**8/8/2023 Meeting Notes with Alec McCullick (**Gas Data Management and Gas Foundry Portfolio)

The goal of the meeting was to learn about what it takes to use Foundry for data management used for TIMP-risk groups, specifically:

1. What’s the current state of things in Foundry with gas data?
   1. What gas data sets are available?
   2. What are future plans?
2. What is Foundry currently capable of:
   1. How well can it work with GIS data?
   2. What does data integration look like?
3. What does a migration process look like
   1. What kind of support can IT and Data Management team provide?
   2. What kind of bandwidth and work is required on the TIMP engineers’ side?
4. How will migration into Foundry help us?
5. What kind of long-term maintenance would be needed:
   1. What kind of long-term commitment do the engineers need to provide?
   2. How will Foundry handle future data and format changes?

**What’s the current state of things in Foundry with gas data?**

Gas Data Management and Gas Foundry Portfolio have been getting the gas distribution and gas transmission GIS datasets into public ontology Level 2. They are looking for new datasets in gas to get into ontology and set up data stewardship for those ontologys and datasets. DIMP and TIMP datasets are currently on the roadmap.

They have a list of TIMP/DIMP datasets to be integrated [here](https://pge.sharepoint.com/:x:/s/GasOperationsDataAssetManagement/ERqOZKreCDZCjMesj427Qq4Bs14spagqqGVX-wk4mLz5wQ?e=pQBfv2), provided by Todd Arnett. They would like to prioritize the data-sets, and **we can provide influence and prioritization to fit our needs, as there haven’t been any specific requests currently.**

**What is Foundry currently capable of?**

Alec show as an example the Gas Valve ontology, which is essentially a pipeline integrating from various data sources, including GeoMART, SAP, and GDGIS datasets, into a final “ontology object” (fancy name for a final cleaned dataset).

This process translates well to our current process used to compile the MarinerDB used for risk modeling (MarinerDB can be though of as a final “ontology object”).

I noted that right now there’s quite a bit of manual data cleaning and validation needed during the data integration process (which everyone hates). The goal in Foundry is to automate as much of this as possible, and in the example Gas Valve ontology, these transformation blocks were written by the pipeline engineers, with some being adapted from logic taken from the data stewards.

* IMO, this is the most intensive part, to be super explicit and clear about data integration procedures such that these transformation blocks can be written and maintained.

**Currently Foundry is able to interact with GeoMART, SAP, GDGIS, Excel, and SQL databases. In addition, it can currently do Geospatial-overlap.**

*\*\*Remark\*\*: Maybe useful to evaluate and compare what GIS/Geomart capabilities that engineers currently use that are NOT available or adequate in Foundry*

**What does migration into Foundry or data integration with Foundry look like?**

Note that previously, Even Lin and Steven Hui have already worked with Jorgen Vos to integrate MarinerDB (just the database tables themselves, not the upstream raw sources) into Foundry. This is considered as a “product ontology”, and while the connection has not been working for a few months (most likely due to the server migrations), it should be enough to run the risk models on it.

The official process for integrating datasets into Foundry, as described by Alec (will illustrate with ILI dataset migrating from GeoMART to Foundry, for example)

* Put in request with his team, specify which datasets we are interested in (e.g. ILI tallys)
* Stakeholders meet to share understanding of this dataset – an engineer SME would be needed (e.g. Satvinder)
* Figure out data columns for the dataset
* Work with SME to build descriptors and metadata (e.g. ILI vendors for each year, etc)
* Work with SME to build quality check rules.
* Select critical data elements, data management team helps to build the actual objects within Foundry, show and test the build

At the end of the process, there will need to be a final data steward for the integrated data set, there’s some kind of “annual certification process” (paused currently).

Bandwidth required of engineer SME: approximately 3 hours per week max, depending on complexities.

\*\*My remarks\*\*: Once Satvinder has finished compiling the huge ILI tally excel dataset, it might be a relatively simple lift to transfer the process he used integrating data sources and GeoMART into Foundry. Even if it doesn’t include ALL the upstream sources, we’ll at least have that available in Foundry, with data quality checks once the process is complete.

On IT/data management’s end: IT done building with electric for the year, so they are looking for next projects! We can be it!

**How will migration into Foundry help us?**

We will have a central data repository location. Analysis on it is fast, and data version and access control are built-in. There’s clearly a lot of company support behind this tool, and centralization of our process in Foundry makes adapting work from Electric and Wildfire convenient.

Additionally, this migration can help push along adoption by upstream teams as well, reducing waste and manual data entry.

**What kind of long-term maintenance will be needed?**

Integration into Foundry doesn’t eliminate future problems due to data or format changes, but the ability to keep track of changes better make it a bit easier.

Alec mentions the current procedures other teams have used include:

* Validation and changes in “foundational” data sources (e.g. ILI reports) are communicated with IT, and they will make updates to these changes.

This method doesn’t seem to scale, however, if all data changes need to be routed through the IT team. Alec mentions right now the scale is small, but they expect to have **specific governance procedures for data change approvals**.

In addition, each team should have a data steward, attached to each ontology (e.g. MarinerDB). For TIMP the data steward is **Steven Liu**.

**Next steps**

Contact **Todd Arnett** and **Susan Minarcin** to tell them we are interested in working with them next to integrate datasets (ILI tally, pipe attributes, etc), and provide our prioritization of the dataset list. Put in request, and go from there.

Timeline starting near end of Aug. is sufficient.