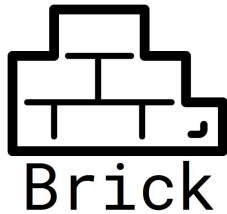# Dataset: An Open Dataset and Collection Tool for BMS Point Labels

Gabe Fierro gtfierro@cs.berkeley.edu

Sriharsha Guduguntla (UC Berkeley), David E. Culler (UC Berkeley)
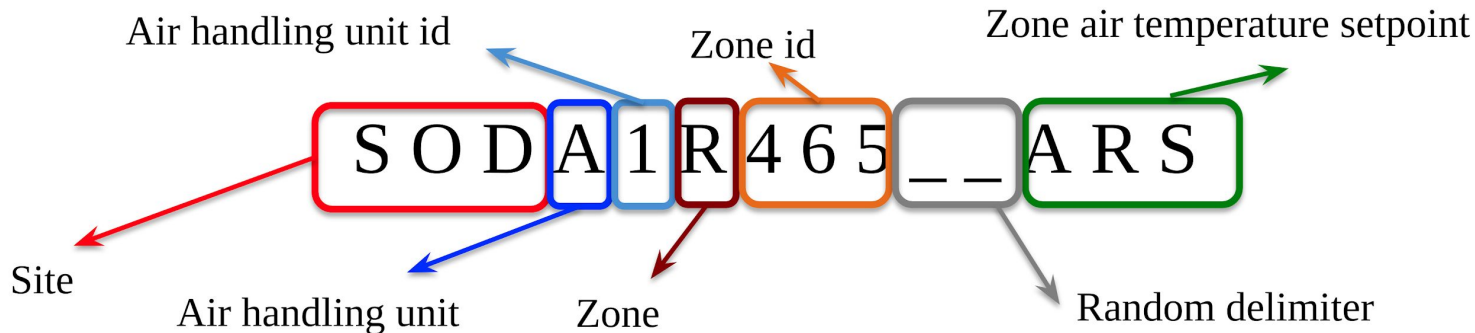
riselab
UC Berkeley

Standardized Building Metadata

- Enables increased penetration of energy efficiency measures, facilitates data analysis at scale, lower commissioning costs

- Growing interest and development on building metadata:

  - Academia, Industry, Government, Standards Bodies

©2017 RISELab

Challenges

- Transformation/standardization of **existing** building metadata is still a problem

- Metadata in built environment is characterized by **extreme heterogeneity**

  – Site-specific idioms and conventions

  – Inconsistent even within an enterprise

Air handling unit id

Zone id

Zone air temperature setpoint

S O D A 1 R 4 6 5 _ _ A R S

Site

Air handling unit

Zone

Random delimiter

Automated Translation

- Ongoing work to enable automated translation of existing, unstructured building metadata

    – Human-in-the-loop active learning (Bhattacharya 2015)

    – Clustering-based classification (Balaji 2015)

    – Transfer learning, NLP techniques (Koh 2018)

- Development, evaluation limited by access to **real-world building labels**

Dataset: Raw Building Metadata

- Dataset of attributes (labels, units, descriptions) for **103,064 points** for **92 buildings**
  - Mostly campus buildings

- Anonymized (no building names), but otherwise untreated

- Available online; periodic releases

- **Open source tool** for scraping and contributing additional data

Open Datasets

Raw Building Metadata

Semantic metadata standards for buildings such as Brick and Project Haystack show promise in enablilng wide-scale deployment of energy-efficiency measures and advanced building management technologies. Converting existing building metadata to these standard forms is an area of active research.

The data available here is a collection of metadata pulled from real building management systems around the world. By making this data available, we hope to facilitate research into the automated conversion of unstructured metadata into standard forms. The dataset currently contains attributes for **103,064 points** from **92 buildings**.

Data is distributed in CSV form:

- Periodic releases of all data: ZIP file
- Data for individual sites: CSV files.

This dataset is not static!. Email CSV dumps of building metadata (or questions) to data [AT] mortardata [DOT] org. We have also produced an open-source tool for scraping and cleaning building metadata from existing buildings.

Dataset: Raw Building Metadata

```
AH1R.SVFD:POWER
AH2R.SVFD:POWER
AH3R.SVFD:POWER
AHU.AH1N.CCO
AHU.AH1N.CCV
AHU.AH1N.COILTEMP
AHU.AH1N.Filter
AHU.AH1N.HALLWAY.Day Night
AHU.AH1N.HALLWAY.Heat Cool
AHU.AH1N.HALLWAY:AUX TEMP
AHU.AH1N.HALLWAY:CTL STPT
AHU.AH1N.HALLWAY:VLV 1 COMD
AHU.AH1N.HCO
AHU.AH1N.HCV
AHU.AH1N.MATSET
AHU.AH1N.Mixed Air Damper Position
```

Some with clear delimiters...

