Meeting Minutes: LDES Working group - Onboarding session

Project:	SEMIC	**Date and Time:**	30/04/2025 14:00 - 16:00
Meeting Type:	Working group	**Location:**	Virtual
Coordinators:	Anastasia Sofou Isabaut Martens Pieter Colpaert	**Issue Date:**	05/06/2025

Agenda of the webinar		
	Welcome & Tour de table	Slides 1 - 6
	The new trajectory & plans for the workshops	Slides 7 - 26
	Q&A	Slides 27
	Technical walk-through current LDES specification & issues	Slides 28 - 67

Meeting Slides
LINK

Participants		
Name	**Initials**	**Organisation**
Adam Arndt	AA	DK
Ana Baptista	AB	UMinho - Portugal

Participants		
Name	**Initials**	**Organisation**
Anastasia Sofou	AS	DG DIGIT - SEMIC
Architect	A	Not Shared
Arthur Vercruysse	AV	UGent - IMEC
Benjamin Degenhart	BG	Not Shared
Bert Van Nuffelen	BVN	Digital Flanders
Dries Moreels	DM	UGent
Dwight Van Lancker	DVL	Movias & UGent
Elf Pavlik	EP	W3C Solid CG
Emiel Dhondt	ED	SEMIC
Faruk Karabulut	FK	SEMIC
Georges Lobo	GL	EC
Gérard Soisson	GS	LU
Igor Tričkovič Rifelj	ITR	SI
Isabaut Martens	IM	SEMIC
Jim Yang	JY	NO
Jitse De Cock	JDC	SEMIC
Josema Alonso	JA	SEMIC
Juan Alvarez	JA	SEMIC
Julian Rojas	JR	UGent - IMEC
Karel Kremer	KK	RedPencil - Flemish Govt. LBLOD/DS4SSC - DECIDE project
Kurt Janssen	KJ	IT CommV

Participants		
Name	**Initials**	**Organisation**
Laurens Ramandt	LR	Digital Flanders
Maarten Duhoux	MD	ERA
Maarten Segers	MS	Amexio
Mantas	M	LT VSSA
Marina	M	ERA
Martynas Mockus	MM	LT
Mila Frerichs	MF	Civic Vision
Nuno Freire	NF	Europeana
Paul Hampson	PH	EURA
Peter Bruhn Andersen	PBA	Danish Agency for Digital Government
Petteri Kivimäki	PK	NIIS
Pieter Colpaert	PC	UGent - IMEC
Ranko Orlic	RO	Digital Flanders
Sander Van Dooren	SVD	DG DIGIT - SEMIC
Terje Sylvarnes	TS	The Norwegian Digitalisation Agency
VirginijaU	V	Not Shared
Vladimiras Desiatnikovas	VD	LT DVMS
Xueying Deng	XD	IMEC

Full Meeting Minutes

Welcome & Tour de table	AS introduced the session by highlighting the importance of achieving semantic interoperability through core vocabularies and application profiles, emphasising that the main goal of the Linked Data Event

Slides 1- 6 **Speaker:** Anastasia Sofou, Isabaut Martens	Streams (LDES) initiative is to bridge the gap between technical and semantic interoperability. IM outlined upcoming standardisation track events. These include this LDES onboarding session of 30 April, followed by the following events for which participants can already register. • LDES-Working group I on 7 May • LDES-Working group II on 12 June • LDES-Working group III on 3 July Additionally, a broader community workshop will take place on 22 May, featuring several invited speakers who will present their experiences with LDES. These speakers will discuss how they have implemented LDES and share their experiences from both the consumer and producer perspectives. IM facilitated a tour de table, guiding participants as they introduced themselves and their roles within their respective organisations. During this tour de table Mantas from LT VSSA has highlighted his request for support in adding to LDES in their Universal API.
The new trajectory & plans for the workshops Slides 7 - 26 **Speaker:** Pieter Colpaert	PC highlighted that the specification has evolved from a simple vocabulary to a comprehensive technical standard with usage notes, emphasising the need to stabilise the specification through the current trajectory, addressing open issues and feature requests and provided an overview of various implementations in toolchains. **Consumer-oriented spec** PC explained that the ambition of the current trajectory is to achieve a consumer-oriented, backwards-compatible, and stable release of the Linked Data Event Streams (LDES) specification. At present, the specification includes usage notes for both publishers and consumers of LDES, but a clear choice needs to be made to streamline the specification. The recent strategy has been to adopt a consumer-oriented view to ensure loose coupling between servers and clients. This approach aims to create clients that work within an ecosystem rather than being tied to a specific server. By doing so, servers can adopt features from the LDES specification without strict normative requirements. The goal is to make the specification testable and ensure that clients can implement certain features if servers follow specific guidelines. PC explained that the new consumer-oriented specification will adhere to a new structure and will be located at the same place as the main specification. The new structure will first provide an overview of the Linked Data Event Streams terms and their intended use. It will detail what a consumer must implement, including aspects as initialisation, traversing the event stream, state management, and handling HTTP error codes. Additionally, there will be a chapter on interpreting retention policies, which is a significant part of what LDES introduces.

