

Centre for
Data Analytics

Insight



Water Management, Web of Things, and Enabling Technologies

Souleiman Hasan, Edward Curry

W3C WoT Interest Group Meeting
Munich, Germany, 2015

WatErnomics



Insight Centre for Data Analytics



<https://www.insight-centre.org/>



Insight Centre @ National University of Ireland, Galway

Formerly DERI

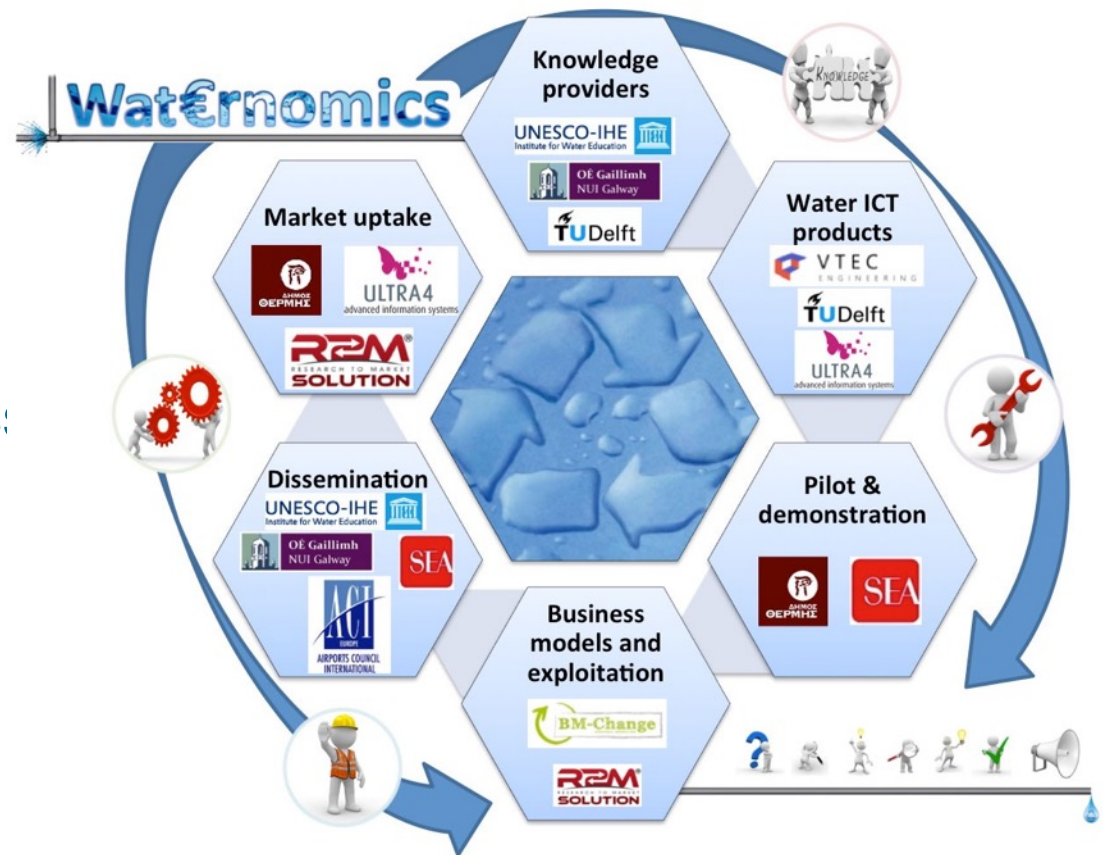


~ 150 researchers

Wat€rnomics



- WATERNOMICS will develop and introduce ICT as an enabling technology to manage water as a resource, increase end-user conservation awareness and affect behavioural changes, and to avoid waste through leak detection.



Overview

- **Use Case**
- **Event Processing Paradigm**
- **Semantics**
- **Approximate Matching**
- **Thematic Event Processing**
- **Building the IoT**

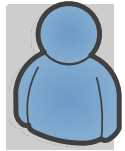
Use Case

- Home Water Awareness and Conservation -

Motivation and Purpose for John

- **Motivation**
 - **Lower personal water footprint**
 - **Water is a scare resource**
 - **Save money on water bill**
- **Purpose**
 - **Set monthly water consumption target**
 - **Keep aware of consumption**
 - **Get recommendations from things or Web**

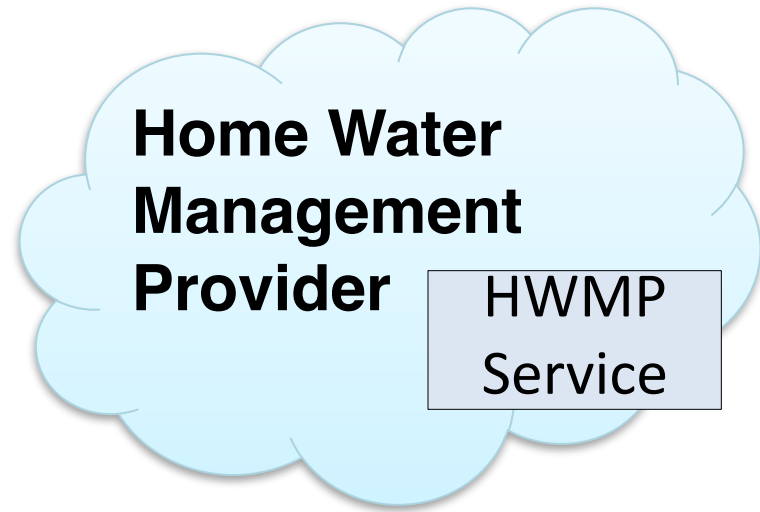
Actors



John



Laptop,
phone,...

