**///** OSLO LDES: Thematische werkgroep 3

**////////////////////////////////////////////////////////////////////////////////////////////////////////**

Datum: 17/03/2022

Locatie:  Teams meeting (virtueel)

Moderators: Thijs Hegge, Arne Scheldeman, Pieter Colpaert, Geert Thijs, Bert Van Nuffelen

**////////////////////////////////////////////////////////////////////////////////////////////////////////**

**Aanwezigen**

* Digitaal Vlaanderen
  + Thijs Hegge
  + Annelies De Craene
  + Geert Thijs
  + Arne Scheldeman
  + Filip Borloo
  + Dwight Van Lancker
  + Bert Van Nuffelen
  + Simon Claus
  + Veerle Beyaert
  + Adriaan Walpot
  + Geraldine Nolf
  + Justine Ottevaere
* IMEC
  + Pieter Colpaert
  + Philippe Michiels
* NSX
  + Frans Verstreken
  + Philip Huysmans
* Vlaams Instituut Voor de Zee (VLIZ)
  + Marc Portier
* Atos
  + Patrick Michels
* Konsolidate
  + Christophe Cop
* Esri Belux
  + Gert Bergers
  + Wim Driessens
* MyCSN NV
  + Jan Geukens
* RedPencil
  + Niels Vandekeybus
* District09
  + Maarten Segers
* Cegeka
  + Joke Lybaert
* Archers NV
  + Joost Vandebosch
* AWV
  + Joris Hoogeboom
* Departement Mobiliteit en Openbare Werken
  + Inge Van Hooydonk
* Andere
  + Alexander Kerckhofs

**Agenda van de werkgroep**

|  |  |
| --- | --- |
| 09u00 – 09u10 | **Welkom & introductie** |
| 09u10 - 09u20 | **Overzicht: wat hebben we gedaan in de vorige werkgroep** |
| 09u20 – 10u35 | **Aanpassingen aan het datamodel** |
| 10u35 – 10u45 | **Pauze** |
| 10u45 – 11u45 | **Metadata** |
| 11u45 –  12u | **Algemene feedback & Volgende stappen** |

1. **Inleiding**

*[We verwijzen naar de slides voor meer informatie.]*

1. **Overzicht: wat hebben we gedaan in de vorige werkgroep?**

*[We verwijzen naar de slides voor meer informatie.]*

Er werd een overzicht gegeven van de activiteiten/discussies uit de tweede thematische werkgroep, dewelke georganiseerd werd op 17 februari 2022:

* Herhaling eerste thematische werkgroep:
  + Herhaling OSLO & LDES
  + Introductie tot SEMIC
  + Eerste poging tot een datamodel

* Versiebeheer en relaties tussen LDESs
  + OSLO Observatie
  + OSLO Generiek: Versie
  + Relaties tussen LDESs: eventually consistent

* Eerste poging tot een datamodel
  + Disambigueren van ‘View’
  + Volledigheid van een LDES
  + Paginatie: hoe nemen we de elementen mee op een node?

De feedback en opmerkingen die werden ontvangen gedurende de tweede thematische werkgroep werden meegenomen en vormden de basis voor de derde thematische werkgroep. Zo werd in deel 3 ‘Aanpassingen aan het datamodel o.b.v. feedback’ dieper ingegaan op de verschillende aanpassingen aan het datamodel. In deel 4 lag de focus op het luik rond metadata en bijbehorende aspecten.

1. **Aanpassingen aan het datamodel op basis van feedback***[We verwijzen naar de slides voor meer informatie]*
   1. **Benamingen van de termen**

In het datamodel zal er zowel gewerkt worden met Nederlandstalige als Engelstalige termen. Dit komt voort uit het feit dat binnen OSLO Engelstalige termen enkel zijn toegestaan wanneer het om technische documentatie of begrippen gaat. Gedurende de implementatie mag men uiteraard wel met Engelstalige termen werken.

