
# LeaPS Project Brief

# Project Summary

- Model predicts whether or not a pipe will burst (1=burst, 0=non-burst)
- Predictions are static and binary
- Data is static and does not get updated
- Model does not give any time scale for predictions
- Model does not predict ‘amount of leakage’

# LeaPS Demo

https://leaps.stackworx.dev/home

# 00 Network Tiers

**1. Pipe Level**

- Individual sections of pipe.
- 1,241,801 assets currently comprise STW network.

**2. Isolation Level**

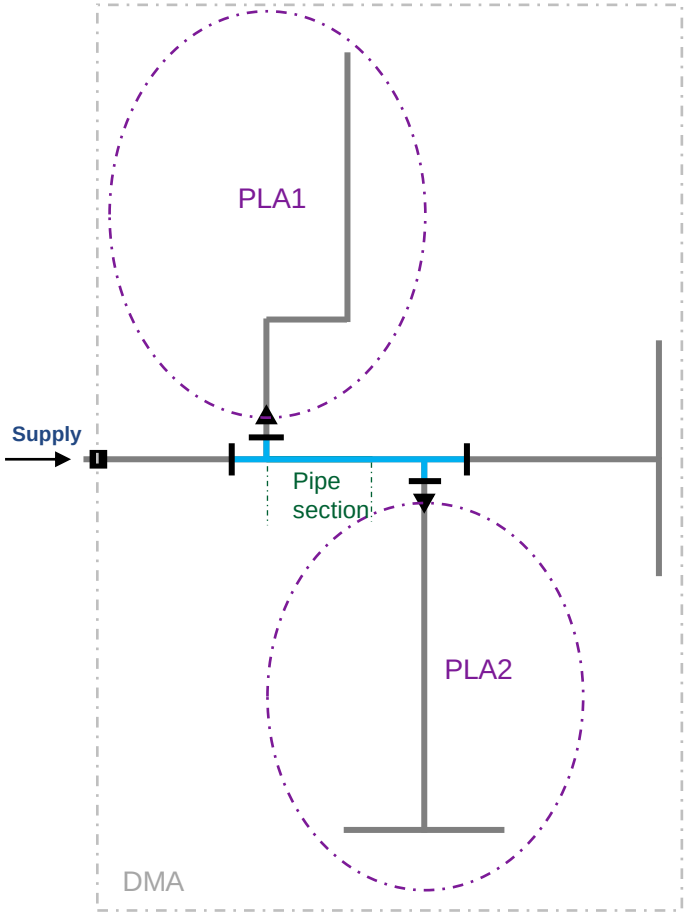
- Valve to valve groups of pipes.
- 346,707 isolations comprise STW network.

**3. PLA Level**

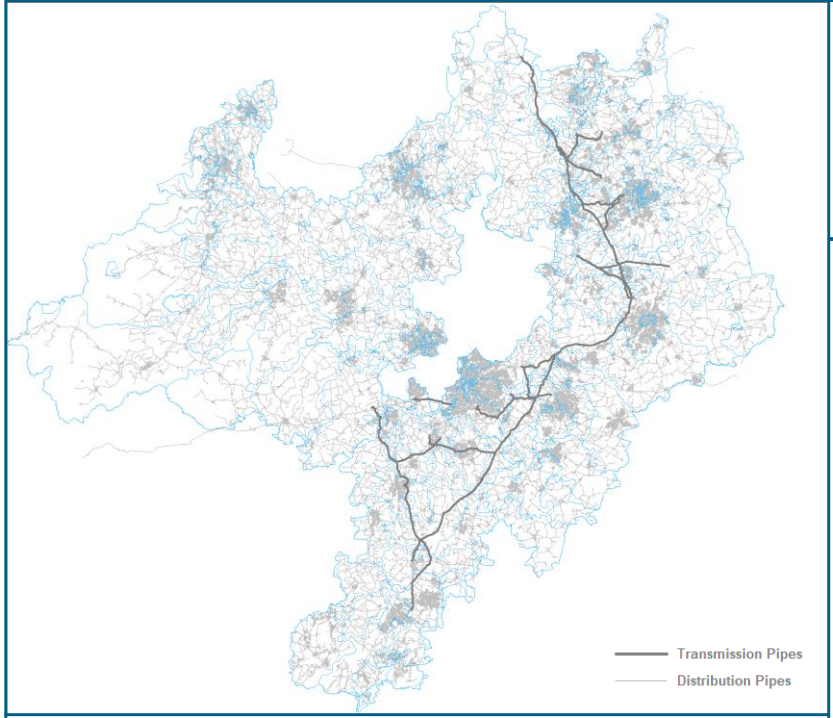
- Pressure Logged Area.
- Independently monitored pressure defined areas (I.e. PRV'd area).
- 5,680 unique PLA's in current STW network.

**4. DMA Level**

- District Metered Area.
- Independently monitored and defined areas of demand.
- 4,349 unique DMA's in current STW network.



# 01 Asset Data – Pipe Extract




## Coverage:

- Contains **1,241,801** pipe asset records comprising **100%** of the STW clean water network to be assessed within LeaPS.

## Summary:

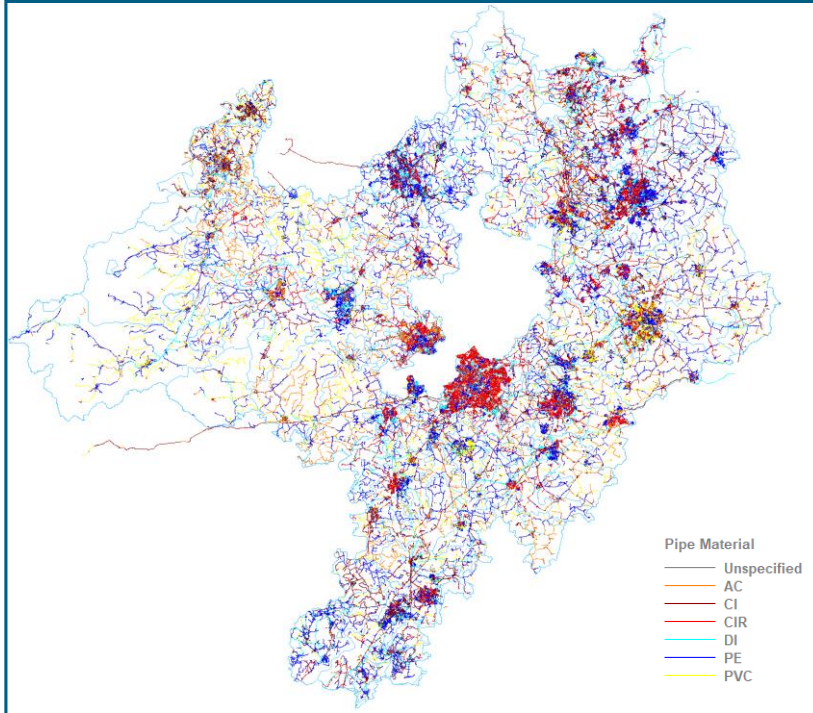
- Direct 'Pipe' extract from STW corporate GIS, reflects network 'as is' within the organisation. (early 2020 version, which best aligns with majority of the data sets used elsewhere in the analysis).
- Majority of 'Pipe Level' data will be associated back to this data set (using "UID" / "Tag" refs).

## Content:

- *36 Columns total, including…..*

| Column | Example | Null% | Description |
|---|---|---|---|
| Tag | 1398010890 | 0.0% | Pipe unique ref |
| Type | D | 0.0% | Distribution (D), or Transmission (T) |
| Measured_l | 231.7 | 0.0% | Pipe length (m) |
| Nominal_di | 90 | 2.6% | Pipe diameter (relative to specified units) |
| Nominal__1 | mm | 2.5% | Pipe diameter specified units (mm or in) |
| Material | HPPEP | 5.6% | Pipe material |
| Water_Type | PW | 0.0% | Water type, potable (PW) or raw water (RW) |
| Id_dma_cod | 7354 | 3.3% | Pipe DMA association |
| Id_control | CGSV26 | 0.7% | Pipe Control Group Association |
| Id_pla_cod | 7365 | 14.1% | Pipe PLA association |
| CWDP_No | 17 | - | Number of customer connections |

# 01 Asset Data - Asset Base




## Coverage:

- Contains **1,241,801** pipe asset records comprising **100%** of the STW clean water network to be assessed within LeaPS.