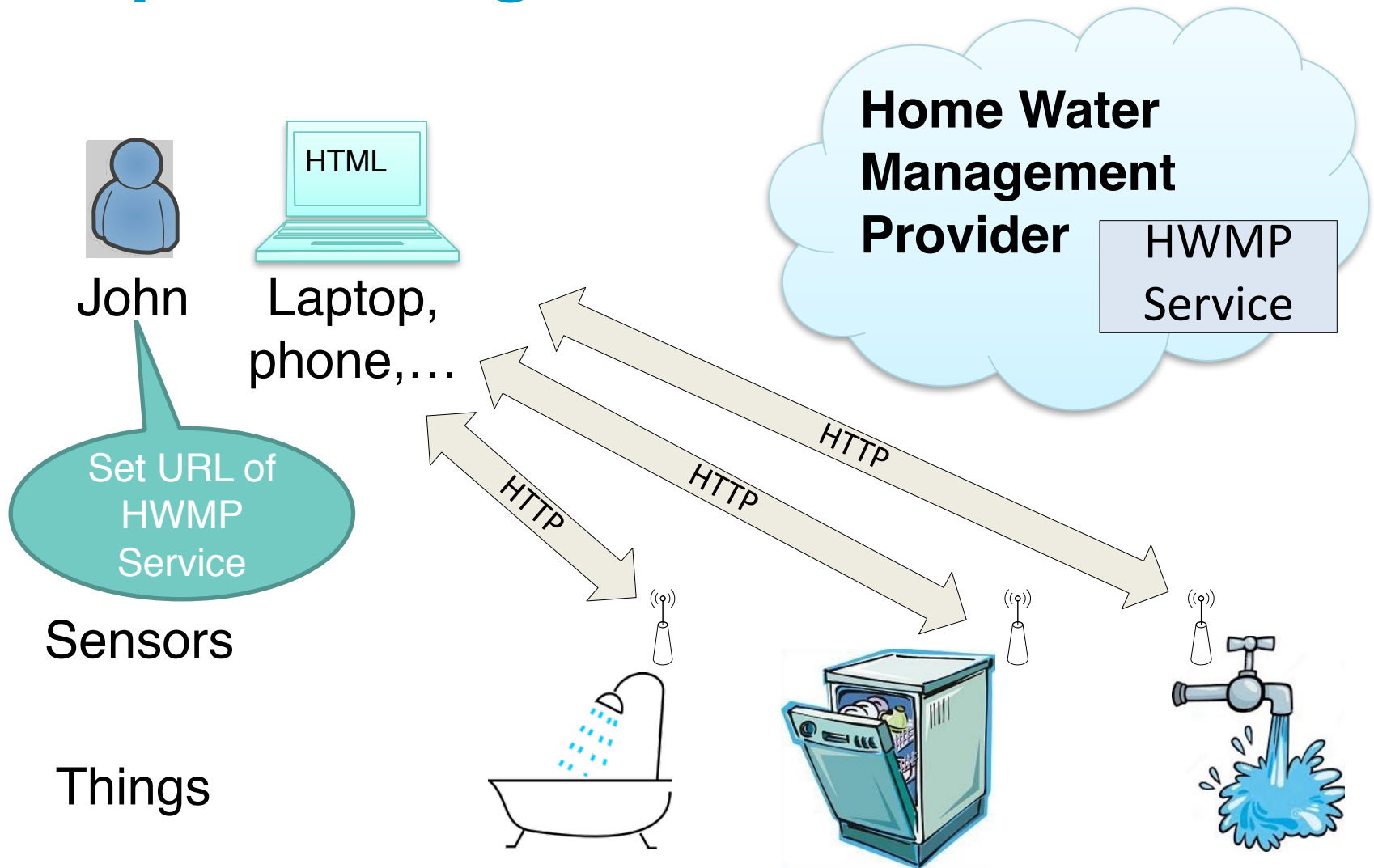


Sensors

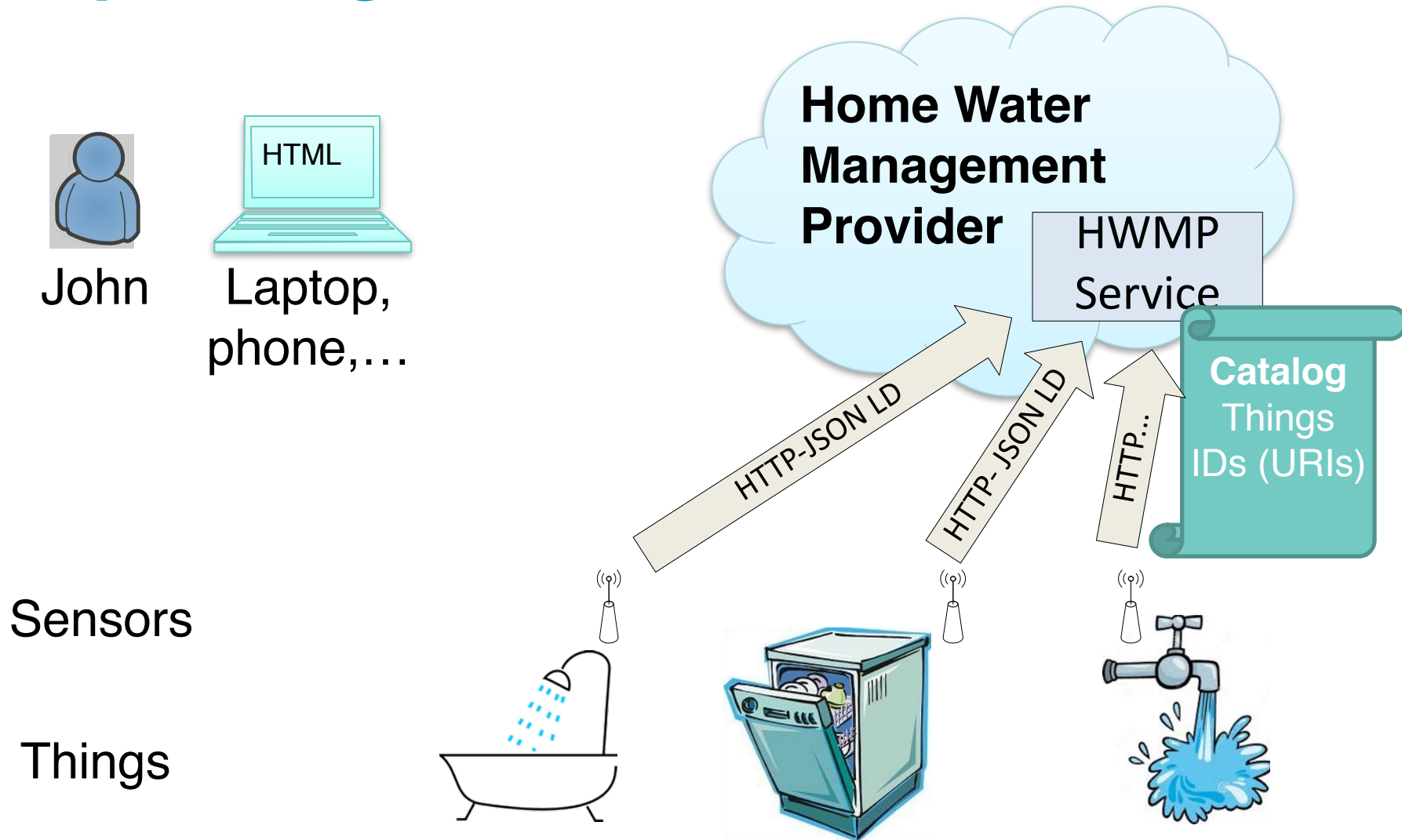
Things



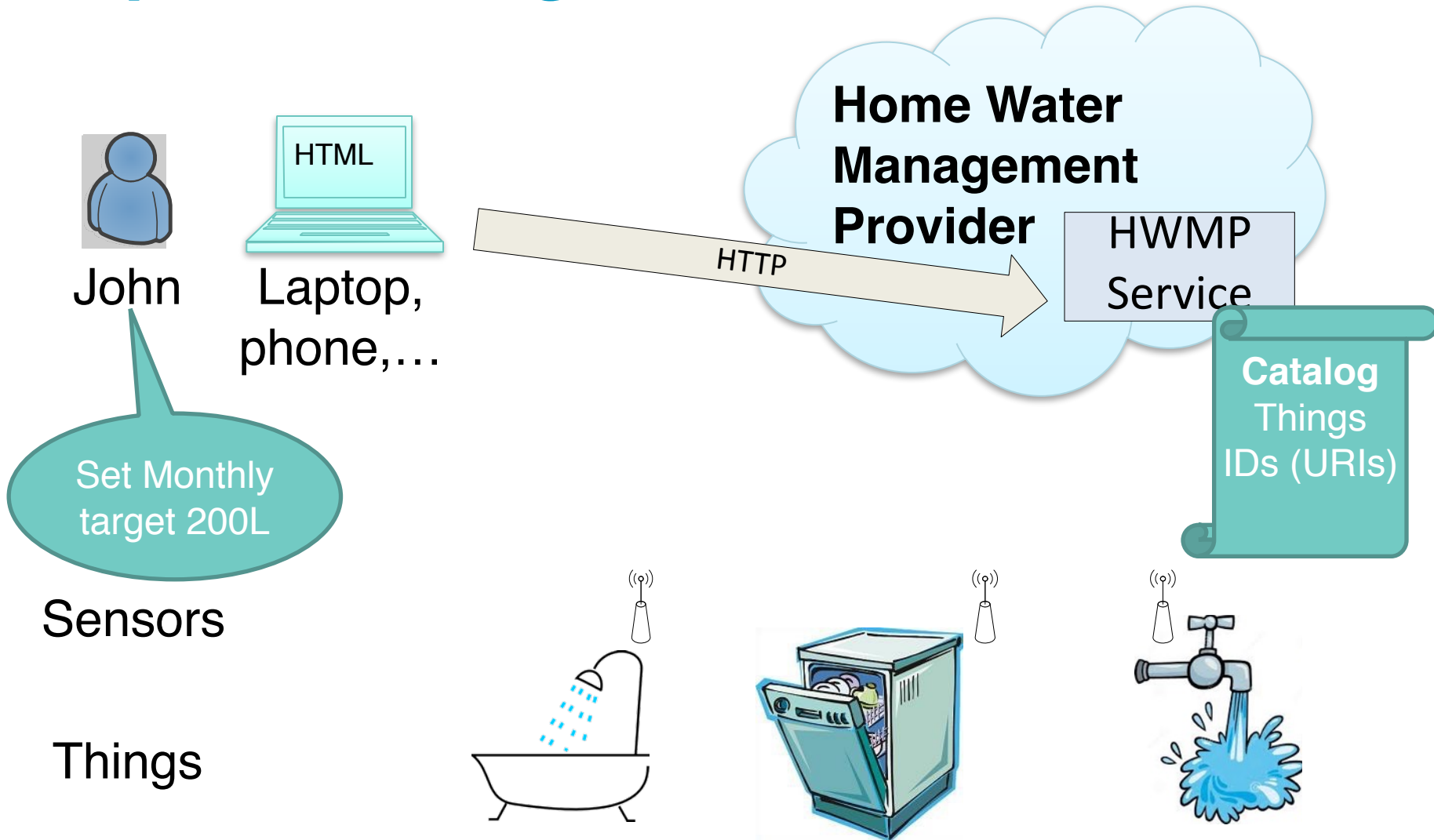
Step 1- Configuration



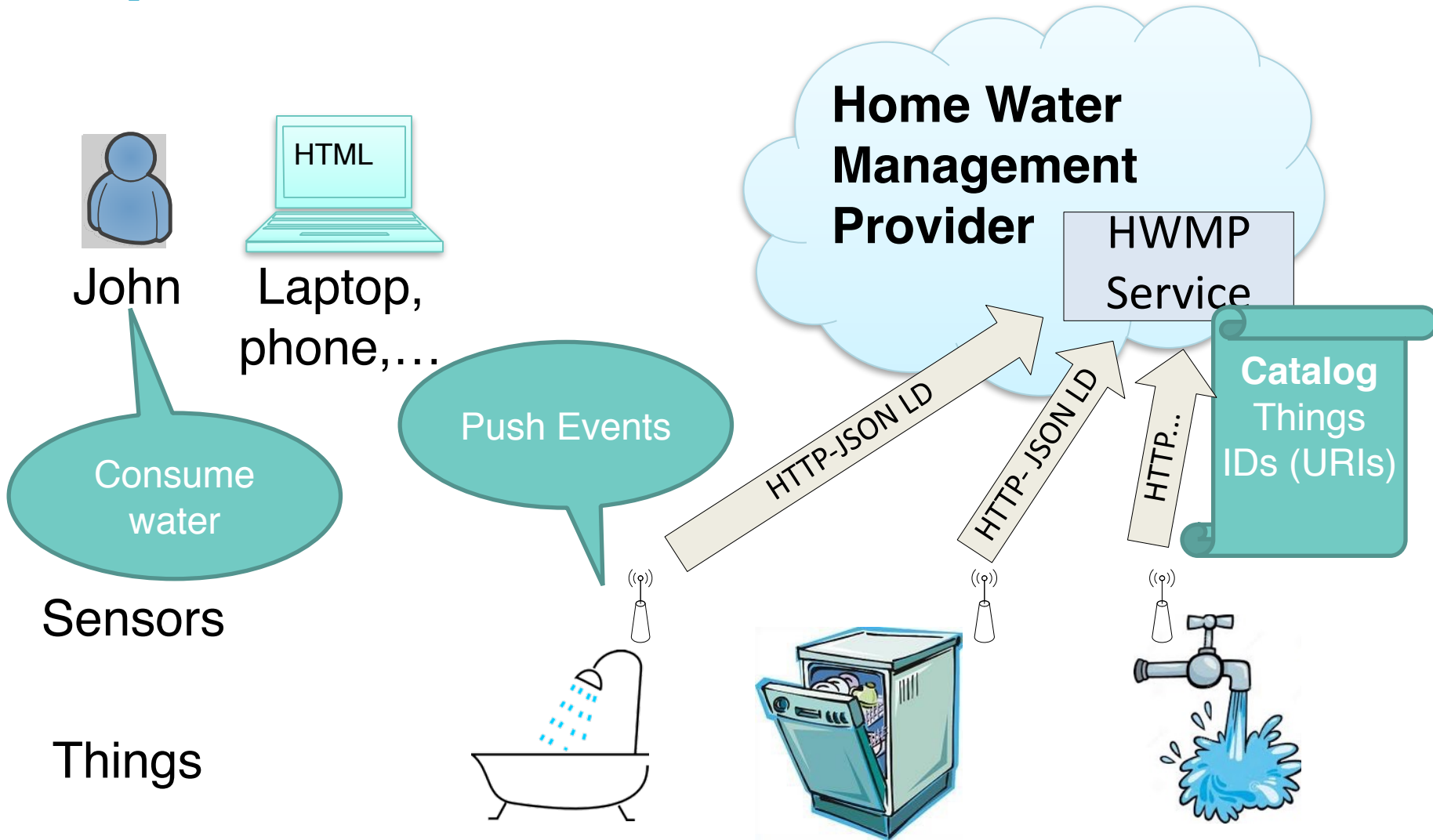
