**10/23/2023 Meeting Notes**

**Invitees**

·         Required: Avery, Christine, Hassen, Laila, Sindhu, Satvinder, Lee, Steven, Kumari Archana, Huber Adrian

·         Optional: Yin Allen, Siegmund, Matt

**Attendees**

·         Avery, Christine, Hassen, Laila, Lee, Steven, Yin Allen, Kumari Archana

**Meeting Minutes**

·         Introductions, Roles & Responsibilities

·         Bring all historical pipe tallies into Foundry

o    It's in Excel formatted sheets on shared network folder - move to ESFT or some network from where OIT can ingest them onto Foundry

·         Pipe Tally sheets up until 2021 are uploaded in Geomart

o    Business has all Pipe Tally sheets stored on their network drive so no need to go to Geomart to source

·         Pipe Tally sheets are scanned for new data on monthly basis

o    Any given month could have a single sheet, multiple sheets, or no sheets [data]

·         Business has total of 287 Pipe Tally sheets from start to 2023

o    OIT to ingest all 287 Pipe Tally sheets onto Foundry

**Introductions, Roles & Responsibilities**

·         **Christine Avery**

o    Business Owner SME and maintains version control

o    Going on maternity leave for 3 months shortly so won't be joining calls when on leave

o    Direct and guide business team with requirements and/or other project tasks as needed

o    Business Ownership of the Ontology object

o    Source System and Ontology Metadata upload onto - <https://pge.collibra.com/dashboard>

·         Connect with **Matt Siegmund** in AKM for guidance on this

·         **Laila Hassen**

o    Product SME for ILI Pipe Tally and wants to make the process routine for rest in the team

o    Provide requirements for the project

o    Complete Business Requirements sheet in the [Ontology Object Template Gas TIMP Pipeline Tally.xlsx](https://pge.sharepoint.com/:x:/r/sites/ODGTeam/Shared%20Documents/Conceptual%20Model/Gas%20Ontology%20Federated%20Team/Gas%20TIMP%20ILI%20Pipeline%20Tally%20Public%20Ontology/Ontology%20Object%20Template%20Gas%20TIMP%20Pipeline%20Tally.xlsx?d=wcebe11d0b4ae44d9bef78fe2babdd1e1&csf=1&web=1&e=8bsUfa)

o    Provide object properties [columns/fields] metadata

o    Help complete requirements toward Object class Public Level 2 [Ontology Ownership and Maturity Tracker](https://damask.palantirfoundry.com/workspace/module/view/latest/ri.workshop.main.module.f7947030-7466-4810-a3e7-24cf4c12afb8)

o    Perform business data integrity, validation, and User Acceptance Testing [UAT] as applicable

o    Ontology Object Ownership assignment within Foundry

o    If useful - [BDS Onboarding.pptx](https://pge.sharepoint.com/:p:/r/sites/dma/Shared%20Documents/DM%26A%20Asset%20Registry/PwC%20-%20Asset%20Registry%20Docs/Asset%20Registry%20CDE%20Inventory/Cross-team%20Collaboration/BDS%20Onboarding%20+%20Kickoff/BDS%20Onboarding.pptx?d=w324055c8973c43c48844a760de4e7a75&csf=1&web=1&e=JpbgZ6)

·         **Steven Lee**

o    Analyst - Will review data integrity, formatting, metadata, etc. as project progresses

o    Complete Business Requirements sheet in the [Ontology Object Template Gas TIMP Pipeline Tally.xlsx](https://pge.sharepoint.com/:x:/r/sites/ODGTeam/Shared%20Documents/Conceptual%20Model/Gas%20Ontology%20Federated%20Team/Gas%20TIMP%20ILI%20Pipeline%20Tally%20Public%20Ontology/Ontology%20Object%20Template%20Gas%20TIMP%20Pipeline%20Tally.xlsx?d=wcebe11d0b4ae44d9bef78fe2babdd1e1&csf=1&web=1&e=8bsUfa)

o    Provide object properties [columns/fields] metadata

o    Help complete requirements toward Object class Public Level 2 [Ontology Ownership and Maturity Tracker](https://damask.palantirfoundry.com/workspace/module/view/latest/ri.workshop.main.module.f7947030-7466-4810-a3e7-24cf4c12afb8)

o    Perform business data integrity, validation, and User Acceptance Testing [UAT] as applicable

o    Ontology Object Ownership assignment within Foundry

o    If useful - [BDS Onboarding.pptx](https://pge.sharepoint.com/:p:/r/sites/dma/Shared%20Documents/DM%26A%20Asset%20Registry/PwC%20-%20Asset%20Registry%20Docs/Asset%20Registry%20CDE%20Inventory/Cross-team%20Collaboration/BDS%20Onboarding%20+%20Kickoff/BDS%20Onboarding.pptx?d=w324055c8973c43c48844a760de4e7a75&csf=1&web=1&e=JpbgZ6)

