HDS Requirements WIP

Working page to draft requirements.

1. **Business Architecture Requirements**
   a.

Use Case Description	Staged Data (HDS)	Request Data	Extract Data
	Data contained in the HDS will conform to OpenIDL data model standards	Data shall be aggregated during extraction	Data extracted is for single defined use based on data request and agreement to access data
	Minimal data attributes to be available consist of the "Day 1" OpenIDL data model fields, other attributes in the OpenIDL data model are populated at the option of the carrier	Data shall be anonymized during extraction	Data to be extracted, aggregated, and anonymized as defined in the data request
	Data shall consist of policy and loss transactions over the course of the policy term and lifetime of any associated claims based on source system activity	Only aggregated and anonymized data shall be transmitted beyond the private analytics node	
	Data shall be current to the Prior Month + 45 days	No Personally identifiable information (PII) data shall be transmitted beyond the private analytics node	
	Companies shall maintain data in the HDS for 5 prior years plus current year	Timeframes for data to be included in aggregation shall be defined	
	Data in the HDS will remain within 5% error tolerance per line and state based on OpenIDL edit package	Attributes to be used in aggregation shall be identified	
	Access to data is via permissioned access	Logic for extracting data shall be defined	
		Calculations to be used in aggregations, analysis, and reporting shall be identified and defined	
		The specific use of the information shall be defined	
		Permitted accessors to the information and users of data shall be defined	
		Proportion of individual carrier information to be used in extraction to the population of data in the extraction shall be communicated	
		Only authorized approvers may commit carrier to a data request	
		No altering or embellishing data including appending outside data is permitted unless approved by carrier	
		Requested data shall be for one time use only. Additional uses for data will require a new request.	
		No changes to request, attributes used, extraction patterns, accessors, users, or specific use of the data is permitted post consent	
		Data request communication shall be through a communications protocol within OpenIDL and archived within OpenIDL	
		Individual carrier contribution to a data request will not exceed X% of the population of premium, losses, exposures, etc.	
		Data calls have an expiration date from which consent is needed, if applicable	

2. **Strategic Architecture Requirements**
   a. General

- Must store and make accessible data elements provided by the data model working group.
- Must conform to the OpenIDL data model standards. (?)

Query Execution

- Must work to maximize the simplicity of the extraction patterns.
- HDS must be able to support arbitrary queries. Includes questions in the future that we may not yet know.
- Extraction patterns are written against a specific version of the model.
- Must support the state DOI, such as auto coverage and auto territory.
- Must have sufficient performance.

Schema Evolution

- May only conform to going-forward models.
- New attributes will not need to be historically filled.
- Should allow for schema evolution with minimal impact.

History

- Must provide current data, defined as month end plus 45 days.
- Must hold history according to regulatory requirements, generally defaulted to 5 years.

Data Quality

- Must have edits to verify data quality before being available for consumption.
- Must have tolerances for the edits that default to 5%.

Structural Concerns

- Must have traceability across all layers.
- Must provide a mechanism for model versioning.

3. **Data Dictionary**

   a. **Policy**

i. Field Type	Field	Day 1 or 2 Req.	Definition	Expected Values	Comments
1. All	Record LOB	1	A code identifying the broader line of insurance the transaction is to be reported as.	Personal Auto	
1. All	Record Type	1	A code identifying the type of transaction being reported.	Premium, Paid, OS, ALAE, Limited Coding	
2. Policy	Company ID	1	A unique 4-digit code assigned by the Statistical Bureaus to identify each insurance company.	Numeric	
2. Policy	Annual Statement Line	1	A code identifying the line of insurance the transaction was written under.	Alpha numeric. NAIC - 19.1, 19.2, 21.1	
2. Policy	Policy Category	1	A code identifying whether the record is Personal or Commercial.	Personal / Commercial / Surplus	
2. Policy	Subline	1	A code identifying the coverage.	Liability, No-Fault, Physical Damage, Med Pay	
2. Policy	Policy Identifier	1	A unique combination of letters and numbers that identifies an insured's insurance policy. This is the foreign key for connecting the claim and policy records.	Alpha-numeric. No Limit to the field length.	
2. Policy	Policy Effective Date	1	The date of when the policy became effective.	https://en.wikipedia.org/wiki/ISO_8601 - YYYY-MM-DD	
2. Policy	Transaction Type	1	A code identifying whether the record is a New Business, Change, Renewal, Cancellation, etc.	New, Renew, New Change, Renew Change, New Cancel, New Audit; Renew Audit	
2. Policy	Transaction Effective Date	1	Date the change becomes effective. Not the booking date.	https://en.wikipedia.org/wiki/ISO_8601 - YYYY-MM-DD	needed to calculate exposure