Step 2- Registration



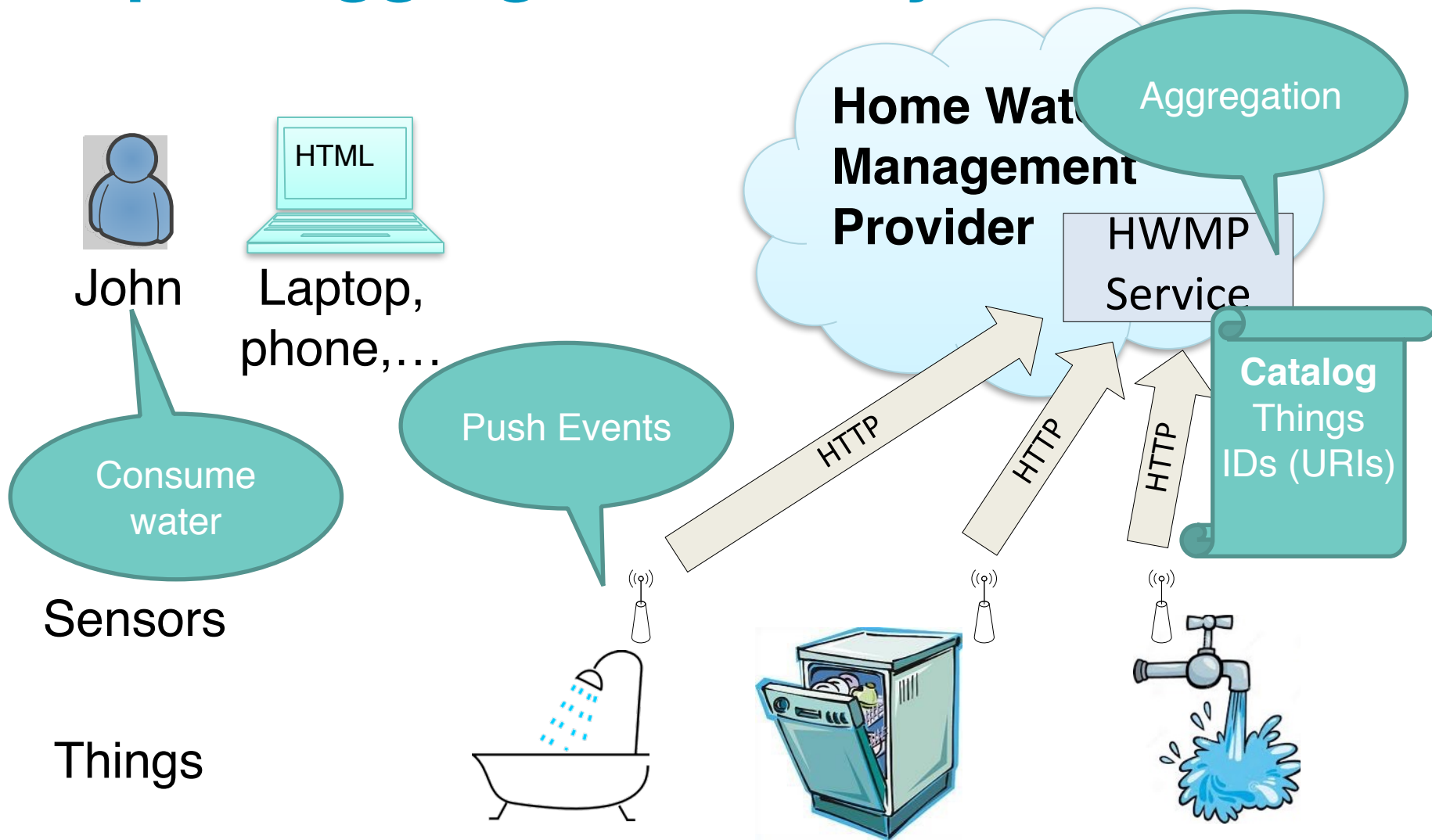
Step 3- Set Target



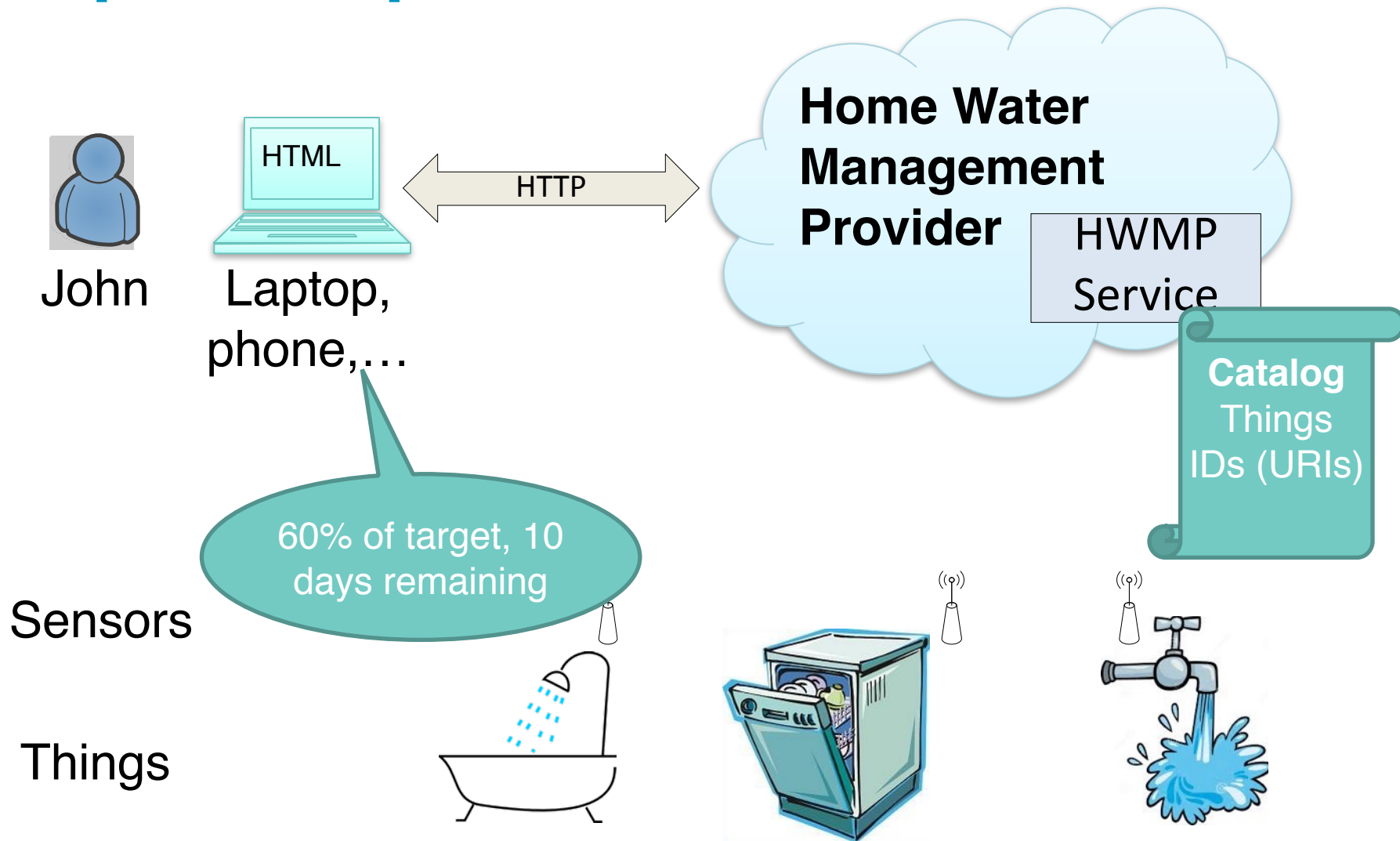
Step 4- Data Transfer



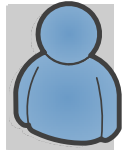
