 2009-09-05T12:41:43.001Z |
| <i>vbin:2006-48-14</i> | architettura_projecttekst_appartementsgebouw hemelsoet.pdf | |

3.3 Relaties tussen objecten

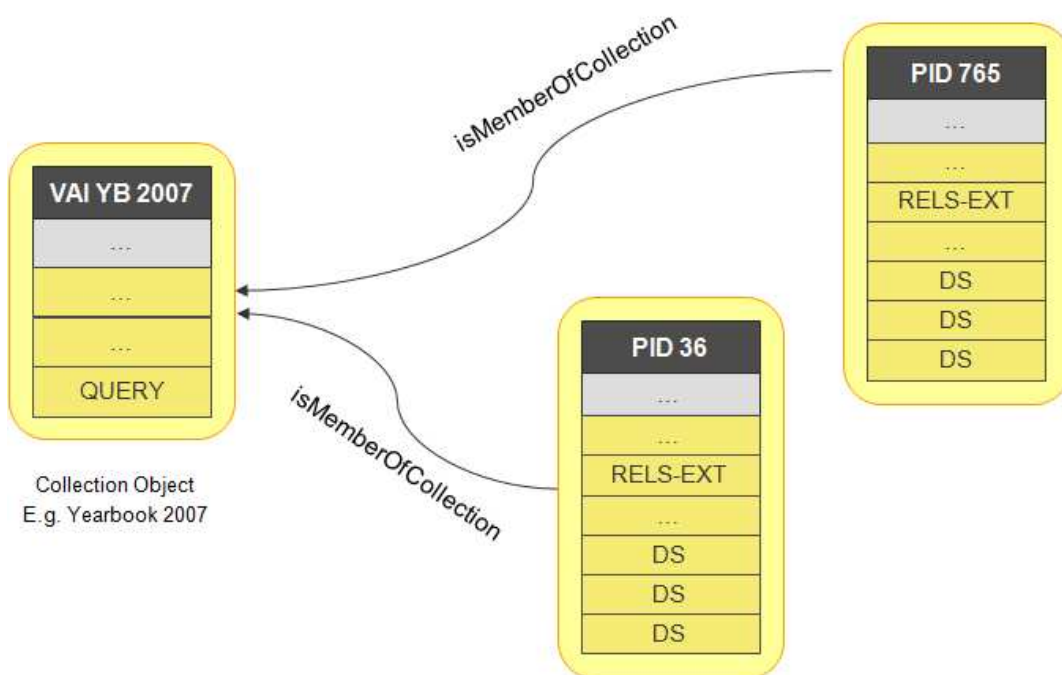
In elk object kunnen vervolgens relaties worden vastgelegd met een onbeperkt aantal andere objecten. Deze relaties worden bewaard als één van de metadata-datastromen binnen het object zelf (zie hoger). Op die manier

kunnen hiërarchische structuren van (geneste) objecten worden opgebouwd. Ook de relatie tussen objecten en zgn. 'collectie-objecten' kan op deze manier worden vastgelegd. Deze relaties worden uitgedrukt door middel van RDF (Resource Description Framework).⁸

Schema:

Alle objecten (lees: ontwerpprojecten/portfolio's) die voor een zelfde jaarboek-oproep werden ingezonden kregen dezelfde relatie mee tot een bepaald collectie-object in Fedora. De collectie heet bijvoorbeeld *Jaarboek_Architectuur_2006-2007*. Alle objecten die hiertoe behoorden kregen de relatie 'isMemberOfCollection *Jaarboek_Architectuur_2006-2007*' mee. Op deze manier worden de objecten dus gelinkt met de verschillende collectie-objecten in Fedora en zijn ze ook opnieuw te bevragen in het e-depot.

Voorbeeld van een isMemberOfCollection statement in RDF:



```
<rdf:RDF>
  <rdf:description rdf:about="info:fedora/vcol:2006-48">
    <fedora:isMemberOfCollection rdf:resource="info:fedora/Jaarboek_Architectuur_2006-2007">
      </fedora:isMemberOfCollection>
    </rdf:description>
  </rdf:RDF>
```

⁸ RDF staat voor Resource Description Framework. Het is een W3C webstandaard om gegevens voor te stellen en uit te wisselen en is opgebouwd uit drie elementen: subject - predicaat - object. Het subject is de *resource* die beschreven wordt. Het predicaat is welk kenmerk of aspect van die *resource* beschreven wordt. Het object tenslotte is wat de waarde van dat kenmerk is. Bijvoorbeeld: foto X – heeft fotograaf – Jan Pieters. Het onderliggende datamodel kent meerdere syntaxes, bijvoorbeeld RDF/XML, N3, Ntriples, enz.

4. Bouwstenen van het VAI/CVAa e-depot

4.1 Server

Uiteraard is naast de keuze van de software ook de hardware belangrijk. De hardware moet immers zo onkwetsbaar mogelijk zijn om de beschikbaarheid en de leesbaarheid van de bestanden te garanderen op lange termijn. Voor het e-depot VAI/CVAa werd gekozen voor het huren van een dedicated linux server met backup.

4.2 Fedora storage layer

De samengestelde objecten in Fedora zijn opgebouwd uit één of meerdere content items (cf. supra) die tot eenzelfde object behoren. Die content items worden afzonderlijk opgeslagen in de Fedora storage layer. In dit project maakten we gebruik van de relationele database MySQL. Fedora ondersteunt evenwel tal van andere storage systemen.

4.3 Fedora repository

Deze layer bevat de effectieve Fedora repository functionaliteit.