Dataset: Raw Building Metadata

```
AH1R.SVFD:POWER
AH2R.SVFD:POWER
AH3R.SVFD:POWER
AHU.AH1N.CCO
AHU.AH1N.CCV
AHU.AH1N.COILTEMP
AHU.AH1N.Filter
AHU.AH1N.HALLWAY.Day Night
AHU.AH1N.HALLWAY.Heat Cool
AHU.AH1N.HALLWAY:AUX TEMP
AHU.AH1N.HALLWAY:CTL STPT
AHU.AH1N.HALLWAY:VLV 1 COMD
AHU.AH1N.HCO
AHU.AH1N.HCV
AHU.AH1N.MATSET
AHU.AH1N.Mixed Air Damper Position
```

Some with clear delimiters...

```
SODA4R731__ASO
SOD31BLD_1_VAR
SOD31BLD_2_VAR
SOD31BLD_RESET
SOD31MAX_X_CVP
SOD31MAX___CVP
SOD31MAX___CVS
SOD31NET___STA
SOD31NET___TMR
SOD32BLD_1_VAR
SOD32BLD_2_VAR
SOD32BLD_RESET
SOD32NET___STA
SOD32NET___TMR
SOD33BLD_1_VAR
SOD33BLD_2_VAR
SOD33BLD_RESET
```

...and a few without

Dataset: Raw Building Metadata

bas_raw	point_type	haystack
CMU/SCSC Gates/Eighth Floor/8126 Machine Room CRAC-9/% Capacity	% Capacity	equip chiller coolingCapacity
CMU/SCSC Gates/Eighth Floor/8126 Machine Room CRAC-9/Fan Run Hours	Fan Run Hours	fan vfd equip run cmd
CMU/SCSC Gates/Eighth Floor/8126 Machine Room CRAC-9/Humidifier Run Hours	Humidifier Run Hours	humidifier run cmd
CMU/SCSC Gates/Eighth Floor/8126 Machine Room CRAC-9/Humidity	Humidity	humidity sensor
CMU/SCSC Gates/Eighth Floor/8126 Machine Room CRAC-9/Temperature	Temperature	temp sensor
CMU/SCSC Gates/Eighth Floor/FCU-15/Fan Coil CHW Vlv	Fan Coil CHW Vlv	fcu fan chilled water valve cmd
CMU/SCSC Gates/Eighth Floor/FCU-15/Fan Start	Fan Start	fan cmd
CMU/SCSC Gates/Eighth Floor/FCU-15/Fan Status	Fan Status	fan curstatus
CMU/SCSC Gates/Eighth Floor/FCU-15/FANCOIL ALARM	FANCOIL ALARM	fcu alarm
CMU/SCSC Gates/Eighth Floor/FCU-15/FANCOIL HAND	FANCOIL HAND	fcu alarm
CMU/SCSC Gates/Eighth Floor/FCU-15/Run	Run	cmd
CMU/SCSC Gates/Eighth Floor/FCU-15/Schedule	Schedule	cmd
CMU/SCSC Gates/Eighth Floor/FCU-15/Zone Temp _ Zone Temp	Zone Temp _ Zone Temp	zone air temp sensor

- Some buildings have more than just labels
- Very few buildings with existing "ground truth"

Contributing to the Dataset

- Point scraping tool:

- Primitive, but open source :)

- Scans network for BACnet devices

  - hoping to expand to other protocols

- Dumps BACnet properties to CSV file

```
$ pip install pointscan
$ pointscan scan
$ ls *.csv
-rw-rw-r-- 1 oski oski  27441 Nov  9 20:32 NAE35-OrindaLib-JCI-10.34.117.160-501.csv
```

Contributing to the Dataset

- Web-based tool for metadata cleaning and anonymization

- Upload CSV produced by `pointscan` or other tool

- Apply cleaning rules:

  - Find/replace substrings

  - Split fields

  - Remove sensitive text

Conclusion

Gabe Fierro ▶

gtfierro@cs.berkeley.edu ▶

https://brickschema.org ▶

- Open dataset of building point labels
  (and other attributes)

- Data and collection tool at
  https://data.mortardata.org

- Want to promote research in metadata
  normalization

riselab
UC Berkeley