Step 5- Aggregation/Analytics



Step 6- Keep Aware



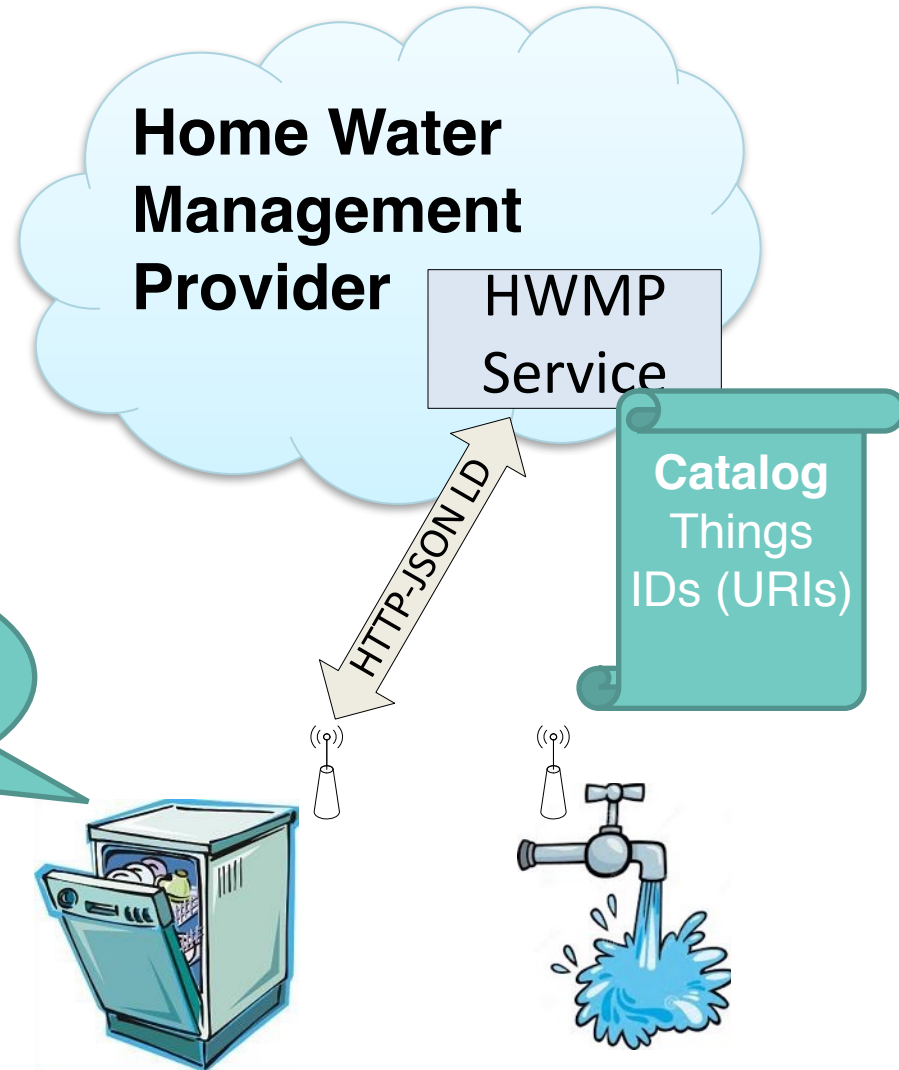
Step 7- Thing Thinks!



John



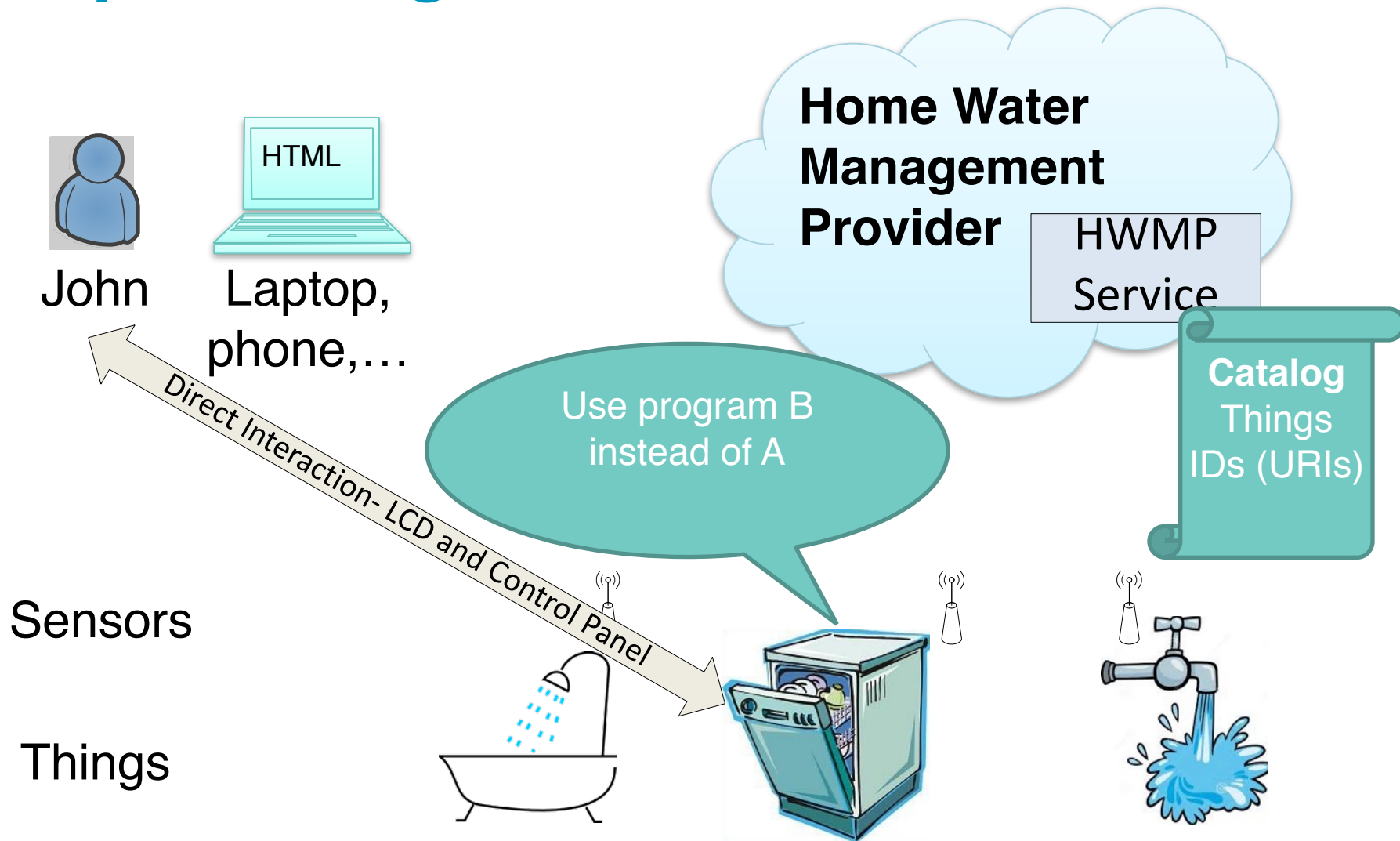
Laptop,
phone,...



