

# 2023-01-13 Meeting Agenda

## Date

13 Jan 2023

## Attendees

### 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.



This is a weekly series for The Regulatory Reporting Data Model Working Group. The RRDMWG is a collaborative group of insurers, regulators and other insurance industry innovators dedicated to the development of data models that will support regulatory reporting through an openIDL node. The data models to be developed will reflect a greater synchronization of data for insurer statistical and financial data and a consistent methodology that insurers and regulators can leverage to modernize the data reporting environment. The models developed will be reported to the Regulatory Reporting Steering Committee for approval for publication as an open-source data model.

openIDL Community is inviting you to a scheduled Zoom meeting.

Join Zoom Meeting

<https://zoom.us/j/98908804279?pwd=Q1FGcFhUQk5RMEpkaVlFTWtXb09jQT09>

Meeting ID: 989 0880 4279

Passcode: 740215

One tap mobile

+16699006833,,98908804279# US (San Jose)

+12532158782,,98908804279# US (Tacoma)

Dial by your location

+1 669 900 6833 US (San Jose)

+1 253 215 8782 US (Tacoma)

+1 346 248 7799 US (Houston)

+1 929 205 6099 US (New York)

+1 301 715 8592 US (Washington DC)

+1 312 626 6799 US (Chicago)

888 788 0099 US Toll-free

877 853 5247 US Toll-free

Meeting ID: 989 0880 4279

Find your local number: <https://zoom.us/u/aAqJFpt9B>

Sean W. Bohan  
Nathan Southern  
peter antley  
Hicham Bourjali  
Jeff Braswell  
Brian Hoffman  
Susan Chudwick  
Dale Harris  
Michael Payne  
peter antley  
Sandra Darby  
Jenny Tornquist  
James Madison  
Allen Thompson  
Ash Naik  
Mason Wagoner  
Lanaya Nelson  
Reggie Scarpa

## Minutes

### I. Introductory Comments

#### A. Welcome

#### B. LF Antitrust Statement read by Peter Antley

### II. Official Business

#### A. PA: last week we looked at commercial auto

1. Tracked down some test data; looked at stat plan
2. Realization: Using same stat plan as personal auto but 2 columns interpreted differently.
  - a. Personal has a column for good student discount and for private passenger/penalty points
  - b. Commercial has two columns for automobile classification/commercial automobile use (line 35)
  - c. Otherwise they are exactly the same tables
3. We're looking specifically at Position 34. We have codes 1-5 coming in for Personal Auto, and Codes 1-9 for Commercial. We interpret 1-5 and 1-9 a bit differently. Key difference: Subline 1 is personal auto and Subline 2 is Commercial.
4. Position 35 - Private Passenger Penalty Points vs Commercial Automobile Use
5. PA key question: do we want to have a separate policy and a separate claim table for commercial? Or do we want to just have some extra columns and put all the auto in together? This discussion may bridge the gap between AWG and business matters. PA asked James his thoughts:
  - a. JM: this is half business and half architecture. The two different lines (personal/commercial auto) are fairly similar, but key differences exist. There are two ways to solve it - and no "free lunch":
    - i. Multiple disparate concepts can be put in the same table (in the same record) but if we do that, we pick up a sparsity problem - which is - if you have the concept of an auto (and it's insured auto, because auto wouldn't know the concept of personal vs. commercial) - i.e., we've embedded two concepts, here. We have the concept of a covered vehicle, and we have to put attributes in it. A covered vehicle will have a VIN - same for both personal and commercial.
    - ii. We have numerous non-sparsed columns. They will all be 15-30-80 - things affiliated with the auto itself. At some point we find ourselves with attributes in one or the other that don't apply to both. We might have two attributes that are basically but not completely the same for personal/commercial, and it needs to be filled in differently for each.
    - iii. We typically say - all things being equal, keep things simple - but embedding in one table is *not* simple. But neither is making lots of tables simple - they need to be connected.
    - iv. Most people prefer flatter, wider tables - and one of our guiding principles on openIDL is that we go in this direction unless there is some reason not to.
    - v. If the sparsity problem is broad, however, we may want to start breaking up tables in that instance.

vi. Here JM invited thoughts/feedback from the group:

1. DH: right now we're getting one auto file right now that has both personal and commercial. He is unsure if this lends to one or two tables. The other consideration: as they went through the commercial/personal auto plan, they ended up with many more fields than are available personally. Longer term we will have many in personal that we don't want in commercial and vice-versa.
2. JM: a business question to ask: -how related are these business concepts, but also: -how many common vs. disparate attributes are there? And architecturally do we spread this out?
3. DH: Confirmed that the stat file they send to AAIS has both personal and commercial - the reports that they send out are not necessarily the same. They break it up. (SD: there is a field that distinguishes personal from commercial).
4. PA: called on JT who confirmed - coverage report does have both the personal and commercial on it. (Submitted to NAIC combined for filings).
5. JM: looking at this philosophically from perspectives of pattern and scale. The problem is that if we look for example at position 34, and only viewed this position, we would have no idea if a given categorization is personal or commercial. Anytime we have this behavior, where we have to look at another column to understand an initial column, this drives business intelligence (BI) tools crazy, because this is an embedded rule. The upshot of an embedded rule is saving technical space, but the presence of BI tools (e.g. Tableau)- looking at one column to grasp another - creates numerous issues for the tools. *We generally want to avoid this, but will pick up table complexity as a downshot.*
6. JM: The best option may be to remove as many embedded rules as possible, and pursue building/adding additional tables as an alternate solution. In theory, numerous tables should not pose a problem.
7. PA: in building extraction pattern, he started off with only personal auto. For a while, they've been talking about doing premium and loss tables. They changed the names however to auto policy table and auto claim table.
8. They currently have an auto policy table and an auto claim table. When PA runs his extraction pattern, the first thing that happens: it creates a temporarily table that merges these two together to establish queriability. If we broke commercial out into its own thing it wouldn't really change the extraction or merge process. PA: very much leaning toward breaking commercial out into a separate table.
9. JM: agreed - his recommendation is pull these out. Avoid embedded rules which will create much bigger problems later w/slicing and dicing data.
10. PA: interpretability would be the same; we are eliminating the sparsity. Because after ETL we are abandoning the If /then/that and we will be decoding. Once he loads into the table it is no longer positional.
11. JM: In other words, PA is eliminating the embedded rule issue by introducing the sparsity behavior. PA: by splitting into columns, we are eliminating the ambiguity problem. Several options: We can embed the rule, or we can introduce sparsity, or break up the tables (these are the three basic options). This is a design decision. from a business perspective: we have to say to the business, *how many of these attributes are you going to have? If it's 3 or 4, you live with the sparsity. We do not want embedded rules.*
12. PA: We have no embedded rules.
13. JM: should we do a sparsity analysis and live with reasonable sparsity? This is a business decision.
14. PA: As we push and start to make this bigger, how many of our Day 2s-Day 3s will include a lot more sparsity?
15. JM: we should have a threshold, however. (No more than a dozen items each? - Rule #2)
16. JM: We need to ask the business, however, how they see this evolving in the years ahead and it scales up. If it's low sparsity, keep it flat. And if it gets high sparsity, we need to break it up and accept the complexity of more tables and joints.
17. PA: wants to revisit this equation next week after we get our Day 2s and Day 3s - doesn't have offhand what we're going to add in future, in a mature state, and what will or won't contribute to sparsity.
18. AN: Should we look at a major overhaul or upgrade to start, and then come back to where we are? Assuming that we take approach 3, then this is a very sophisticated (involved) solution - breaking tables appropriately, etc. we have to consider the fact that there are multiple columns, entities, sparse information btwn tables over a period of time. We want to get there in the future, but for now, we need a solution that fits in-between.
19. JM: We did this early on, but ran into concerns about the complexity of the model. If you take the time to brainstorm a full domain model out, you won't miss something - paint yourself down a path and wish you'd thought of something a year ago.
20. Review of previous path done early in the project (along the lines of Ash's proposal)
  - a. Dale Harris was the business contact here
  - b. We started with the notion of a policy (foundational notion in insurance) - and categorized into personal policy and homeowner policy, except they have a commonality as an <agreement> and a commonality of a <package> - agreements can be packaged with other agreements. Personal has a vehicle, which hangs off the policy. Same doesn't apply to homeowners.
  - c. One might have a claim against any type of policy. We hang the notion of claim off the notion of agreement, so that any claim and policy can be connected.
21. AN: If I look at stat reporting, the purpose is to figure out how are policies priced and what does pricing look like, and how does it look in the market currently? This is different than the way policies are structured. What are the factors influencing pricing? e.g., passenger, driver and geography characteristics - then build all these variables and take them into account for pricing, and figure out how these two merge. Abstracting it out with C++ will lead to a covered entity. AN: interested in having this discussion.
22. PA: wants to ensure that we have the proper raw tables to fulfill our own business needs. If we break out commercial auto into its own entity now, that's an easy move, and will help us avoid problems later on.
23. JM: agrees that separating them out (bottom up - as we encounter it, we do it) for "cleaner" results. However wants to have the discussion in the case of each line "where do we see this going?" In other words, having a vision discussion (anticipatory) - we should look 1, 2, 3 years out if possible. PA on the same page.
24. DH: wrt commercial vs personal - on commercial side we aren't necessarily tying a driver to a vehicle. Ensuring vehicle not the driver.
25. Vehicles Identified through VIN, not DOT #. On personal side, they assign a driver but this is not necessarily how rating is determined/applied.
26. JM: the notion of a driver interesting - what if we have multiple drivers per vehicle (can be true of personal or commercial). DH: Underwriting based on driver records, but not necessarily rating based on the driver record. JM: yes, but this is where we start to break the driver out into a separate table.
27. DH: on the PI side there is the assignment of driver. JM: feels like this is worth breaking out, because any number of drivers can be based on any number of vehicles.
28. RS: on the PI side every vehicle gets assigned a driver, and it's typically the highest rated.
29. SC: Yes, they are assigning a driver but this isn't necessarily the way they are rating it.