2. Policy	Transaction Expiration Date	1	Date the change ends.	https://en.wikipedia.org/wiki/ISO_8601 - YYYY-MM-DD	needed to calculate exposure
2. Policy	Market	1	Market in which the policy is being written through.	Voluntary / Residual	
2. Policy	Accounting Date	1	The date the transaction was financially booked.	https://en.wikipedia.org/wiki/ISO_8601 - YYYY-MM-DD	
2. Policy	Program	1	A code that indicates whether the basic rules and forms used in writing the policy are using certain Bureau's rules or form or if they are independent.	ISO, AAIS, Independent, State Standard, Other	
2. Policy	Multi-Car Discount Code	1	A code that indicates whether there is a multi-car discount or just a single rated car.	Multi, Single, Not Applicable	
2. Policy	Package Code	1	A code identifying whether the policy is standalone or part of a package.	Standalone, Package	
2. Policy	Pool Affiliation	1	Identifies business written in a pool such as an assigned risk facility or joint underwriting association, and business not written in a pool.	Pooled, Direct	
2. Policy	Pricing Tier	1	Company assigned pricing mechanism.	Standard, Sub-Standard, Preferred, Modeled	
2. Policy	NC Program Enhancement Ind	1	Based on North Carolina Session 2015 House Bill 288, premiums and losses resulting from program enhancements must not comingle with basic data. This code indicates the use of enhanced endorsements.	Yes / No	
2. Policy	NC Reinsurance Facility	1	A code that distinguishes between business written under the NC Reinsurance Facility and that which is not.	Ceded, Retained, Not Written	Reinsurance fund business but financially recorded as direct written premium
2. Policy	SC Reinsurance Facility	1	A code that distinguishes between business written under the SC Reinsurance Facility and that which is not.	Ceded, Retained, Not Written	SC State Exception A
2. Policy	Multi-Car Risks	1	A field that indicates whether the multi-car risk is subject to the PIP Premium Discount and whether the Principal Auto is not Discounted.	Multi Car Subject to PIP Discount - Principal Auto Not Discounted, Multi Car Subject to PIP Discount - Additional Auto Discounted, Multi-Car Risks Not Subject to PIP Discount; All Other Autos	NJ State Exception B; Applies to PIP Multi-Car Risks only.
3. Driver	Gender	1	Gender of highest rated operator of each insured vehicle.	Male, Female, X	
3. Driver	Driver Age	1	Driver's age when policy was written.	Provide raw age.	
3. Driver	Marital Status	1	Indicates whether driver is single or married.	Single, Married, Other	
3. Driver	Safe Driving /Defensive Driver Discount	1	Discount for the practice of using driving strategies that minimize risk and help avoid accidents, as by predicting hazards on the road	Discount Applied, Does Not Qualify, Discount not Offered on Coverage	
3. Driver	Good Student Discount	1	Discount for student getting good grades in school.	Discount Applied, Does Not Qualify, Discount not Offered on Coverage	
3. Driver	Number of Penalty Points	1	Number of motoring conviction codes, fixed penalty notices, endorsements against the driver.	Numeric. Provide number of points	
3. Driver	Drivers Training Discount	1	Discount for Driver taking Driving School lessons.	Discount Applied, Does Not Qualify, Discount not Offered on Coverage	

