This document contains a list of the Analysis Exchange (AE) ontology elements[[1]](#footnote-1) that are relevant to the different analytic capabilities in the VA Travel Voucher Fraud use-case as provided by IBM, SAS, NetOwl, Esri, and Thomson Reuters.

# Seed Data

The seed data, provided by IBM, will be represented as a row per record with different columns for the features, some of which can translate into the AE elements and some that may be processed downstream. This represents the content of the VA Travel Voucher. The following elements should be found in there.

* Classes
  + VATravelVoucher – *our principal object that attaches to people and events*
  + Human
  + Travelling – *The event of travel, both to and from the facility, which might have different end points*
  + Building
  + PostalAddress
  + Name
  + SocialSecurityNumber
  + CurrencyMeasure
* Object Properties
  + hasVoucherTravel – *subject: VATravelVoucher, object: Travelling*
  + hasVoucherBeneficiary – *subject: VATravelVoucher, object: Human*
  + hasVoucherClaimant – *subject: VATravelVoucher, object: Human*
  + hasVAVoucherTreatingFacility – *subject: VATravelVoucher, object: Building*
  + hasAddress – *subject: Building, object: PostalAddress*
  + hasAgent – *subject: Travelling, object: Human*
  + hasDestination – *subject: Travelling, object: Building*
  + hasOrigin – *subject: Travelling, object: Building*
  + hasSocialSecurityNumber – *subject: Human, object: SocialSecurityNumber*
  + hasName – *subject: Human, object: Name*
  + hasVoucherExpense – *subject: VATravelVoucher, object: CurrencyMeasure*
* Data Properties (default objects are strings)
  + hasDateOfBirth – *subject: Human*
  + hasFirstName – *subject: Name*
  + hasLastName – *subject: Name*
  + hasTime – *subject: Travelling*
  + hasSignatureDate – *subject: VATravelVoucher*

Some information can also be supplemented from the known, more-or-less fixed information about VA treatment facilities which are also part of the seed data. Some elements drawn from these (such as addresses) are covered above and in the NetOwl section. Others that could be drawn from this information are below:

* Classes
  + TelephoneNumber – *subclass of Identifier*
* Object Properties
  + hasTelephoneNumber – *subject: Building, object: TelephoneNumber*
* Data Properties (default objects are strings)
  + hasVISN – *subject: Building (this is an identifying number for the facility)*

There are also some built-in RDF properties that will be used. Among these is “rdfs:label” which will be used for cases where there are textual labels that do not point to an instantiated object, but represent the text associated with an entity. Examples of this include “PostalAddress” for the text field as drawn from the voucher (before extraction of its subcomponents as performed below) and Building for the text field that indicates a treatment facility’s name. The property “rdfs:comment” is also used to provide further information on how to use certain classes in the ontology.

Additionally, there are some other fields that can be applied to people which are not directly provided in the VA Travel Voucher. These include:

* Classes
  + DriversLicenseNumber – *subclass of Identifier*
  + EmailAddress – *subclass of Identifier*
* Object Properties
  + hasResidence – *subject: Human, object: Building*
  + hasDriversLicenseNumber – *subject: Human, object: DriversLicenseNumber*
  + usesTelephoneNumber - *subject: Human, object: EmailAddress (maybe we can reuse the hasTelephoneNumber here, but there might be a subtle distinction)*
  + usesEmailAddress – *subject: Human, object: EmailAddress*

# NetOwl

The content of the VA Travel Vouchers will contain addresses, but these addresses are all part of a single field. We will break this down into the constituent parts of the address using NetOwl. This fits the following properties

* Object Properties
  + hasCity – *subject: PostalAddress, object: City*
  + hasState – *subject: PostalAddress, object: StateOrProvince*
  + hasCountry – *subject: PostalAddress, object: Nation*
* Data Properties (default objects are strings)
  + hasAddressUnit – *subject: PostalAddress*
  + hasStreetNumber – *subject: PostalAddress*
  + hasZipcode – *subject: PostalAddress*

There also may be elements that can be drawn from the fields on itemized reimbursement expenses. This may be a good use for NetOwl as well, but we need to create examples of these fields to see what relevant pieces can be pulled from them. This includes parking, tolls, lodging, and meals. Locations and values can be extracted from these to provide further properties for the VATravelVoucher, but some examples are needed to clarify this. This is likely to need the following:

* Data Properties (default objects are strings)
  + hasUnitOfMeasure – *subject: CurrencyMeasure*
  + numericValue – *subject: CurrencyMeasure*

# Esri

There are multiples analytic capabilities that Esri might supply. First, we can generate paths between locations (providing Esri the addresses for sources and destinations for Travel events) to establish the distance and time taken to travel. Second, the voucher forms also allow the claimant to simply put in the name of the facility, leaving the address as optional. In these cases, being able to specify an address and geo-coordinate candidates given a facility name would be a useful analytic.

