

# PETR

INTERNET OF THINGS  
RESEARCH HUB



## IoT Observatory: Enabling a Data Ecosystem for IoT

Aastha Madaan

University of Southampton





# Sharing holds the Value!

- Emergence of IoT devices and systems has empowered people and organizations
- Not all things in IoT capture everything, data needs to be reused and shared
- Improvements in governance, research, IoT data processing and analytics are driven by data and application sharing
- Moving from sharing scientific data to sharing sensor data
  - Correlations with social media data, weather forecasts, timetables required for data insights
- **Concerns for sharing IoT data and tools** – Ownership, Interoperability, Querying, Access Control

# IoT Observatory



Inspired by the [Web Observatory](#) initiative of [Web Science Trust](#)



Anyone can share a dataset/stream (publisher)



Anyone can locate a dataset/stream (user) for access across nodes



**Public datasets:** anyone can access



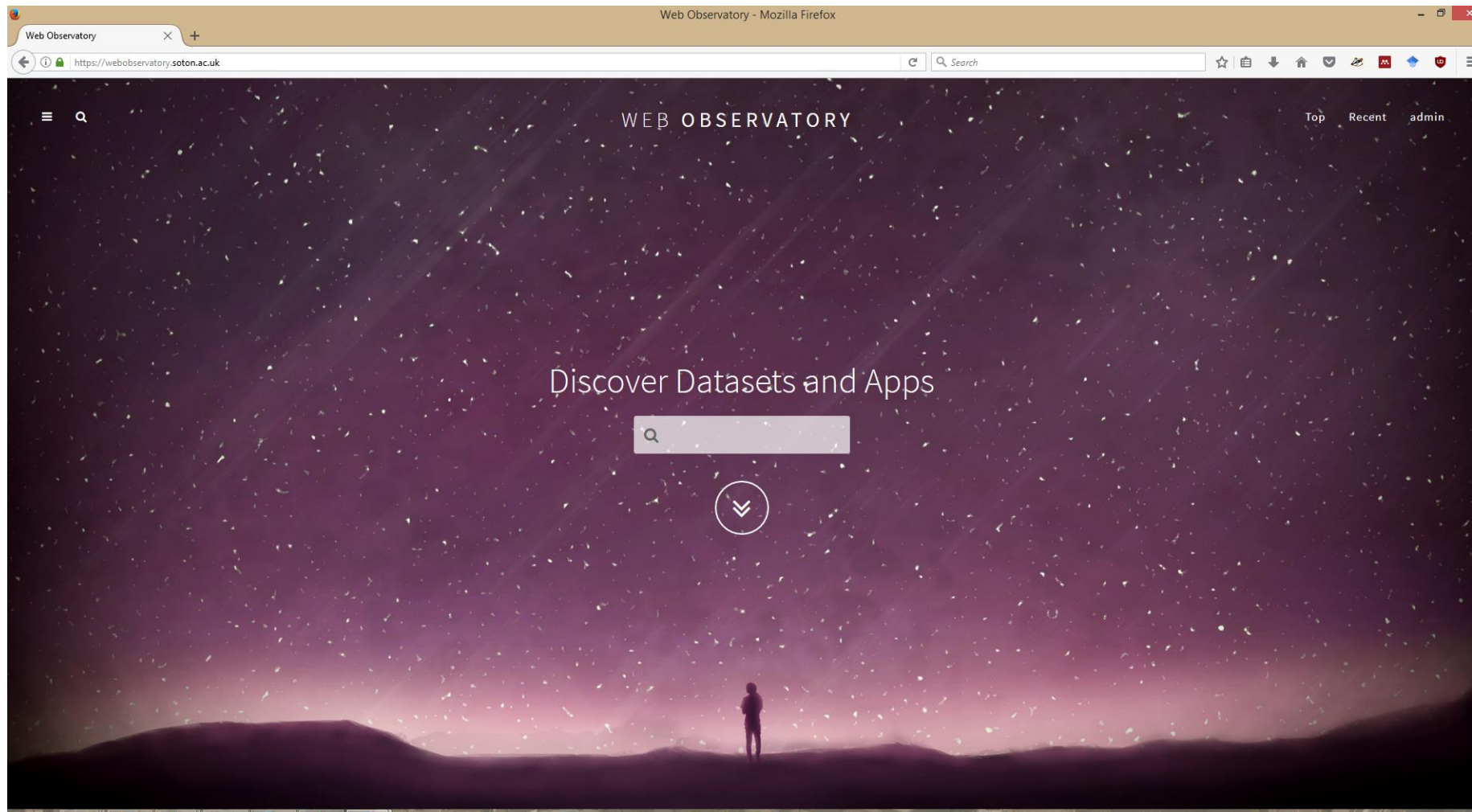
**Private datasets:** access can be granted by publisher