3. Driver	55 & Over Discount	1	Discount for 55 & over years olds for successful completion of the PA certified training course.	Discount Applied, Does Not Qualify, Discount not Offered on Coverage	PA State Exception B
3. Driver	Accident Prevention Credit	1	Indicates whether the vehicle has an accident prevention credit.	Discount Applied, Does Not Qualify, Discount not Offered on Coverage	NY State Exception A
4. Vehicle	Body Style	1	The shape and model of a vehicle.	** See Body Style tab	
4. Vehicle	Body Size	1	Type of Vehicle	Private Passenger, Sports Car, Oversized Car or Limo, All Other	
4. Vehicle	Engine Size (Motorcycle)	1	Size of Engine in Motorcycles or Motor Scooters	Provide actual engine size.	NJ State Exception C (Applies to Motorcycle only)
4. Vehicle	VIN	1	Unique combination of numbers and letters that identify each vehicle.	Vehicle Index - Alpha numeric. No Limit to the field length.	
4. Vehicle	Garaging State	1	State where the vehicle is principally garaged.	NAIC Standard	
4. Vehicle	Garage ZIP5	1	First 5 digits of Zip Code of where the vehicle is principally garaged.	Alpha numeric 5 bytes.	
4. Vehicle	Garage ZIP4	1	Last 4 digits of Zip Code of where the vehicle is principally garaged.	Alpha numeric 4 bytes.	Optional Reporting field
4. Vehicle	Vehicle Year	1	Model Year	4 byte numeric. YYYY	
4. Vehicle	Vehicle Use	1	Primary use of the vehicle.	Personal / Commute / Business / Farm	
4. Vehicle	Vehicle Symbol	1	Abstraction of the vehicle rating symbols. Not company rating symbols.	Use current code values from the AAIS Stat Plan.	
4. Vehicle	Vehicle Performance	1	Vehicle Performance characteristic	Standard, Intermediate, High, Sports Car, Sports Premium (NJ only), All Other	
4. Vehicle	Per Commute Miles	1	Number of miles driven per commute.	Numeric. Raw numbers	
4. Vehicle	Annual Vehicle Miles	1	Number of miles driven annually.	Numeric. Raw numbers	
4. Vehicle	Anti-lock Brakes Discount	1	Code indicating whether Anti-Lock Brakes discount was applied to the vehicle.	Discount Applied, Does Not Qualify, Discount not Offered on Coverage	
4. Vehicle	Safety Restraint Discount	1	Code indicating whether Safety Restraint discount was applied to the vehicle.	Discount Applied, Does Not Qualify, Discount not Offered on Coverage	
4. Vehicle	Anti-theft Device Discount	1	Code indicating whether Anti-Theft Device discount was applied to the vehicle.	Discount Applied, Does Not Qualify, Discount not Offered on Coverage	** NJ and NY require additional information. Applies to Comp Physical Damage coverage
5. Coverage	Coverage Code	1	Code representing the coverage of the vehicle the premium on the record applies to.	** See Coverage Code Values tab	
5. Coverage	Per Person Limit	1	Per Person Limit for the Coverage	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.	
5. Coverage	Per Accident Limit	1	Per Accident Limit for the Coverage	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.	