* 1. **Toevoegen ‘RootNode’**

Aan het datamodel uit de tweede thematische werkgroep (zie slide 13) werd de klasse ‘RootNode’ toegevoegd. Deze RootNode staat voor een beginpagina van een Linked Data Event Stream, die zelf al leden kan bevatten, en via wiens gelinkte pagina’s alle leden kunnen gevonden worden.

**Discussie/beslissing**: Vanuit de werkgroep werd de vraag gesteld welke URI (Unique Resource Identifier) er best kan (her)gebruikt worden voor deze klasse. Zal er binnen Data.vlaanderen een URI aangemaakt voor deze klasse of vragen we aan de eigenaar van de [SEMIC LDES specificatie](https://semiceu.github.io/LinkedDataEventStreams/) om hier een URI voor te voorzien?

* Om te bepalen welke URI wordt toegekend aan een RootNode dient het model eerst afgewerkt te worden. Nadien zal ook de impact op de [SEMIC LDES specificatie](https://semiceu.github.io/LinkedDataEventStreams/) bekeken worden. Op deze manier kunnen de nodige issues aangemaakt worden via de juiste github repository en kunnen de juiste leden uit de LDES community betrokken worden bij de discussies.
  1. **Relatie tussen Eventstream en Node**

In dit onderdeel werd de klasse ‘Node’ voorgesteld. Een node is een pagina die de relatie tot andere pagina’s beschrijft via een relatie. Bovendien kunnen via deze relatie(s) andere nodes gevonden worden.

Naast de klasse ‘Node’ werd ook de klasse ‘TREE::(View)’ toegevoegd aan het datamodel. Deze klasse werd overgenomen uit de TREE specificatie en representeert een specifieke indexering/fragmentatie van een dataset.

**Discussie/beslissing**: Waarom worden er haakjes gebruikt in de klasse TREE:(View)?

* Dit komt omdat deze klasse niet expliciet voorkomt in de TREE specificatie. Echter wordt er in de specificatie wel impliciet verwezen naar deze klasse door te vermelden dat het domein van het attribuut retentiebeleid een TREE:View is. Binnen het OSLO LDES traject zijn er twee mogelijke oplossingen die zullen besproken worden binnen het kernteam:   
  + De klasse expliciet opnemen in de TREE specificatie;
  + De klasse publiceren binnen het OSLO LDES traject als ‘View’.

**Discussie/beslissing**: Vanuit de werkgroep werd de vraag gesteld of TREE:(View) de juiste klasse is om het attribuut retentiebeleid op te nemen.

* Dit zal binnen het kernteam bekeken worden. Er werd wel vermeld dat TREE:(View) de meest logische klasse is om retentiebeleid op te nemen. Dit omdat het retentiebeleid dan geldt voor alle elementen van de LDES in deze specifieke view, en niet voor één element/node. Hierdoor vermijden we de nood aan het definiëren van een apart retentiebeleid voor elk element.