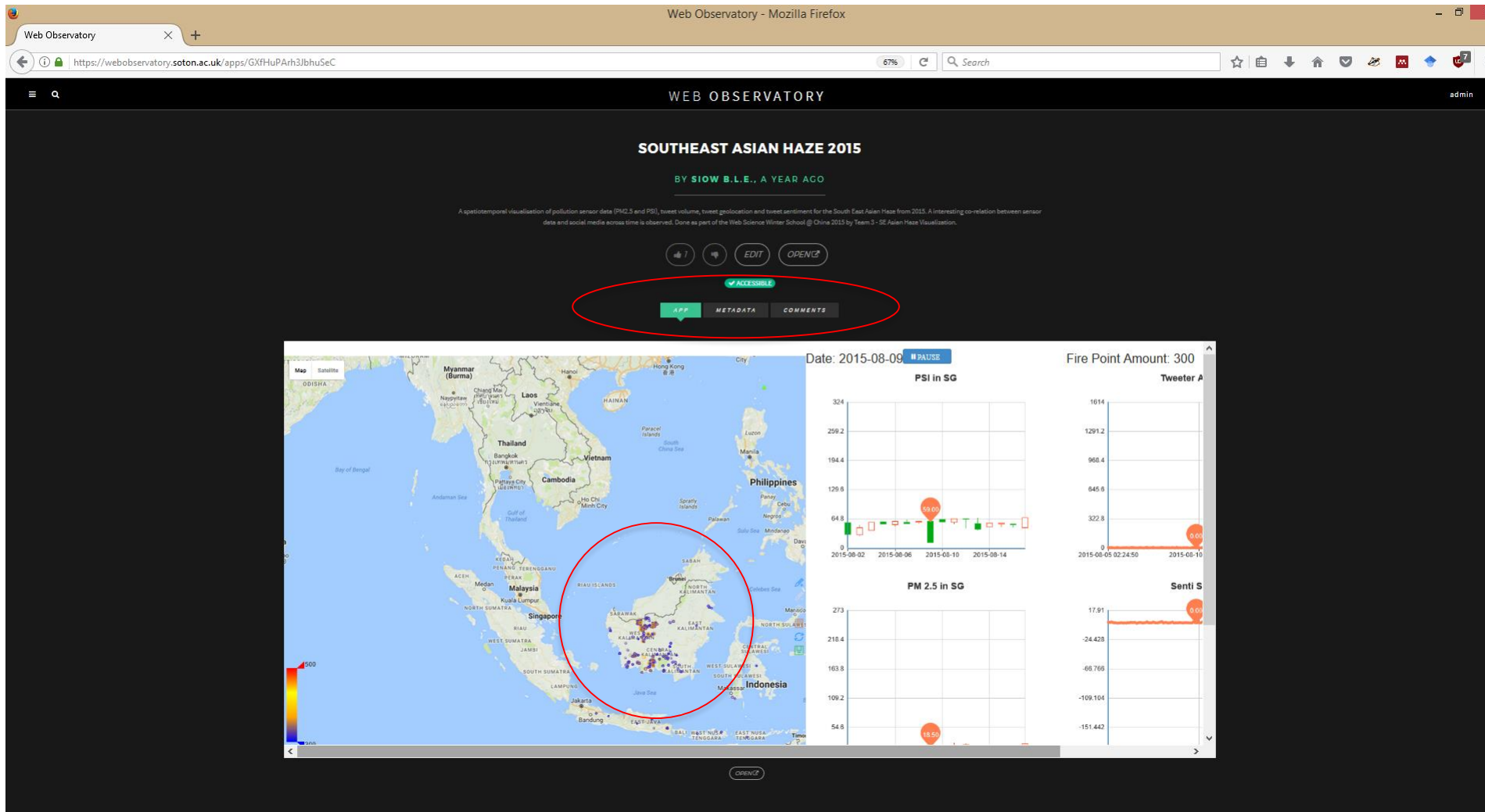
Datasets can be kept on Publishers' site

