**///** OSLO LDES: Thematische werkgroep 2

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

Datum: 17/02/2022

Locatie:  Teams meeting (virtueel)

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

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

**Aanwezigen**

* Digitaal Vlaanderen
  + Thijs Hegge
  + Annelies De Craene
  + Geert Thijs
  + Arne Van Der Stuyft
  + Filip Borloo
  + Dwight Van Lancker
  + Bert Van Nuffelen
  + Simon Claus
  + Veerle Beyaert
  + Adriaan Walpot
  + Geraldine Nolf
  + Mathieu Beirlaen
  + Justine Ottevaere
* IMEC
  + Pieter Colpaert
  + Philippe Michiels
  + Bart Matthys
* Vlaams Instituut Voor de Zee (VLIZ)
  + Milan Lamote
  + Marc Portier
* Atos
  + Patrick Michels
* Konsolidate
  + Christophe Cop
* Esri Belux
  + Gert Bergers
* MyCSN NV
  + Jan Geukens
* RedPencil
  + Niels Vandekeyb

**Agenda van de werkgroep**

|  |  |
| --- | --- |
| 09u00 – 09u10 | **Welkom & introductie** |
| 09u10 - 09u30 | **Overzicht: wat hebben we gedaan in de vorige werkgroep** |
| 09u30 – 09u40 | **Aanpassingen aan het datamodel** |
| 09u40 – 10u10 | **Versiebeheer en relaties tussen LDESs** |
| 10u30 – 10u40 | **Pauze** |
| 10u40 – 11u50 | **Navigatie doorheen een LDES** |
| 11u50 – 12u | **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 eerste thematische werkgroep, dewelke georganiseerd werd op 20 januari 2022:

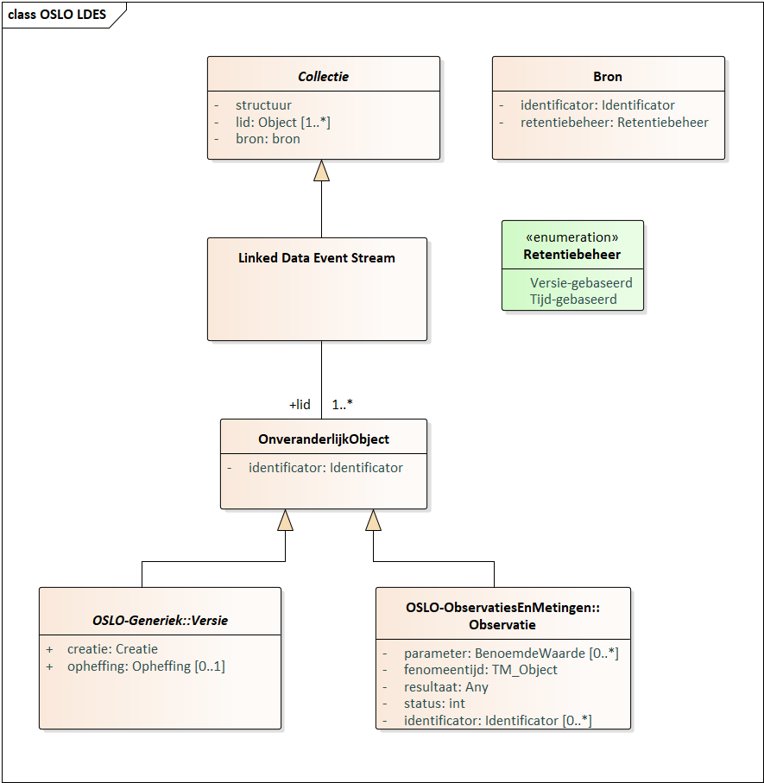
* Herhaling business werkgroep:
  + Introductie tot OSLO
  + Introductie tot Linked Data Event Stream
  + Brainstorm oefeningen

* SEMIC specficatie
  + Introductie tot de specificatie
  + Retentiebeleid(en)
  + Versiematerialisatie
* Eerste poging tot een datamodel
  + Basisdefinitie van een LDES
  + Collectie/structuur/lid/bron
  + OnveranderlijkObject

De feedback en opmerkingen die werden ontvangen gedurende de eerste thematische werkgroep werden meegenomen en vormden de basis voor de tweede thematische werkgroep. Zo werd in deel 3 ‘Aanpassingen aan het datamodel o.b.v. feedback’ dieper ingegaan op de benamingen van de elementen. In deel 4 lag de focus op observaties en versiebeheer en in deel 5, 6 en 7 lag de focus de relaties tussen verschillende Linked Data Event Streams en het navigeren doorheen een enkele Linked Data Event Stream.

