White Paper

Managing the proliferation of data generated by the Internet of Things, and transforming that data into useful information for the Smart Cities of tomorrow.

Abstract:

The increasing number of sensors and other infrastructure devices being installed to facilitate Smart City initiatives results in huge amounts of data being generated. Smart Cities are not unitary but are federations of cooperating organisations where the data generated is often, at present, consigned to data silos with little interaction between those islands of data. The amount and distribution of data generated makes it difficult and time-consuming to present meaningful information in sufficient time to make it useful in a Smart City environment.

This short white paper discusses the way that data brokering technology integrated with federated service agreements, as implemented by Agility DataBroker, can provide a cost-effective way to manage and distribute that data, transforming it into usable information in a timely and efficient manner.

www.arjuna.com

Contents:

Executive Summary

"Organizations staring down the daunting task of IoT data management seek features that enable them to process, store, and analyze the crush of information they can now generate." – Aberdeen Group, February 2015

The data generated by the Internet of Things (IoT) is growing at something like 30% a year according to the Aberdeen Group. The sheer volume of data generated makes it difficult for centralised data management facilities to extract usable information from the data within a meaningful timeframe. The intelligence has to move further out towards the periphery to ensure that only pertinent data is consolidated centrally to facilitate timely data analysis.

There is a further complication when considering the data generated from IoT in a Smart City context, which is that the data generated across the city is typically held in silos by different organisations who have legislative, process, moral, political and other constraints on how they can share data. The success of a Smart City undertaking is based on joined-up thinking which requires joined-up data. It is necessary to break down the barriers that exist between the different data owners while ensuring that security, licensing, privacy and other constraints are adhered to.

This white paper explores a potential solution to this problem in the form of Agility DataBroker, a software framework authored by Arjuna Technologies Ltd., a spin-off from Newcastle University in England. As its name suggests the software acts as a data broker, allowing multiple, heterogeneous sources of data to be consolidated, cleaned, analysed and distributed across a network. The framework can be deployed out towards the periphery of the IoT infrastructure so that only the appropriate data ends up where it needs to be in a timely manner. It allows for the exchange of metadata, defined by the data owners in machine readable format, so that data can be automatically stored, analysed and converted for use by the data beneficiaries.

Agility DataBroker is unique, as far as we are aware, in its use of federated service agreements between data owners to share their data in a policy-led manner. The resultant combination of multiple data sources produces an exponential benefit to the Smart Cities of tomorrow.

This white paper investigates the aforesaid problems in more detail, and demonstrates how Agility DataBroker can help provide a solution.

Proliferation of data

"*In God we trust, all others bring data*". - Dr. W. Edwards Deming in The Elements of Statistical Learning,

Deming's words, repeated famously by Michael Bloomberg, still hold true today, but there are many who believe that too much data is as bad as too little. According to IBM, every day society generates 2.5 billion gigabytes of data in organizations, via the

Internet of Things, and as a result of our personal digital activities, and that less than 1% of that data is mined for valuable insight. Aberdeen Group believes that organisations using IoT will double the amount of data they generate every three years.

The Smart City of tomorrow will be dependent on the machines, sensors and devices connected together into the Internet of Things, but with an almost exponential increase in the amount of data each generates, extracting meaningful information from that data using existing technologies becomes an almost impossible challenge. The challenge is hard enough for a single organisation extracting data from a single network of similar devices, but the Smart City of tomorrow will have additional challenges.

The sensors and other devices currently deployed tend to serve a specific purpose for a specific organisation. Energy companies providing city-wide services deploy sensors to monitor flow across their grid, traffic sensors on traffic lights monitor traffic for the Traffic Management Centre, weather sensors provide data for the forecasting services. The Smart City of the future will need to have data from all services to be able to run effectively. If not then how can they know when storms might damage power lines that affect the control of traffic through the city? The fact is that with data the whole exceeds the sum of the parts, as adding more data sets increases exponentially the number of ways in which that data can be combined. What makes the scenario more challenging is that the most valuable IoT applications are probably not yet recognised, and when recognised will probably change rapidly over time. As a result it is likely that the combinations of data required will have to change in step.

Centralised data analysis engines just can't keep up with the amount of data being thrown at them, with half of the organisations canvassed by Aberdeen Group admitting to failure in improving time-to-decision in the last year. The key is to distribute the intelligent data gathering process, making sure that each of the stakeholders in the data receives only that data which is meaningful to them and in a timely manner. This, however, introduces an additional layer of complexity in that there is usually only one data owner for a particular dataset but multiple, potential data beneficiaries. The data owner may be happy to share all or part of their data but will likely only be able to do so under certain constraints. To facilitate the Smart City we need to provide them with the capability to extract meaningful data from multiple data sources, combining that data in ways that allows joined up thinking to occur.

Need for federated service agreements

If we look at the different types of organisation involved in serving a city, it is easy to see how the data available may be used in different ways to facilitate more intelligence-based services. Traffic monitors placed at intersections provide information that could be used by traffic controllers to manage traffic flow, by police services to target speeding vehicles or track criminals, by ambulance services to determine the quickest route to an accident, or by members of the community to determine their quickest route to work or back home. Each beneficiary would be best served by receiving only a subset of the data available e.g. the police only getting data on speeding vehicles or specific number plates, and the data owner would be best served by not having to propagate the whole

of the data generated. In an increasingly commercialised world there may also be licensing considerations, costs, security and privacy implications in the provision of the data to external groups.