*More on: [IoT Observatory @PETRAS](#)*

# IoT Observatory : An Example



# IoT Observatory : An Example (Contd.)



# Discover and Search



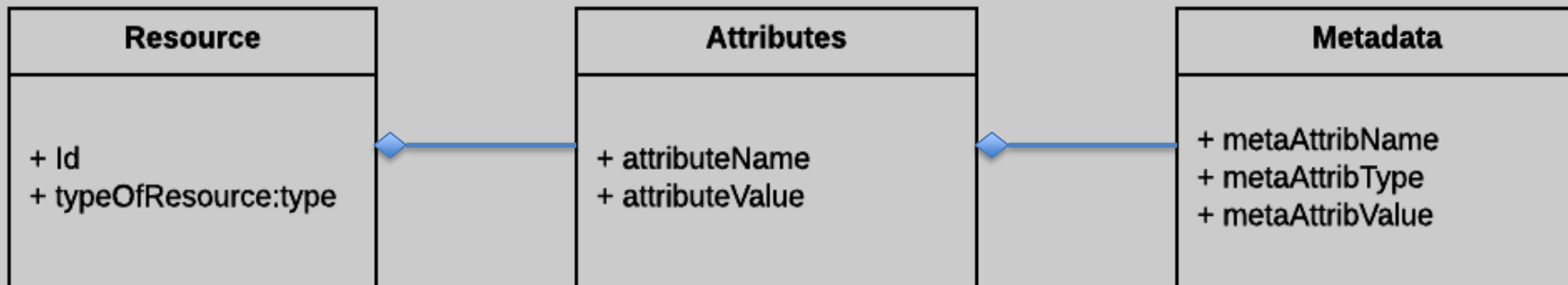
Searching across IoT/Web Observatory sites

- <http://search.webobservatory.org/>

The screenshot shows the Web Observatory Search interface. At the top, there's a navigation bar with 'HOME' and 'WEB OBSERVATORIES'. Below it, a search bar contains the text 'twitter'. The results are displayed in a table with columns: Dataset Name, Description, Provider, and Link to Resource.