	Finally, the vocabulary and their terms and semantics, which were previously scattered throughout the specification, will be consolidated into their own respective section at the end of the specification. **Server primer** PC addressed concerns about the lack of a server specification if only a consumer specification exists, explaining that having clear, technically interoperable clients adhering to the same technical specification will simplify compliance. PC proposed creating a document, a server primer, that provides best practices for building servers and producers of LDES. This primer will not be normative but will offer guidance based on the best practices observed from various pilots and implementations. PC highlighted that the primer would focus on use cases related to data harvesting, use cases involve centralising data around specific themes while maintaining the authoritative source at the member states level. Although harvesting is the primary focus, the primer will provide clarity on various aspects of LDES implementation **Implementation reports** In addition to the consumer-oriented specification and the server primer, a third component will be the LDES implementation reports. These reports will be hosted in a newly created GitHub repository, with a folder for each domain. This setup allows for clear documentation on how to build an LDES for specific domains, such as railway infrastructure or cultural heritage. The implementation reports will address domain-specific questions, such as how entities change and what SHACL shapes they adhere to. PC mentioned that the repository already contains an implementation report from a previous pilot and a pull request for cultural heritage feeds. PC encouraged others to contribute their implementation reports by opening pull requests with the relevant documentation **Standardisation track** PC outlined again the schedule for the upcoming LDES working groups and events. Highlighting that after these sessions, there will be a public review period for implementations from July 3rd until the SEMIC 2025 conference. • Working group 1: Restructuring the spec and extending retention policies. • Working group 2: Consumer algorithm, iteration, and state management. • Working group 3: Server primer with best practices. Issues in the LDES GitHub repository are tagged with the relevant working group number, allowing to filter by label. The approach suggested being delving into the technical discussions

	about what to add or reject from the specification. Each workshop will last only two hours, so to maximise efficiency, there will be extensive pre-discussions on GitHub. Participants can weigh in on the specification, and PC will prepare the pull requests and issues in advance. During the workshops, the most viable solutions and open pull requests will be presented. With three possible decision that can be made during the call: • Accept and merge • Conditional accept: merge when conditions are validated • Full reject Currently a pull request has been opened for the upcoming working group, ready for review. This pull request serves as a basis for further documentation and advancing the specification with more detailed issues, PC encouraged the participants to review and comment on the pull request.
Q&A Slide 27 **Speaker:** Pieter Colpaert	Mantas from LT VSSA raised the question regarding LDES on top of Kafka streaming, with Kafka channels in the backend and LDES messages on top. PC confirmed that it is possible and mentioned that there have already been implementations done in this way. SVD elaborated that building an LDES on top of a Kafka would be a natural fit, explaining that LDES can be a very effective way of publishing a Kafka topic containing semantic data anonymously on the web. Noting that a limitation of Kafka clusters is that they are typically not published anonymously on the web, allowing people to subscribe to a topic. SVD provided an example from Flanders, where the address registry uses Kafka in the backend and an LDES server on top of it to listen to Kafka and publish the stream.
Technical walk-through current LDES specification & issues Slides 28 - 67 **Speaker:** Pieter Colpaert	PC explained that LDES provides a long-term solution by continuously integrating updates and historical data, ensuring data remains interoperable over its lifetime. An LDES client starts with an entry point URL, traverses multiple URLs, emits history, and stays synchronised as new members are published. The client should run on a schedule or as a daemon to continuously process updates, either daily or as a constantly running process. **TREE** The LDES Vocabulary is built on top of TREE hypermedia specification. The LDES client requires to implement the TREE hypermedia specification. However, currently it is not yet defined which parts of the specification are most important for LDES, as implementing the entire specification is not the goal. PC provided a simple example to help understand the Tree hypermedia specification. PC explained that the example, which is also documented in the specification, involves a collection of members, such as Subject 1