## Summary:

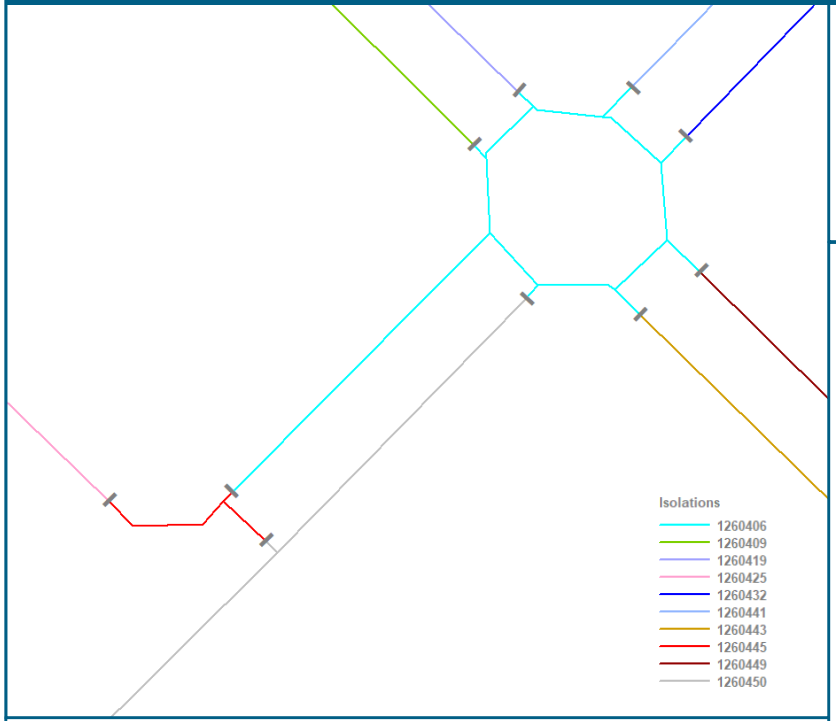
- Data set aligned to the “Pipe Extract” to provide infilled values relating to pipe material, age and diameter characteristics.
- Majority of infilled ‘Model_’ values completed by STW, however any new assets >2018 infilled by MWH (due to limited maintenance of the dataset within STW).

## Content:

- *35 Columns total, including…..*

| Column | Example | Null% | Description |
|---|---|---|---|
| Pipe_ID | 1398080549 | 0.0% | Pipe unique ref |
| Environment | Rural | 13.7% | Urban/rural designation |
| Original_Yearlad | | 57.6% | Raw 'Year Laid' value |
| Original_Material | CI | 5.6% | Raw 'Material' value |
| Original_Diameter | 1 | 3.0% | Raw 'Diameter' value |
| Soil_Type_Des | shallow loam over limestone | 9.6% | Soil description (raw STW built data) |
| SoilType_CG | Top Soil | 14.6% | Soil type (raw STW built data) |
| Model_Yearlaid | 1905 | 3.5% | Processed/infilled 'Year Laid' value |
| Model_Material | CI | 0.5% | Processed/infilled 'Material' value |
| Model_Diameter | 75 | 3.0% | Processed/infilled 'Diameter' value |

# 01 Asset Data - Isolations




## Coverage:

- Derived from **1,241,801** pipe asset records comprising **100%** of the STW clean water network to be assessed within LeaPS.
- Results in **346,707** 'Isolation' data elements

## Summary:

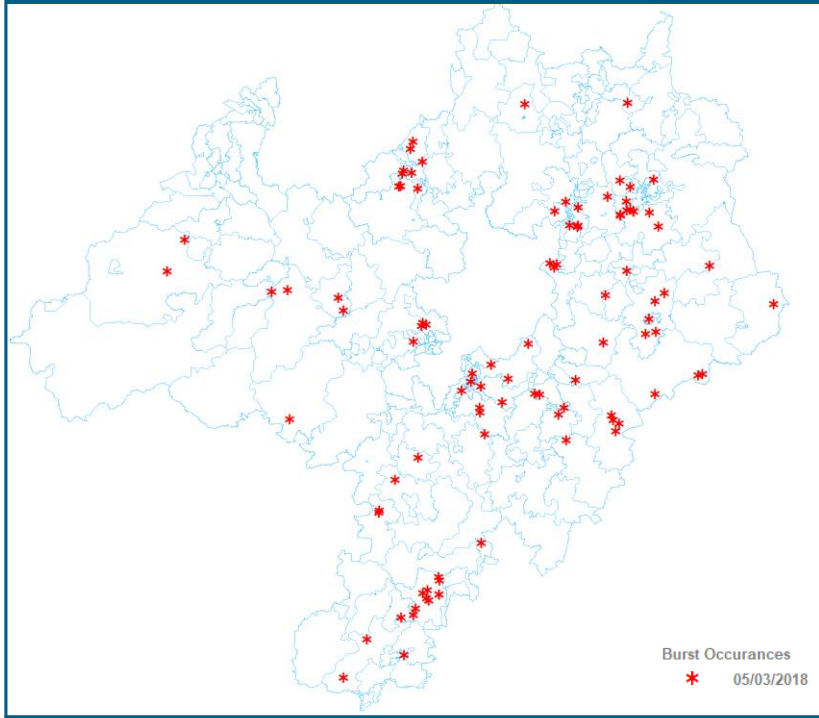
- The 1.25 million pipe assets that comprise the STW network are separated into groups by valves (424,865 valves in STW).
- These groups of pipes are known as 'isolations' and 346,707 unique isolations exist in the current STW network.

## Content:

- *12 Columns total, including…..*

| Column | Example | Null% | Description |
|---|---|---|---|
| Isolation_No | 1000057 | 0.00% | Isolation unique ueference |
| Pipe_No | 9 | 0.00% | Individual pipe sections within isolation |
| Tot_Len | 652.19 | 0.00% | Total length of isolation (m) |
| Valve_No | 3 | 0.00% | Total number of valves comprising isolation boundary |
| Closed_Valve_No | 1 | 0.00% | Number of closed valves comprising isolation boundary |
| CG | CGSV03 | 0.80% | Isolation Control Group association |
| DMA | 2731 | 4.40% | Isolation DMA association |
| CWDP_No | 9 | 0.00% | Isolation number of customer connections |
| Circumference | 564.996 | 0.00% | Circumerence of all valves associated with isolation (m) |
| Volume_m3 | 2.42 | 0.00% | Total volume of isolation (m3) |

# 01 Mains Repair – SAP Burst Data (example set)




## Coverage:

- **Example data** set from whole STW region:
  - Containing 49,855 burst records
  - Time period 07/06/2010 to 30/04/2019

## Summary:

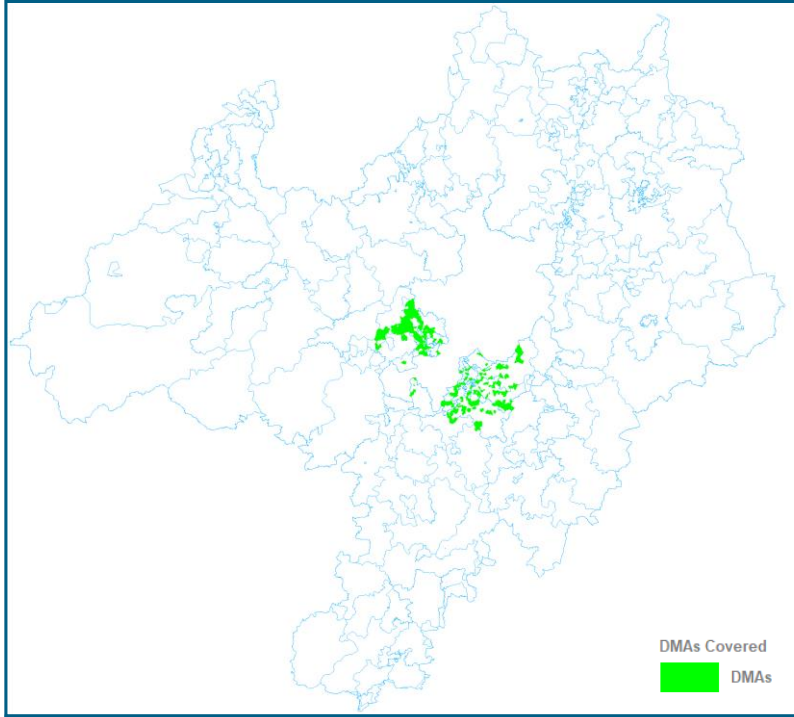
- SAP Mains repair data, contains all 'confirmed' burst occurrences in the STW network.
- 'Validation' routines applied to data to identify such things as duplicate entries (as per "Valid…" columns in data).