* Entities
  + Transitway – *This is a SUMO construct for a path (which we might adopt)*
  + TimeDuration – *For the length of time estimated for travel events*
  + NumericValue – *Encodes the magnitude of distance*
* Object Properties
  + hasTime - *subject: Transitway, object: TimeDuration*
  + hasDestination – *subject: Transitway, object: Building*
  + hasOrigin – *subject: Transitway, object: Building*
  + hasDistance – *subject: Transitway, object: SUMO:LengthMeasure*
* Data Properties (default objects are strings)
  + hasLatitude – *subject: Building*
  + hasLongitude – *subject: Building*
  + hasUnitOfMeasure – *subject: LengthMeasure*

# Thomson Reuters / Enrichment

We have not fully specified what analytic output we expect from Thomson Reuters so this is more speculative and subject to change, but generally it is expected that we will receive enrichment information on the people involved in the claims. These will likely become object properties and data properties with Human as its subject. Some important features could include verification of their home address.

It is also unclear whether there will be a direct adaptation of Thomson Reuters content, but nevertheless, we can hypothesize about the kinds of information we expect to enrich with and attach to the Human entities. Right now we limit these to binary descriptions of whether some quality is present or not for an entity.

* Data Properties (default objects are 0 for false, 1 for true)
  + hasDeclaredBankruptcy – *subject: Human*
  + hasLiens – *subject: Human*
  + hasArrestRecord – *subject: Human*
  + hasCriminalRecord – *subject: Human*

# SAS Statistical Analysis

We have some initial ideas laid out as to what results might derive from SAS. These are based on the VA Vouchers and enrichment from the other analytics above going in and produce various information from these records and also historical information that might be available. Some of the expected attributes produced in this analysis are listed below:

* Data Properties (default objects are 0 for false, 1 for true)
  + accordsWithHistoricalMileage – *subject: VATravelVoucher*
  + accordsWithHistoricalExpenses – *subject: VATravelVoucher*
  + accordsWithHistoricalTravelFrom – *subject: VATravelVoucher*
  + accordsWithHistoricalTravelTo – *subject: VATravelVoucher*
  + accordsWithHistoricalTreatmentCenter – *subject: VATravelVoucher*
  + accordsWithHistoricalClaimant – *subject: VATravelVoucher [this might be false when there are past results for the Veteran but with different claimants]*
  + accordsWithHistoricalVeteran – *subject: VATravelVoucher [this might be false when there are past results for the claimant but with different veterans]*
  + hasRiskProbability – *subject: VATravelVoucher, object: double, 0…1 [this could be inverted to so that 0 is bad and 1 is good, but the opposite is assumed for now]*

# ESRI - Geoclustering

Given addresses, ESRI can provide clusters of activity associated with locations on the map.

# i2 – Link Analysis

Given entities and relationships (likely principal entities include Human, VATravelVoucher, and possibly Building). This returns enriched relationships that are flagged as problematic. These can either be new relationships that supplement existing relationships (with a flag among the provenance) or they can be brand new relationships added to the exchange model. These will be defined as we move forward.

# ESRI – Visualization

With these sources, enrichment, and analyses from SAS and i2, we can provide ESRI information that can be visualized on a map. This would be a good final state to demonstrate the overall impact of the capability.

# Additional – Provenance

We expect to preserve provenance statement by statement. This will be a set of fields to include:

* label – has a record of any provenance provided by the analytic itself (not mapped into any particular representation, but held in case needed downstream)
* source – the analytic producing the result or data source if data
* date – the creation date for the statement
* version – the version of the source analytic if applicable.

1. Some elements are in **red**. These are not currently present in the data model, but will be added. [↑](#footnote-ref-1)