

Summary of webinar on LDES for the Base Registries Community - Jan. 2023

Webinar details

Project:	SEMIC: LDES	Webinar Date/Time:	17/01/2023 13:00 - 15:00 CET
Meeting Type	Webinar	Meeting Location	Cisco Webex Meetings
Meeting Coordinator:	Arne Van Der Stuyft	Issue Date:	25/01/2023

Meeting Agenda

<ol style="list-style-type: none">1. Welcome & Introduction2. What is a linked data event stream (LDES)?3. LDES and Access to Base Registries (ABR)4. LDES as a solution - uses cases5. LDES implementations – success stories6. Wrap up and next steps
--

Full meeting minutes

Session	Summary
Welcome & introduction Speaker: Anastasia Sofou (DIGIT)	Welcoming of participants and practicalities of the webinar. Context and objectives of the webinar were disclosed, namely: <ol style="list-style-type: none">1. Familiarise with Linked Data Event Streams2. Demonstrate the business value Linked Data Event Streams can bring for base registries3. Show Linked Data Event Streams in practice

<p>What is a Linked Data Event Stream (LDES)?</p> <p>Speaker: Arne Van Der Stuyf (SEMIC Team)</p>	<p>General introduction of a Linked Data Event Stream (LDES) focusing on:</p> <ul style="list-style-type: none"> ● Introduction on current data strategies; ● Advantages and disadvantages of current data sharing; strategies from the perspective of both data owners and data users; ● Definition of LDES, based on advantages and disadvantages described previously; ● Summary of responsibilities for the different roles relevant for LDES. <p>- <i>See presentation for more information</i> -</p> <p>Question and Answer</p> <p>Question: There were many disadvantages about the current data sharing strategies in the initial intro (slide 11 and slide 22), but based on on the presentation of LDES-related implementation steps (slide 14 in combination with slide 19), it seems that considerable additional effort is needed by public administrations to use a LDES.</p> <p>Answer: A LDES implementation represents additional effort to be set up, but it facilitates the operational aspects on the long run by automating/standardising the synchronisation of multiple systems, resulting in a more efficient data sharing strategy than the current use of APIs or data dumps.</p>
<p>LDES and Access to Base Registries (ABR)</p> <p>Speaker: Bert Van Nuffelen (SEMIC Team)</p>	<p>Analysis of the adoption of LDES as a data sharing technology in the context of base registries, providing:</p> <ul style="list-style-type: none"> ● A clear definition of a base registry, including the interconnected aspects of a base registry; ● A review of the main interoperability challenges for base registries on all levels of interoperability (legal, organisational, semantical and technical); ● An explanation of how LDES could be used as a core element for base registries and the benefits associated with its use. <p>- <i>See presentation for more information</i> -</p> <p>Questions and Answers</p>

Question: What is the role of base registries in the context of data spaces?

Answer: In line with the shown pilots, data spaces will be applicable in cross-border scenarios. Additionally, both data sharing technologies are performing an interoperable data sharing exercise which implies similarities. Nevertheless, SEMIC has not been contacted to elaborate on the use of LDES in the context of data spaces, even if there are definitely synergies.

Question: Which intermediaries can be involved in the issues around data sharing? Are there still complexities in place?

Answer: Most data intermediaries will be neutral, i.e. it is most likely that they do not have specific legal obligation on what data can or cannot be shared via their LDES. Although, when it comes to history (i.e. storage of the data), the data intermediary can define a data retention policy as part of the LDES environment.

Question: Are there any economical aspects related to the role of a data intermediary (i.e. the role responsible for cleansing, harmonisation and the publication of data through a relation dataset and/or a REST API)?

Answer: They will play an increasingly important role. You can see nowadays many companies that bring together multiple datasets and create APIs out of it. Often there is manual data integration on top of which products are sold. With LDES this costly data integration becomes way cheaper and faster. This implies that one can expect that there will be more data intermediaries offering data APIs through subscriptions thanks to increased efficiency (e.g. with LDES).

Also data consumers could come together collectively and pay together for certain amounts of queries across a data space.

By doing so, LDES creates sustainability around open data in general.

Question: How should one find the balance between Linked data and sustainability (i.e. It is digital, but there are energy costs).