## Content:

- *16 Columns total, including…..*

| Column | Example | Null% | Description |
|---|---|---|---|
| FINAL_ASSET | 1200051092 | 7.8% | STW assigned burst pipe reference |
| FWR_SIZE | 4" | 0.0% | Pipe size recorded against the repair |
| FWR_MATERIAL | CAST IRON | 1.2% | Material recorded against the repair |
| FDB_TECHNIQUE | UNDER PRESSURE | 1.2% | technique of repair |
| WORK_ORDER | 11831009/20 | 1.2% | order reference assigned to the repair |
| DMA_NAME | 4567 | 1.2% | DMA reference assigned to the repair |
| OPERATION_ STD_TEXTKEY | MRC4 | 1.2% | repair code reflecting type/diameter |
| FWR_CREATE _DATE | 2.01803E+16 | 1.5% | date record/repair job created |
| PIPE_ID | 927917 | 0.5% | MWH assigned pipe ID (associates back to UID) |
| Valid | | N/A | record validity (Blank or "N") |
| Valid_Class | 0 | N/A | Record validity classification code |

# 01 Time Based Data – DMA Flow (example set)

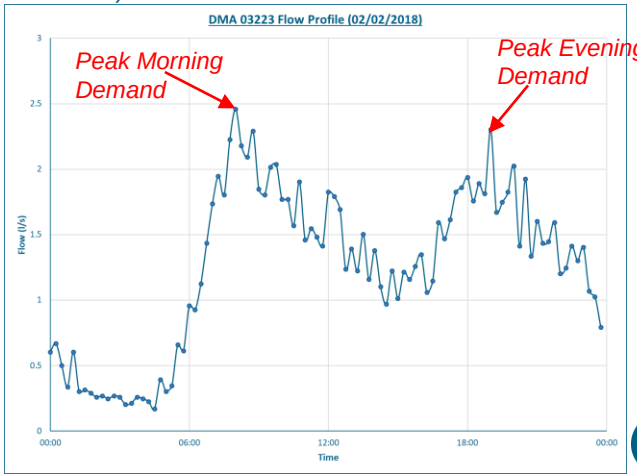



## Summary:

- DMA daily flow data
- .csv rows contain each time step (15 minute increments) and columns contain each DMA observed flow value.
- Each value reflects the observed flow rate at that time (l/s)

## Content:

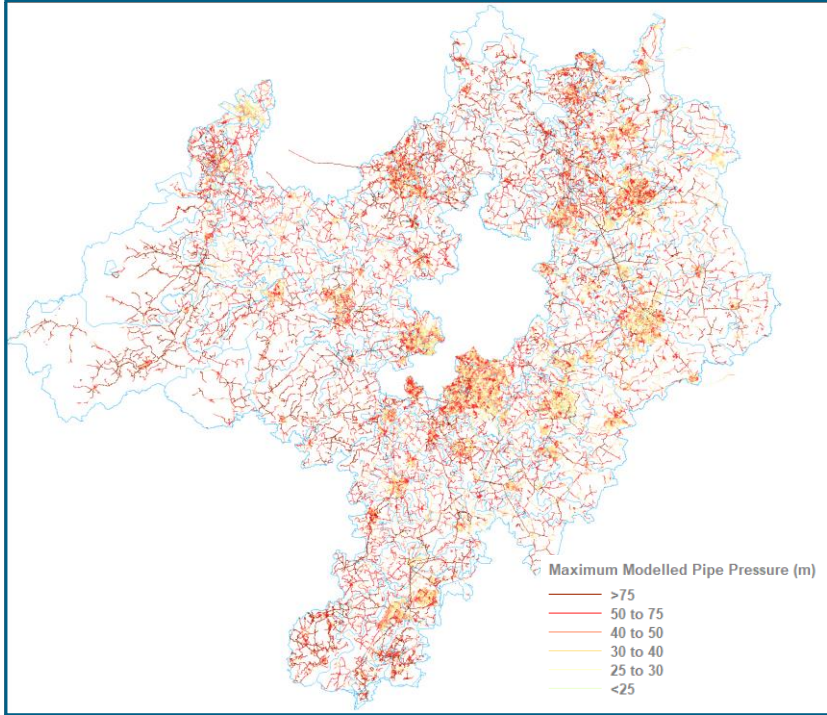
- 15 minute observations provide flow/demand change information across any given period (below chart reflects 24 hour period)
- Values can be used to also calculate Total daily flow (TDF) for DMA using the incremental values. (DMA 03223 TDF = 0.11Ml)




## Coverage:

- **Example data** set from Central region:
  - Containing 125 DMAS
  - Time period 01/02/2018 to 01/09/2018

# 02 Hydraulic Models – Performance (Pressure)




## Summary:

- Pipe pressure data derived from the STW hydraulic model library.
- Each hydraulic model reflects 24 hour performance of each designated STW (Control Group) system.
- Such things as pressure, flow and head loss can be derived from the models and associated back to individual STW pipe assets.

## Content:

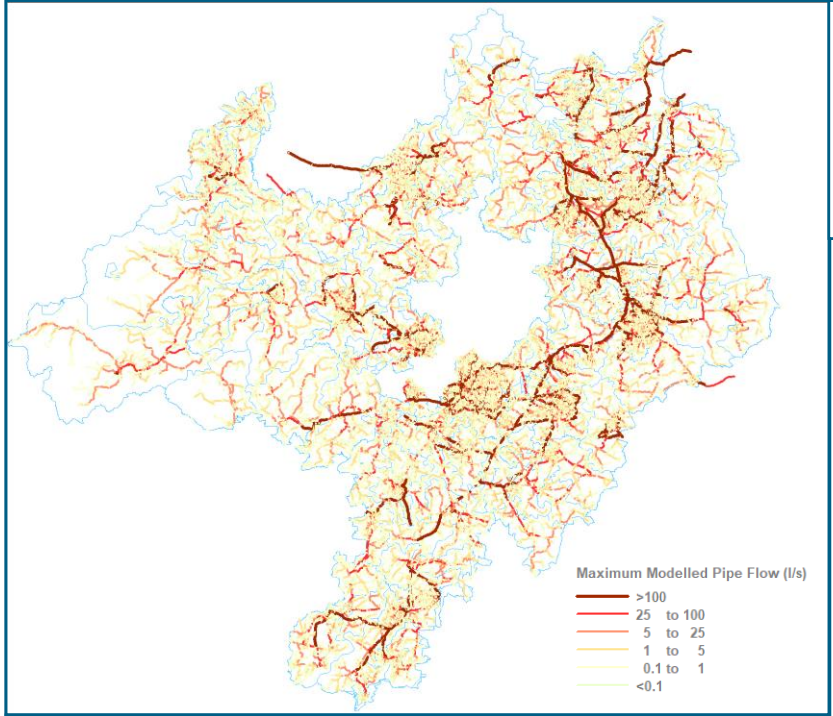
- *7 Columns total*

| Column | Example | Null% | Description |
|---|---|---|---|
| Tag_text | 1398010890 | 0.0% | Pipe unique ref |
| Isolation | 1000002 | 0.0% | Pipe isolation association |
| Pres_Min | 69.2 | <0.1% | Minimum modelled pressure (m) - across 24 hour run |
| Pres_Min_Time | 9.5 | <0.1% | Time of modelled minimum pressure (e.g. 09:30) |
| Pres_Avg | 78.7 | <0.1% | Average modelled pressure across 24 hour model run |
| Pres_Max | 82.6 | <0.1% | Maximum modelled pressure (m) - across 24 hour run |
| Pres_Max_Time | 3.75 | <0.1% | Time of modelled maximum pressure (e.g. 03:45) |

## Coverage:

- Contains **1,148,477** pipe asset records comprising **92.5%** of the STW clean water network to be assessed within LeaPS.
- Assets not included will comprise such things as, Strategic Grid mains, new assets not in models (>2019) and mains not operationally necessary in the hydraulic model.