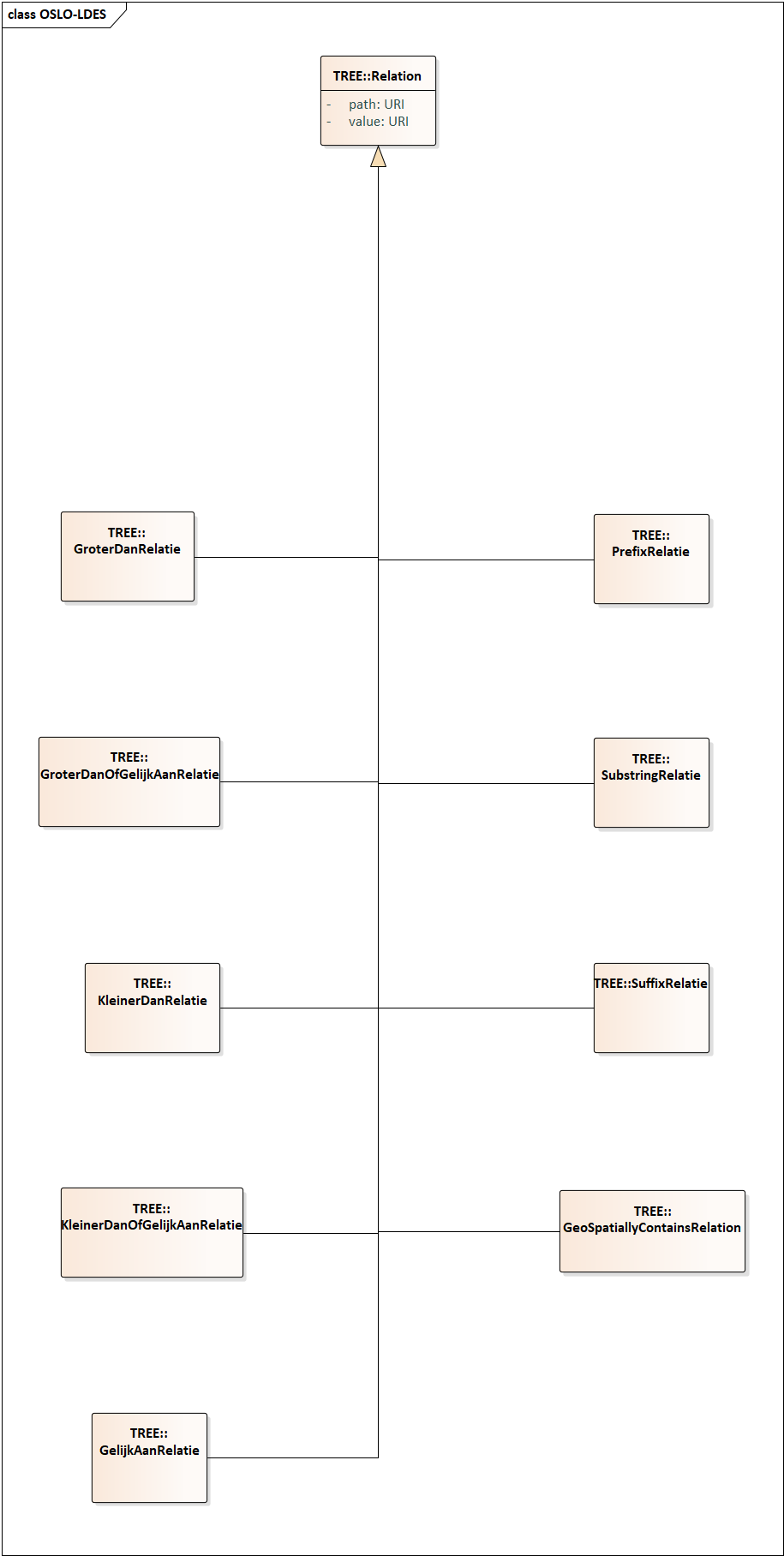
**Discussie/beslissing**: Kunnen er verschillende retentiebeleiden voorkomen op één view?

* Het is mogelijk om verschillende retentiebeleiden te definiëren op eenzelfde view. De fragmentaties kunnen hierbinnen ook verschillen. Telkens als je een eventstream anders fragmenteert, creër je een andere view en telkens als je een ander retentiebeleid definieert, creër je ook een andere view.

Dit wordt ook aangegeven door de rolnamen op de associatie tussen Eventstream en Node. Omdat TREE:(View) een associatieklasse is worden de rolnamen mee overgenomen op deze relatie.

* 1. **Relaties tussen Nodes**

De verschillende relaties (zie figuur 1)komen voort uit de [TREE specificatie](https://treecg.github.io/specification/#Relation) en worden binnen het datamodel van OSLO LDES opgenomen om de relaties tussen verschillende nodes te kunnen beschrijven.