Sensors

Things

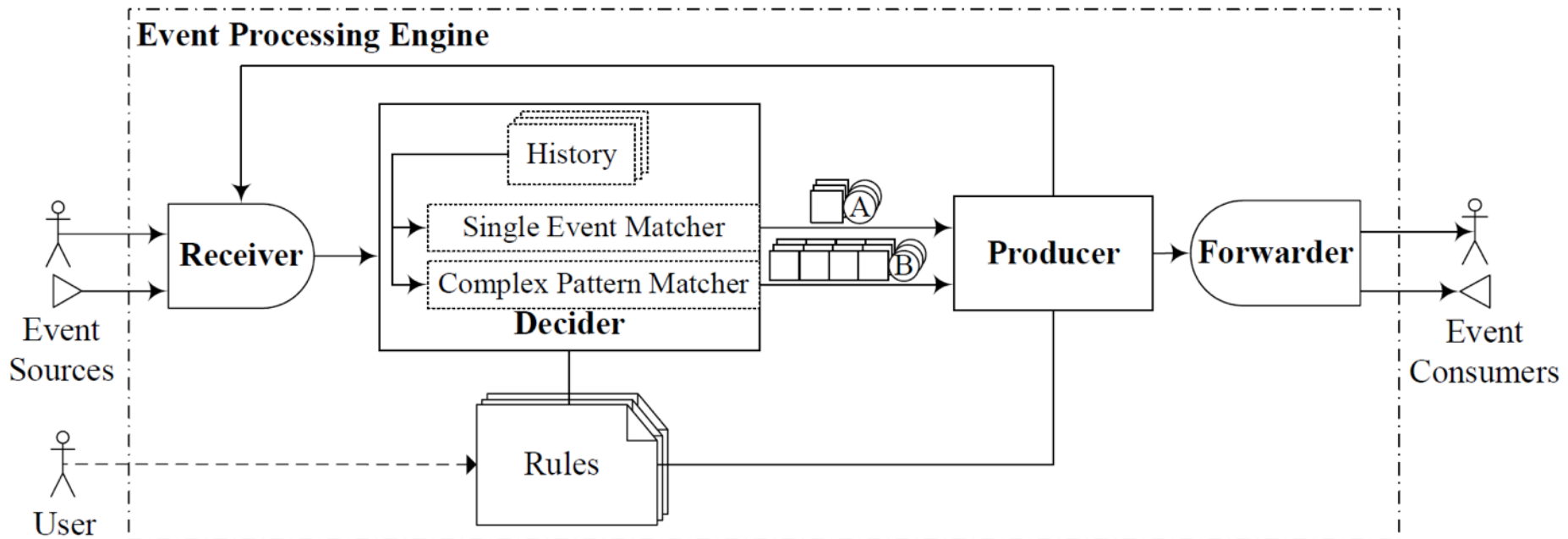
Step 8- Thing Recommends



Going Large Scale

Event Processing and Semantics

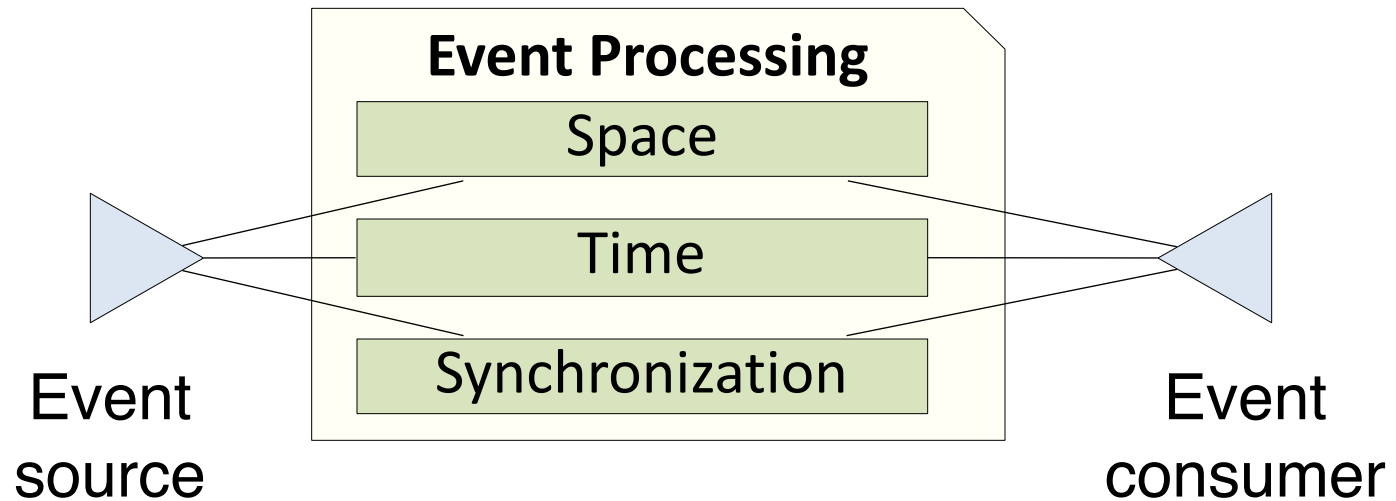
Complex Event Processing Systems



(ADAPTED FROM CUGOLA AND MARGARA)

SOULEIMAN HASAN AND EDWARD CURRY. 2014. APPROXIMATE SEMANTIC MATCHING OF EVENTS FOR THE INTERNET OF THINGS. ACM TRANS. INTERNET TECHNOL. 14, 1, ARTICLE 2 (AUGUST 2014), 23 PAGES.

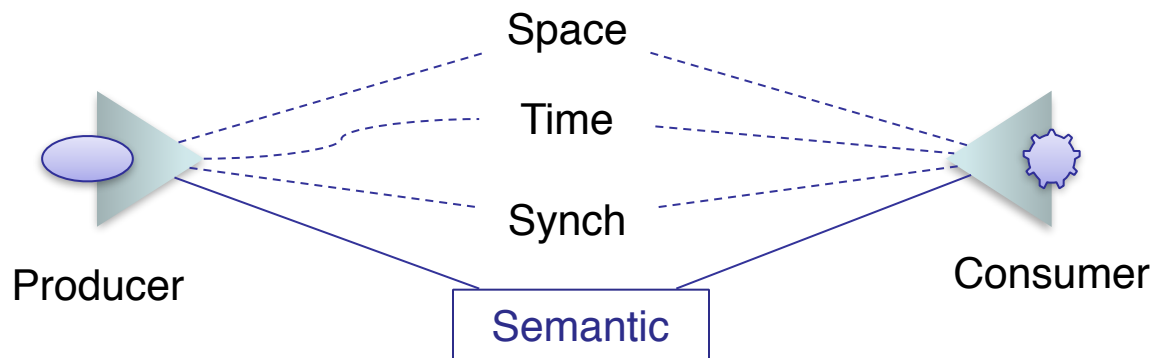
The Principle of Decoupling



Patrick Th. Eugster, Pascal A. Felber, Rachid Guerraoui, and Anne-Marie Kermarrec. 2003. The many faces of publish/subscribe. *ACM Comput. Surv.* 35, 2 (June 2003), 114-131.

Problem

- **Event producers and consumers are semantically coupled**
 - Consumers need prior knowledge of event types, attributes and values.
 - Limits scalability in heterogeneous and dynamic environments due to explicit dependencies
 - Difficult development of event processing subscriptions/rules in heterogeneous and dynamic environments.



Statistical Vector Space Semantics and Approximate Event Processing

- *Distributional hypothesis*: the context surrounding a given word in a text provides relevant information about its **meaning**.
- Simplified semantic model.
- Associational and quantitative.
- Explicit Semantic Analysis (ESA) is the primary distributional model used in this work.

A **wife** is a **female partner** in a **marriage**. The term "**wife**" seems to be a close term to **bride**, the latter is a **female participant** in a **wedding ceremony**, while a **wife** is a **married woman** during her **marriage**.

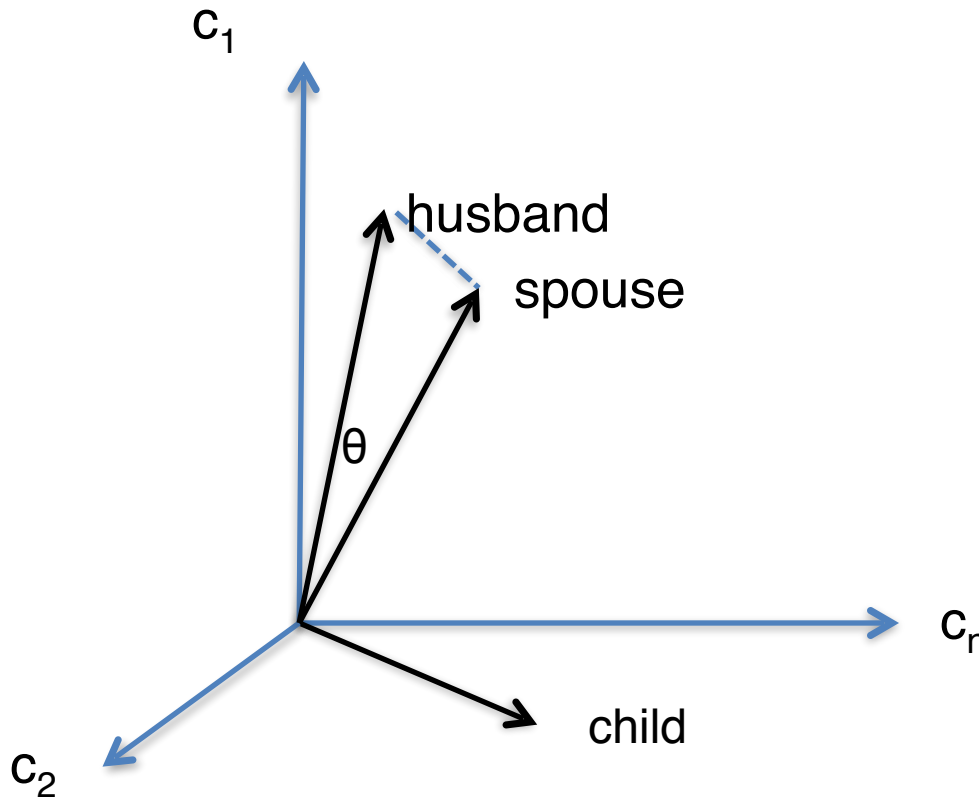
Slide Credits: Andre Freitas (<http://andrefreitas.org/>), [Freitas et al., 2013]

Works as a semantic ranking function

ESA interpretation vector

spouse

Spouses of the Prime Ministers of Canada (0.6558)
Adultery (0.4153)
Widow (0.4095)
Alimony (0.3751)
Spousal abuse (0.3467)
Domestic partnership (0.3292)
Rights and responsibilities of marriages in the United States (0.3258)
First Lady (0.3206)
Common-law marriage (0.2919)
Family (0.27550)
Princess consort (0.2705)
Divorce (0.2383)
...



E.g. $esa(\text{room}, \text{building}) = 0.099$

E.g. $esa(\text{room}, \text{car}) = 0.009$

Slide Credits: Andre Freitas (<http://andrefreitas.org/>) [Freitas et al., 2013]