# 02 Hydraulic Models – Performance (Flow)




## Summary:

- Pipe flow data derived from the STW hydraulic model library.
- Each hydraulic model reflects 24 hour performance of each designated STW (Control Group) system.
- Such things as pressure, flow and head loss can be derived from the models and associated back to individual STW pipe assets.

## Coverage:

- Contains **1,148,477** pipe asset records comprising **92.5%** of the STW clean water network to be assessed within LeaPS.
- Assets not included will comprise such things as, Strategic Grid mains, new assets not in models (>2019) and mains not operationally necessary in the hydraulic model.

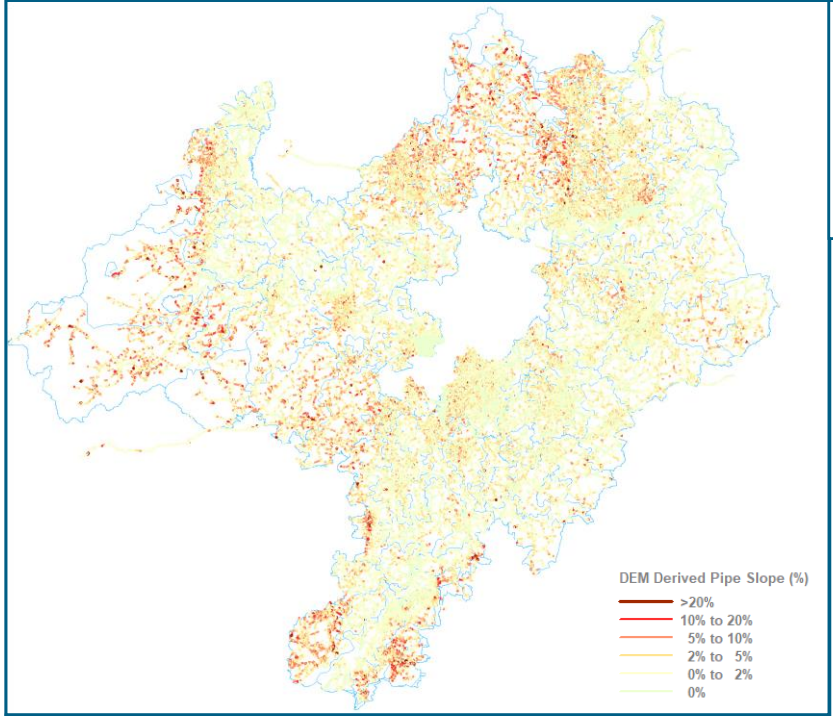
## Content:

- *8 Columns total*

| Column | Example | Null% | Description |
|---|---|---|---|
| Tag_text | 1200507317 | 0.0% | Pipe unique ref |
| Isolation | 1001395 | 0.0% | Pipe isolation association |
| Flow_Min | 0.3 | <0.1% | Minimum modelled flow (l/s) - across 24 hour run |
| Flow_Min_Time | 3 | <0.1% | Time of modelled minimum flow (e.g. 03:00) |
| Flow_Max | 0.9 | <0.1% | Maximum modelled flow (l/s) - across 24 hour run |
| Flow_Max_Time | 9.0 | <0.1% | Time of modelled maximum flow (e.g. 09:00) |
| Headloss_Max | 0.0 | <0.1% | Maximum modelled headloss (m/km) - across 24 hour run |
| Headloss_Max_Time | 9 | <0.1% | Time of modelled maximum headloss (e.g. 09:00) |

- Approx. 3,000 records have “0” against all attributes however these appear to be genuine zero flow pipe assets at the extremities of the network.

# 02 Network Analysis – Slope




## Coverage:

- Contains **1,241,801** pipe asset records.
- **92.5%** of these have a model derived slope value.
- **98.3%** of these have a DEM derived slope value.
- **99.9%** data coverage when considering model or DEM data for each pipe asset.

## Summary:

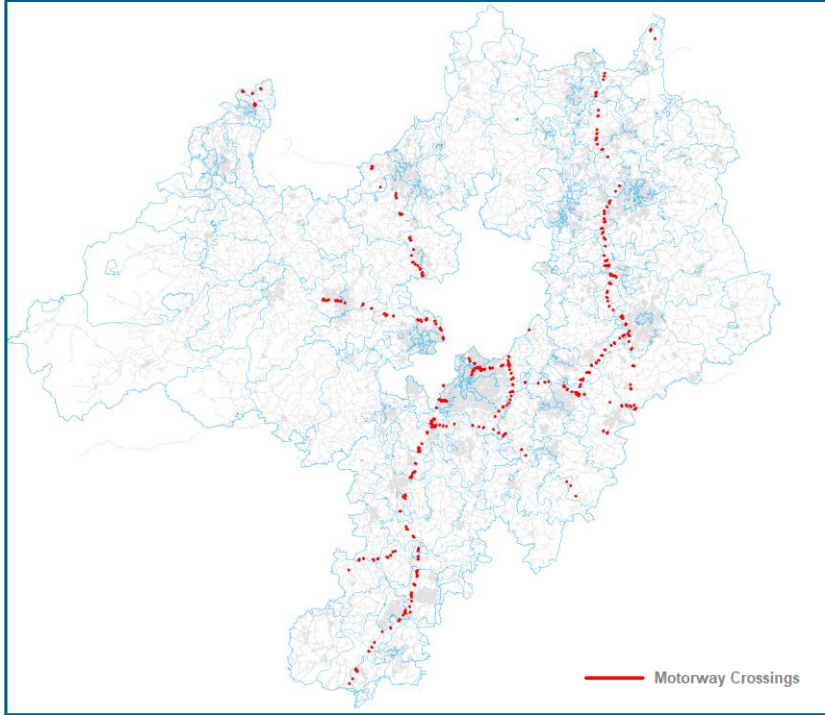
- Every pipe in hydraulic model comprises a start node and end node each with an assigned elevation that can be used to derive the pipe slope.
- Contours have also been used to build a Digital Elevation Model (DEM) of STW region, from which each pipes min and max elevation can be derived to calculate pipe slope.

## Content:

- *14 Columns total, including…..*

| Column | Example | Null% | Description |
|---|---|---|---|
| Tag_text | 1398010890 | 0.0% | Pipe unique ref |
| Isolation | 1000002 | 0.0% | Pipe isolation association |
| Mod_S_Elevation | 171.976 | 7.5% | Model derived pipe start node elevation |
| Mod_E_Elevation | 173.586 | 7.5% | Model derived pipe end node elevation |
| Mod_Slope_m | 1.61 | 7.5% | Model derived slope (m drop) |
| Mod_Slope% | 0.694965 | 7.5% | Model derived slope (slope %) |
| DEM_Min_Elevation | 172.3 | 1.7% | DEM derived pipe minimum elevation |
| DEM_Max_Elevation | 173.7 | 1.7% | DEM derived pipe maximum elevation |
| DEM_Slope_m | 1.4 | 1.7% | DEM derived slope (m drop) |
| DEM_Slope% | 0.624349 | 1.7% | DEM derived slope (slope %) |

# 02 Network Analysis – Location Classification




## Coverage:

- Contains **1,241,801** pipe asset records comprising **100%** of the STW clean water network to be assessed within LeaPS.

## Summary:

- Each pipe analysed geospatially against OS data to identify relationships to specific features such as Motorways.
- Resulting in a matrix of proximity factors for each pipe asset.

## Content:

- *37 Columns total, including…..*

| Column | Example | Null% | Description |
|---|---|---|---|
| Tag | 1398010881 | 0.0% | Pipe unique ref |
| Water_Crossing | | N/A | Pipe association with specified geographical feature:<br>Y = association<br>Null Or "N" = No association |
| Rail_Crossing | | N/A | |
| Road_Crossing | Y | N/A | |
| Motorway_Crossing | | N/A | |
| Tree_Prox | Y | N/A | |
| Natural_Land | Y | N/A | |
| Private_Land | | N/A | |
| SSSI | N | N/A | |

- Above contains some of the factors in data set, full data set contains >20 geographically defined pipe location considerations.

