Input paper for IAB Semantic Interoperability Workshop

San Jose, 17-18 March 2016, see https://www.iab.org/activities/workshops/iotsi/

Dave Raggett (W3C) and Soumya Kanti Datta (EURECOM)

We introduce the Web of Things and the W3C Interest Group, and outline a number of challenges for discussion at the workshop.

The Web of Things as a platform of platforms

There is widespread agreement that the current Internet of Things (IoT) is fragmented with lots of non interoperable platforms, data silos and a bewildering variety of technologies and standards. W3C is seeking to counter this fragmentation through introducing an abstraction layer that sits on top of existing platforms and standards. This has two main aspects: the first focuses on application developers with a simple scripting model based upon software objects that stand for local or remote things (physical or abstract entities). The second is a uniform framework for naming and describing things, building upon Tim Berners-Lee's vision for linked data. In essence, we want to expand the Web from a web of pages to a much bigger web of things. The core Web architecture is based upon names (URIs for addressing), resources, and protocols for accessing these resources. For the Web of Things, URIs are used to name things and to access their descriptions. Servers use these descriptions to create software objects with properties, asynchronous actions and events for use by applications. The servers automatically manage communication along proxy chains for remote things.

Applications can be written using dynamically typed languages such as JavaScript, statically typed languages like Java and C#, and state machine languages such as SCXML. The application logic is decoupled from the underlying protocols, data formats and encodings, as well as the communication patterns (push, pull, pub-sub, peer to peer). These can be chosen by the platform developer according to the context. Real-time requirements can be addressed through placing control close to the sensors and actuators involved. In some cases, it makes sense to multiplex data from many sensors, and to buffer readings so as to optimize use of the network. A further complication is where devices are ambient or battery powered, and consequently need to spend most of their time asleep to conserve power. Web of Things servers can be implemented on a wide range of devices from micro-controllers to edge computing platforms to cloud based server farms.

The clean separation of application logic from the lower layers in the communication stack is made possible through extensive use of rich descriptions. The associated vocabularies can be divided into vertical vocabularies for specific application domains (e.g. smart homes, smart healthcare, smart cities, and smart industry), and horizontal vocabularies that are applicable across many domains. W3C is seeking to standardize horizontal vocabularies which can be grouped into terms for describing things (e.g. the data model exposed to applications), security and communications metadata. Domain specific metadata is needed to express the semantics, e.g. that this is a pressure sensor and the units are in Pascals. Additional metadata can be provided to indicate just what the pressure sensor is measuring, when the sensor was installed, the last time it was calibrated, its owner, and so forth. We anticipate the need for many different kinds of metadata vocabularies.

Semantic Interoperability across Platforms

Interoperability requires a shared understanding of exchanged information along with an agreed expectation for the response to this exchange. The Web of Things introduces a "things" layer between the application layer and the underlying layers in the communication stack as shown in the following table.

Application	Scripts that define thing behavior in terms of their properties, actions and events, using APIs for control of sensor and actuator hardware.
Things (I4.0 Components)	Software objects that hold their state. Abstract thing to thing messages including life cycle events. Semantics and Metadata, Data models and Data.
Transfer	Bindings of abstract messages to mechanisms provided by each protocol, including choice of communication pattern, e.g. pull, push, pub-sub, peer to peer, ...
Transport	REST based protocols, e.g. HTTP, CoAP. Pub-Sub protocols, e.g. MQTT, XMPP. Others, including non IP transports, e.g. Bluetooth.
Connectivity	Underlying communication technology with support for exchange of simple messages (packets). Many technologies designed for different requirements.

Platform developers are responsible for implementing the layering. Application developers are freed from having to deal with the lower layer challenges and can verify compatibility in terms of the metadata for the data models and semantics. Different platforms may use different formats for data and metadata, and this needs to be handled automatically to preserve the clean separation of the application logic from the transfer and transport layers. This in turn requires standard ways to map platform specific representations to the Web of Things. For instance, to map metadata into the <subject, predicate, object> model used by the W3C Resource Description Framework (RDF).