Answer: Nowadays, everybody processes data over and in parallel on their own machines. With Linked data, it is possible to perform some operations once only and save costs (also energy). This is due to the persistence of the elements, which is key to go back to the source and take benefit of everyone's work

	<p>through a higher networking effect. If data is published in data dumps and new identifiers need to be created every time, in order to reprocess and re-link data.</p>
<p>LDES as a solution - uses cases</p> <p>Speaker: Julian Melendez (SEMIC Team)</p>	<p>Presentation of a practical use case* on geospatial information where base registries from multiple member states were disclosed via a LDES.</p> <p>A live demo was conducted to concretely show the added value for a Member State of publishing their base registry as a LDES, providing examples across Europe. In addition, a use case on professional migration patterns was sketched as a further potential LDES implementation.</p> <p>- See presentation for more information -</p> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;"> <p>* Additional information: the shown use case works on the basis of a substring which is based on the fragmentation of the LDES. This implies that one can create his/her own fragmentation of the LDES using the substring bucketizer</p> </div> <p>Questions and Answers</p> <p>Question: How are the indexes maintained? Are there services that continuously monitor the (immutable objects in) LDES and keep the indexes (according to the TREE specification, published as Linked Data) up to date? Are those services somehow reusable among source datasets or do they need to be tailored to the individual LDES?</p> <p>Answer:</p> <p>What could happen in practice, is that the data publisher or the data owner or an authorised party by the data owner would need to create and allow the connection of the regional data source to expose this LDES and the related modules. This would allow anyone to create indexes and keep them up to date.</p> <p>It's something that has already been created for different projects. For example, in the case of the Flemish Smart Data Space, we already have models that we just can plug in in front of a link database stream that can.</p> <p>These indexes, and as it was discussed before, are completely agnostic or completely neutral of the domain. The LDES doesn't tie to any type of vocabulary on top of domain specific data. These modules can produce these indexes by which you</p>

basically have a pipeline of data flowing and keeping you up to date with the latest updates in these modules. And that's what a LDES aims at, namely reusable models that can be plugged and, of course, all open source.

Question: It was stated during the presentation that the LDES is maintained by the data publisher, but as in this use case, the index on top of the LDES is maintained by a data intermediary. Does this imply that the data intermediary must check for new LDES entries and update the indexes accordingly?

Answer: An authorised party of the data owner needs to allow access to the data in order to create a LDES and thus an indexation. By doing so, it obtains a pipeline of data flowing which allows a data user to stay up-to-date. The idea is that when an event happens (i.e. a new address is added or modified), a new object is created with a version URI and published into the LDES as a new object. This gets added to the LDES and by doing so the indexes will be updated. Every change will be logged and by doing so, historical data can be kept and is accessible.

Question: The shown pilots are not exactly streams of events. For example, updates to addresses or new addresses or for example streams of Sensor observations. So, would an LDES be useful also outside an event-based context?

Answer: A LDES does not need to end up in a view that exposes events. The view itself however will always be kept in-sync through events that happen at the source. So yes, it would be useful outside an event based context.

Question: Publishing sensor observations will need a way to 'forget' observations and objects in the tree. Does LDES provide options for that?

Answer: Yes, a retention policy can be defined.

Question: As an implementer (or when harvesting the infrastructure), it is quite hard to update the LDES when you get the graph, e.g. for a data catalogue, as it contains dataset distributions, contact points, data service etc. As an implementer, you need to decide whether updating a specific LDES distribution means that you will need to consider if the datasets need modified as well. So, implementers can just look at this from a pure link data perspective. To conclude, you need to define what

	<p>the independent objects are and the independent events that caused the existence of these objects in the LDES.</p> <p>Answer: Indeed, the boundaries of what you will add to the stream have to be defined. Nevertheless, there is not only a version-based approach to this discussion, there also is a time-based approach. For the base registries, CRUD notification can be built on top of it.</p> <p>To give an example, if an object is deleted, then this can be used as a trigger to delete a certain entity on a server. In that sense, different management techniques of event sources exist; such techniques just need to be described properly in the applicable retention policy.</p>
<p>LDES implementations – success stories</p> <p>Speakers: Marina Aguado (ERA), Marc Van Andel (Kadaster)</p>	<p>Presentations on the tangible business value that LDES could bring to an organisation. Two guest speakers shared insights on how they successfully implemented LDES as data sharing technology in the the following organisations:</p> <ul style="list-style-type: none"> ● ERA - the European Railway Agency ● Kadaster - The Netherlands’ Cadastre, Land Registry and Mapping Agency shared his insights. <p>- <i>See presentation for more information</i> -</p>
<p>Wrap up and next steps</p>	<p>Webinar conclusion and next foreseen steps by the SEMIC Team in relation to LDES and ABR:</p> <ul style="list-style-type: none"> ● Further implementation of the ERA pilot ● Pilots being developed in the LDES community: <ul style="list-style-type: none"> ○ Pilot with marine data in Europe ○ Internet of Water (IoW) in Flanders ● Second webinar on LDES and Base registries: 'Webinar dedicated to the review of Data Catalogue Application Profile for Base registries (BregDCAT-AP)' ● Survey on the adoption of LDES and their challenges related to base registries: link to EU Survey.