# 02 Network Analysis – Pipe Intersections


## Coverage:

- Contains **1,241,801** pipe “UIDs” comprising **100%** of the STW clean water network to be assessed within LeaPS.
- This is further divided into **2,548,078** individual records, reflecting each unique pipes association to individual geographical features.

## Summary:

- Some individual pipes in STW can be >1km in length.
- “one” pipe can intersect numerous geographical elements by varying degrees.
- This data set sub divides each pipe across the various geographical OS elements it is associated with.

## Content:

- *8 Columns total*

| Column | Example | Null% | Description |
|---|---|---|---|
| tag_text | 1200087520 | 0.0% | Pipe unique ref |
| Length_m | 2 | 0.0% | Intersection Length (m) |
| os_id | SP08 | 0.0% | OS tile relating to features |
| THEME | Roads Tracks And Paths | N/A | Feature theme |
| DESCRIPT1 | Roadside | N/A | Feature description 1 |
| DESCRIPT0 | | N/A | Feature description 2 |
| MAKE | Natural | N/A | surface "make", Natural/Manmade classification |
| Location_Criteria | Roadside Intersection | 2.0% | Assigned Criteria as per theme/description/make combination |

- Each Pipe asset (“tag_text”) may exist multiple times in the dataset reflecting the number of OS features that pipe intersects.
- Each row in the dataset reflects each individual intersection and represents the length of that intersection only.

# 03 NERC Land Cover 2015

## Summary:

- LCM2015 was produced at the Centre for Ecology & Hydrology by classifying satellite images from 2014 and 2015 into 21 Broad Habitat-based classes.
- Dataset comprised of 25m grid land use classifications
- Centroid of each pipe asset used to assign the LCM2015 land use classification at that location.

## Content:

- *4 Columns total, including…..*

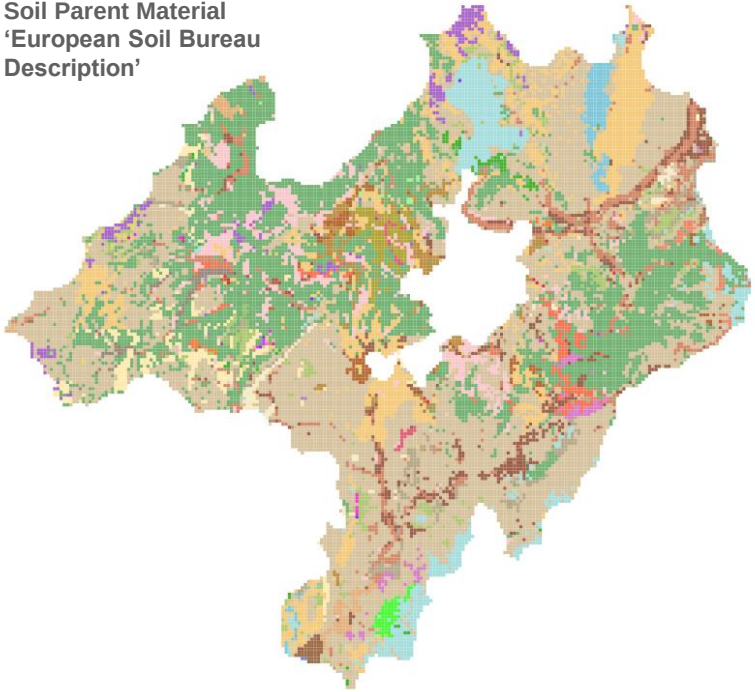
| Column | Example | Null% | Description |
|---|---|---|---|
| Tag_text | 1398080549 | 0.0% | Pipe unique ref |
| Band1 | 21 | 0.0% | Band1 land use code |
| Band1_ Classification | Suburban | 0.0% | Band 1 land use classification |
| Band2 | 152 | 0.0% | Band 2 land use code (RGB classification) |

## Coverage:

- LCM2015 full UK coverage.
- **100%** of STW assets can be assigned relevant values and classifications.

# 03 BGS Soil Parent Material

Soil Parent Material
'European Soil Bureau Description'



## Coverage:

- BGS full UK coverage.
- **100%** of STW assets can be assigned relevant values and classifications.

## Summary:

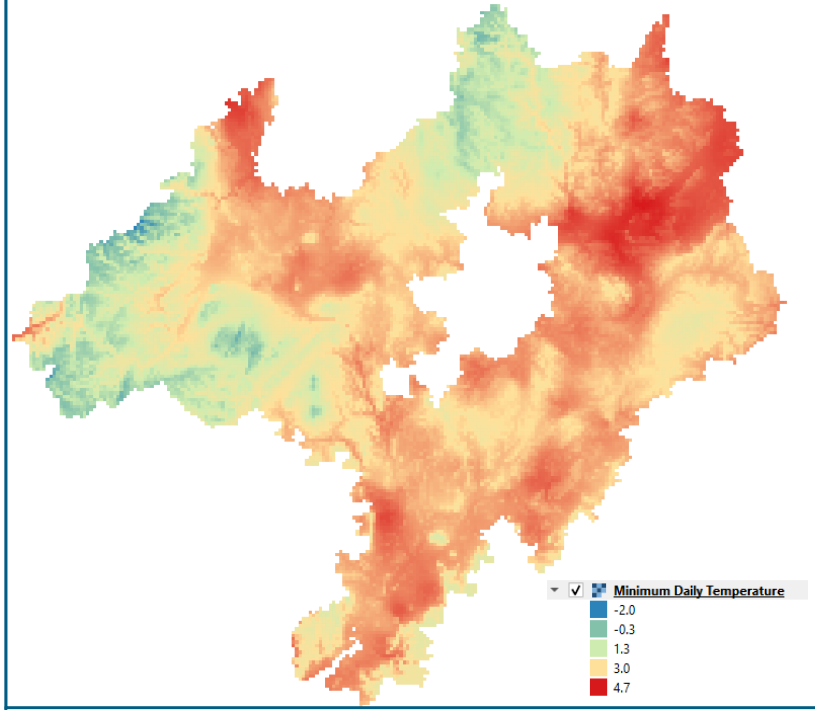
- British Geological Survey derived "Soil Parent Material" information.
- 1km tile resolution of 6 criteria relating to soil, including description, grain size and texture.
- Centroid of each pipe asset used to assign the BGS classifications at that location.

## Content:

- *10 Columns total, including…..*

| Column | Example | Null% | Description |
|---|---|---|---|
| Tag_text | 1398080549 | 0.0% | Pipe unique ref |
| Esb_desc | LIMESTONE | 0.0% | European Soil Bureau Description |
| Carb_cntnt | HIGH | 0.0% | Carbonate Content |
| Pmm_grain | ARGILLIC - ARENACEOUS | 0.0% | Parent Material Grainsize |
| Soil_group | MEDIUM TO LIGHT(SILTY) TO HEAVY | 0.0% | Soil Group |
| Soil_tex | CLAYEY LOAM TO SILTY LOAM | 0.0% | Soil Texture |
| Soil_depth | SHALLOW | 0.0% | Soil Depth |

# 03 HadUK – Grid Climate Observations




## Coverage:

- HadUK full UK coverage.
- **100%** of STW asset coverage for 2010 to 2019 environmental observations.