*Figuur 1: Relaties tussen nodes*

**Discussie/beslissing**: In de tweede thematische werkgroep werd er besloten om van de klasse ‘TREE::Relation’ een associatieklasse te maken. Echter werd het bij nazicht van de TREE specificatie duidelijk dat deze klasse om zowel semantische en praktische reden geen associatieklasse moet zijn. Het voordeel van de klasse niet als associatieklasse op te nemen, is dat de verschillende subklassen (zie figuur 1) expliciet mee opgenomen kunnen worden in datamodel.

**Discussie/beslissing:** Finaal zullen de subklassen (i.e. verschillende soorten relaties) als document of als codelijst worden toegevoegd om niet het beleid te moeten volgen van het gehele datamodel bij een wijziging (bv. een extra relatie).

* Het doel is om een thesaurus te bereiken waar iedereen aan kan meewerken, maar zodanig dat ook iedereen naar dezelfde URI’s kan verwijzen.
* De complete lijst zal aangevuld en beheerd worden op binnen de [TREE specificatie](https://treecg.github.io/specification/#Relation). Naar deze standaard zal verwezen worden in de usage note.

**Discussie/beslissing**: Vanuit de werkgroep kwam de vraag waarom net deze relaties (zie figuur 1) aan het model toegevoegd moeten worden.

* De relaties zorgen ervoor dat je als client weet hoe je door de LDES moet navigeren en hoe je dus ook de andere elementen van de LDES kan terugvinden.

**Discussie/beslissing**: Kunnen relaties gebruikt worden om gelaagdheid, i.e. hiërarchie, toe te voegen aan je dataset?

* Dit is niet het primair doel van relaties. Relaties dienen om binnen één collectie (i.e. één Linked Data Event Stream) de link tussen de verschillende nodes weer te geven. Als je het schema gaat veranderen (i.e. de relaties) maak je een nieuwe collectie of een nieuwe Linked Data Event Stream. De relaties tussen verschillende LDESs komt neer op de relaties tussen afgeleide datasets.

**Discussie/beslissing**: Kunnen de verschillende relaties gebruikt worden om links tussen verschillende business entiteiten te beschrijven?

* Het is niet het doel om deze business links te beschrijven a.d.h.v. bovenstaande relaties. De links tussen business entiteiten vallen binnen de generieke scope van een OSLO model en zal worden weergegeven aan de hand van een connector in UML. Hier gaan het echt om relaties tussen nodes, niet tussen objecten.

**Discussie/beslissing**: Kan de fragmentatie strategie expliciet gemaakt worden?

* Dit zal na de werkgroep bekeken worden door het kernteam en mee opgenomen worden tegen de vierde thematische werkgroep.

**Discussie/beslissing:** Wat is het verschil tussen een TREE::View en TREE:Relation?

* TREE::VIew geeft weer hoe de EventStream gefragmenteerd is. TREE::Relation geeft weer hoe je ten gevolge van de specifieke fragmentatie informatie kan terugvinden op de verschillende nodes.
  1. **Retentiebeleid als deel van ‘View’**

Het retentiebeleid wordt mee opgenomen als deel van de view. We verwijzen hier ook naar de discussie bij C ‘Relatie tussen Eventstream en Node’. Er werden twee retentiebeleiden expliciet mee overgenomen vanuit de [LDES specificatie](https://semiceu.github.io/LinkedDataEventStreams/#retention), namelijk ‘[DurationAgoPolicy](https://semiceu.github.io/LinkedDataEventStreams/" \l "time-based-retention)’ en ‘[LatestVersionSubset](https://semiceu.github.io/LinkedDataEventStreams/" \l "version-subsets)’.

**Discussie/beslissing**: Waar zit het juridische beleid? Hoort deze hier (i.e. het retentiebeleid) thuis? Valt access control hier ook onder?