Dataset Name	Description	Provider	Link to Resource
Bangladesh Tweets	A collection of geotagged tweets located within the geographic boundaries of Bangladesh. See query info for the boundary box used for tweet location. Collected from the 100% Twitter firehose.	Southampton Web Observatory	undefined
Comedy Tweets	A filtered collection of tweets related to comedy/fun words. {"rs": "twitter_comedy_tweets.tweets", "count": 869405, "size": 333020944, "avgObjSize": 383.4857514638907, "storageSize": 396632064, "numExtents": 11, "nindex": 1, "lastExtentSize": 112091136, "paddingFactor": 1, "systemFlags": 1, "userFlags": 0, "totalIndexSize": 52367280, "indexSize": [{"_id": 52367280}], "ok": 1}	Southampton Web Observatory	undefined
DD2015_jot	Citypulse Smart City Data From: 2013 Till: 2015 Multiple datasets (Traffic, Pollution, Weather, Twitter, Cultural Event, Parking, Library Event) in the city of Aarhus, Denmark.	Southampton Web Observatory	undefined
EDL Tweets	Tweets related to the #EDL hashtag. Data collected during June 2013 and August 2013. This was collected using the 100% Twitter Firehose	Southampton Web Observatory	undefined
ego-Twitter	Directed 81,306 1,768,149 Social circles from Twitter	SNAP - Stanford University	<a href="http://snap.stanford.edu/data/egonets-Twitter.html">http://snap.stanford.edu/data/egonets-Twitter.html</a>
Geoparse Twitter benchmark dataset	Open free to access geoparsing benchmark dataset containing 1000's of tweets during several natural disasters with manually labelled location entries. Set of 4 JSON files each with manually annotated labels for location added. Location annotations provide an extra 'entities' field type for 'mentions'. Each JSON file covers a different natural crisis event (Hurricane Sandy 2012, Milan Blackouts 2013, Turkish Earthquake 2012, Christchurch Earthquake 2012). The BSD license includes a clause mandating that derived works acknowledge the University of Southampton IT Innovation Centre and cite the original paper that describes this dataset: Middleton, S.E. Middleton, L. Modafferi, S. "Real-time Crisis Mapping of Natural Disasters using Social Media", Intelligent Systems, IEEE, vol.29, no.2, pp.9,17, Mar-Apr. 2014, DOI:10.1109/MIS.2013.126	Southampton Web Observatory	undefined
Higgs-twitter	Tweets 456,631 14,855,875 Spreading processes of the announcement of the discovery of a new particle with the features of the Higgs boson on 4th July 2012.	SNAP - Stanford University	<a href="http://snap.stanford.edu/data/higgs-twitter.html">http://snap.stanford.edu/data/higgs-twitter.html</a>
Iphone Tweets	A collection of tweets containing the string iPhone. {"rs": "twitter_iphone5s.tweets", "count": 76604, "size": 32257008, "avgObjSize": 421.08777609524304, "storageSize": 37318656, "numExtents": 4, "nindex": 1, "lastExtentSize": 19243008, "paddingFactor": 1, "systemFlags": 1, "userFlags": 0, "totalIndexSize": 3106880, "indexSize": [{"_id": 3106880}], "ok": 1}	Southampton Web Observatory	undefined

# Data/Information Model



Applications



Datasets/  
Data-streams

Intermediate  
Statistics



(DCAT)



**HYPERCAT**

**Schema.org**





# Enabling Legitimate Access



Fine-grained access control

- Formal description of permitted processing in metadata of datasets



Negotiating access

- Evaluation of User credentials – access permissions, datasets, requests, purpose