- Top-1 and Top-k most *probable* mappings

{**type**: increased energy
consumption event,

measurement unit: kilowatt per hour,

device: computer,

desk: desk 112c,

office: room 112,

floor: ground floor,

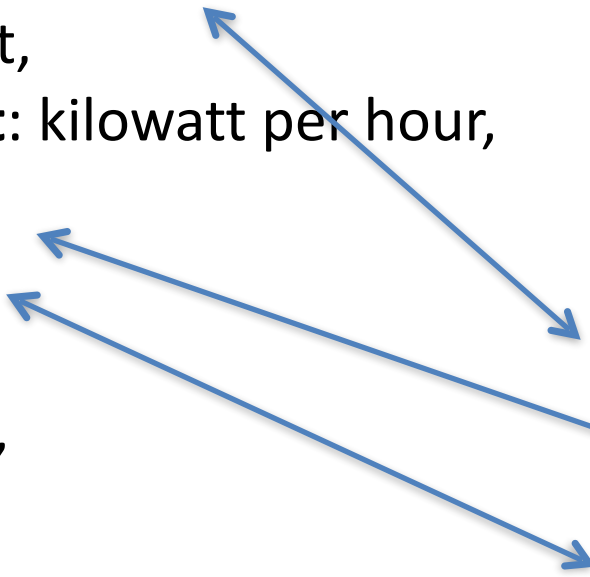
zone: building,

city: Galway,

country: Ireland,

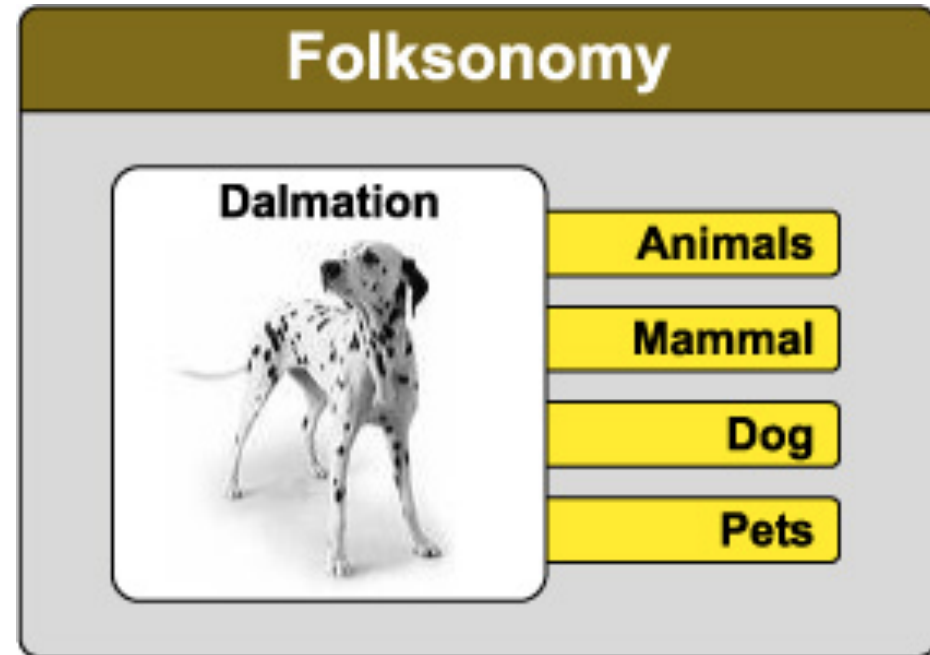
continent: Europe}

{**type** = increased energy
consumption event,
device = laptop~,
room~esa = room 112}



Thingsonomies and Thematic Event Processing

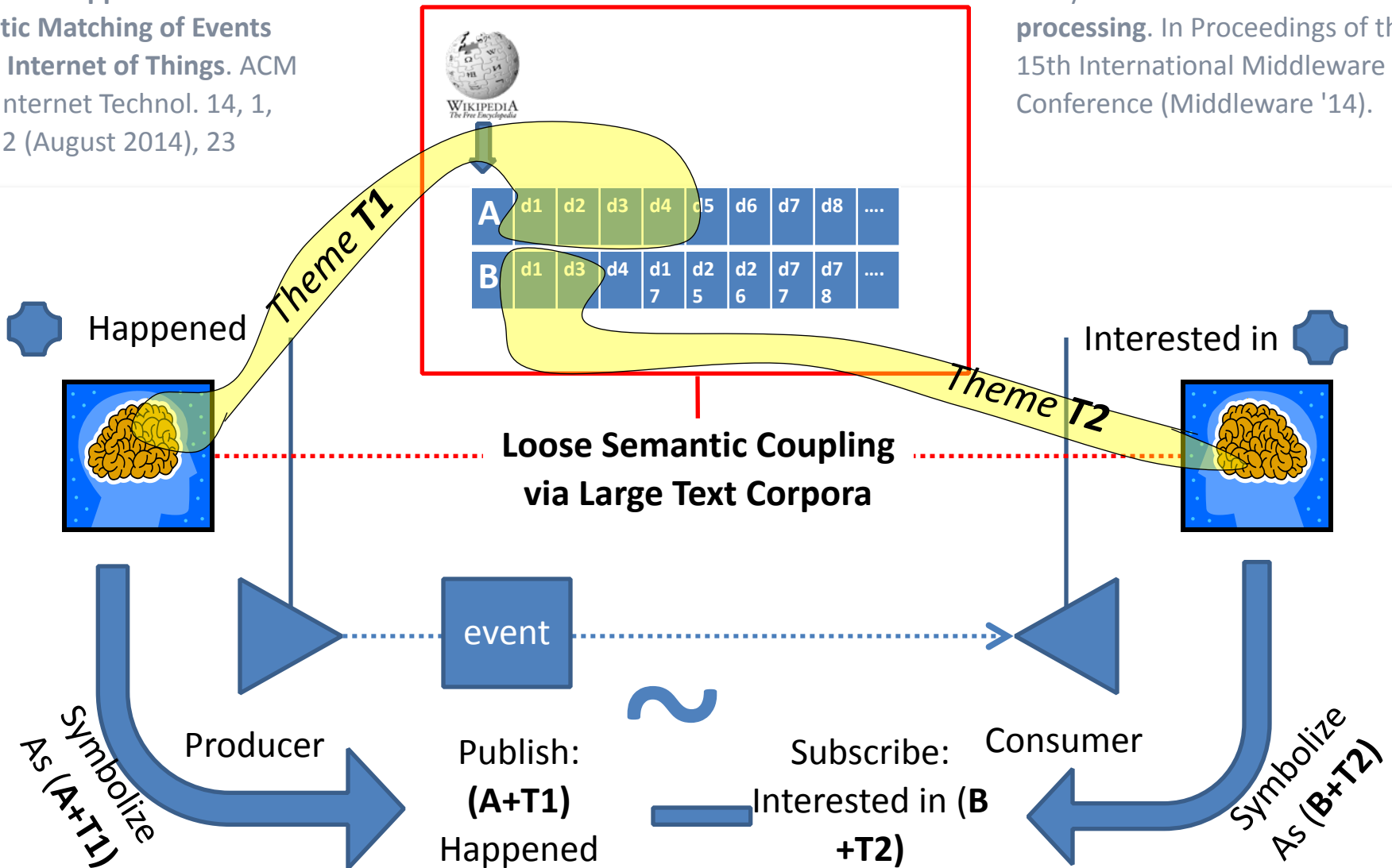
- Inspired by works in social tagging, i.e. **folksonomies**
- Folksonomies are bottom up approach to semantics
- Free words



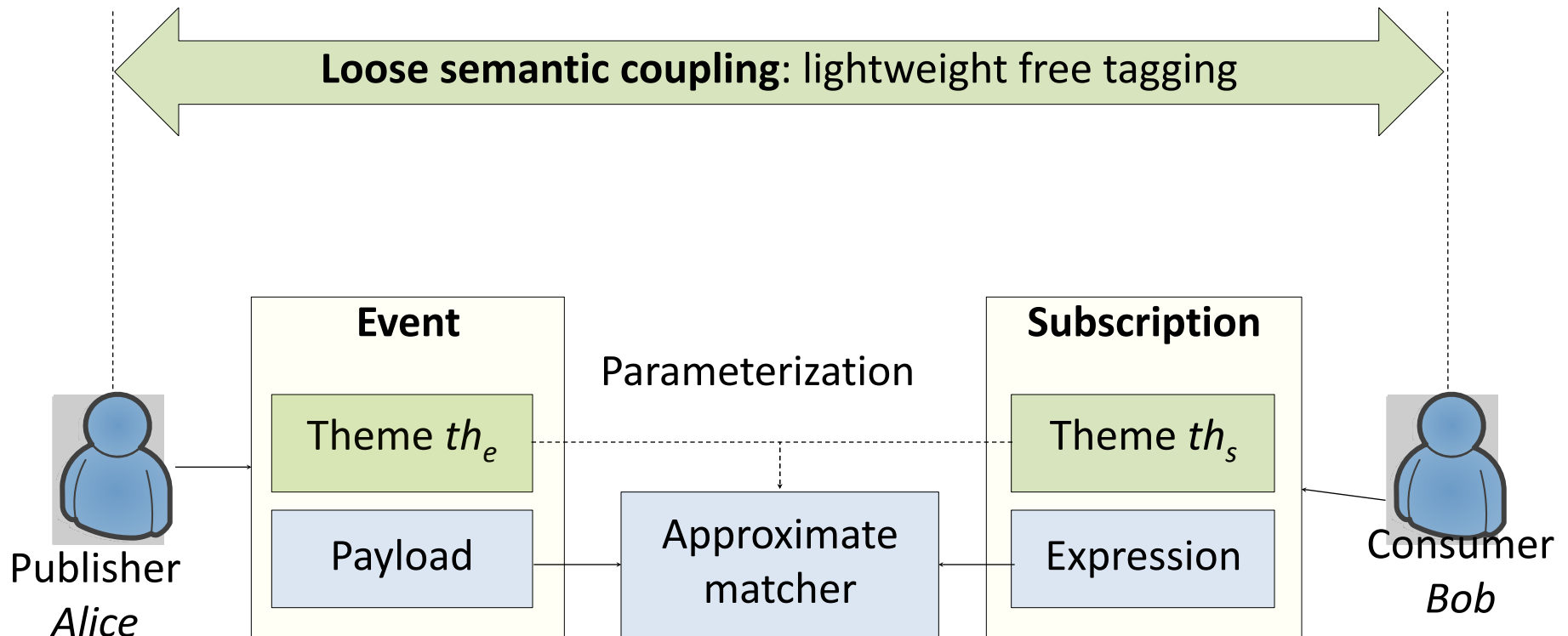
Insight Approach 5: Thematic Event Processing

Souleiman Hasan and Edward Curry. 2014. **Approximate Semantic Matching of Events for the Internet of Things**. ACM Trans. Internet Technol. 14, 1, Article 2 (August 2014), 23 pages

Souleiman Hasan and Edward Curry. 2014. **Thematic event processing**. In Proceedings of the 15th International Middleware Conference (Middleware '14).