	and Subject 2, each containing more triples. This setup allows a client to find the collection and understand that the current page contains a fragment of the collection, with members like Subject 1 and Subject 2, any RDF serialisation can be used in this case. Features of the TREE spec • Initialisation • Member extraction • Traversing a search tree BVN raised the question regarding the handling of inverse relationships in membership calculations, asking for clarification on how to handle cases where there are pointers to a subject, rather than just assuming a downward direction from the subject. PC explained that inverse relationships are not included by default but can be enabled through annotated SHACL shapes, emphasising the importance of using named graphs for clarity. RO raised a concern about the need for guidance on how to map LDES in DCAT to avoid confusion, RO emphasised that providing such guidance is advisable. JR working on the report within the Tree community group mentioned that this report will elaborate on the mapping of LDES in DCAT. **LDES Specification** LDES does not use the concept of a tree collection directly but subclasses it, creating what is called an event stream. The significant addition in an LDES event stream is that all members are immutable; they cannot be edited or deleted once they are part of the stream. This makes the event stream an append-only log, conceptually similar to a ledger or a Kafka topic, where new items are continuously added, and the first item remains part of the stream forever. A question was raised by Mantas regarding the plans to possibly add UML diagrams according to the SEMIC Style guide. clarified that there are no plans currently, but a note will be taken to investigate the added value of this. PC explained the significance of two powerful properties: the timestamp path and the version path. • ldes:timestampPath: Used to determine order of events in the even stream, ensures that events are published in the correct chronological order. ○ A producer cannot publish events out of order when using the timestamp path. • ldes:versionOfPath: IRI that indicates the version of, helping in tracking changes and maintaining the version history of elements within the event stream.

	As the number of items grows, it becomes inefficient to check for new items by scanning the entire list repeatedly. To address this, it is suggested to keep the pages small to ensure efficient fetching for clients.
	• Fragmentation: To improve efficiency, the event stream can be fragmented into multiple pages instead of a single page. This involves creating links or relations between these pages.
	• Retention policies: Importance of retention policies in managing the history of event streams. Three types of retention policies currently specified.
	  ○ ldes:DurationAgoPolicy
	  ○ ldes:PointInTimePolicy
	  ○ ldes:LatestVersionSubset
	
	Current status of LDES specification
	PC has reviewed the current Linked Data Event Streams (LDES) specification and identified several areas for improvement. These updates aim to enhance clarity, focus on consumer needs, and address retention policy issues.
	• Abstract revision: Proposed revision aims to emphasise the importance of reading and implementing the specification, highlighting the consumer-oriented benefits.
	• Introduction: Suggested that the mention of compatibility with other specifications is not critical and may be better introduced later in the document
	• Fragmenting and pagination: Section on fragmenting and pagination may be more fitting for a server primer. Examples from this section are recommended to be moved to a separate document that provides a deeper dive into server implementations.
	• Retention policies enhancement: New pull requests will be opened next week to introduce more advanced retention policies. The existing three policies do not encompass all scenarios. Changes include:
	  ○ Transactions: Additional text will explain how retention policies impact transaction boundaries.
	  ○ Deletions: A specific retention policy for deletions is necessary. The current policy retains deletion activities indefinitely, which may not align with user needs.
	• Chapter revisions:
	  ○ Chapter 2: Consider removing the chapter since the updated specification will require an LDES client to incorporate certain chapters from the tree specification. This will offer a more comprehensive explanation of traversing and pruning a search tree. Instead, reference the tree specification.
	  ○ Chapter 4: Proposed to be dropped and reworked into the server primer.

	Communication Channels Participants were reminded of the importance of active communication channels. • The Matrix channel is increasingly active, providing direct communication options. • Participants are encouraged to review the issue lists and pull requests on GitHub. PC expressed optimism for participant engagement in the upcoming standardisation working groups.