Market Drivers and High-Level Architecture for IoT-enabled Data Marketplaces

Omar Elloumi chair AIOTI WG Smart cities[1]

Tom De Block chair AIOTI Distributed Ledger Technologies

Natalie Samovich chair AIOTI WG Smart Energy [2]

Disclaimer: this paper does not suggest an AIOTI-agreed position. It is a proposal by the authors with the intent of soliciting industry-wide feedback. Official AIOTI deliverables on the subject will be published during the course of 2019.

[1] Nokia Bell-Labs and serves as advisor to startups incl. datapace.io
[2] Head of Research and Innovation at Enercoutim, co-founder Resilientenergy

Data Marketplaces enabling the exchange of data sets and data streams are analogous to Digital Marketplaces such as eBay, an analogy made interesting by the fact that few predicted the meteoric rise of eBay and its compatriots. Stories told about eBay's launch in 1995 quote prominent venture capitalist David Cowan as saying**: "Stamps? Coins? Comic books? You've GOT to be kidding. No-brainer – pass".**

It is now estimated that Digital Marketplaces will impact 40 per cent of worldwide retail by 2020.

Data Marketplaces, in particular dealing with IoT, face scepticism similar to that seen at the launch of e-commerce. Some have questioned their commercial viability. And much like the early days of eBay, existing Data Marketplaces have focused initially on niche opportunities. However, we believe there is huge potential for Data Marketplaces to succeed as first-movers establish themselves or scale-up their efforts by consolidating adjacent opportunities.

Today we have a range of reasons to believe that the market drivers for IoT-enabled Data Marketplaces are solid, especially with the Internet of Things (IoT) becoming pervasive.

1. What are the market drivers for selling and buying IoT data?

Data as a Service (DaaS) offerings (consumption data, geographic information, etc.) have been available for a number of years and have come to form a multi-billion dollar industry. But IoT cannot be considered a mere extension of DaaS. IoT brings together Information Technology and Operational Technology, often in relation to critical industrial processes, which complicates the task of making IoT data available to third parties, given concerns around potential threats to privacy, security, safety and the confidentiality of commercial intelligence.

Transacting IoT data must be different in many respects in order to build much-needed trust in IoT-enabled Data Marketplaces, trust that will be key to their sustainability. But before we address the specifics of IoT data transactions, what are the main market drivers of an IoT data economy?

1.1. Selling IoT data dramatically improves the business case for organisations' digital transformation based on IoT

The newfound success of IoT technologies and applications is the result of recent advances in information and communication technology (ICT), advances that are making IoT affordable for a large number of use cases where previously the cost was prohibitive. Selling data is expected to become an integral part of any IoT business case, and this should result in IoT being deployed on a larger scale thanks to quicker return on investment. In 2016, a publication quoting Gartner, advocated 80 per cent of companies would fail to monetize IoT data[3]. However, the situation has changed favourably since. Industry now understands how 'IoT plumbing' works and is increasing its focus on the monetization potential of IoT data.

[3] https://www.gooddata.com/blog/80-percent-of-companies-will-fail-to-monetize-iot-data-according-gartner

1.2. External and fresh IoT data complement data generated internally

Data generated internally to an organisation is usually not enough to remain competitive, enhance customer experience, and improve strategic decision-making. Looking at IoT-enabled mobility, for example, a car equipped with LIDAR (light detection and radar), a gyroscope and an accelerometer can accurately detect bumps and potholes on the road. Those data sets could be extremely useful for municipal governments as well as companies in fields such as car insurance, navigation applications, and road maintenance. This would however rely on incentive for this data to be shared. In this example, the entities to benefit from access to the data are not in a position to collect this data on their own.

Efforts are also ongoing to predict the development of potholes even before their formation. Provided that a sufficient number of cars are equipped with adequate sensors to generate the necessary data, local authorities could improve road safety, decrease road maintenance costs, and limit the need to compensate road users for damage following insurance claims.

The demand for external data arises from continuous innovation and related value creation. IoT-enabled environments entice 'innovation adaptation' as projects can gradually generate additional value based on marginal improvements to existing infrastructures, provided the high scalability potential of related architectures especially if using open standards for IoT (e.g. oneM2M).

New infrastructure installations, such as smart electric vehicle charging stations that leverage data from energy markets and other sources, are good examples where cross-domain data will be crucial for commercial sustainability.

Most organisations find themselves going through four stages of IoT project implementation. Each stage poses challenges and requires organisational "mind share".

The first stage aims to achieve optimal infrastructure selection based on a limited set of internal or inward-focusing use cases. It is defined in the graphic below as the "infrastructuralisation" stage. It allows for the implementation of the first value-generating use cases with well-defined expectations pertaining to the associated data.