5. Coverage	Property Damage Limit	1	Property Damage Limit for the Coverage	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.	
5. Coverage	Combined Single Limit	1	Combined Single Limit for the Coverage	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.	
5. Coverage	Premium	1	Transactional Premium applicable to the coverage	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.	
5. Coverage	Deductible	1	Deductible for the Coverage	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.	
5. Coverage	UM /UIM Stacking Indicator	1	Indicates whether UM/UIM Limits are stacked (per vehicle) or non-stacked (per policy)	Limits Stacked, Non-Stacked	UM/UIM Motorist field & UM/UIM Indicator
5. Coverage	Medical Expenses Ded Amt	1	Deductible amount applicable to Medical Expenses.	Provide whole number	Applies to NJ only
5. Coverage	NJ Threshold /Tort Limitation	1	Code indicating whether the insured has a Threshold and Tort Limitation on the policy.	Liability - No Threshold/Tort limitation, Liability - Verbal Threshold/Tort Limitation, Liability with Unknown Threshold /Tort Limitation; PIP - Medical Exp Benefits as Primary Cov - No limitation, Medical Exp Benefits as Primary Cov - Verbal limitation, Medical Exp Benefits as Secondary Cov - No limitation, Medical Exp Benefits as Secondary Cov - Verbal limitation	NJ State Exception A
5. Coverage	Primary No-Fault Health Plan	1	Determines what type of no-fault health plan the driver has.	Employers' primary, PIP primary, Medicare primary, No basic PIP medical (i.e. private health plan), Employers' work loss plan & medicare coordination both primary, Employers' work loss primary w/o basic PIP med exp	NY State Exception B
5. Coverage	Combined First Party Benefits	1	Code describing the combination of Medical Benefits Limit, Loss of Income $, Accidental Death $, Funeral Expense $	Day 1 Stat Plan values. Will be replaced by Day 2 fields.	PA State Exception C

b. **Claim**

i. Field Type	Field	Day 1 or 2 Req.	Definition	Expected Values	Comments	
1. Claim Identifier	Policy Identifier	1	A unique combination of letters and numbers that identifies an insured's insurance policy. This is the foreign key for connecting the claim and policy records.	Alpha-numeric. No Limit to the field length.		
1. Claim Identifier	Policy Eff Date	1	The date of when the policy became effective.	https://en.wikipedia.org/wiki/ISO_8601 - YYYY-MM-DD		
1. Claim Identifier	Claim Number	1	A unique combination of letters and numbers that identifies an insured's claim.	Alpha-numeric. No Limit to the field length.		
1. Claim Identifier	VIN	1	Unique combination of numbers and letters that identify each vehicle.	Vehicle Index - Alpha numeric. No Limit to the field length.		
1. Claim Identifier	Accident Date	1	Date the loss occurred	datetime: https://www.iso.org/iso-8601-date-and-time-format.html	datetime: https://www.iso.org/iso-8601-date-and-time-format.html	
1. Claim Identifier	Cause of Loss	1	Reason the loss occurred.	BI, Death Claim, Property Damage, Loss of Income, Med Pay, Collision, Flood, etc	geocode: https://en.wikipedia.org/wiki/ISO_6709	
1. Claim Identifier	Claim Status	1	A code showing whether the claim is open or closed.	Open, Closed, Reopened		

2. Claim Activity	Claim Number	1	A unique combination of letters and numbers that identifies an insured's claim.	Alpha-numeric. No Limit to the field length.		
2. Claim Activity	Claimant Identifier	1	Code identifying the various claimants for a single loss.	Alpha-numeric. No Limit to the field length.		
2. Claim Activity	Coverage Code	1	Code representing the coverage of the vehicle the premium on the record applies to.	BI, PD, CSL, UMBI, UMPD, UMBIPD, UMUIMBI, UMUIMPD, UMUIMBIPD, UIMBI, UIMPD, UIMBIPD, PIP, Comp, Coll, Medical Benefits, Med Pay.		
2. Claim Activity	Accounting Date	1	The date the transaction was financially booked.	https://en.wikipedia.org/wiki/ISO_8601 - YYYY-MM-DD		
2. Claim Activity	Paid Loss	1	Loss Amount paid on the claim for this transaction.	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.	Use to derive the Transaction Code	
2. Claim Activity	Paid LAE	1	Loss Amount paid on Loss Adjustment Expenses for this transaction.	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.	Use to derive the Transaction Code	
2. Claim Activity	Outstanding Loss	1	Loss Amount reserved for possible future payments for this claim, claimant and coverage.	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.		
2. Claim Activity	Outstanding LAE	1	Loss Amount reserved for Loss Adjustment Expenses for this claim, claimant and coverage.	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.		
2. Claim Activity	Subrogation - Loss	1	Loss Amount recovered from a 3rd Party.	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.		
2. Claim Activity	Subrogation - LAE	1	Loss Amount for Loss Adjustment Expense recovered from a 3rd Party.	Numeric with a decimal and a negative sign when applicable. Decimal is required when including cents.		
2. Claim Activity	Tort Limitation	1	A code that indicates whether the tort limitation applies to the driver.	Rejected / Accepted	State Exception for KY & PA	

