

2022-05-09 Architecture WG Meeting Notes

Date

09 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)
- Satish Kasala (Hartford)
- Adnan Chourhury (Chainyard)
- Isaac Kunkel (Chainyard)
- Megan Ebling (AAIS)
- Truman Esmond (AAIS, TSC Chair)
- Jeff Braswell (openIDL ED)
- Ken Sayers (AAIS)
- Greg Williams (AAIS)
- Joan Zerkovich (AAIS)
- Nathan Southern (openIDL)

Call Recording



050922_openIDL...rchWG_Call.mp4

Discussion notes:

- Data Model WG Calls are suspended while Architecture works on progress
- Truman:
- As applications emerge, Architecture will evolve
- Architecture that is good enough for primary use case
- From TSC call on 5/5: can't address all use cases up front, address 1 & 2
- JeffB
- Right to focus on plan to address infrastructure
- what are the other types of patterns they MAY exists that may effect it in some fashion
- Truman - first one being
- JeffB - stat reporting
- Truman - suggest we create WG to target that, another for North Dakota DMV use case
- are there openIDL facilities to do that, helpful to illustrate how we are doing it, stateless type of use case (not stat reporting)
- Stateless: every day, does this vehicle have insurance - yes or maybe or no
- Only historical if you save it
- David wants to get to maintain state of every policy as good as you can
- Ken
- David agreed that history is part of the deal
- Truman - for stat reporting it must be
- Jeff
- North Dakota shouldn't be a heavy lift, important to have knowledge, just doing something
- Truman
- Suggest, Mississippi - addresses, kept private, to solve problems
- USAA - if you can do MISS, you can do stat reporting which is automatic
- Jeff
- purpose for setting out those use cases
- Truman - what we got to, DMWG isn't considering anything but stat reporting
- ND DMV project is very simple
- Jeff - whats done could be used for a variety of purposes
- Truman - whats the extraction pattern of the DMWG that would provide the report
- what are the data fields the report extracts to
- did the data model - didn't get to the report
- data in data store, so this query can run into each. node
- what you deploy with one policy vs large carriers
- diff arch, diff choices
- any datastore you want
- spreadsheet, policy admin system, whatever an org trusts
- then need to attest to the level of qual to the network
- What arch are we going to solve for (examples)
- Ken - similar opinion for this group
- evolved to the point where we need travelers and Hartford want to do POC solves reporting, know it doesn't solve all
- spin off - if we retile this to be stat reporting Arch WG, we know focus, but shouldn't be the only focus
- need to have a focus that says, this is the focus of this group, can get feeds from other groups
- HHT want to be distracted by those other things
- hosted node, adapter running
- Truman
- whats that architecture going to do?
- Ken
- regulators (Eric, George) - we assert process not the actual data
- think about the process
- Truman - challenge our own internal processes
- Regulators want to go faster, but requirements are we reconcile data, which we don't get until just now
- we don't have the checksum to deliver faster unless they say "we will take it faster and lower quality bar"
- JeffB
- this WG should be oversight and have other WGs for use cases
- take into account other considerations
- if you say this group is only stat reporting and other groups
- Truman - we need to have room for other use cases
- JeffB - to use the example, we are all building with gravity but need plumbing and sewers, etc.
- Truman - data privacy and information quality, provenance of the answer being provided
- JeffB - dont we have to deal with the infra and technology
- Truman - done
- Ken - no we don't, we have. a picture, no hosted node - no adapter - now lets make them real
- Truman - not just define data model in harmonized data store but also how it operates
- walk it all the way through - one report for 52 states, history is there, looking at queries run against stat data
- Ken - ND can be done with existing arch with some security mitigations
- different thread, discussion not really architecture based
- Truman - transaction patterns lightweight,
- Satish - stat reporting and other use cases (doesn't know about the ND use case)
- expand on use case for this is how it will fit
- what is the real scope? Stat reporting, ND, Mississippi?
- is this (Arch WG) beyond stat reporting?
- Truman - wanted to focus on stat reporting in the last meeting, deal with other projects as they come, not yet mature, and open to other members
- Joan - questions - openIDL Arch Working Group is looking at, has a role to accept request for info, extraction and maybe transformation of that data, sent back to openIDL platform to receiver
- in the case of stat reporting, platform needs to receive the request, process on carrier's side, get answers back to the requestor
- North Dakota Drivers Licenses - similar: request for information, processed, report sent back out
- describe use case in terms of core functionality, see if applications use of openIDL
- doesn't seem to be wildly different functional requirements between the 3 use cases
- Ken - didn't mention non functional requirements