- Exchange approximations of meanings



Hasan, S. and Curry, E., 2014. Thematic Event Processing. *Middleware 2014*.

- Thematic tags added to events

Event

energy, appliances, building

type: increased energy consumption event,
measurement unit: kilowatt per hour,
device: computer,
office: room 112

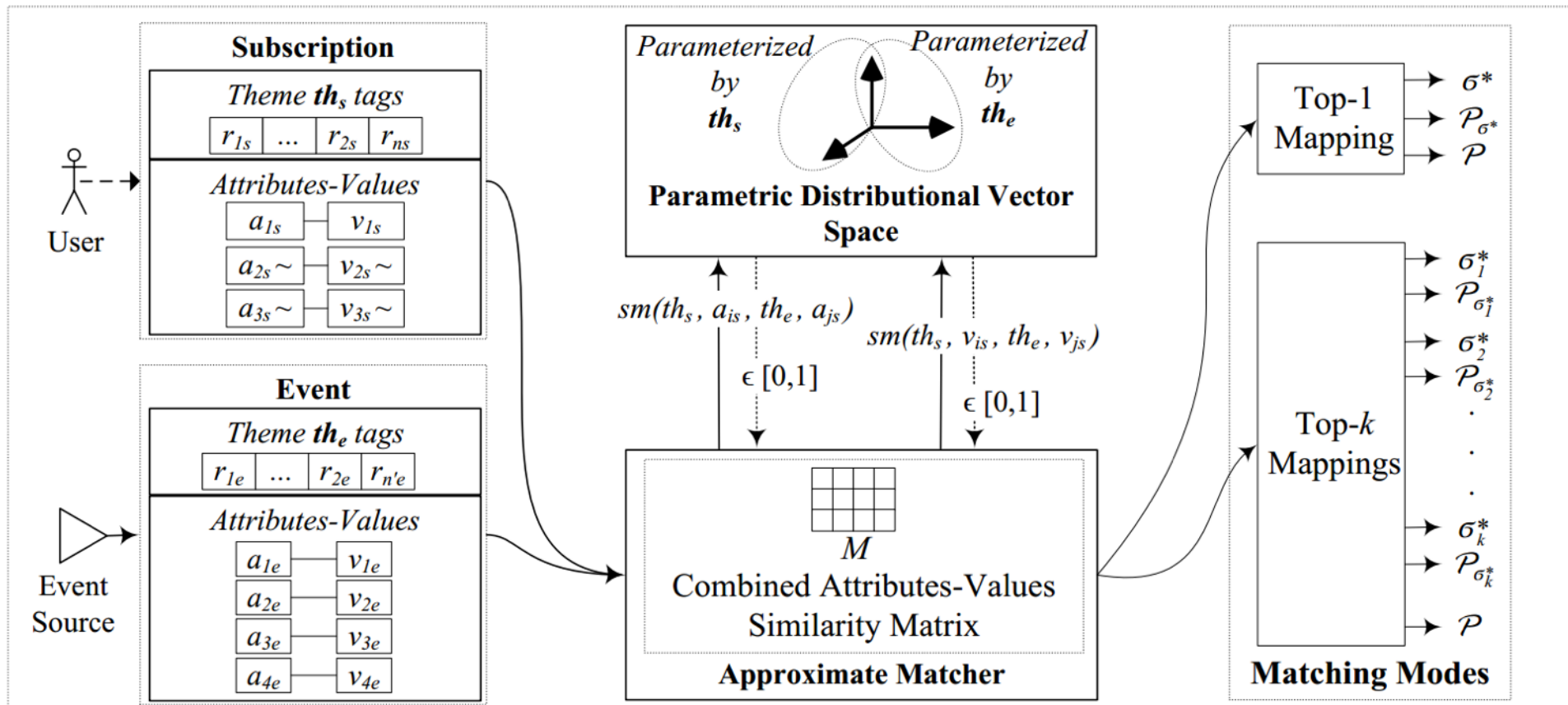
- Thematic tags added to subscriptions

Subscription

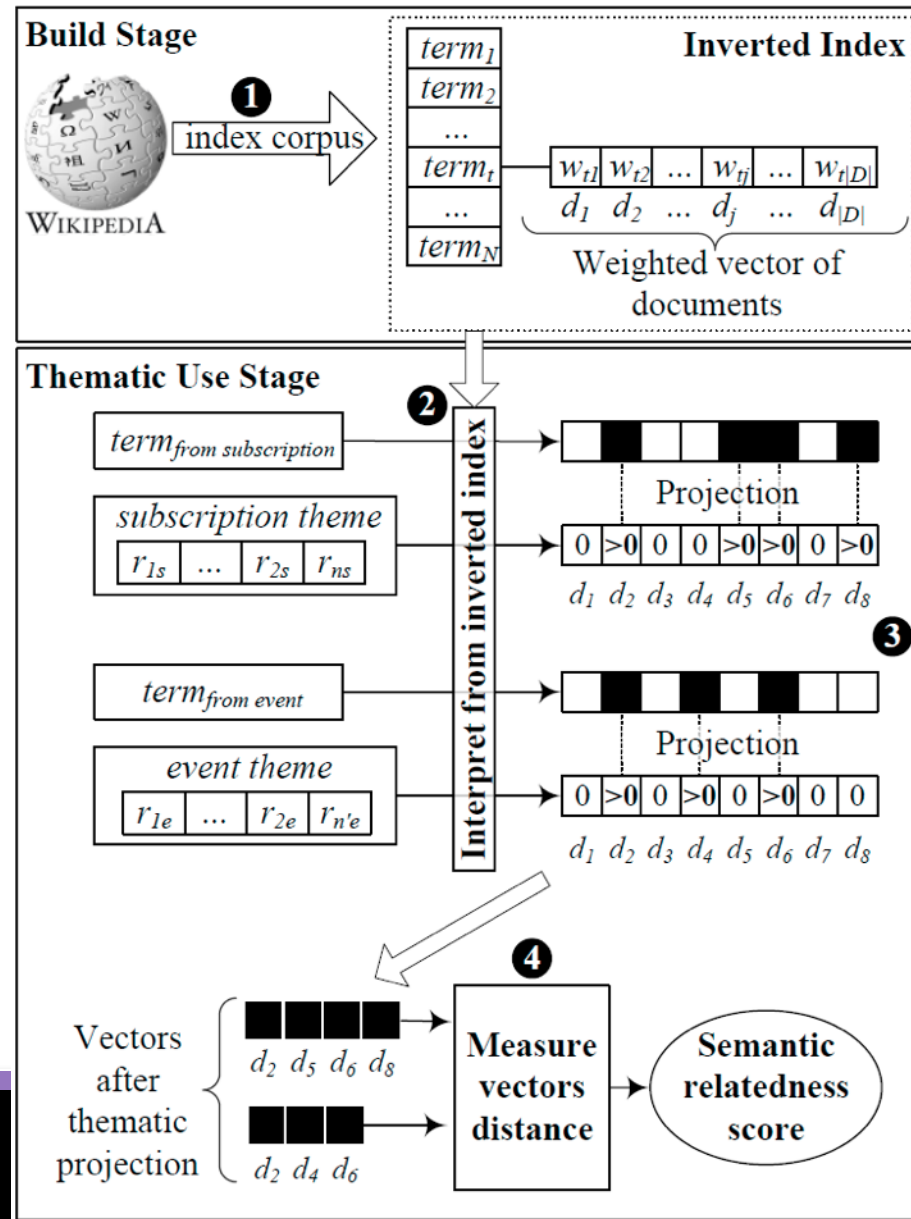
power, computers

type= increased energy usage event~,
device~= laptop~,
office= room 112

- Thematic tags used to parameterize the semantic measure



- Project vectors in a distributional semantic vector space
- Thematic projection