* Metadata beschrijft de data die aanwezig is in de dataset. Hierdoor moet de metadata de informatie over het retentiebeleid of de acces control kunnen uitlezen vanuit de dataset, i.e. de collectie. Dit zal verder bekeken worden tegen de vierde thematische werkgroep. Wel moeten er op verschillende views verschillende beleiden kunnen gelden.
  1. **Interpretatie ‘Onveranderlijkheid’**

De interpretatie van onveranderlijkheid wordt overgenomen uit de onderliggende OSLO klassen. Dit gaat over objecten die ofwel zelf een event/gebeurtenis zijn (bv. een observatie) of het resultaat zijn van een gebeurtenis.

* 1. **Ruimte voor afgeleide datasets**

De voorgestelde klassen (AfgeleideCollectie, AfgeleideEventStream, VersieMaterialisatie) komen reeds impliciet voor in de [SEMIC LDES specificatie](https://semiceu.github.io/LinkedDataEventStreams/). In het datamodel van OSLO LDES werden bovengenoemde klassen expliciet opgenomen. Wat we hier voorstellen geeft een houvast om afgeleide datasets en afgeleide collecties te kunnen incorporeren in een verder datamodel.

* **AfgeleideCollectie** = De klasse voor de verschillende types van afgeleide datasets. Enkele voorbeelden zijn:   
  + Versie-materialisaties
  + Snapshots
  + Tijdsreeksen
  + Aggregaties
  + Combinaties
  + Schema mappings en alignering
  + Anonimiseringen
  + Reconciliaties
* **VersieEventStream** = Een eventstream die verder bouwt op een collectie na een versie-materialisatie.

**Discussie/beslissing**: Er is nood aan een onderscheid tussen een versie-materialisatie en snapshots. Het is bovendien enkel mogelijk om een snapshot van een LDES te versie-materialiseren.

* Snapshot = een gefilterde versie van je LDES op een specifiek moment in de tijd. Dit blijft nog wel steeds een LDES. De elementen in een snapshot kunnen nog steeds versie-indicatoren bevatten.
* Versie-materialisatie = een collectie die enkel een specifiek versie (bv. de laatste versie) van de elementen bevat die deel uitmaken van de respectievelijke collectie.

Tegen de volgende werkgroep zullen de definities, labels en voorbeelden besproken worden binnen het kernteam.

* 1. **Shape**

Gedurende de werkgroep werd er vermeld dat de shape, gedefinieerd door een URI, opgelegd moeten worden en coherent moet zijn aan elementen van de LDES. Op dit moment wordt er binnen de LDES-community quasi uitsluitend gewerkt met SHACL. Stel dat een data-leverancier ShEx wenst te gebruiken, dan moet hij/zij dit coherent doen doorheen de gehele LDES. Het is dus zeker niet zo dat er de facto een SHACL shape gebruikt moet worden.

**Opmerking**: Er werd reeds geëxperimenteerd met de ShEX shape, maar puur ter validatie van de source selectie. Dit vormde geen probleem.

**Discussie/beslissing**: Kan je in een EventStream verschillende manieren gebruiken om dezelfde shape te beschrijven?

* Dit kan als je een representatie meegeeft, i.e. men geeft de URI met een beschrijving van de inhoud.

**Discussie/beslissing**: Vanuit de werkgroep kwam de vraag of de mogelijkheden van shapes vastgelegd moeten worden.

* Dit moet zeker niet vastgelegd worden. Enkel moet een dataleverancier de shape coherent gebruiken doorheen heel zijn Linked Data Event Stream (LDES).

**Discussie/beslissing**: Wat is de inhoud van een shape?

* Hiervoor bestaan twee mogelijke definities:
  + Welke minimale informatie is aanwezig om de fragmentatie te beschrijven?
  + Wat is de structuur van de onveranderlijke objecten in een LDES?
* Deze informatie zal toegevoegd worden in de gebruiksnota van het attribuut ‘Shape’. Het doel is om de definitie en gebruiksnota generiek te houden zodanig dat deze dekkend is voor alle thema’s.