30. JM: stat layout is very constraining. These are the same constraints we would face ad infinitum if we built them in. Are the insurance companies the ones limiting themselves to 1 driver/vehicle, or is this just the product of a stat layout?
  31. DH: we're limiting ourselves because of stat.
  32. JM: Normalized model provides much greater freedom - potential to find bizarre, obscure affiliations that will give us a major lift.
  33. DH: The other reason to keep personal and commercial separate is the notion of composite rating in commercial - where we don't have a list of individual vehicles, just a list of many grouped together that are similar. So we don't get all the details.
  34. JM: A more robust model will have broader functionality, but is harder to do. Version 2 of model might need more normalization.
  35. JT: We're just working with stat because this is what the NAIC wants, and it's very restrained. Will there be compliance issues re: including fields that fail to comply with certain states?
  36. JM: Even if we use stat, we may want to shred the data into normalized tables later on. - They did this at the Hartford, but it took considerable effort and was expensive.
  37. PA: If we normalize, then we can load with Json, and load multiple tables very easily.
  38. JM: Yes - it will add considerable power, but may encounter friction with large # of regulatory bodies that have been stat for decades and do not wish to change.
- b. JM: Conclusion: if we truly want to use openIDL as a major platform for deep insight at the industry level, we will need to break forward with a more complex model.
  - c. JM: Right now we still don't have an NVP - we agreed to goout with stat, but perhaps we should return to the Dale Harris spreadsheet that groups large #s of attributes. If we can start doing this, spreadsheets are business-y enough that they can capture all the attributes and what entities do they go with (i.e., Day 1, Day 2, broken into policy claim and vehicle) - JM: that approach is a highly effective one. If we do this and avoid embedded, and can chunk them together, we can go behind the scenes and ask 'How much do we normalize it'
  - d. DH: Day 1 is really our stat plan. But Day 2-3 we're looking at adding 29 new variables - e.g., policy address. The notion of driver from a commercial side goes away. Most of these will stay with the commercial.
  - e. JM: We will make a case-by-case basis for when we break things out. Believes we should break commercial and personal auto apart, and break driver out of both of those.
  - f. JM: Vision discussion. (PA asked does this mean breaking out across all lines and having person and policy separate?)
    - i. Whenever we talk about people, we get into very ambiguous fields.
    - ii. There is a legal constraint - you have to know who you book with even when it's wrong
    - iii. If you embed a homeowner concept into homeowner policy, you meet your legal compliance and you are done
    - iv. However, if you want to do something w/homeowner beyond this, you'd break it out, so you can have multiples and because you want to know something about this entity.
    - v. Can lead to entity analysis
    - vi. JM to PA: yes, it's the right question to ask but it's a messy case, because people have a lot of ambiguities that vehicles don't.
    - vii. Key question: looking ahead, down the road, are we going to want openIDL to function as an entity analysis?



GMT20230113-1...1920x1080.mp4

## Discussion items

| Time | Item | Who | Notes |
|------|------|-----|-------|
|      |      |     |       |
|      |      |     |       |

## Action items