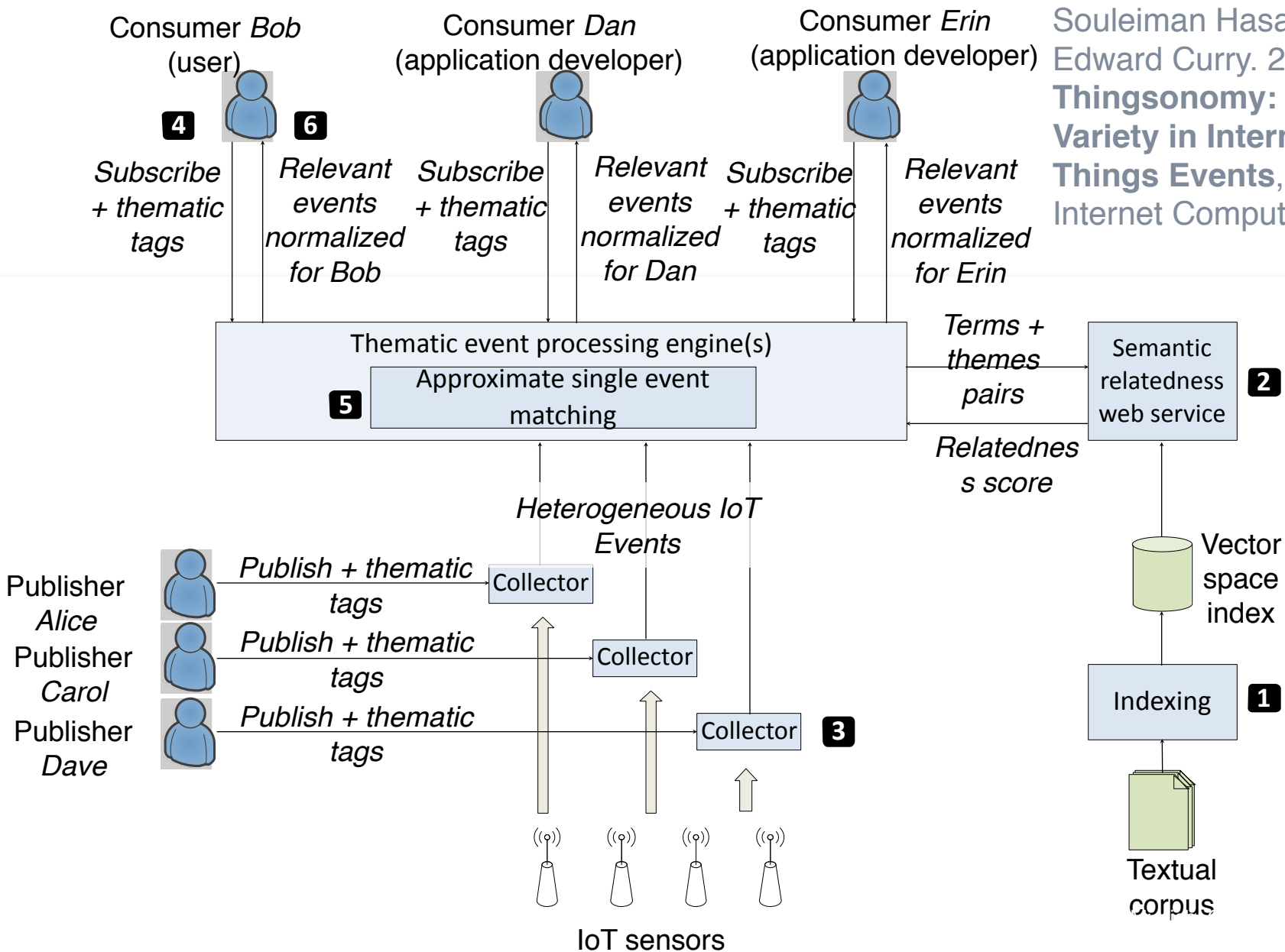


Building the IoT

Challenges for Building IoT Systems

- **Vastly heterogeneous, decoupled, and distributed nodes**
- **Lack of central coordination, reference requirements, or data model**
- **High overhead associated with software design associated with establishing agreements between parties**

Building IoT Event System



Souleiman Hasan and Edward Curry. 2015. **Thingsonomy: Tackling Variety in Internet of Things Events**, IEEE Internet Computing

For WoT IG

- To have *minimal description* on format/ semantics
- To enable *free tags* in things/events description
- To *externalize the semantics* model outside of the WoT framework to allow more possible semantics model: domain specific, ontologies, statistical semantics, non-symbolic semantics

Conclusions

- **Coupling necessary for crossing boundaries**
- **Decoupling necessary for scalable software**
- **Event-based systems need extension to address the coupling/decoupling tradeoff for semantics**
- **Approximate and thematic event processing exchange approximations of meaning with loose semantic coupling**

References

- CUGOLA, G. AND MARGARA, A., 2011. Processing flows of information: From data stream to complex event processing. *ACM Computing Surveys Journal*.
- EUGSTER, P.T., FELBER, P.A., GUERRAQUI, R. AND KERMARREC, A.M., 2003. The many faces of publish/subscribe. *ACM Computing Surveys (CSUR)*, 35(2), pp.114–131.
- Carlile, Paul R. "Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries." *Organization science* 15.5 (2004): 555-568.
- SOULEIMAN HASAN AND EDWARD CURRY. 2015. TACKLING VARIETY IN INTERNET OF THINGS EVENTS, IEEE Internet Computing (In Press)
- SOULEIMAN HASAN AND EDWARD CURRY. 2014. APPROXIMATE SEMANTIC MATCHING OF EVENTS FOR THE INTERNET OF THINGS. *ACM TRANS. INTERNET TECHNOL.* 14, 1, ARTICLE 2 (AUGUST 2014), 23 PAGES. DOI=10.1145/2633684 HTTP://DOI.ACM.ORG/10.1145/2633684
- HASAN, S., O'RIAIN, S. AND CURRY, E., 2013. TOWARDS UNIFIED AND NATIVE ENRICHMENT IN EVENT PROCESSING SYSTEMS. IN THE 7TH ACM INTERNATIONAL CONFERENCE ON DISTRIBUTED EVENT-BASED SYSTEMS (DEBS 2013). ARLINGTON, TEXAS, USA: ACM.
- HASAN, S., O'RIAIN, S. AND CURRY, E., 2012. Approximate Semantic Matching of Heterogeneous Events. In *6th ACM International Conference on Distributed Event-Based Systems (DEBS 2012)*. Berlin, Germany: ACM, pp. 252–263.
- SOULEIMAN HASAN AND EDWARD CURRY. 2014. THEMATIC EVENT PROCESSING. IN PROCEEDINGS OF THE 15TH INTERNATIONAL MIDDLEWARE CONFERENCE (MIDDLEWARE '14). ACM, BORDEAUX, FRANCE, 109-120. DOI=10.1145/2663165.2663335 HTTP://DOI.ACM.ORG/10.1145/2663165.2663335
- HASAN, S., CURRY, E., BANDUK, M., AND O'RIAIN, S. TOWARD SITUATION AWARENESS FOR THE SEMANTIC SENSOR WEB: COMPLEX EVENT PROCESSING WITH DYNAMIC LINKED DATA ENRICHMENT. THE 4TH INTERNATIONAL WORKSHOP ON SEMANTIC SENSOR NETWORKS 2011 (SSN11), (2011), 60–72.

Dataset and Software

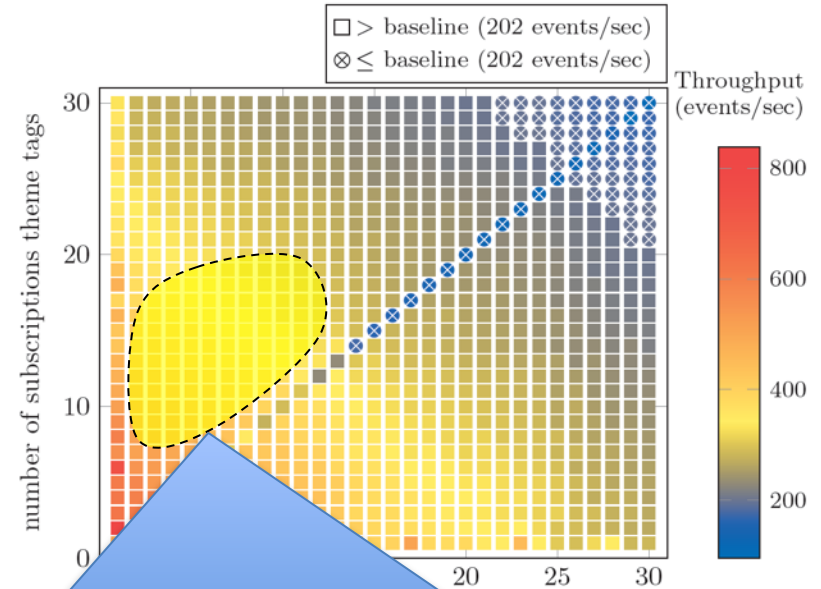
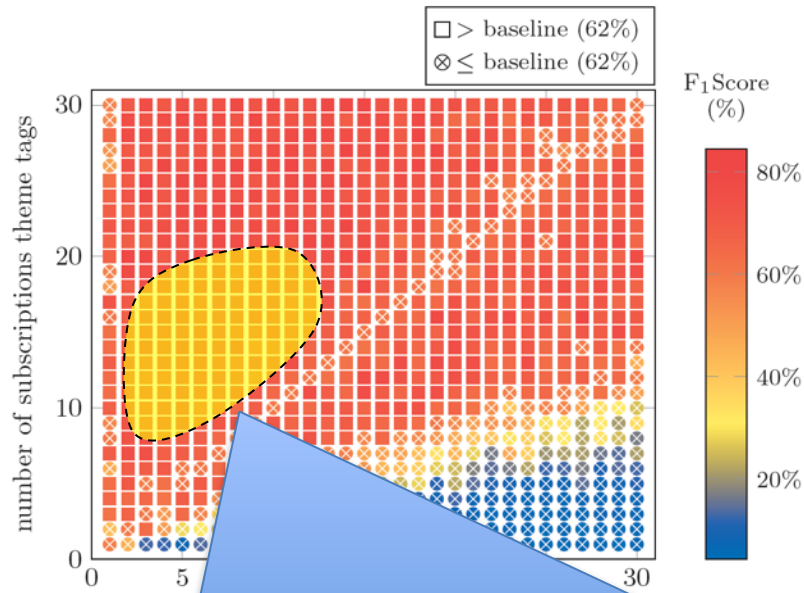
- **Dataset**

- Souleiman Hasan, Edward Curry, Thematic event processing dataset, DOI: 10.13140/2.1.3342.9123
- Available at http://www.researchgate.net/publication/263673956_Thematic_event_processing_dataset

- **Collider**

- Souleiman Hasan, Kalpa Gunaratna, Yongrui Qin, and Edward Curry. 2013. Demo: approximate semantic matching in the collider event processing engine. In Proceedings of the 7th ACM international conference on Distributed event-based systems (DEBS '13). ACM, New York, NY, USA, 337-338. DOI=10.1145/2488222.2489277 <http://doi.acm.org/10.1145/2488222.2489277>

- **Alternative cost:** number of exact rules needed to compensate heterogeneity
- **Precision, Recall, F₁Score:** reflects ratios of true/false negatives and positives
- **Throughput:** # of matched events per time unit
- **Standard error:** in performance of sample of thematic tags associated with subs and events



The use of less terms to describe events, around 2-7, and more to describe subscriptions, around 2-15, can achieve a good matching quality and throughput together with less error rates.

Lightweight amount of tags.

F₁Score (%)

throughput (events/sec)