1. **Aanpassingen aan het datamodel op basis van feedback**

Gedurende de eerste thematische werkgroep werd er een eerste poging tot een datamodel voorgesteld aan de werkgroep (zie figuur 1: datamodel TW1). De feedback en opmerkingen die werden gegeven door de deelnemers werden verzameld in het [verslag van de eerste thematische werkgroep](https://github.com/Informatievlaanderen/OSLOthema-ldes/raw/standaardenregister/reports/Verslag%20thematische%20werkgroep%201%20-%2020%20januari%202022.docx) en op basis van deze opmerkingen werden er enkele aanpassingen gemaakt aan het datamodel. Het aangepaste datamodel wordt getoond in figuur 2: datamodel TW2.

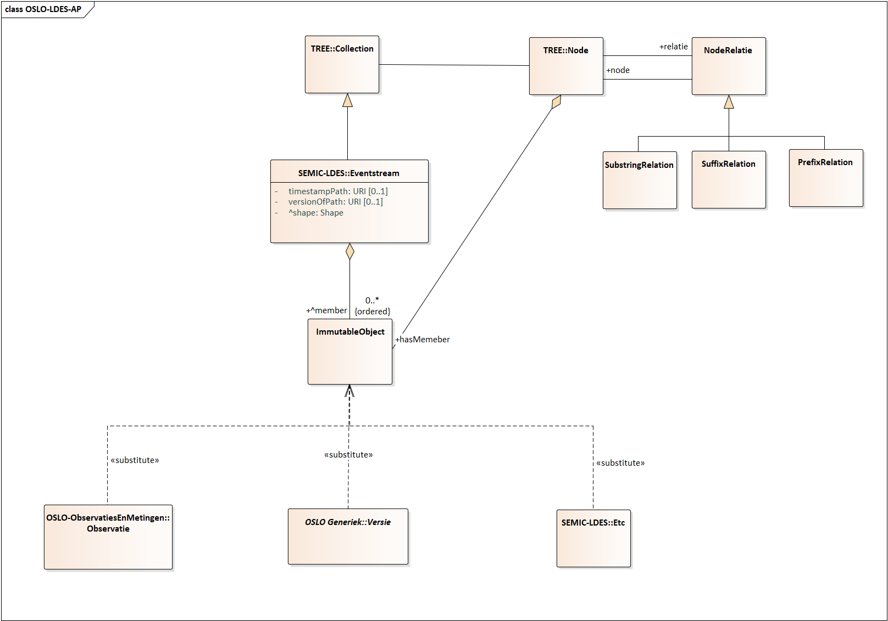
****

***Figuur 1: datamodel TW1***

Omdat het te ver zou leiden om alle opmerkingen hier te herhalen, hebben we deze samengebundeld tot drie thema’s:

1. Definities en benamingen
2. Versiebeheer
3. Observaties

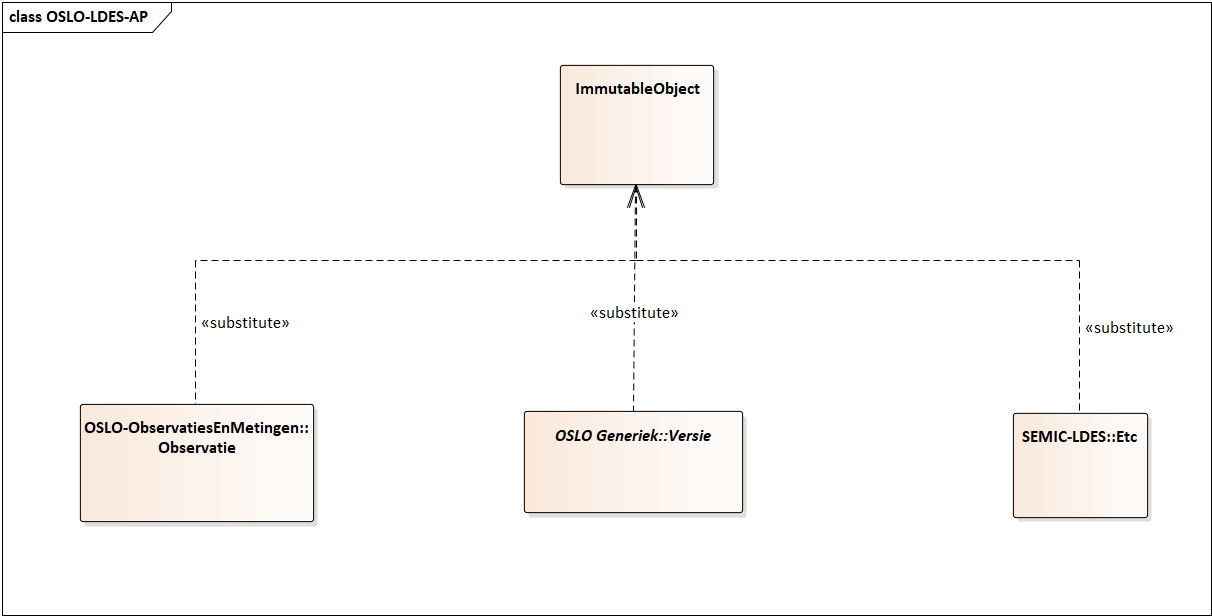
Versiebeheer en observaties werden besproken in deel 4 ‘Versiebeheer en relaties tussen LDESs’. Rond definities en benamingen werd er beslist om te werken met de Engelstalige termen indien mogelijk. Echter is het verplicht om de OSLO specificatie in het Nederlands te ontwikkelen en is Engels slechts toegelaten wanneer het om technische documentatie gaat. Dit zal nader bekeken worden met experten binnen het kernteam om te bepalen hoe en wanneer we zullen werken met Engelse termen.