- when we put the whole node (as previously recommended) found out it was a no-go
- arch that allows data to be private, functionally we can do the ND POC without refactoring, but do not have data privacy at the level we want
- Joan - lets not get into details of ind use cases
- if you want to separate where data stored and processed, if we take on that issue, and resolve it, and then we do desktop exercise of anything not be done for MS or ND or stat reporting
- still doesn't think there is a hinderance
- hoping group can focus on separating the data processing from the openIDL node - focus and test arch against the 3 applications we have in front of us
- Ken - think what we heard from RR WG and TSC - lets not put datamodel in front of that, do it with Stat Reporting to get agreement
- JoanZ - can do ND today without this, if you reach you still can do ND DL or Mississippi Crisis track projects
- Ken - diff is the data model, can't support with new architectures yet
- we want to do the ND with the new data model, doesn't see reason why we can't put new data model in, said to use current stat plans as data model
- very easy to support multiple data models, nuance but it is there
- technically as soon as we have stat reporting working with old stat plan, new use cases are trivial
- Truman - could do stat reporting, accelerate timeframe, (2021 homeowners on Friday), would make them happy to make it faster
- simply stat plan and same reports
- JoanZ - for this Arch WG - if we focused on splitting the platform in 2 - data repository and processing of that data, secure and private in carriers data center and other part of open IDL is participating in network receiving requests and sending responses
- what are the steps that need to be done to split the platform in two to meet the needs
- unaware of any application that would be stopped, could still move forward with three applications in front of us
- Satish
- trying to document what steps
- Stage data (carrier), request data (regulator), process the data (carrier, AAIS, etc.), extract data, combine data, publish data
- structure the thoughts and arch as well
- ken - need to meet non-functional requirements of timeliness, quality
- JoanZ - multi-function platform, network, with nodes, deliver functions, don't tailor to any one use case, provide platform that meets needs and can support multiple
- Truman - lay out each use case - what data needed, extraction, publish: by staging 1 kind of data set can work with multiple use cases
- one address applies to multiple use cases
- this one needs monthly, this one annual, this one quarterly
- JeffB
- data pipeline is right - needs to have "communicate or transmit" added -
- Truman - need to connect openIDL pipelines, now from the data owner/node delivers to analytics node, current use case analytics node is run by AAIS
- we run compilation manually - regulators, thats not how they get stat reporting data, they dont have an analytics node yet, can combine data in tableau report, give Regs access to that
- JeffB - extraction - extract data from HDS based on request, needs to be collected and x before combined
- Truman - request, extraction pattern goes to node and they consent, they say "yes" and then extraction occurs, sent to analytics node
- JeffB - receipt of request locally
- Truman - transmission baked in
- after extraction data piped from data owner node to analytics node
- current model is manual process
- JeffB - messaging arch of sorts, diff types of data, particular report, streams of data, hash written onto chain, one of the things to address: what is the architecture/schema for sending data
- Ken - that's private data collection
- Truman - payload of what these things are,
- JeffB - diff types, our responsibility of how to distinguish that data
- Truman - each use case does, yes
- perform function of the functions
- put whatever payload in gossip protocol, gigs of data, stream vs batch
- Jeff - suggesting clarity that aspect should be documented
- Truman - pushed that description to each use case (privacy, access, granularity, freshness, etc.) what does that use case support? what does the arch needs to support (or enhance) to support new use case
- is it good enough for what I need
- purpose defines arch
- ken - if we can say time series and format are orthogonal to primary architecture, define use case and constraints - this group (Architecture WG) focuses on the infrastructural arch (HDS, extract patterns etc are use case specific) but movement of data
- Truman - whatever is necessary for purpose each node is participating in
- ways we can make this very simple
- solve problems
- regardless of use case - biggest issue - is time-series
- what openIDL does is provide assurances of data owner, staged according to use the data will be inc into
- whatever network gets passed a token for the data = "the data is this good"
- some companies giving data quarterly or monthly - if we simply had check that says all companies met the rules
- Ken - verification can change, reporting and combo steps can change
- Satish
- data pipeline - staged-etc. - that would be on the x-axis
- y axis - do we need time series in this stage, do we need in combine? (example)
- data privacy?
- grid
- Truman - transparency - first is control
- if data owner agrees to do that, select-all, publish - may or may not give them privacy
- privacy from anonymity of each pub of data
- analytics node owned by AIS (right now) - could know who reported what/when
- Satish
- next meeting
- data pipeline steps are agreed upon
- list the qualities which could be a use case for each step in the pipeline

Use Case	Use Case Description	Stage Data	Request Data	Extract Data	Transmit Data	Combine Data	Analyze Data	Publish Data
Stat Reporting								
ND Drivers License								
Data Privacy								
History								
Non-repudiation								
Data Consent								

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

Recording: (on cloud - auto transcription)