

5/31/2022 HDS Task Force Meeting

Date

24 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)
- Peter Antley (AAIS)
- Jeff Braswell (openIDL)
- Nathan Southern (openIDL)
- Greg Williams (AAIS)
- Ken Sayers (AAIS)
- Megan Ebling (AAIS)
- Rajesh Sanjeevi
- Satish Kasala (Hartford)
- Truma Esmond (AAIS)
- James Madison (Hartford)
- Allen Thompson (Hanover)
- Tsvetan G (Senofi)
- Joan Zerkovich (AAIS)
- David Reale (Travelers)
- Dale Harris (Travelers)

Agenda:

1. Discuss HDS needs, requirements (see notes below)

Time	Item	Who	Notes

Action items

Notes:

[HDS REQUIREMENTS WIP IS HERE](#)

- What happened last week
- First Meeting under the name "HDS Task Force"
- first formal

- prior had been watercooler discussion around data modeling
- going forward - what is the HDS
- start back and not spend time the group has done prior - how do we move HDS forward
- James
- work on requirements (ask to send)
- Useful to look at a couple models, instead of one
- risky to try to do all in one layer
- balance
- useful to have notions of the b-model at highest level, persistence model, loading model
- business, presentation, persistence, loading
- b-model
- largely like list of elements, meaningful chunks
- presentation - optimized for writing queries
- some business but not concerned about system performance
- Persistence - preserving history and truth
- ultimate definition of truth, not necessarily easy to use
- Loading - get the data into the system
- compromise
- loaded data into persistence, presentation asks "what is easier to consume rather than truth"
- tried to build dimensional model that preserves all history is hard
- Business Model
- implementation and perf agnostic
- one step down from data dictionary
- Ken - excited, loading model is different
- James - entitles, attributes and relationships understood by business users
- biz user willing to dive in, but understandable
- no regard for implementation, tech, or processing req to load it
- Element - single piece of info
- entity - set of elements
- business entity that is not recognizable to business users and tends to occur in conversations - policy, claim, vehicle, and home are typical examples
- business model has to be careful about not getting crazy about abstract entitles
- relationship - use of entity names "user has a" "policy is a"
- model must not repeat the attributes of an entity in more than one place
- model must make it difficult to over-count the elements and values
- prevent excessive de-normalization
- if you take premium, and repeat premium for every claim, could over-sum
- reasonable level of denormalization
- business model should make it difficult
- sparsity is the measure of empty values to total number of values
- put homes and autos on the same row
- totally different attributes for each
- silly key, but bunch of elements filled out for one, others filled out, bunch of blanks
- humans love flat and wide
- put things together that are unrelated get sparsity problem
- not defining 8 normal forms right now
- will define in 3 min
- no entity when populated will have sparsity greater than 10%
- if occurs, entity reviewed
- entity that is highly covariant with another entity may be embedded in it
- Peter
- when get 10% for #11 - how do you come up
- James - throwing darts, consensus is what we decide
- "should be reviewed " - conversation trigger
- rationalization for normalizing, not just theory
- can make address its own domain
- most b people - its policy address (or garaging address)
- saying - when some entity is highly covariant, doesn't do much sep from it,
- denormalization argument
- #13 - laundry list, worth enumerating entity dictionary
- seed list
- Policy, Vehicle, Driver, Coverage, Claim, Claim Event, MORE
- list of what they are and what they need
- grouping data dictionary elements
- intro entities - intro relationships
- policy has one of more vehicles
- vehicle has one or more drivers
- policy has zero or many claims
- many claim events
- MORE
- Kudos from all
- Ken - great guidepost for coming up with a good model but not necessarily requirements
- needs to... <examples>
- Dale
- is that the requirements for HDS
- see this as a lower level than HDS
- how the HDS would be created but not necessarily the requirements
- will be helpful in guiding development
- still need bullet list requirements