## Summary:

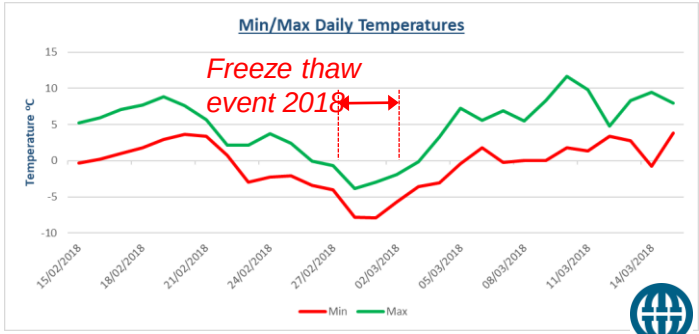
- HadUK 1km Grid climate observations for whole of UK
- Each observation metric (i.e. minimum daily temperature) available as a .NetCDF file containing the location, date and observed value for that metric.
- Post processing will be necessary to relate data back to pipe or DMA level asset association. (>350 individual .NetCDF files)

## Content:

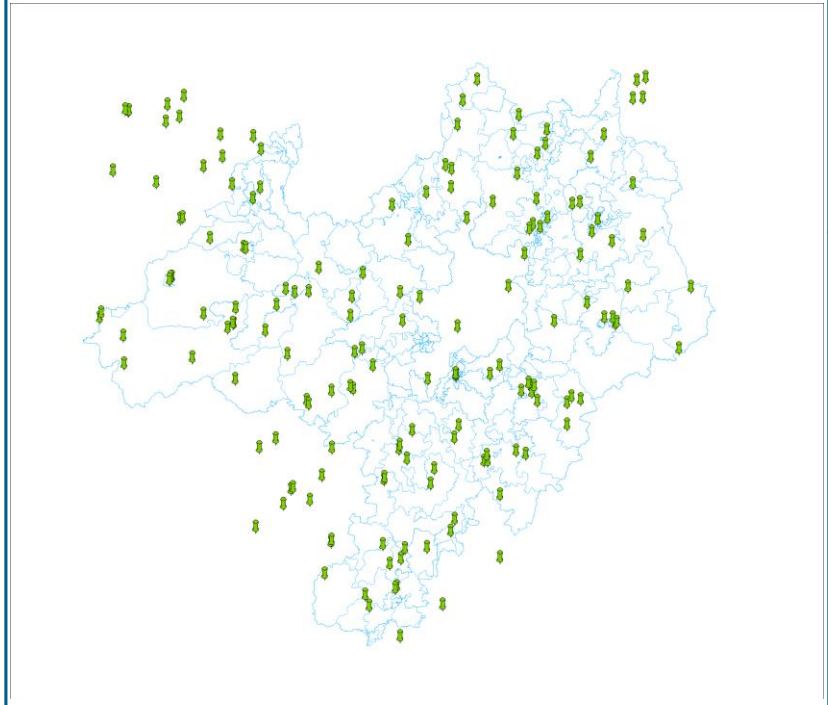
- Observations collected:
  - Rain (daily)
  - Minimum air temperature (Daily)
  - Maximum air temperature (Daily)
  - Ground Frost (Monthly Nº of days)
  - Snow Lying (Monthly Nº of days)

## Example:

- Buxton Feb/Mar 2018 (Air temperature)



# 03 MIDAS – Weather Stations



## Coverage:

- Contains **164** weather stations.

## Summary:

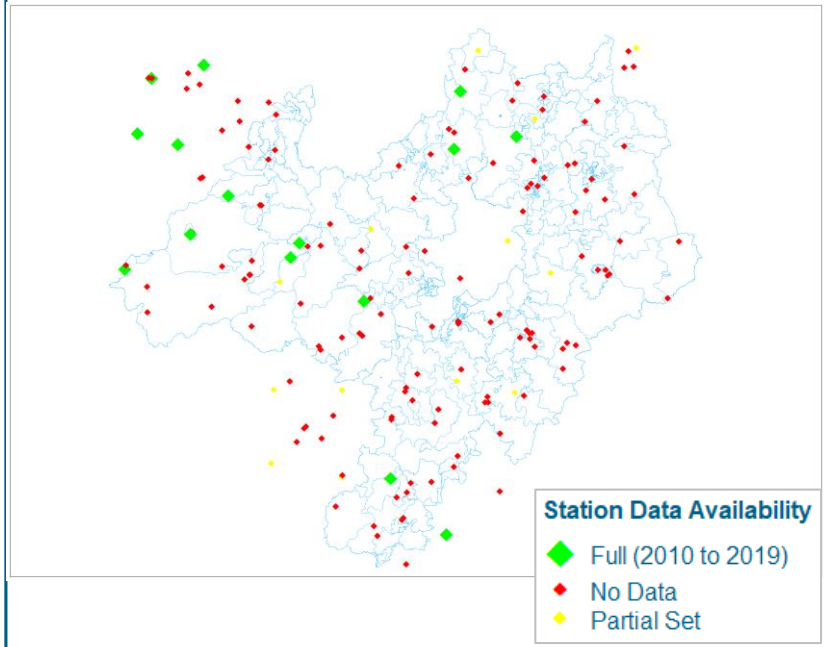
- Weather stations across the STW region, data sets include daily rain, daily temperature, soil temperature and hourly weather observations
- Each station has varying degrees of data completeness
- Full sets contain data from 2010-2019

## Content:

- *6 Columns total,*

| Column | Example | Null% | Description |
|---|---|---|---|
| Station Name | Rhyl | 0% | Name of weather station |
| Station Code | 1139 | 0% | Unique code of weather station |
| Region | Clwyd | 0% | County |
| Lat | 53.314 | 0% | Latitude of weather station |
| Long | -3.501 | 0% | Longitude of weather station |
| Elevation (m) | 9 | 0% | Elevation of weather station |

# 03 MIDAS – Daily Rain




## Coverage:

- 14 stations with full data coverage (2010 to 2019).
- 14 stations with partial coverage.
- 136 stations with no relevant ‘rain’ data.

## Summary:

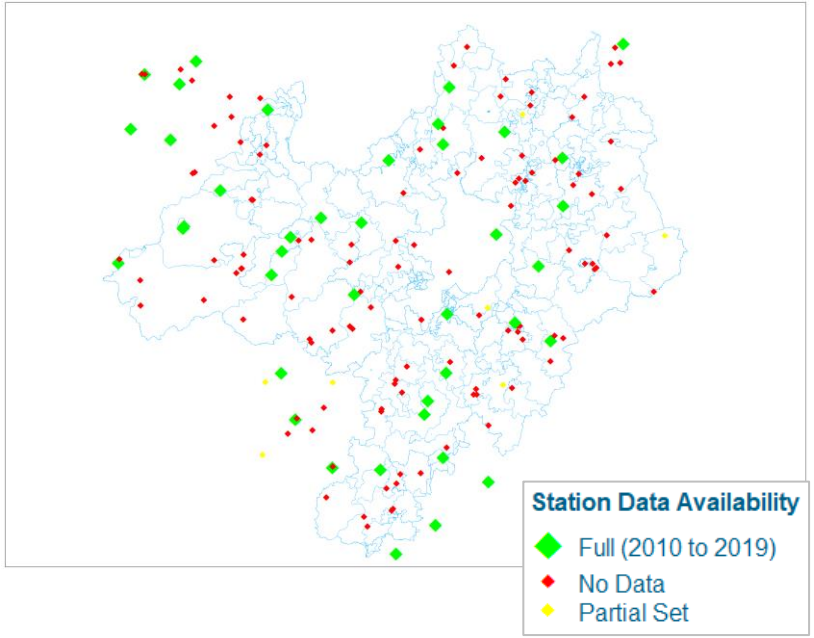
- Each Stations data within individual .csv files
- Rows 1 to 62 contain column details
- Row 63 onwards contains the timeseries data outlined below

## Content:

- *15 Columns total, including…..*

| Column | Example | Description |
|---|---|---|
| Ob_date | 01/01/2019 00:00 | Date/time of reading |
| Prcp_amt | 0.6 | mm of precipitation per day |
| Src_id | 539 | Unique code of weather station |

# 03 MIDAS – Daily Air Temperature




## Coverage:

- 39 stations with full data coverage (2010 to 2019).
- 9 stations with partial coverage.
- 116 stations with no relevant ‘air temperature’ data.