It is followed by organisational adaptation, the second operationalisation stage. Mastering the initial set of use cases creates opportunities to augment the initial scope and seek external data sources, with a view to the creation of additional value beyond the initial scope of a particular use case.

Interoperability platforms and evolving open-source 'plug-and-play' solutions allow for the creation and growth of an ecosystem. This leads to the next stage after core competencies and value creation are established. The need for new cross-domain data sources as well as the drive to leverage the original data streams in additional ways will lead organisations to look for external data to leverage. This process allows the establishment of a bidirectional push and pull of internal and external data

The first three stages lead to the fourth stage of marketplace participation. The pathway towards the external data demand grows organically based on the experiences and adaptation of infrastructures and organisational support and business processes.

Stage 1 IoT Infrastructuralisation.

Stage 2 Data generation and integration of other internal data sources

Stage 3 Analytics and the need for external data

Stage 4 Cross-leverage buy, sell and provide data and services leading towards Marketplaces participation

Figure 1 Stages towards IoT marketplaces

1.3. Artificial Intelligence (AI) and Machine Learning (ML) will yield greater value if algorithms are trained on large volumes of representative data

There is widespread industry agreement that AI and ML algorithms, if trained with the largest possible volume of high-quality data, can create **new business opportunities and revenue streams**. **Acquiring high-quality IoT data is thus comparable to acquiring the raw materials essential to production in conventional industries.**

Recognizing that a strong business case for IoT Data Marketplaces is emerging, elaborated by reports such as Western Digital and Accenture's report on the 'Dawn of the Data Marketplace'[4], this paper describes a ***possible*** **'High-Level Architecture for an IoT-enabled Data Marketplace'** under development in the Alliance for Internet of Things Innovation (AIOTI).

We aim to provide a snapshot of what this reference architecture could look like to invite feedback and encourage new contributions to this work.

2. High-Level Architecture and key concepts of an IoT Data Marketplace

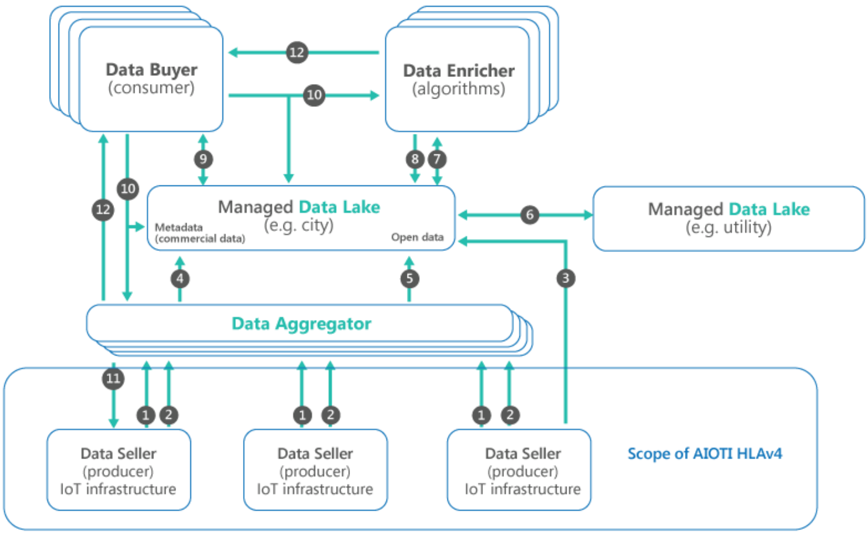

Figure 2: A possible high-level architecture for an IoT Data Marketplace

Figure 2 provides a possible high-level architecture for an IoT Data Marketplace. This proposed reference architecture includes functions that could be mapped to different stakeholders, and **multiple functions can be implemented by the same administrative stakeholder** in a given operational deployment.

- **Data Sellers** are entities that deploy an IoT infrastructure, for example smart energy meters. These entities are interested in selling the collected data or subsets of that data. This sale must be in accordance with privacy regulations and data owners' consent. The Data Seller would typically publish both commercial data (1) and open data (2) using a Data Aggregator. Alternatively, the open data may be contributed directly to a Managed Data Lake (3).

- **Data Aggregators** are programmed to aggregate mostly 'dumb' data streams from different sources, merging these data streams to create more valuable sources of information. The Data Aggregator would typically contribute both open data (5) and metadata pertaining to commercial data sets (4) to a Managed Data Lake. Metadata would provide a **semantic description** of the data as well as the **terms of contractual agreements** governing data transactions. The Data Aggregator would be responsible for transacting data on behalf of data producers in exchange for a portion of associated revenue streams.