How to facilitate re-use of vocabularies

In an open market of services, the ability to compose services from different vendors will be easier if they share the same vocabularies (a more formal term is *ontologie*s). What can we do to make it easier for people to re-use vocabularies rather than reinvent the wheel? One approach is to provide repositories such as schema.org, which provide the means to developers to browse for existing vocabularies and to upload their own. Another approach is to create and maintain the vocabularies together with the vendors to ensure that their wishes can be incorporated, and that they accept the vocabulary from the start. Vocabularies will inevitably vary according to the specific assumptions about how they will be used. This makes it important for these assumptions to be documented along with the vocabularies. What is the role for sharing best practices for vocabulary design, and best practices for repositories. Should W3C provide such a repository?

Federated approaches for data and metadata

We are seeing interest in federated alternatives to centralized approaches that hold data and metadata on specific servers. An alternative is to distribute computation, data and metadata across a large number of servers based upon a combination of distributed hash tables and block chains, where the servers are provided by a wide range of companies/organizations. This could be important as a way to avoid the domination of a single company in the way that has happened for web search. A vendor independent approach could be important for national security.

Interoperability in the context of evolving services

There are huge challenges for maintaining interoperability on a web scale marketplace of services. Weakly coupled communities will naturally diverge over time. How can a service consumer be confident about continuing interoperability with the service supplier? Good design practices can help, e.g. enabling consumers to ignore fields that they don't understand. However, that only gets you so far. We can learn from the experience with Linux package management solutions, which resolve dependencies automatically, and allow for multiple versions of libraries to be installed. What is the role of version numbering? What about MUST-UNDERSTAND attributes?

Security, trust and privacy

Security is one of the biggest challenges for the IoT. There will be a need to control who or what can access data and metadata. This also applies to discovery, as knowing what devices or services are available can breach people's privacy, or likewise harm the reputation of businesses. Having discovered a "thing", the security metadata can be used to determine the requirements for accessing it. In principle this can be done when designing a composition of services. Trust pertains to expectations as to the accuracy and quality of data and consistency in quality over time. Trust can in part be based upon brands from organizations who have established a reputation for safeguarding privacy and for providing accurate information. By analogy with the banking world's legal requirements for "know your customer" when setting up new accounts, there will be a need for attestation by trusted third parties who can provide assurances as the validity of an identity's attributes.

W3C Web of Things Interest Group

Following a workshop in 2014, W3C launched the Web of Things Interest Group at the start of 2015. The group has been working on use cases and requirements and a technology landscape survey, with the work split across four task forces as described below. The Interest Group is now preparing the way for chartering a Working Group to drive selected work items along the standardization track. We are also reaching out to industry alliances and SDOs, e.g. IETF/IRTF, ETSI, oneM2M, GSMA, OMA, OIC, IIC, Industrie 4.0 and OPC, with a view to integrating a wide variety of platforms as part of the Web of Things.

Thing Descriptions

This task force has so far focused on the data models that things expose to applications and ways to serialize this with JSON. Open topics include integrity constraints, streams as first class data types and late bound things for properties, responses to actions, and events. We're starting to look at semantics, security and communications metadata.

APIs & Protocols

This task force has focused on the APIs exposed to applications and the binding of abstract messages at the "things" layer to specific protocols such as HTTP and CoAP. Further work is anticipated on expanding this to other protocols, such as MQTT, MQTT-SN, XMP, Web Sockets and AMQP.

Discovery & Provisioning

This task force has focused on discovery, and the categorization of different techniques into several broad categories

- *Things near me* - e.g. based upon Barcodes, Bluetooth Low Energy, ZigBee etc.
- *Things on my network* - e.g. based upon zero-conf protocols and UPnP.
- *Things listed on a registry* - e.g. hosted on a home gateway or cloud based server
- *Semantic based discovery* - based on metadata and relationships including social networks.

Future work is anticipated on provisioning in collaboration with the security task force, see below.

Security, Privacy and Resilience

This task force has focused on current approaches to security for the IoT. Further work is anticipated on privacy and resilience, e.g. to faults and cyber-attacks. Today each platform defines its own approach to security. This makes it harder to ensure end to end security when developing services that span multiple platforms. Collaboration with IoT industry alliances and SDOs is needed to arrive at a common framework that can be adopted as part of open marketplaces of services on a global scale.