- David
- this is great - good way to once we know we want to build, loves most
- should be sanity checks and guideposts
- most
- to answer any one question - what is the HDS supposed to do
- level of sparsity or normalization vs... if I don't have a stated biz req for what HDS is meant to do is touch
- might not need to, business requirements are more lax
- not sure what HDS is supposed to do
- Peter
- must support state DOI such as auto coverage and auto territory
- Truman - can't do "all DOI" must define each one
- Ken - short cut in the meeting
- Truman - cant skip that - critical to answering each question
- any two data calls
- annual, last year, year/half over course of year or a point in time
- cant go backwards
- Joan - is it possible that theres types of question, not answer every one, but fall into simple classes
- annual data call, one set of data 100s of questions answered from that set of data, very static
- other type - whats the state today - only a few classes of questions or sets that are queried
- simplify
- Truman - agree , not each state report, couple diff ones, returning over historical period that exists or a snapshot
- David - decouple it
- really important
- HDS should not be referencing specific use cases
- should be ref by the DMWG
- define DM that fits known use cases and predict future use cases
- implement and expose data elements
- any other requirements - easy and extensible
- don't define specific things
- requirement - HDS must implement data models developed by DMWG
- Dale - data must conform to DMWG data model standards
- David - it shall store and expose, make it accessible, make it real
- gist of it
- Dale - data should be current through 45 days
- David - Dales closer
- then we need to go one layer deeper
- these are decision points
- middle layer missing - perf of HDS
- bridging gap of what Dale wrote and this
- Peter - thinks DMWG will need more - very far from technical
- David - closes to business requirements
- user stories, non functional requirements, business requirements
- thing that must be done
- data that falls out of that - what we need to implement in HDS
- one layer deeper and above
- what are the requirements
- we don't need to know every change, just know and turned into a req
- req. - data is current as of X Date
- shall be exposed by HDS
- discuss them
- or could say "performance - what is acceptable - do we need instantaneous reads, what level of adhoc query do we need
- to figure out
- then decide how to build DM based on that
- Satish - traceability of the data
- trace thru HDS and back to source system
- implementation - could be keys
- others
- how easy it would be for the data calls to discover the data in hds
- do we have a req where data call or extraction pattern, need to know what is in HDS
- what attributes do we have
- what is the level of relationship of those attributes
- Ken - model we all know
- Satish - discoverable in realtime?
- do we have discoverable data requirements? confirming to a model
- David - a business requirement - what the user experience is for the data callers
- what the expectations are
- start from the user experience and user stories and work our way down
- how will it be used and how will carrier load into it
- two user stories not terribly well defined today
- what will that look like, how structure
- not defined
- Satish - must provide data current to 45 days prior - hinting at historical data? do we want to explicitly define history we maintain?
- Dale - 5 years + current year
- Peter - want to say CGL and NJ have weird reqs - closer to 7
- David - HDS will store data as mandated by regulatory statutes
- Truman - does it matter if apply to each co differently?
- David - not every req needs to be enforced by openIDL, on carrier to und that
- requirement - HDS must be current and have hold back period of x years, then the carrier can decide how they want to implement
- carrier can decide how to implement, req "it must be there"