1. **Metadata**

Het doel is om LDESs te kunnen plaatsen in de brede context van metadata om ze vindbaar te maken, enerzijds op dataportalen en anderzijds op machineleesbare catalogi. Het is de balans vinden om beide communities te kunnen ondersteunen. Wel loopt er al lang een traject om mensen te laten kiezen wat de beste manier is om metadata te ontsluiten, namelijk metadata DCAT met twee specifieke subtrajecten. Het eerste subtraject is DCAT-AP-VL voor open data, het tweede subtraject kadert binnen geodata, namelijk GeoDCAT-AP VL.

*[We verwijzen naar de presentatie voor meer informatie]*

* 1. **Eventstream: Een dataset of een dataservice?**

Is een Linked Data Event Stream een dataset of een dataservice? Verstaan we onder een LDES een dataset of een dataservice en hoe kan een client deze LDES dan terugvinden in een datacatalogus? Kunnen we dezelfde mapping gebruiken om machines te ondersteunen?  
  
Hiervoor is momenteel de datavindplaats (voor mensen) in ontwikkeling en het zou een doel kunnen zijn om ook LDESs te kunnen terugvinden op deze datavindplaats. De vindplaats voor machines en voor mensen moet wel steeds elkaar ondersteunen.

* Een **dataset** is een verzameling gegevens, gepubliceerd of samengesteld door één agent, die beschikbaar is voor toegang of download in één of meerdere voorstellingswijzen.​
* Een **dataservice** is een verzameling van operaties die ofwel toegang bieden tot één of meer datasets ofwel een dataverwerkingsdoel hebben.

**Discussie/beslissing**: Binnen dit traject moeten we LDESs zo compatibel mogelijk maken met DCAT om de LDESs die we aanbieden ook op Europees niveau vindbaar te maken.

**Discussie/beslissing**: Kunnen we enkele user stories beschikbaar maken op de datavindplaats om de client te ondersteunen? Enkele voorbeelden zijn:

* Ik wil via de datavindplaats een overzicht vinden van alle tree:Views op een ldes:EventStream.
* Ik wil via de datavindplaats alle afgeleide datasets vinden van een bepaalde dataset.
* Ik wil alle eventstreams vinden die kunnen voldoen aan een bepaald shape snippet.

**Discussie/beslissing**: Er is geen andere mogelijkheid dan LDESs te documenteren. Het is niet aan de databeheerder om ervoor te zorgen dat de client alle data makkelijk kan raadplegen. Het is aan hem/haar om alle data aan te bieden ongeacht de granulariteit (in de ISO wereld wordt hiernaar verwezen met het begrip ‘serie’).

* Het is zoeken naar het evenwicht tussen enkel de EventStream beschrijven door metadata of ieder element van je EventStream te beschrijven door metadata. Wel moet er op een bepaald niveau van granulariteit metadata beschikbaar zijn.
* **Conclusie**: Het is aan de dataleverancier om het laagste niveau van granulariteit te bepalen en dit te beschrijven. Anders kan men de data niet terugvinden. Als een client vindt dat alle views belangrijk zijn en dat deze ook beschreven dienen te worden, dan is het aan de front-end om hier goed mee om te gaan. Deze discussie mag dus open gelaten worden, er moet enkel een minimum aan metadata beschikbaar gesteld worden.

**Discussie/beslissing:** De interpretatie in de context van metadata DCAT rond een LDES heeft voorrang op andere afspraken die ooit in relatie met TREE en DCAT opgenomen werden. Heeft dit een impact op het model?

* Zoals het nu (zie slide 42) wordt voorgesteld, is een LDES een dataset en kan de metadata DCAT conform gesteld worden.

**Discussie/beslissing**: Een LDES wordt als een dataset gezien (zie slide 42) en kan op deze manier teruggevonden worden op de datavindplaats. De vraag is echter of de discussie die hierboven gevoerd werd rond het niveau waarop we metadata gaan aanbieden binnen de scope van dit traject valt.