******

***Figuur 2: Datamodel TW2***

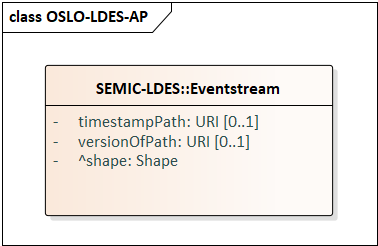
1. **Versiebeheer en relaties tussen LDESs**
   1. **OSLO Observatie en OSLO Generiek**

Aangezien er meerdere types van onveranderlijke objecten kunnen bestaan, valt de effectieve invulling van een onveranderlijk object buiten de scope van dit traject (bijvoorbeeld alle objecten gedefinieerd onder eender welk OSLO traject zouden als een onveranderlijk object binnen een LDES gepubliceerd kunnen worden). In het model wordt dit weergegeven door de ‘substitute’ relatie (zie figuur 3: Substitute relatie). Deze relatie impliceert dat onveranderlijk object (‘ImmutableObject’) invulling zal krijgen als een instantie van een specifieke klasse in een Linked Data Event Stream.



*Figuur 3: Substitute relatie*

Echter blijft de vraag hoe men een dergelijke specifieke invulling kan realiseren. Het antwoord hierop is dat dit configureerbaar is door de data-leverancier. Hij/zij kan door een URI toe te kennen aan de attributen ‘timestampPath’ en/of ‘versionOfPath’ zelf configureren hoe hij/zij versiebeheer wilt realiseren en de relaties tussen objecten wenst voor te stellen.



*Figuur 4: Configureerbaarheid van een Linked Data Event Stream*

Gedurende de thematische werkgroep werden er enkele voorbeelden getoond van mogelijke invullingen door gebruik te maken van OSLO Observaties en Metingen en versiebeheer. De voorbeelden werden uitgewerkt in JSON-LD, en zijn raadpleegbaar in de presentatie.

* 1. **Relaties tussen LDESs**

Om een relatie tussen verschillende Linked Data Event Streams te kunnen weergeven, is het belangrijk om te weten hoe een Linked Data Event Stream is opgebouwd. Een dataset, bijvoorbeeld het adressenregister, zal een set van event streams worden, die verbonden zijn met elkaar. Meer specifiek wordt het adressenregister onderverdeeld in vier Linked Data Event Streams, namelijk: een LDES van adresobjecten, een LDES van straatnamen, een LDES van gemeenten en een LDES van postinformatie.

In deze context kwam de term ‘eventually consistent’ aan bod. Deze term impliceert dat de verschillende Linked Data Event Streams niet altijd consistent zijn, maar dit wel op een zeker moment in de tijd zullen zijn. Dit is het basisprincipe voor relaties tussen Linked Data Event Streams. Hoe deze relaties zelf (op technisch niveau) weergegeven kunnen worden, zal in de derde thematische werkgroep aan bod komen door middel van metadata.