- Risk assessment and trust computation for sharing datasets

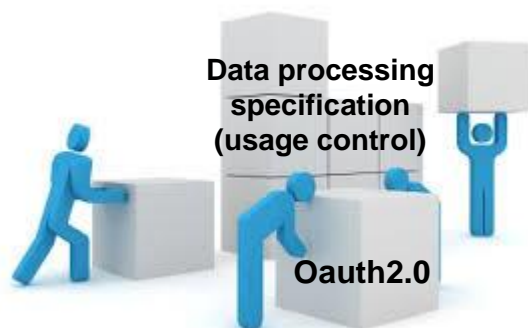


Legal and Ethical challenges

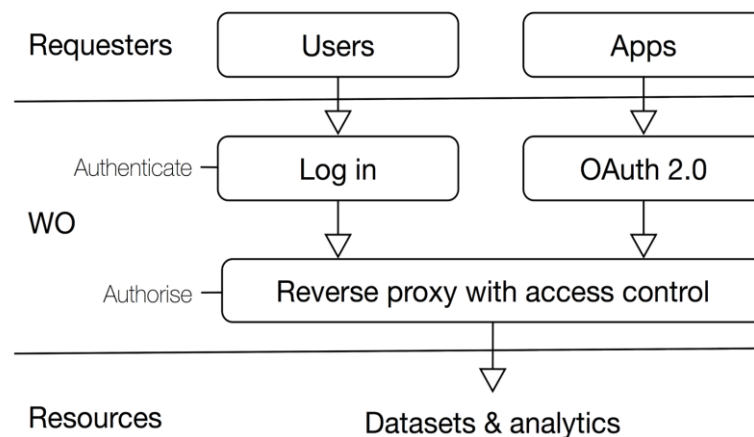
- Terms of Data Use



# Enabling Legitimate Access (Contd.)

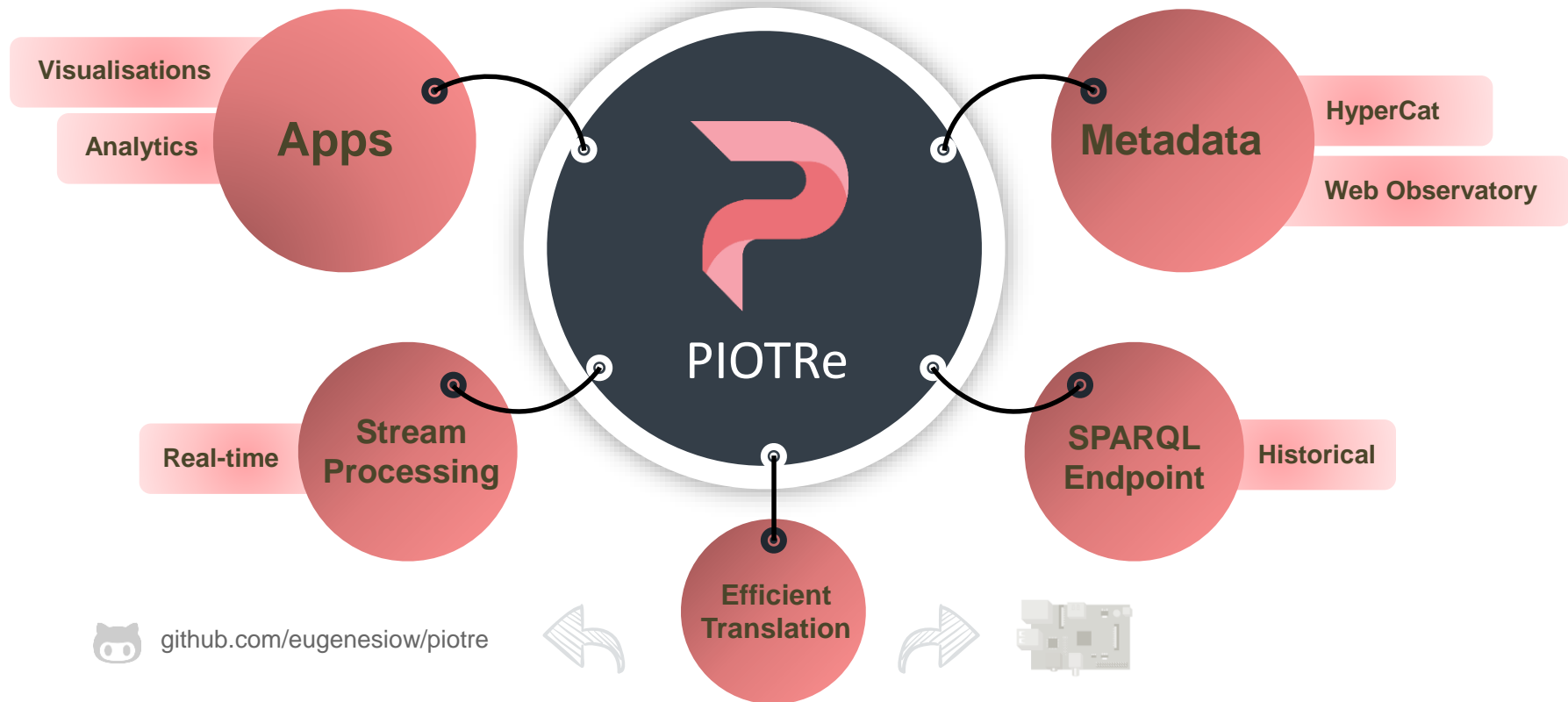


- Reverse proxy and URL obfuscating technologies to protect private and sensitive data
- Data Publisher controls – location, visibility & access to datasets and streams
- Access control model based on intended usage (data analytic function or statistics)
- Protocol for specification of the model within metadata of datasets



# Personal IoT Repository

PIOTRe means “rock” - the foundation (for lightweight computers) on which real-time IoT applications can be built. Interoperable, efficient, valuing data ownership and locality.