·         **Jayna Thanki**

o    Ontology Public Object Designer

o    Gather requirements and complete the Design-Data Mapping Base, Design - Data Mapping Curated, and Design - Data Mapping Augmented sheets within [Ontology Object Template Gas TIMP Pipeline Tally.xlsx](https://pge.sharepoint.com/:x:/r/sites/ODGTeam/Shared%20Documents/Conceptual%20Model/Gas%20Ontology%20Federated%20Team/Gas%20TIMP%20ILI%20Pipeline%20Tally%20Public%20Ontology/Ontology%20Object%20Template%20Gas%20TIMP%20Pipeline%20Tally.xlsx?d=wcebe11d0b4ae44d9bef78fe2babdd1e1&csf=1&web=1&e=8bsUfa)

o    Gather object properties/columns metadata

o    Gather object properties/column CDE [Critical Data Element] identification whenever possible

o    Spot check object property data as applicable for general data validity

o    Guide business toward Object Public class Level 2 maturity

o    Ensure project follows appropriate access model per [[02.21.2023] Access Model Migration Reader's Digest.pptx](https://pge.sharepoint.com/:p:/r/sites/DigitalCatalystAnalytics/Shared%20Documents/Foundry%20Operations/Ontology%20Maturity%20Implementation/Ontology%20Access%20Model%20Migration/%5B02.21.2023%5D%20Access%20Model%20Migration%20Reader%27s%20Digest.pptx?d=wcabaa026c92742119f165f837e429f7a&csf=1&web=1&e=fxbNsu)

o    Provide OIT [Ontology Implementation Team] public ontology consultancy

·         **Archana Kumari**

o    Ontology Gas Product Owner

·         Archana to fill in details as appropriate

·         **Adrian Huber**

o    Ontology Data Engineer

·         Build data pipelines per Ontology standards

·         Implement object ontology design on Foundry and create appropriate object layers

·         Set up appropriate datasets and object schedules, health checks, fidelity checks, as applicable

·         Demo object upon completion and when ready for business overview

·         Provide overall OIT and data engineering technical consultancy as needed

**Action Items**

·         Jayna to request access for business teams on Ontology Design artifact SharePoint – done

o    [Gas TIMP ILI Pipeline Tally Public Ontology](https://pge.sharepoint.com/:f:/r/sites/ODGTeam/Shared%20Documents/Conceptual%20Model/Gas%20Ontology%20Federated%20Team/Gas%20TIMP%20ILI%20Pipeline%20Tally%20Public%20Ontology?csf=1&web=1&e=kDH9Sw)

·         Jayna to send link to Ontology Design template for business review and input as applicable - done

o    [Ontology Object Template Gas TIMP Pipeline Tally.xlsx](https://pge.sharepoint.com/:x:/r/sites/ODGTeam/Shared%20Documents/Conceptual%20Model/Gas%20Ontology%20Federated%20Team/Gas%20TIMP%20ILI%20Pipeline%20Tally%20Public%20Ontology/Ontology%20Object%20Template%20Gas%20TIMP%20Pipeline%20Tally.xlsx?d=wcebe11d0b4ae44d9bef78fe2babdd1e1&csf=1&web=1&e=8bsUfa)

·         Jayna to follow up within OIT team on options for Pipeline Tally Excel sheet ingestion onto Foundry

·         **Business to complete necessary Foundry trainings to get familiar with Foundry system and use**

o    Use below link to get guidance on Foundry access and various trainings that are available

·         <https://wiki.comp.pge.com/display/ARAD/Palantir+Foundry>

o    Business to review BDS Onboarding PPT if useful. Provides information on CDE [Critical Data Element] identification process as well as many other data steward and SME related tasks as they apply to Ontology

·         [BDS Onboarding.pptx](https://pge.sharepoint.com/:p:/r/sites/dma/Shared%20Documents/DM%26A%20Asset%20Registry/PwC%20-%20Asset%20Registry%20Docs/Asset%20Registry%20CDE%20Inventory/Cross-team%20Collaboration/BDS%20Onboarding%20+%20Kickoff/BDS%20Onboarding.pptx?d=w324055c8973c43c48844a760de4e7a75&csf=1&web=1&e=JpbgZ6)