- Ken - must meet retention requirements
- Dale - prior + 45 days
- processing, get thru it, makes sure it is right, dont get feeds every day, some monthly
- David - hold off period to make sure it is right
- what we need + req
- Peter - trailing 24 months right now, 45 days will make all real happy
- Dale - edit package, open IDL sponsored and maintained edit package the info goes through prior to landing in HDS
- must maintain min of 5% tolerance at state line level
- similar to SDMA
- Peter - can do based on SDMA
- Joan - part of NAIC req
- talk about what data needs to be avail for RR - 5% tolerance as well
- David - edit "before being in HDS" - when we load HDS could conform to 3 packages and fail 7 based on data call, could conform perfectly but their could be a new data call that doesn't
- Ken - example - provide VINs, I don't provide vins which would fail that call
- David - run against data in the HDS
- Peter - wont have NAIC data to get to the current reconciliation wont have long after 45 day period
- Ken - are we conflating 2 reqs - put data in with quality but might not meet reqs of data call
- 2 diff reqs
- Satish - is there any req on backwards compatibility
- DM will change over time
- are we mandating HDS is backwards compatible with previous vers of the DM
- Peter - in the camp of extraction patterns written against current iterations of model
- extraction model should tie to one model
- David - prob DMWG will update data model, will happen, all my old data when it updates might
- might not have existed on a policy 7 years ago
- some issues there where - needs to be define
- Ken - data model needs to change over time
- Satish - how does HDS need to behave
- Dale - even though conform to data model, not every attribute in the data model
- minimum required elements provided for everyone, all others at the option of the carrier
- Ken - ND example of being very sparse and still participating
- Dale - if doing stat reporting, minimal amount of data required
- David - almost think those reqs can be handled by the queries, HDS for simplicity only has to adhere to DM going forward
- if DM changes, anything in the HDS can optionally be updated
- Dale - even the fields in the dm today, not all are required - just Day 1s
- David - query tells us if it is needed for a specific use case
- simple req on HDS - required fields only mandated to be updated when a new DM is released
- could always go backwards
- example - middle of 2023, DM updated, we might just re-run all 2023 against the new DM, but unlikely go back to 2019
- query checked - what data is there
- HDS shall only conform to DM going forward not backwards compatible
- Ken - how does new data relate to old data
- worries about migration of the model
- informs how we build the model
- ways to do it
- extensibility - lead to - less normalized more extensible kind of format
- sparsity but you could have struct of db not change with new data elements
- 2 very diff architectures
- Jeff - more of a graph approach with facts vs schema/tables
- David - more JSON like
- document-like
- easy to start loading new data in year 5
- query old data from year one
- ignores if not there
- whats there passes the check
- less performant
- with a performance requirement - would let us know if an acceptable solution
- Satish - carrier - change to the DM, v 2 - carrier 2 is still on V1, don't / wont migrate - how does the extraction pattern work?
- David - 2 EPs or 1 that checks for both versions, decouples DB from use case better
- if I am VA, running same data call for five years, only some carriers have new data, write EP that that hits both
- Ken - SLA approach - everyone shall conform by X time
- more consent, fewer dark corners
- David - carrier concern - whole value prop, set data, load HDS, takes churn off of us
- updating dMs every year, req fields added every year, looks less attractive, sla requirement
- very hard, erodes value prop
- Ken - aspire to migrations for those things, easier to run
- some of these will be trivial
- most part - once solved model
- adding fields not reorganizing
- David - governance and guidelines "not requesting carriers add more data every couple months"
- know there will be iterative
- needs to be sensible, but cant be constantly asking carriers to add more mandatory data
- Dale - there will be a governance process inside each carrier
- new item doesn't happen overnight
- David - unfunded mandate
- cant commit to mandated field from openIDL, internal gov processes for data David cannot guarantee success on
- mandate - these extra fields are mandatory, cant go back to. a line of business "put this in its mandatory" - they will say no
- even if you make it optional, too much drift/variation of data in there, it will be a mess

- major concerns here
- Jeff - standard model doesn't change too much, when they do it is thought out and structured
- high level reqs, not HDS but inform "how we make it extensible"
- Peter - if we add a new element shouldn't be backfilling for new model
- new attributes
- Jeff - policy - should you do it or not
- not just add a field and have a diff iteration of the model every month
- Peter - auto guy wants "yes/no" for usage based insurance
- Jeff - anticipation of structure vs availability of the data
- update history - - policy question
- more work
- concept of extraction request - federated and exists within diff carriers
- not a simple topic
- David - goes to why concerned about moving forward, if this is mechanism by doing data calls and asking for elements outside of standard - needs to be mechanism
- right now don't need to work about data call not matching stat reporting requirements
- needs to be easy and extensible to respond to data calls
- unless we expect regs to only be choosing data from regulated list of data models
- Dale - info not in HDS< may fall for carrier to do directly rather than openIDL
- never going to have all data calls going thru openIDL
- David - if it was really seamless, regulators look and say "i need this one element" and could easily sqy "we can add that" it would make our lives really easy,
- used for adhoc data calls
- if it is hard, can go around openIDL, less attractive
- hard to predict
- consider - might be beneficial for long term, 99% of data for data call, easy for carrier to add an element would make it easy for carriers
- Dale - picked up a lot of things for regulator in Day 2 and Day 3
- David - covid 19 - no one saw that exact data call coming
- always bespoke?
- Dale - might have been able to do it thru statistical data
- might have needed to get to coverage codes
- Peter - influenza of 1920 - 100 year disease
- David - req - we aren't going to cover every possible data call, trying to get the vast majority
- Peter - homework - getting regulators back together Friday
- short call
- altogether in the room on Friday
- talk about queries, sample questions
- agenda?
- Dale - working on what he views as business reqs
- 3 buckets
- data, data integrity, access and security and comms
- vetting with his folks at Travelers

Recording:



openIDL_Arch_H..._TF_053122.mp4