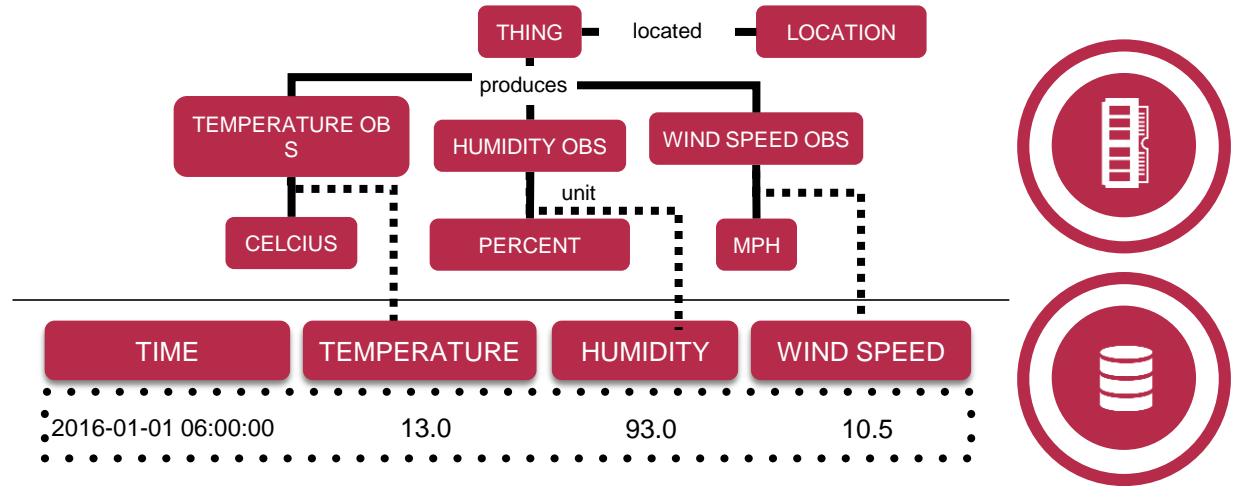
Siow, E., Tiropanis, T. and Hall, W. (2016) PIOTRe: Personal Internet of Things Repository: The 15th International Semantic Web Conference P&D

# Querying Time-series IoT Data

**20K**  
UNIQUE DEVICES  
dweet.io

**99.5%**  
FLAT SCHEMATA

**92.8%**  
WIDE SCHEMATA



Siow, E., Tiropanis, T., Hall, W. (2016). "Interoperable and Efficient: Linked Data for the Internet of Things." The 3rd International Conference on Internet Science