**10/30/2023 Meeting Notes**

* TIMP ILI Current Process: Geomart could not accommodate request to update fields from the new pipe tally format, so that prompts the switch to using Foundry/Ontology as the substitute
  + geomart not prioritizing the request due to limited team/funding
  + pipe tally format will continue to be updated so it becomes a recurring long-term challenge
* Pipe Tally Status:
  + Pipe tallies up to 2021 were uploaded in Geomart (only traditional; Non-traditional & EMAT pipe tallies are not included)
  + Pipe tallies after 2021/2022 are in a new format therefore, are not uploaded into Geomart
  + To date, there are 354 total pipe tallies (including traditional, NT, and EMAT). Only approx. 287 pipe tallies are 'converted' into a standardized format (traditional & non-traditional can use the **ILI Pipe Tally Conversion Template\_v4 )**
    - using the .csv files will be all the historic files
  + Note: change management for Version 5 of the traditional & non-traditional pipe tally format.

**Action Items**

* Create a standard EMAT pipe tally template using a recently delivered file:
  + Laila Hassen, Steven Lee
* Upload EMAT pipe tallies into Foundry sharepoint:
  + Laila Hassen, Steven Lee
* Prioritize 2001-2022 converted pipe tally ingestion; disregard 2023 for now:
  + Laila Hassen, Steven Lee, Sharan Patil

11/13/2023 Meeting Notes

* ILI Pipeline is inspected every 7 years
* Vendor sends one cumulative ILI Pipeline Tally, however, business team maintains each version in a respective folder structure to preserve each pipeline tally sent by the vender over a period of time to understand and maintain historical to present state of the given pipeline
* We can possibly use **Route and ML Survey date** as compound primary key - confirm with Laila/ILI Pipeline engineers
* Steven Lee is waiting to get permission in order to set up a new SharePoint location, which OIT engineer can use to create a permanent ILI Pipeline Tally dataset ingestion pipeline from Foundry
* Satvinder is interested in creating PowerBI reports using ILI Pipeline Tally datasets /object in Foundry at some point in future once the object is productionlized – live Foundry to PowerBI connectivity is possible - other Foundry users currently have similar projects, i.e. Gas Corrective PowerBI report
* Satvinder can create his own PowerBI report - does not need GFP or another team to create it for him - he just wanted to know if PowerBI and Foundry connectivity was possible

Action Items

* Business SharePoint Location for ILI Pipeline Tally files:
  + Steven Lee

11/20/2023 Meeting Notes

* Column A-D in Pipetel\_Pipe\_Tallly... excel file are added in addition to columns vendor sends
* Ingest all traditional and non-traditional tallies using template 3 from 2001 through 2022
* Ingest 2023 tallies after Steven converts them to newest template 4
* Route, Year, and ILI Survey Distance (ft) may make the row unique
* Sharan will analyze it further once the data is ingested into Foundry
  + We may need to just add an arbitrary primary key because there is no best option for a row to be unique using existing pipeline tally sheet columns that vendor sends
* Follow-up tasks

| Task | Assigned to | Due date |
| --- | --- | --- |
| Move ALL pipeline tally data CSV with 2001 to 2022 on to newly created ILI Pipeline SharePoint | Thanki, Jayna |  |

11/27/2023 Meeting Notes

* Vendor sends one cumulative ILI Pipeline Tally. Business adds each new Pipeline Tally file to historical Pipeline Tally file - they are always appended and never replaced due to Pipeline Tally Project Name being different for each respective Pipeline Tally file the vendor sends
  + OIT data engineer to create one dataset to ingest all historical pipeline tally files and then set up a monthly schedule to ingest each respective month's Pipeline Tally files into the same single dataset
* Business normally gets 20-50 Pipeline Tally files a year
* Foundry Schedule - Monthly - end of 2nd week of each month
  + Set up to run on 15th of each month
* Business to create an Excel file with file name and row count and put it in same location as the monthly ingestion location so OIT can use that to do Fidelity checks for each Pipeline Tally file ingested into Foundry
  + If any discrepancies found, OIT / OPS team to look into the errors and follow up with business as appropriate, if needed
* Satvinder, Steve, Allen, and Laila's replacement will do data QA for first 3 months
* We need a Business Data Stewards name to assign as primary contact in Foundry for Pipeline Tally BDS for Object Public Maturity Level 2 classification
  + Shared BDS Onboarding PPT again in chat for Satvinder since he joined after the initial meetings where this was shared
  + Referred Satvinder to Matt for further questions on BDS onboarding, Foundry training and MEA access process