1. **Navigatie doorheen een LDES**

Naast de mogelijkheid om relaties tussen verschillende Linked Data Event Streams te kunnen weergeven, is het ook noodzakelijk om te kunnen navigeren doorheen één Linked Data Event Stream. Hiervoor zijn de concepten paginatie, view, node en relatie cruciaal. Om deze concepten te verduidelijken werd er vertrokken vanuit de [TREE specificatie](https://treecg.github.io/specification/). Door het feit dat deze TREE specificatie niet altijd sluitend is, werden er enkele mogelijke interpretaties en voorstellingen gepresenteerd en besproken gedurende de werkgroep om hierover een gedragen beslissing te kunnen nemen. De opties en beslissingen bij de verschillende concepten worden hieronder weergegeven.

1. Een View kan in de TREE specificatie op twee manieren geïnterpreteerd worden. Het is daarom belangrijk om het concept van een View te disambigueren (zie slide 45).   
   * Interpretatie 1: *een View is de index/invariant/organisatie waarmee doorheen de collectie (LDES) kan worden genavigeerd​.*
   * Interpretatie 2: *een View is een beginpagina, die zelf al leden kan bevatten, via wiens gelinkte pagina’s alle leden kunnen worden gevonden.​*

**Discussie/beslissing:** Het gaat hier over twee verschillende interpretaties van ‘View’. Beide interpretaties zijn nodig. Interpretatie 1 gaat over een view tussen objecten terwijl interpretatie 2 gaat over de view om de beginpagina.

Het zou goed zijn om aan beide interpretaties een andere naam te geven om verwarring tussen beide concepten te vermijden. Voor de namen werden door de deelnemers enkele mogelijkheden gegeven, namelijk ‘BeschrijvingVanIndex’ voor interpretatie 1 en ‘rootNode/entryNode/startNode/homeNode/entryPoint’ voor interpretatie 2. De verschillende opties zullen bekeken worden en het datamodel zal aangepast worden voor de derde thematische werkgroep.

1. Een bijkomende vraag gaat over interpretatie 1. Moet men steeds alle leden van een Linked Data Event Stream kunnen terugvinden via de View?

**Discussie/beslissing:** Over dit punt was er geen eensluidendheid onder de leden van de werkgroep. Verschillende meningen werden geuit, namelijk:

* + Pro volledigheid (alle leden kunnen terugvinden via de View):
    - Het basisidee van een Linked Data Event Stream is volledigheid.
    - Binnen een Linked Data Event Stream staat het retentiebeleid standaard op oneindig: dit impliceert dat alle elementen behouden zullen blijven binnen de LDES. Echter kan een een data-leverancier hiervan afwijken en bijvoorbeeld een [tijdsgebaseerd retentiebeleid](https://semiceu.github.io/LinkedDataEventStreams/#time-based-retention) introduceren.
  + Contra volledigheid:
    - Het tijdsvenster van je Linked Data Event Stream verschuift continu waardoor volledigheid niet kan bereikt worden. Hiermee wijst men op het feit dat je moet synchroniseren met de authentieke bron en dat er een tijdsverschil is tussen het moment dat het element in je Linked Data Event Stream komt en het moment dat de data-afnemer synchroniseert met je Linked Data Event Stream.

Binnen de discussie rond (on)volledigheid moeten steeds de basisprincipes van een Linked Data Event Stream in het achterhoofd gehouden worden, namelijk replicatie en synchronisatie. Voornamelijk door het principe van synchronisatie dient een data-gebruiker zekerheid te hebben dat hij/zij synchroniseert met de authentieke databron. Vervolgens is het aan de data-afnemer om te configureren of hij/zij de volledige databron wenst te gebruiken of slechts een deel hiervan. Hoe hij/zij dit kan doen, komt neer op het principe van fragmentering; op dit principe zal dieper ingegaan worden in een apart OSLO traject.

1. Wat zijn de verwachtingen rond een node? Moet de node zelf weergeven welke elementen zij bevat? In de TREE specificatie wordt momenteel enkel weergegeven welke elementen terug te vinden zijn op de volgende node (zoals op slide 42).

**Discussie/beslissing:** Om te kunnen nagaan welke elementen er deel uitmaken van de node, dient er een onderscheid gemaakt te worden tussen het praktisch nut en het semantisch nut:

* **Praktisch nut** (slide 51): er is geen enkel praktisch nut om op elke node telkens specifiek te vermelden welke elementen er deel van uitmaken.
* **Semantisch nut** (slide 52): het is interessant vanuit een semantisch standpunt om weer te geven dat die relatie bestaat, echter moet in een gebruiksnota gespecificeerd worden dat dit geen enkel praktisch nut heeft. De achterliggende reden dat de relatie geen praktisch nut heeft komt voort uit het feit dat wanneer de relatie er toch zou zijn, je via twee verschillende paden kan raadplegen welke elementen deeluitmaken van de node. Dit leidt aldus tot redundantie wat een impact kan hebben op de performantie.