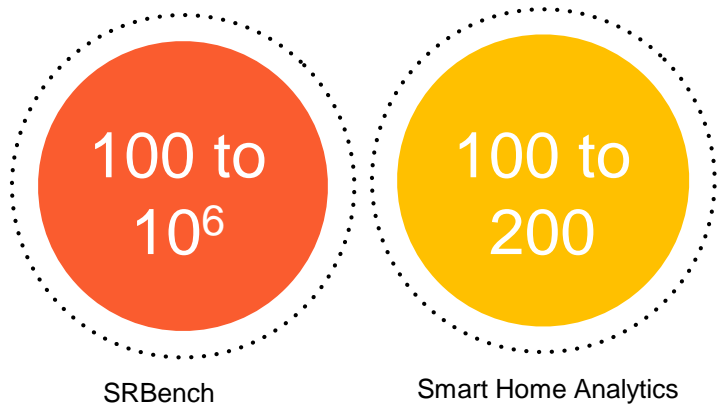
## Queries on Historical Data

From 3 to 3 orders of magnitude improvement



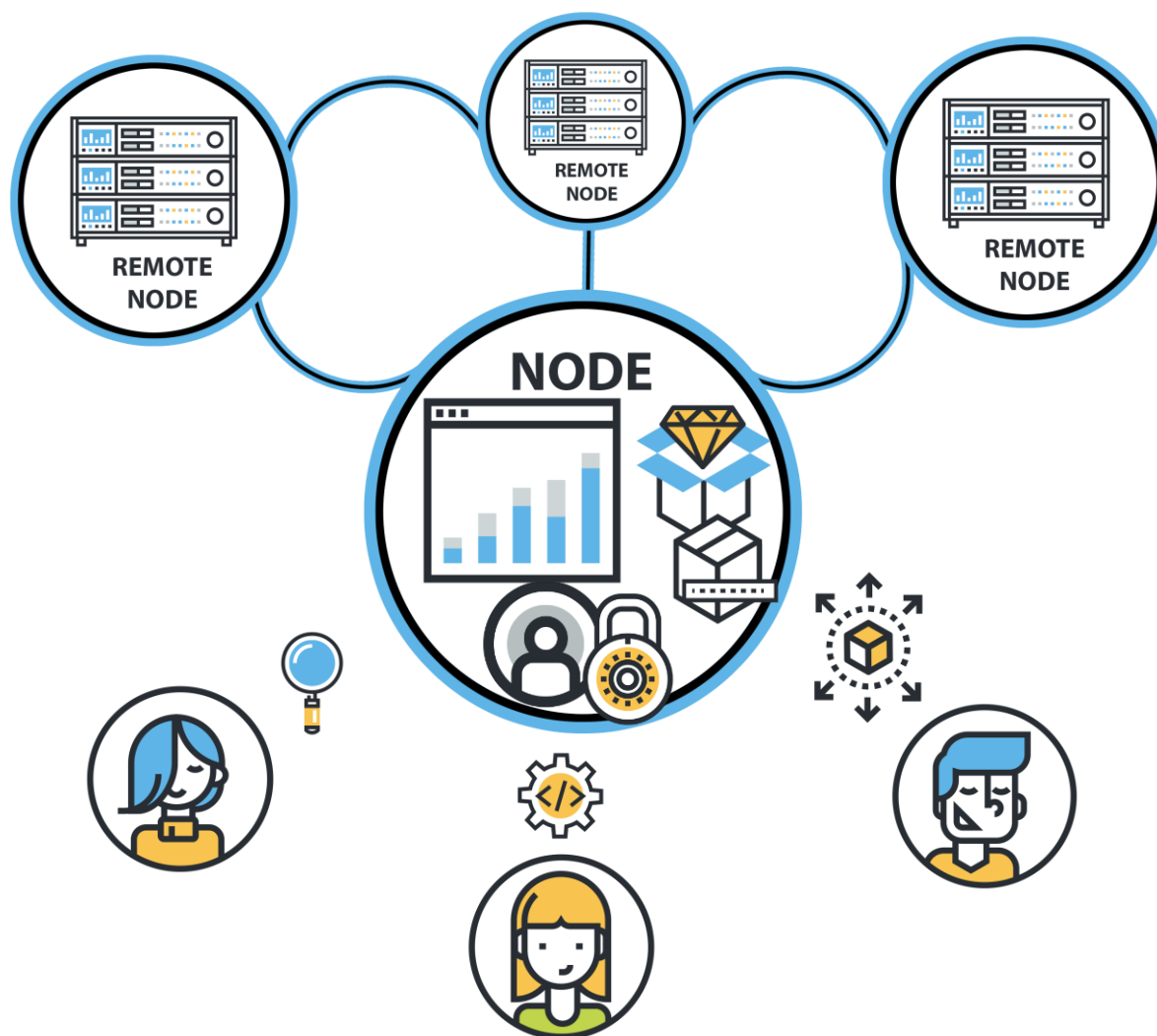
From x9 to x1352 times more efficient in terms of storage

## Streaming Performance Improvement against CQELS



Siow, E., Tiropanis, T., Hall, W. (2016) "SPARQL-to-SQL on internet of things databases and streams." ISWC2016: The 15th International Semantic Web Conference

# Sharing data, tools and analytics



- ✓ Ownership
- ✓ Interoperability
- ✓ Discovery
- ✓ Access Control
- ✓ Querying

@wo\_team

[hello@webobservatory.org](mailto:hello@webobservatory.org)

<http://iotobservatory.io/>

[IoT Observatory @PETRAS](#)