4.4 Islandora module

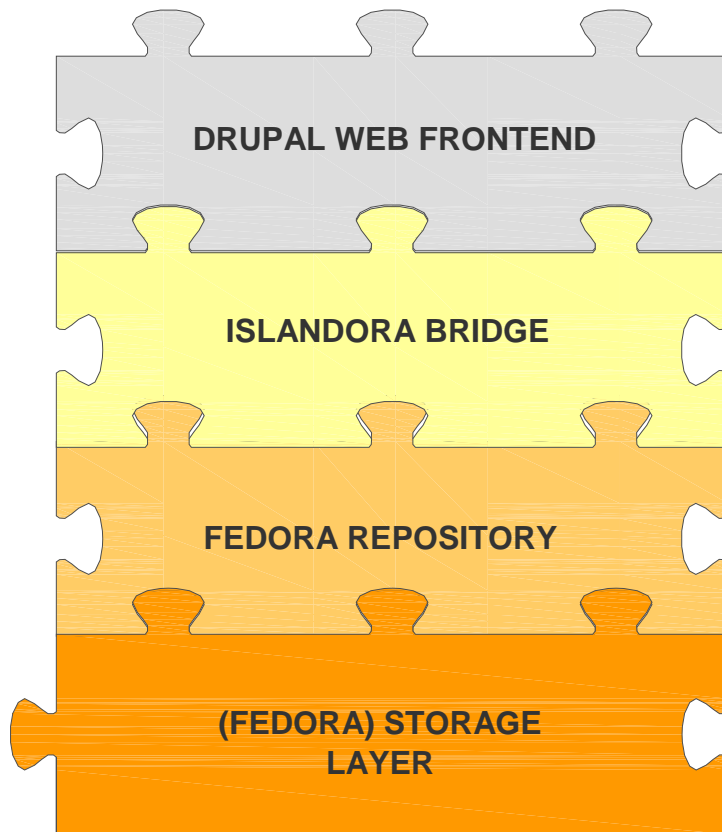
Om de objecten in de Fedora repository te zien en te bewerken wordt in het e-depot gebruik gemaakt van een Drupal webinterface. Opdat Drupal zou kunnen opereren als een front end module voor de Fedora Repository was het noodzakelijk om de Islandora module te installeren als intermediaire module. Islandora vormt de schakel tussen onze uiteindelijke Drupal Front end en het Fedora archiefsysteem.

4.5 Drupal web frontend

Als front end werd voor Drupal gekozen. Dit is een open source content management systeem voor websites.

Net als Fedora, kan ook Drupal rekenen op een stevige gebruikerscommunity.

4.6 Schem



5. Conclusies

5.1 Algemeen

Het ontwikkelen van het e-depot heeft er voor gezorgd dat de medewerkers beter vertrouwd zijn met de concepten, terminologie en eigenschappen van digitale objecten en e-depots. Op die manier kunnen ze de sector beter vertegenwoordigen en opent het ook bepaalde wegen naar samenwerkingsverbanden met andere/grotere organisaties. In plaats van aan de zijlijn te staan als toeschouwer, kunnen we nu meediscussieren en samenwerken met andere partners. Zo maakt het CVAa ondertussen onder meer deel uit van de stuurgroep van het project Archipel.⁹

De focus lag bij de ontwikkeling in de eerste fase op het structureren, bewaren en beheren van objecten en bestanden. Het duurzaam bewaren van de data is namelijk een eerste voorwaarde voor het ontsluiten en leesbaar houden ervan op lange termijn.

Aangezien het om een testomgeving ging, en het niet de bedoeling was om een publieke toegang te creëren, werd de front end slechts in beperkte mate ontwikkeld. Bepaalde mogelijkheden voor (publiek) gebruik van de bestanden werden wel voorzien in de opbouw van het e-depot, maar deze functionaliteiten werden niet verder ontwikkeld.

Uit de use-case werden niet meteen antwoorden ontwikkeld op alle prangende vragen die specifiek zijn voor architectuurarchieven, zoals bijvoorbeeld de duurzaamheid van bepaalde bestandsformaten voor grafische bestanden of metadatasets voor architectuurarchieven. Wel werd een aanzet gegeven om deze problematiek in kaart te brengen. In de testfase werd er bewust gewerkt met een beperkt aantal minder complexe bestandsformaten om de complexiteit in deze fase te beperken.

5.2 Toekomstperspectieven

Op basis van de eerste ontwikkelingen van het e-depot wordt momenteel de tweede ontwikkelingsfase voorbereid. Deze zal over een langere termijn lopen dan de eerste fase omdat er complexere vragen en problemen inzake digitale architectuurarchieven zullen worden behandeld. De verschillende stappen in de tweede fase werken naar concrete antwoorden op vragen over aspecten zoals metadata voor architectuurobjecten, datamodellen voor architectuurobjecten, versiebeheer, identificatie van objecten, interoperabiliteit van architectuurarchieven, beheer van bestandsformaten, copyright, migraties,... Deze vragen komen zowel vanuit de ervaringen uit de eerste ontwikkelingsfase als van beheerders van digitale (architectuur)archieven. Niet al deze vragen zijn specifiek voor de context van architectuurarchieven. De antwoorden hierop zullen dan mogelijk ook vanuit andere projecten rond digitale archieven (nationaal en internationaal) naar voor komen.

Voor het bepalen van de focus van de tweede fase wordt een platform georganiseerd met vertegenwoordigers uit de sector architectuurarchieven en digitaal archiefbeheer.

⁹ Voor meer informatie over het project Archipel, zie website <http://www.archipel-project.be>.