* De impact op het model is dat in de TREE specificatie een LDES wordt beschreven als een dataservice en een LDES als dataset wordt voorgesteld in deze werkgroep. Dit leidt tot een conflict en dient nader bekeken te worden.
  + Het equivalent van een dataservice binnen de TREE specificatie is dat een dataservice meerdere views en meerdere datasets kan surfen (zie ook figuur Geraldine).
* **Conclusie:** Een Linked Data Event Stream kan binnen de context van DCAT gezien worden als een dataset.

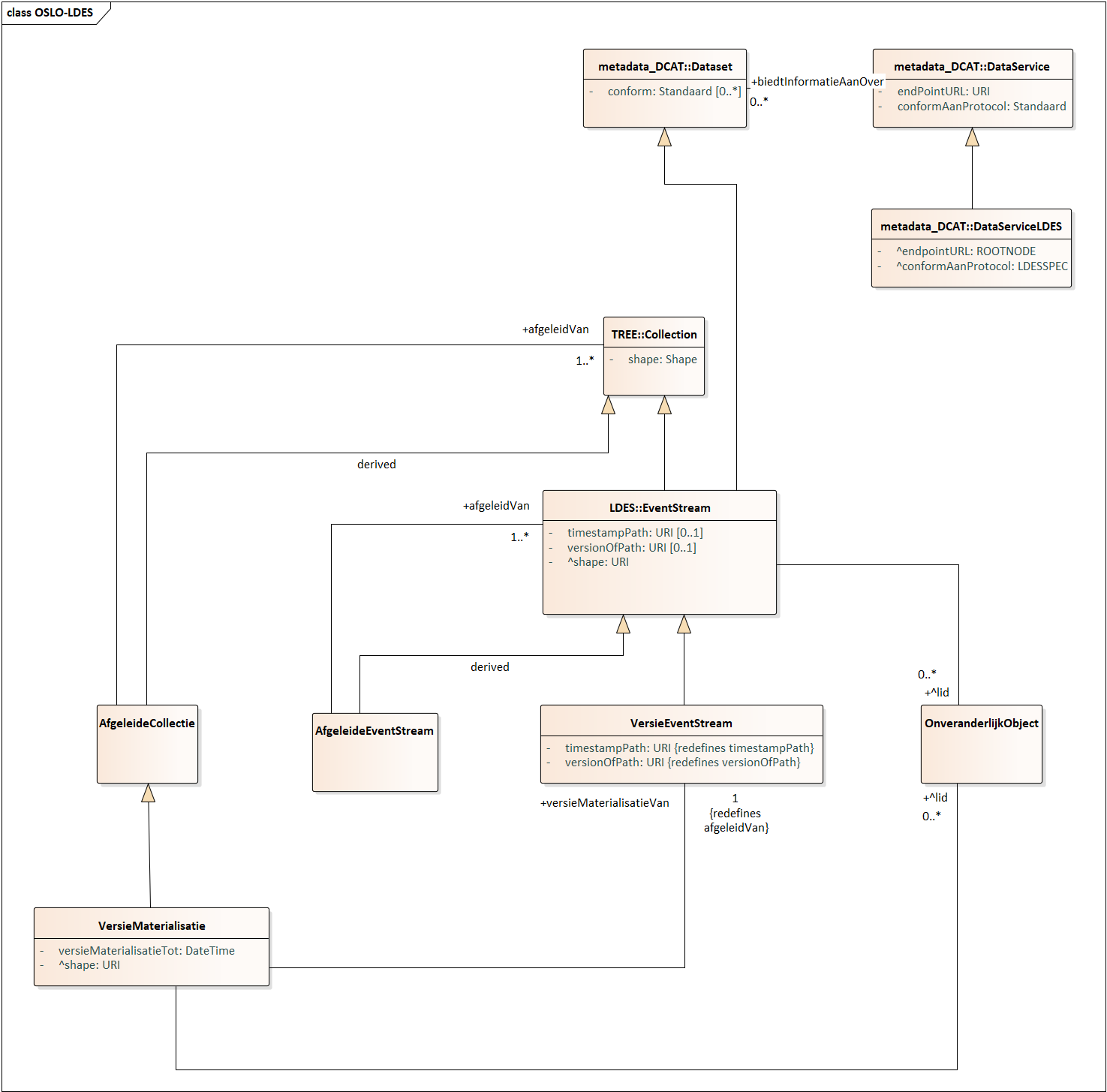
* 1. **RootNode: Een dataservice**

Naar de werkgroep werd de vraag gesteld of een RootNode gezien kan worden als een dataset of als een dataservice.

**Discussie/beslissing**: De RootNode kan gezien worden als een dataservice. Dit is de service op de Eventstream.

* **Conclusie**: Een LDES is een dataset en van zodra er gefragmenteerd of gepagineerd wordt zien we een node als een dataservice. Dit aangezien het doel van een LDES is om de kern API te zijn voor replicatie en synchronisatie.

Volgende structuur werd tot slot voorgesteld aan de werkgroep:

  
*Figuur 1: Voorstel om metadata te betrekken in het LDES datamodel.*

**Discussie/Beslissing**: Wat met alle andere attributen die binnen het applicatieprofiel metadata\_DCAT ook thuishoren (acteur, datum, …)?

* Hiervoor verwijzen we binnen OSLO in de gebruiksnota naar het respectievelijke applicatieprofiel.

**Discussie/beslissing**: Stel dat iemand een LDES wilt delen, dan moet de metadata reeds aanwezig zijn op de eventstream.

* Technologisch gezien is dit geen probleem. Welke request er moet gedaan worden om metadata te verkrijgen vanuit de LDES moet verder worden bekeken.
  1. **Access control**

*[We verwijzen naar de slides voor meer informatie]*

