

2022-05-16 Architecture WG Meeting Notes

Date

16 May 2022

Antitrust Policy

Antitrust Policy Notice

Linux Foundation meetings involve participation by industry competitors, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws.

Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at <http://www.linuxfoundation.org/antitrust-policy>. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrove of the firm of Gesmer Updegrove LLP, which provides legal counsel to the Linux Foundation.



Attendees

- Sean Bohan (openIDL)
- Allen Thompson
- Dale Harris
- Satish Kasala
- David Reale
- Dhruv
- Greg Williams
- Jeff Braswell
- Joan Zerkovich
- Ken Sayers
- Nathan Southern
- Padma Govinda
- Peter Antley
- Truman Esmond
- Tsevetan
- Yanko

Call Recording

Discussion notes:

Use Case	Use Case Category	Use Case Description	Stage Data	Request Data	Extract Data	Transmit Data	Combine Data	Analyze Data	Publish Data
Stat Reporting	Functional								
ND Un-Insured Motorist	Functional								
Data Privacy	Non-Functional								
Time Series History	Functional	<ul style="list-style-type: none">• HDS to maintain time series of premium and loss changes overtime• Ability to collapse time series data to create snapshot at a given point in time	Applicable. What are the constraints?	??	??				

Non-repudiation	Non-Functional	<ul style="list-style-type: none"> • Making HDS Non-immutable • Support audit function 	Applicable. What are the constraints?						
Data Consent	Functional			Applicable, Constraints??	Applicable				
Regulators Dashboard	Functional								

Add a row for each use case and requirement (assumes multiple use cases with multiple requirements for each - what is displayed are examples)
For each requirement, include the details/context for that requirement where it impacts the following stages (columns)

Time	Item	Who	Notes

Action items

Notes:

- Introductions (see attendees above)
- Question about nomination of Vice Chair for the Architecture WG
- Presentation of [Use Cases/Requirements working document](#)
 - Use cases are the first column, requirements next and details by "phase" follow - data pipeline
 - Concern from Travelers and AAIS that this model isn't working for some on the call
 - Breaking up the major steps of the process by how the data moves through the system
- Staging the data correctly starts with the loading of the Stat Plans
- The phases of the data pipeline are not overlapping
- Dale (Travelers) - as long as the data is in the harmonized data store it doesn't matter where it is staged
- Truman (AAIS) - how does the system know a member's data is staged and the subsequent request - how do we know that the data you've staged is good enough for this request?
- Dale - if it is staged it is already in the harmonized data store
- Requirement: We need a method to load data into the harmonized data store
- Truman: how do we know the data is there to respond to a request ?
- DavidR: you could query the system to see if relevant data is there or as a response to a request
- Truman: there needs to be a mechanism to assert data is available in the data network (requirement?) and that it possess the degree of quality when it comes time to answer the request, proves it hasn't been tampered with
- Truman: as the data is staged the network needs to know there is data in stage, its degree of quality, frequency, freshness, etc.
- KenS (AAIS) that's the consent
- DavidR - the requirement doesn't imply that I need to make an attestation when I load data
- Joan & Truman (AAIS): describe role AAIS plays in the current Stat Reporting model and how openIDL should make that irrelevant in the future
- Joan: data is submitted to regulatory reporting according to an approved stat plan and there are quality assurance rules that go around the data and regulators look to the stat reporters to validate and attest to the quality of the data that was used in reporting to regulators. When AAIS gets audited, they and the regulators verify data sourcing, did it meet x standards, what was done if there were issues, how issues were resolved, how they finally got to the right level of data quality before being used in reporting. Data at quality of of attestation
- DavidR - attestation of quality can be in the consent - doesn't understand why every time we write we have to that - seems like requirement should be "the data is accurate". Super hard
- Joan: There is accuracy from the Carrier POV and then there is accuracy from how the regulators view accurate - the purpose of the harmonized data store was to create not only the repository but a standard way of storing data with the right quality that meets the needs of the regulators
- David: how does anyone outside of Travelers attest to the quality of our data?
- Joan: <broad description of the star reporting process at AAIS>
- DavidR: so we write the data to the HDS, and if someone wants to query it they send an extraction pattern and we consent and that's our attestation
- Dale: Stage Data is really the pre-harmonized data store? Should there be another column for pre-HDS?
- Satish - if it is in the HDS is it ready to be pulled
- David - we keep the HDS accurate, up to date, some quality assurance - once it is in the HDS we can put explicit meta data that says its fresh as of this date or good as of that date, be explicit about it. Wouldn't be making an attestation every day. If the data is in there, they know their data is good, as of a certain date, and then thats all they should have to say.. If the data is there or not, can do something like a polling query - shallow - keeps the network up to date without having to get into complicated testing
- Truman: something like a simple hash to the ledger that allows the network to know you are an org that has provided data quarterly and the data needed for quarter by quarter for these lines of business?
- Ken - questions scaling, if someone uses an extraction pattern they haven't thought of but still uses the same data
- David: why not say "if request comes in, these are the data elements needed for this period, if they are in the HDS and we agree that aspect is YES, the data is there. Fits the needs of the call, you either do or do not have permission. All we have to do is keep the HDS updated with data that is accurate and in the right format - keep it updated in the right format as fresh as they can and accurate as possible. The data we have is the best data as of this date
- Joan: there is a clear set of requirements about how data must be processed and made available to the regulators for responding to requests for info. These requirements need to be included. There needs to be an audit trail that this carrier provided data according to the stat plan requirements and the handbook and all these other things by the date they are required to by law
- DavidR: these are requirements that we need to have written here - every single one, so we can design a system around those. Auditability?