4. **Business Data Model**

Define…

   a. The business data model represents the entities, attributes, and relationships as understood by business users.
   b. The business data model has no regard for the implementation technology nor the processing required to load it.
   c. A data element is a singular piece of information, such as one number, one string, one date, etc.
   d. An entity is a set of data elements.
   e. A business entity is an entity that is recognizable to business users and tends to occur in conversations with them, such that the number of elements is not small enough (typically one to five) to be trivially embedded inside another entity. Policy, claim, vehicle, and home are typical examples.
   f. A relationship occurs when business users can understand the use of entity names in a sentence of the form "has a" or "has many" such as "a policy has many claims", where "many" can include one or none, and where more semantically rich terms for "has" may be used such as "belongs to" or "contains".
   g. The model aspires to not repeat the attributes of an entity in more than one place.
   h. The model must make it difficult to over-count elements and values.
   i. Sparsity is the measure of empty values to total number of values. For example, an entity of 100 records and 100 attributes has 10,000 cells. If 1,000 are empty, the sparsity is is 10%.
   j. No entity when populated should have a sparsity level exceeding 10%. If that occurs, the entity should be reviewed for the opportunity to create more than one entity. For example, an auto policy and a home policy are both policies, but have sufficiently different attributes that they could have high sparsity and warrant being two different entities.
   k. Any entity that is highly covariant with another entity may be embedded in it. g. policies have addresses. Address can be broken out for robust handling, or it can be left inside the policy.
   l. The following are entities in the business data model:
      i. Policy – An agreement between a carrier and an insured.
         1. Auto Policy – A policy for a set of vehicles.
         2. Home Policy – A policy for a home and its related risks.
      ii. Vehicle – An insurable object operated by a driver on road.
      iii. Driver – A person who operates a vehicle.
      iv. Coverage – A transfer of a certain type of risk.
      v. Claim – A set of events under a policy that results in payments.
      vi. Claim Event – An event under a claim.
      vii. More! Not exhaustive…
   m. The following are relationships:
      i. A policy has one or more vehicles.

        ii. A vehicle has one or more drivers.
       iii. A policy has many coverages.
       iv. A policy has zero or many claims.
        v. A claim belongs to exactly one policy.
       vi. A claim has many claim events.
      vii. A claim event belongs to exactly one policy.
      viii. More! Not exhaustive…

5. **Presentation Data Model**

   The presentation data model optimizes the ease of writing queries, and to do so, is willing to give up some business precision of the business data model, but is not particularly concerned about system performance, though some adaptation to the implementation model may occur.

      a. TODO: Enumerate the properties of this model.

6. **Persistence Data Model**

   The persistence data model is obsessed with precise preservation of truth and having flexibility into the future. It is willing to be somewhat unrecognizable to business users, and takes into account both system performance and implementation concerns.

      a. TODO: Enumerate the properties of this model.

7. **Loading Data Model**

   The loading data model seeks to make getting data into the system as easy as possible. It is very willing to compromise in favor of making feeds from sources easy, and TODO.

      a. TODO: Enumerate the properties of this model.