## Summary:

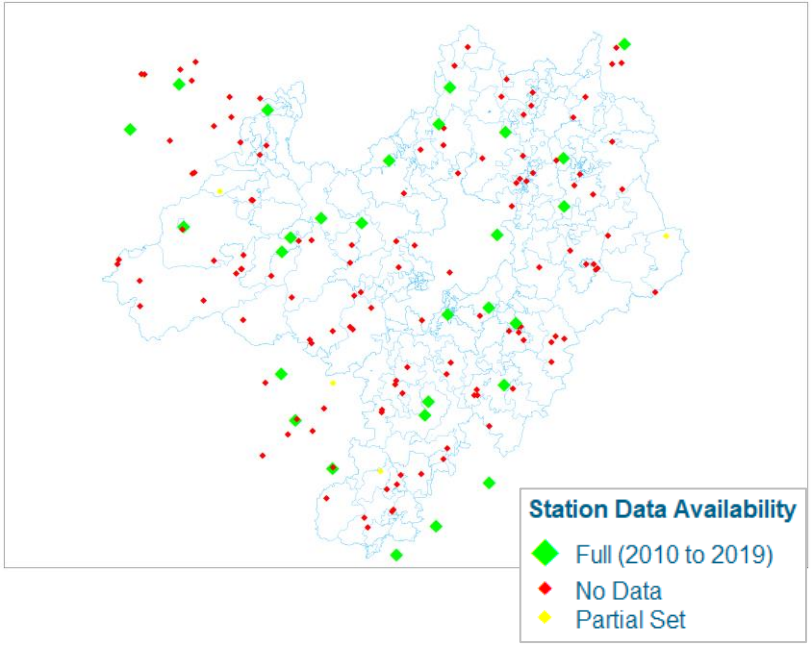
- Each Stations data within individual .csv files
- Rows 1 to 91 contain column details
- Row 92 onwards contains the timeseries data outlined below

## Content:

- *22 Columns total, including…..*

| Column | Example | Description |
|---|---|---|
| Ob_end_time | 01/01/2019 09:00 | Common periods include 12 hour observations ending at 0900 or 2100 or 24 hour observations ending at 0900 |
| Src_id | 17210 | Unique code of weather station |
| Max_air_temp | 8 | Maximum air temperature in day (degC) |
| Min_air_temp | -5.4 | Minimum air temperature in day (degC) |
| Min_grss_temp | -10.5 | Minimum grass temperature (degC) |
| ob_hour_count | 24 | number of hours the observation period spans |

# 03 MIDAS – Soil Temperature




## Summary:

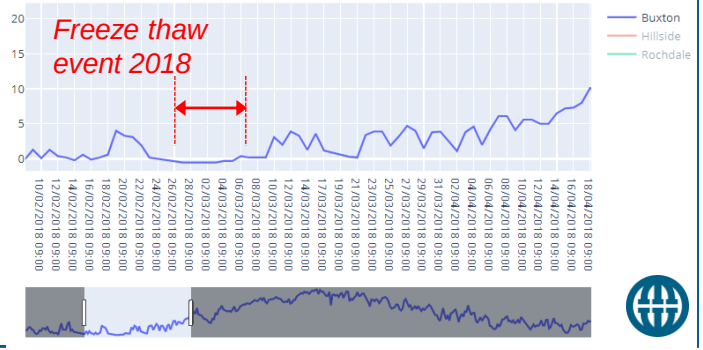
- Each Stations data within individual .csv files
- Rows 1 to 86 contain column details
- Row 87 onwards contains the timeseries data outlined below

## Content:

- *27 Columns total, including…..*

| Column | Example | Description |
|---|---|---|
| Ob_time | 01/01/2019 09:00 | Date and time of observation |
| src_id | 539 | Unique code of weather station |
| q10cm_soil_temp | 5.8 | Temperature of soil at 10cm depth (degC) |
| q30cm_soil_temp | 6.5 | Temperature of the soil at 30cm depth (degC) |

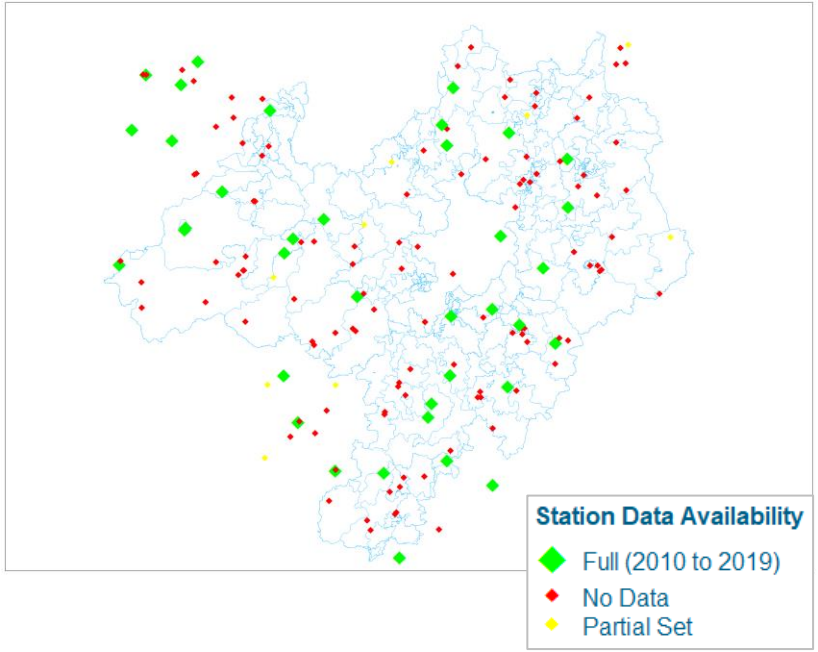
- Buxton Feb/Mar 2018 (Soil temperature)




## Coverage:

- 27 stations with full data coverage (2010 to 2019).
- 7 stations with partial coverage.
- 130 stations with no relevant ‘soil temperature’ data.

# 03 MIDAS – Hourly Weather Observations




## Coverage:

- 35 stations with full data coverage (2010 to 2019).
- 12 stations with partial coverage.
- 117 stations with no relevant ‘weather observations’ data.

## Summary:

- Each Stations data within individual .csv files
- Rows 1 to 281 contain column details
- Row 282 onwards contains the timeseries data outlined below

## Content:

- *104 Columns total, including…..*

| Column | Example | Description |
|---|---|---|
| ob_time | 01/01/2018 00:00 | Date and time (YYYY-MM-DD HH:MM) of the observation |
| src_id | 657 | Unique code of weather station |
| wind_speed_unit_id | 4 | 4=wind speed from anemometer, knots<br>0=wind speed estimated,metres per second<br>1=wind speed from anemometer,metres per second<br>3=wind speed estimated,knots |
| wind_direction | 230 | Wind direction in true degrees |
| wind_speed | 10 | Wind speed, knots |
| cld_ttl_amt_id | 7 | Total amount of cloud, Oktas |
| msl_pressure | 996.6 | mean sea level air pressure, hPa |
| air_temperature | 5.4 | Air temperature, degC |
| dewpoint | 1.8 | Dew point temperature, degC |
| wetb_temp | 3.7 | Wet bulb temperature, degC |
| rltv_hum | 78.1 | Calculated relative humidity, % |
| stn_pres | 992.5 | Station air pressure, hPa |