- **Managed Data Lakes**[5] would typically store a massive amount of data and metadata to enable data discovery, as shown in arrows (7) and (9). This proposed reference architecture assumes that a Managed Data Lake does not store commercial data.
  Following a Data Buyer's discovery of data of interest to them, that Data Buyer would subscribe to an automated smart contract (10) for the agreement and immediate pay-out of the Data Seller's expected price (11). In other scenarios it would remain possible for the Data Seller to receive a revenue stream in a periodic manner, for example once a month. The service provider responsible for a Managed Data Lake would automatically receive a commission on every transaction facilitated, a key requirement for the financial sustainability of the data lake.

    1. After the settlement of the payment, the actual data would be exchanged peer-to-peer (12) between a Data Buyer and Data Aggregator.
    2. A Managed Data Lake could also contain mirrors of metadata from other lakes. The mirroring process is shown in (6).

- **Data Enrichers** are entities buying commercial data or consuming open data (7) with the intention of applying algorithms to enrich data and resell new data sets as a value-added service, typically to provide analytics yielding new insights and predictions. A Data Enricher would contribute its metadata back to a Managed Data Lake (8).

- **Data Buyers** consuming data streams or downloading data sets (12) are interested in the additional value that external data can bring to their internal data.

---

[5] Data lakes have been covered in this blog : https://news.itu.int/what-will-keep-smart-cities-busy-2019/

3. Important concepts in an IoT Data Marketplace

Certain concepts are fundamental to the successful deployment of IoT Data Marketplaces adopting the high-level architecture proposed in Figure 2.

- **Metadata** provide descriptions of the data assets up for sale by different stakeholders as well as the methods to transact in these assets. It is important that data sellers and buyers share a **common understanding of what the data is about**. Reaching this common understanding would only be possible with a **standard** or agreed ontology. The ITU Focus Group on data processing and management and the Open Geospatial Consortium could be the two initiatives to consider this standards gap.

- **Mirroring metadata** is the concept of exposing metadata in a third-party data lake. This mechanism allows for cross-domain data discoverability.

- **Cross-domain data discoverability** facilitates the distributed, collaborative development of data-driven solutions in line with the principles put forward by the EU Digital Single Market, for example.

- **Blockchain and distributed ledger technologies** provide means to build trust into every transaction without the need for central authorities. They are capable of enabling micropayments without transaction fees. They are also valuable in providing proof-of-origin for data sets as well as proof-of-integrity for data lakes.

- **Decentralized, yet federated:** the proposed reference architecture describes a data economy without need for a central entity or centralized powers, which could offer a foundation for a fair distribution of revenue streams. The federation is achieved through the mirroring process.

- **Governance** presents some of the most complex problems in this space. It is difficult to define sustainable governance models for new technology solutions when new models appear continuously and the oldest model is only a few years' old. The governance challenge is two-fold:
    - Keeping up with evolving models and technology, such as blockchain and distributed ledger technologies,

A guide developed by McKinsey on the creation of Data Marketplaces lists **six key enablers**. We see one of these enablers – "achieving consistent data quality"[1] – as perhaps the most important. Auditable and adequate service-level agreements that can ensure that marketplaces deliver data of consistently high quality will become a defining feature of sustainable Data Marketplaces. Quality assurance will come at an additional cost but would deliver a more sustainable model.

including "**their potential to transform and even reinvigorate the governance of cities**"[6]

- Ensuring a fair distribution of revenue streams and avoiding the creation of new monopolies

4. Relationship and synergies with the energy sector

The energy sector is undergoing significant transformation with the aim of achieving 2030 and 2050 decarbonization goals and driving the development of new levels of interconnection, flexibility and decentralization. Decentralization, decarbonization, democratization, deregulation and digitalization are transformational paradigms giving rise to new marketplaces expected to scale across Europe and beyond. It is anticipated that the European energy sector will produce ecosystems of several marketplaces while concurrently aiming to avoid fragmentation.

Working in isolation, the energy sector may not be capable of transforming energy grids and bringing new services to consumers. This transformation will call for the energy sector to interact and learn from experiences of successful cross-sector marketplaces, create interfaces with smart city marketplaces, and collaborate extensively with technology providers and connectivity providers. The energy sector's transversal relevance is a source of both the demand for and supply of fresh data. Energy and industrial platform interfaces are under development to address related requirements.

To be continued

The authors plan to cover IoT Data Marketplaces and related subjects in future publications, recognizing the considerable depth of discussions around the readiness of the enabling technologies and associated implications for technology adoption, differences in the priorities of the public and private sectors, and issues pertinent to the governance of Data Marketplaces (including for IoT). Please stay tuned.

Acknowledgment

The authors would like to thank Matthew Dalais (ITU) and Inês Fonseca (Enercoutim) for their help in producing this article.

6. Sarah Barnes, Smart cities and urban data platforms: Designing interfaces for smart governance. City, Culture and Society