- Joan: if you could validate the data in the HDS at this point of time you write the hash on chain that it was good enough to prove that the carrier had delivered the data according to the standard in the attestation on this date
- DavidR: can be accomplished in a way I suggested so it is much simpler - doesn't understand the hash and what it does
- Joan: you can reproduce what you had at a specific point in time, maybe your data reporting team can respond to this but copies of data for regulatory reporting are kept in the archives
- DavidR: changing to an HDS, are they being asked to keep snapshots every time someone pulls a report?
- Joan: AAIS had to reproduce what you had at that point you delivered a report, in your harmonized data store there was this data and it matches this hash on this date and it was used for this report
- David: Do I need to maintain that data in my HDS at all times? Not a requirement that we heard from Eric in VA - monster requirements change. Introduces need to maintain HDS in a stateful way as opposed to naive way, maintain ANOTHER system of record based on complicated ELT pipeline - it is a non starter. No desire to maintain/manage state in HDS
- Ken: <example of error handling and policy states - state of the policy at a certain date and corrections>
- David: if it is in a HDS you are doing complicated state management (as opposed to flat file)
- Dale: <example of a 4000 vs 400 error on a policy>
- David: There will be errors, 11 systems data is being fed from, going to do normalization, there will be problems.
- Dale: Not sending new transactions but changing the record
- Joan: Lot of discussion over the years in regards to making corrections which lead to SDMA ? allowing a carrier to correct the data themselves and they have an audit trail that says "Dale sent an email request..." and we change their data / make an adjustment after they have sent it - why Carriers have those capabilities to do editing before AAIS accepts because once AAIS accepts this is the snapshot of the data that met these quality standards
- David: the thing that gets archived is the snapshot - we don't archive and freeze the HDS
- Joan - we will not be creating snapshots, we will be creating reports
- JeffB (openIDL): many kinds of data quality checks
format checks to make sure the data in certain fields pizza for meets the format and the definitions for the data elements there's there's also if you have a requirement that you have at least three months of data for a quarter before you actually generate a quarters worth of reporting. That is a different type of data quality check this as I now have all the data needed for that time period because I have i've added all the elements up to that point in time. David's point in certain amount of Meta data saying well you know when was the last time it was the most recent data I have for a certain type of data. might be used for those types of checks and then the type of checks to Joe that you're referring to is like having produced this data distributing down the sort of reasonable tests that look at the total package and might say. everything looks Okay, or it doesn't look Okay, and what you're saying is that, instead of running those centrally I kind of pre checked can be run prior to a certain type of report that requires consistency on a quarterly or annual basis. That would that could be run as a separate type of data quality check with these are all different types of data quality checks that can and probably should be performed and it's not just any one simple statement saying, everything is fine, because it's a number of different factors.
- Joan: Lot of trend analysis comes out of reports, various agencies request information, 1500+ types of data calls out there. dozen agencies ask for data in different ways,
- Satish: are we saying HDS needs to maintain snapshots? Similar to what happens today in the carrier-sense? Data goes thru SDMA process and gets accepted and thats the snapshot available
- David: What i'm hearing from what you just said to teachers anytime someone takes that data call or a Stat report from my hdfs, I need to make some provision, so that if someone were to make that same request years later, or months later, they get the same thing that's a very that's a very big change.
- Jeff - don't think thats what we are saying
- Satish - however we solve it, it could be as simple as tagging that data set or row of data which could be used in data extraction with date and timestamp
- David: our concern is two things duplicating data right and maintaining an extra state full system are the two concerns right which are impacted by this and that's why i'm still confused as to what we just decided on
- Peter (AAIS) - share this with regulators, understand their context
- David: Those are two as you guys know we start architects to colossally different types of patterns for how to build this if you know independent depend on the onset offset style verse you know basically atl and dump it in there, as long as accurate fits the quality checks with two very different types of efforts.
- Satish: DO we really need a snapshot? (after suggesting some form of freshness dating)
- Ken: not doing a snapshot but we do need to reproduce a report
- Jeff: tremendous difference in types of requests. with the the the the correctness of this has to do with passing you know, meeting the requirements for the data definitions for the formats for the. amount of information is contained, and then the check to make sure that prior to submitting something is going to be used for quarterly or annual report that it meets certain over overarching consistency report, you know requirements just edit checks that that nature.
- Dale: those can be conditions of the request
- David: lots of these types of requests, distinct and ad hoc ones. The idea that we could pre determined everything we write in fits all these use cases is going to be hard, if not impossible, to do the best we can do is say "This data fits all the checks that are universal it's formatted correctly it's you know of the right, you know period it everything you tell us that this data must be". Now if it fits all the specific bespoke requirements of a given data call that's really going to have to be a test that's run on extraction pattern or the right or the extraction pattern can make a test pattern for the carrier to run against ahead of time - but we are not going to be able to work that for every pattern as we load it is going to have to be on request
- Jeff: templates or profiles , specific needs that you say that "this type of request requires that this data exists in the system in general and has been supplied and if it's not then you you won't. You will be able to to fulfill that particular request and so some some form of metadata management for the types of requests as a general rule in the future to to know that you can supply or not, as opposed to coming back with blank columns, for example, and just try to do something.

Recording:



openIDL_ArchWG...1622_video.mp4