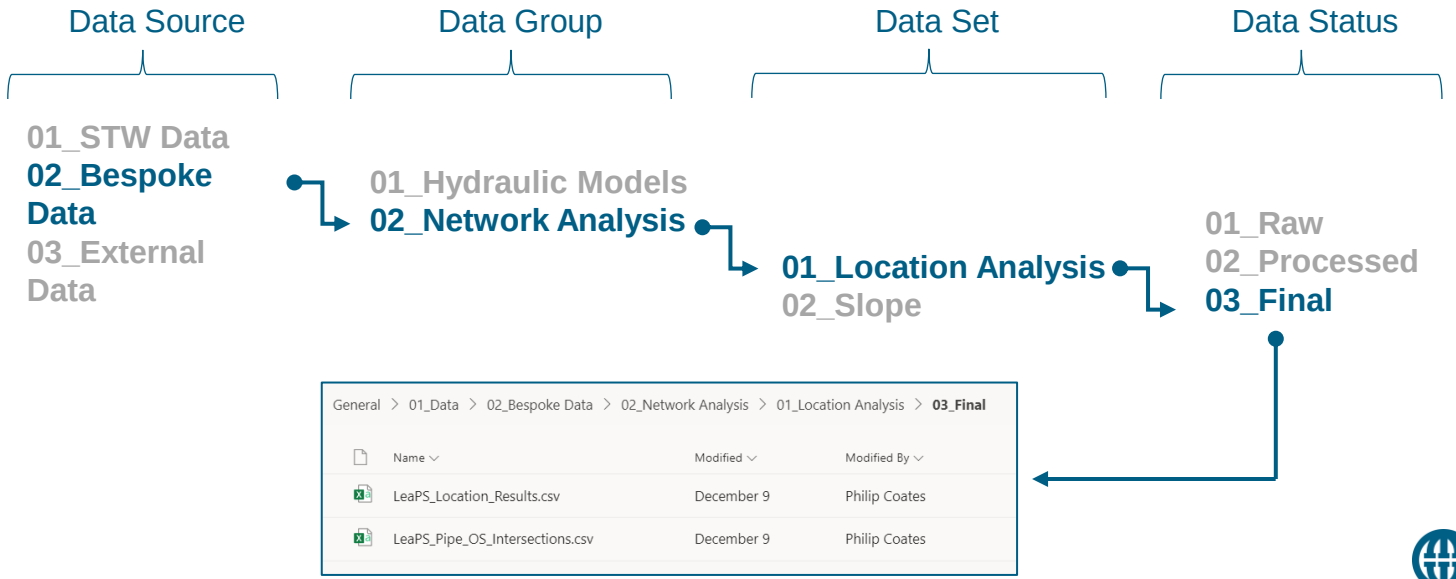
# Microsoft Teams Data Directory

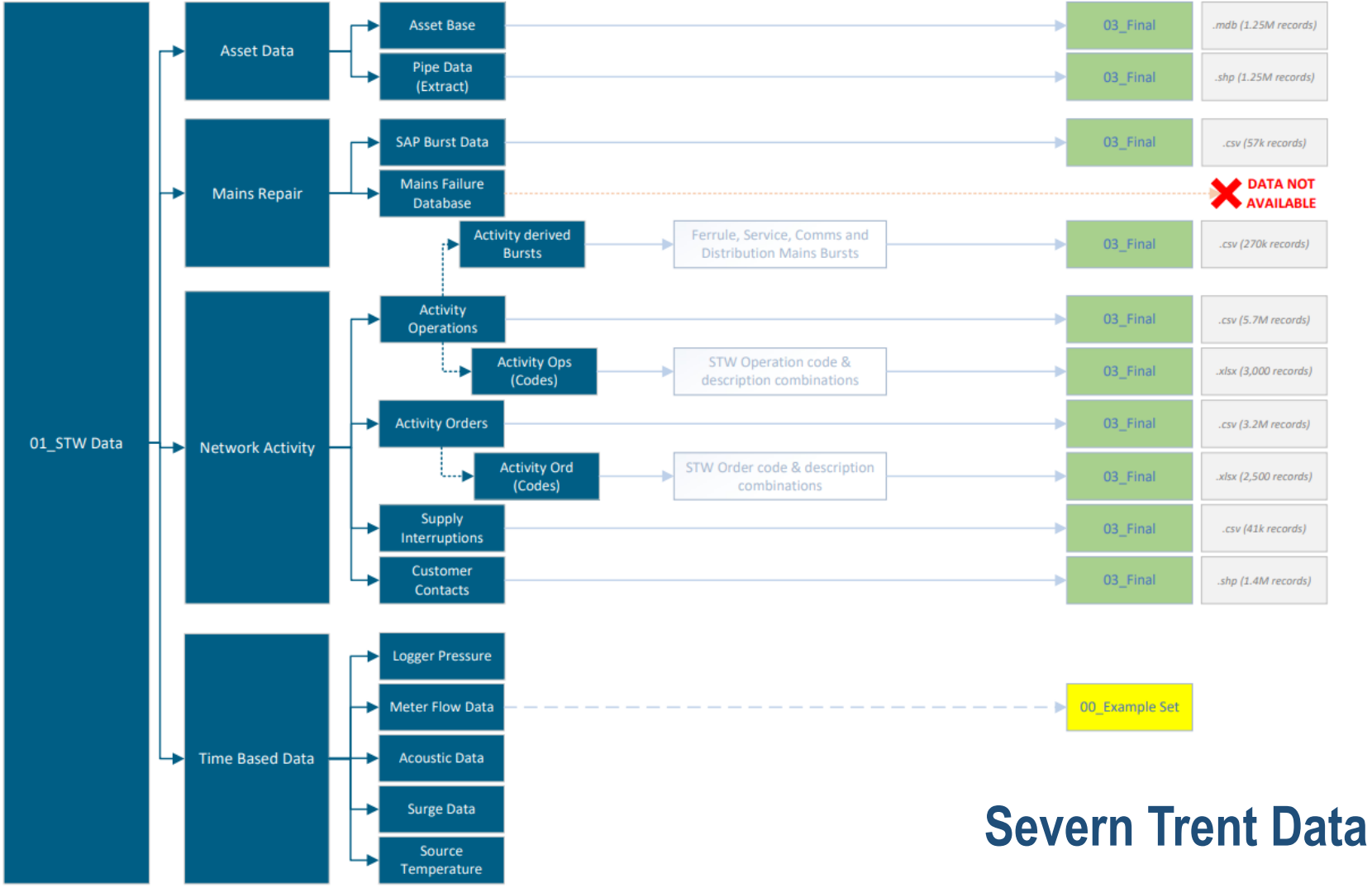
## Data Location:

- All Data referenced within these slides is available at the below Microsoft Teams location.
- All additional data sets still being built/collated will be available in the relevant sub-directory when completed.

https://teams.microsoft.com/_#/files/General?threadId=19%3A7ea032eeb72640dcb55f1792918106e6%40thread.tacv2&ctx=channel&context=01_Data&rootfolder=%252Fsites%252FMWHEnginsoftCollaboration%252FShared%2520Documents%252FGeneral%252F01_Data

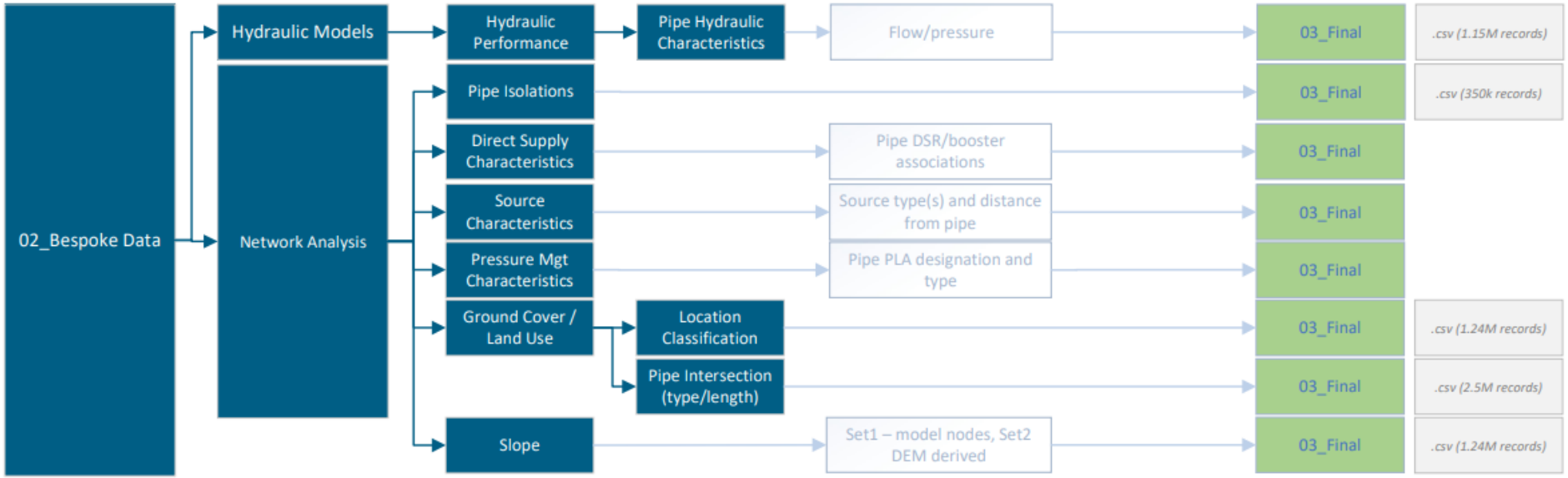
## Directory Structure:






# Severn Trent Data

# Bespoke Data



# External Data