**Discussie/beslissing**: Zijn er technische verwachtingen (zoals http authenticatie tokens) die afgesproken moeten worden of laten we dit vrij en volgen we de best practices van het moment? ​

* Geen reactie. Dit zal later opgenomen worden in de community werking en wordt buiten de scope van dit traject geacht.
  1. **Data machtigingen**

*[We verwijzen naar de slides voor meer informatie]*

**Discussie/beslissing**: Men dient binnen een LDES wel aan te geven waar men data machtigingen kan bekomen. De verdere afhandeling van deze machtigingen valt out of scope.

* 1. **Data encryptie**

*[We verwijzen naar de slides voor meer informatie]*

Data encryptie worden voor de rest van het traject als out of scope beschouwd.

1. **Algemene feedback**

Aan het eind van de thematische werkgroep werd aan de deelnemers gevraagd om feedback te geven op de onderdelen van het datamodel en het algemene datamodel. Om praktische redenen wordt er via dit verslag nogmaals een oproep gedaan om feedback te geven in [Mural](https://app.mural.co/t/beadvtc7549/m/beadvtc7549/1646663791012/b872cbcf01051b28d26cf285c8bffd1087177b14?sender=u048a1117151baed084666519).

1. **Volgende stappen**

Indien u graag zou willen deelnemen aan één van de aankomende werkgroepen, kan u via de volgende [link](https://overheid.vlaanderen.be/opleiding/oslo-ldes) een overzicht van de sessies terugvinden en u ook zo inschrijven. De ‘thematische werkgroep 4 OSLO LDES’ is de eerstvolgende werkgroep voor het OSLO-traject en zal plaatsvinden op 28 april 2022 om 9u00 via Microsoft Teams waarvan de link wordt doorgestuurd naar de deelnemers.

Inschrijven voor de volgende werkgroep (28 april 2022) kan via onderstaande link: <https://overheid.vlaanderen.be/informatie-vlaanderen/agenda/thematische-werkgroep-4-oslo-ldes>