**Follow-up tasks**

| Task | Assigned to | Due date |
| --- | --- | --- |
| Set up Foundry Data Pipeline Service Account Access to business SharePoint | Lee, Steven  Patil, Sharan | Tue, Nov 28 |
| Set up Data Ingestion Pipeline to ingest ILI Pipeline Tally history file into Foundry Raw and Clean layer, set up data fidelity checks | Patil, Sharan |  |
| Analyze ingested Pipeline Tally History dataset for unique row identification, i.e. Route, Year, and ILI Survey Distance (ft) make up a unique row or need to add an arbitrary Index/ID column as Primary Key? | Patil, Sharan |  |
| Foundry Object Business Data Steward Assignment, Business Data Steward General Responsibilities, High Level Foundry Applications / Usage Navigation & Overview | Sidhu, SatvinderLee, Steven Siegmund, Matt Kumari, Archana |  |
| Pipeline Tally Metadata & CDE Identification | Siegmund, Matt |  |

12/4/2023 Meeting Notes

**Agenda**

* What to use as ILI Pipeline Tally Dataset Primary Key
  + **Route-Year-Survey Distance combined do not make a row unique**
  + We tried to create a **hash key** using each row and use that as the primary key for each given row, however, found 700 duplicate rows, which would result in duplicate hash key
    - Need these duplicates removed if we proceed with using hash key as primary key
  + Can vendor send any other key to qualify data as unique?
  + Can we use any other combination of existing fields to make a unique row?
* How to manage ***718 duplicate rows*** discovered in ILI Pipeline Tally dataset in Foundry after ingestion using [All\_2001-2022\_ILIData.csv](https://pge.sharepoint.com/:x:/r/sites/ODGTeam/Shared%20Documents/Conceptual%20Model/Gas%20Ontology%20Federated%20Team/Gas%20TIMP%20ILI%20Pipeline%20Tally%20Public%20Ontology/ILI%20Data%20Traditional%20and%20Non-Traditional%20.csv/All_2001-2022_ILIData.csv?d=wa77bd7deacdc4541b6140b3fdf895a5a&csf=1&web=1&e=Ri4ZFy)
  + Can we delete these?
  + Will there be duplicates coming in future data ingestions?
    - How do we manage duplicates in future data ingestion jobs?

# Meeting notes

* Survey distance values are blank, which causes the row to be not unique
* ILI Pipeline Tally 2001 - 2022 ingested dataset in Foundry has about 718 duplicate rows
  + Had data quality issues with blank rows from vendor in 2018, which caused duplicate data issues
  + Vendor has been educated on data expectations to avoid data quality issues going forward

## **Team Decisions**

* + Bring in all ILI Pipeline Tally data from [All\_2001-2022\_ILIData.csv](https://pge.sharepoint.com/:x:/r/sites/ODGTeam/Shared%20Documents/Conceptual%20Model/Gas%20Ontology%20Federated%20Team/Gas%20TIMP%20ILI%20Pipeline%20Tally%20Public%20Ontology/ILI%20Data%20Traditional%20and%20Non-Traditional%20.csv/All_2001-2022_ILIData.csv?d=wa77bd7deacdc4541b6140b3fdf895a5a&csf=1&web=1&e=mTXkni) as is into Foundry Raw and Clean layers including any duplicate rows
  + Bring in same dataset from Clean Layer on to Base Layer including Duplicate rows
    - Business wants ability to see data as it was sent by Vendor including duplicates
  + Use Hash Key as Primary Key for ILI Pipeline Tally object in Foundry starting from Curated Layer
    - Starting from Ontology Curated layer, remove any duplicate rows found in [All\_2001-2022\_ILIData.csv](https://pge.sharepoint.com/:x:/r/sites/ODGTeam/Shared%20Documents/Conceptual%20Model/Gas%20Ontology%20Federated%20Team/Gas%20TIMP%20ILI%20Pipeline%20Tally%20Public%20Ontology/ILI%20Data%20Traditional%20and%20Non-Traditional%20.csv/All_2001-2022_ILIData.csv?d=wa77bd7deacdc4541b6140b3fdf895a5a&csf=1&web=1&e=mTXkni) to ensure Hash Keys are unique and suitable to be used as Primary Key for the object
  + ILI Pipeline Tally dataset for year 2023 will be ingested into Foundry sometime in 2024 - as prioritized by ILI Pipeline Tally business team