1. Hoe geven we best de relaties tussen verschillende nodes weer? Is dit met behulp van een associatieklasse of door middel van een wederkerende relatie (slide 57)?

**Discussie/beslissing:** Er was eensgezindheid om de relaties tussen verschillende nodes door middel van een wederkerende relatie weer te geven (aangezien ook de TREE specificatie deze aanpak hanteert).

1. Welke relaties kunnen er bestaan tussen verschillende nodes?

**Discussie/beslissing:** Tijdens de werkgroep werden reeds enkele mogelijke relaties gepresenteerd en uitgelegd aan de deelnemers. Deze relaties komen voort uit de [TREE specificatie](https://treecg.github.io/specification/#Relation). Het is de bedoeling om een dergelijke codelijst op te nemen binnen de OSLO-werking. Dit zal een onderdeel uitmaken van een eventueel vervolgtraject inzake fragmentering. Indien dit gevraagd wordt door leden van de werkgroepen, kunnen we binnen het OSLO LDES traject URI’s voorzien voor de verschillende soorten relaties, en voor twee specifieke relaties (tijds- en geo-gebaseerd) kan er reeds een praktische invulling gegeven worden.

1. **Afgeleide datasets**

Afgeleide datasets zijn collecties die afgeleid zijn van een LDES. Enkele voorbeelden zijn:

* Snapshots
* Versie-materialisaties
* Tijdsreeksen
* Aggregaties
* Combinaties
* Schema mappings en alignering
* Anonimiseringen
* Reconciliaties

Voor de derde thematische werkgroep zal er een manier voorzien worden om dergelijke afgeleide datasets te kunnen incorporeren in het datamodel. Iedere individuele afgeleide dataset gaan modelleren zou ons te ver leiden.

1. **Brainstorm oefening**

Door tijdsgebrek is het niet gelukt om de brainstorm oefening gedurende de thematische werkgroep te organiseren. In de brainstorm oefening in [Mural](https://app.mural.co/t/beadvtc7549/m/beadvtc7549/1644223871728/ce828415d3ea921bbfd1f745cab81928966ce9ee?sender=u048a1117151baed084666519) worden er vier vragen gesteld:

1. Heeft u algemene feedback op het datamodel?
2. Kent u nog andere types van retentiebeleiden?
3. Welke andere types van afgeleide datasets kent u?
4. Heeft u nog andere opmerkingen?

Om praktische redenen wordt er via dit verslag nogmaals een oproep gedaan om deze vragen te beantwoorden in [Mural](https://app.mural.co/t/beadvtc7549/m/beadvtc7549/1644223871728/ce828415d3ea921bbfd1f745cab81928966ce9ee?sender=u048a1117151baed084666519).

1. **Overige opmerkingen**

* Onder de deelnemers is er een discrepantie in wat er verstaan wordt onder het begrip ‘onveranderlijkheid’. Meer specifiek zijn er twee verschillende interpretatie van het begrip:

1. Voor een deel van de leden van de werkgroep wordt met onveranderlijkheid een http cache: permanent bedoelt. Het element kan dus absoluut niet mee wijzigen.
2. Een ander deel van de deelnemers vindt dat een element nog steeds onveranderlijk is zelfs wanneer een (schrijf)fout wordt gecorrigeerd.

**Discussie/beslissing:** De opmerking die hier wordt gemaakt zal als vraag meegenomen worden naar het kernteam. In de derde thematische werkgroep zal het antwoord hieromtrent meegedeeld worden.

* Tijdens de werkgroep werd de vraag gesteld of het mogelijk is om de semantiek en/of het gebruik van de verschillende soorten relaties die werden voorgesteld op slide 58.

**Discussie/beslissing:** Voor meer informatie inzake de verschillende soorten relaties verwijzen we naar de [TREE specificatie](https://w3id.org/tree/specification/#Relation).

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 3 OSLO LDES’ is de eerstvolgende werkgroep voor het OSLO-traject en zal plaatsvinden op 17 maart 2022 om 9u00 via Microsoft Teams waarvan de link wordt doorgestuurd naar de deelnemers.

Inschrijven voor de volgende werkgroep (17 maart 2022) kan via onderstaande link: <https://overheid.vlaanderen.be/informatie-vlaanderen/agenda/thematische-werkgroep-3-oslo-ldes>