Today these exchanges of data, if they occur at all, are governed by paper-based service agreements which require significant time and effort to implement and change. What is required is a method of federating the data sources and the data beneficiaries so that the sharing of data is controlled through service agreements implemented in software which carries out the policies determined by the data owner in consultation with its users.

Because of the ever changing nature of anything driven by humanity the city will also need to be able to dynamically change those service agreements in concert with the data owners and consumers, rather than relying on lawyers to argue and reargue changes which may result in the opportunity or emergency having expired.

In short, the existence of any available dataset should be published along with the policies which dictate the service agreement that governs its access, and the metadata that describes their data in machine-readable format. Authorised users can then access that data provided that they conform to the policies without unnecessary delay.

Potential solution

The Agility DataBroker framework allows an organisation to unlock the value of their own data in conjunction with federated data through its controlled consolidation, cleaning, analysis and distribution. It allows for data to be extracted from any combination of data sources, whether they sit in the Cloud or locally, under strictly enforced, but dynamically changeable service agreements which govern its use and deployment. Agility DataBroker can be used to clean and filter data close to source and distribute it in different forms to different recipients so that the appropriate analysts can extract the information they require in an optimal manner.

Agility DataBroker is implemented as an overlay that is deployed as a stand-alone server at each organisation, or part of an organisation, which is to interact in relation to shared data. It provides a means to capture existing and future relationships with other organisations (Note that an organisation might be an independent enterprise, a department or even an individual) and enables organisations to be connected together into a 'Federation' in which cooperation occurs through mutual agreement, but within which each party retains independence.

The service agreements which control the exchange of data between organisations is defined as a set of policies that can be changed dynamically if necessary and which specify all critical access criteria. Within an Agility DataBroker Server, Policy is structured as one or more Policy Modules. Multiple Modules can represent different aspects of an organisation's policy, or represent the interests of different actors within the organisation, or be concerned with particular events. For example one Policy Module might be concerned with legal aspects, another with financial auditing, and yet another with green issues. Policy interacts with the Agility Server through a Policy API. By this

means Policy Modules can control which Service Agreements are to be entered into, how they may be modified over time and under which conditions they may be terminated.
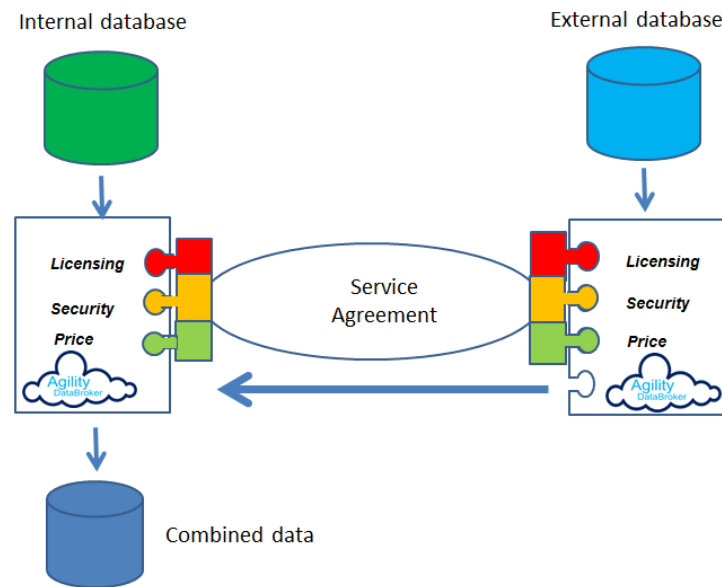
Policy can interact with the organisation in order to report on the progress of the organisation with respect to its Service Agreements and/or to modify the behaviour of the organisation in order to ensure conformance with its Service Agreements. The implementation and behaviour of Policy Modules are not restricted, beyond the requirement that the Policy API be supported. Policy Modules can interact with other Policy Modules, processes, human users and data in whatever way they see fit. The overall behaviour of the organisation will however need to be directed by the organisation's requirements to meet their Service Agreement obligations and the understanding of the business, financial and legal implications of failing to do so. Agility DataBroker does not in any sense 'enforce' Service Agreements or define how the organisation should behave. It simply records and reports on changes to Service Agreements as they occur and, if appropriate, permits those changes to be responded to automatically.

Data can therefore be extracted from the IoT, be cleaned close to source, be consolidated, analysed and distributed in accordance with these policies to one or more data beneficiaries who, in turn, may need to consolidate that data with data from other sources to provide the meaningful information required to run the city. By defining their data in machine-readable format, the data owner is able to share the metadata, facilitating easier, automatic conversion for use by the data beneficiaries.

The end result is that data owners retain control and visibility of to whom their data is consigned and the policies that dictate its access and use. Data beneficiaries only receive the datasets that are meaningful to them. The federation between the parties, implemented via the service agreements, allows for the policies under which data is exchanged to be dynamically altered where necessary.

www.arjuna.com
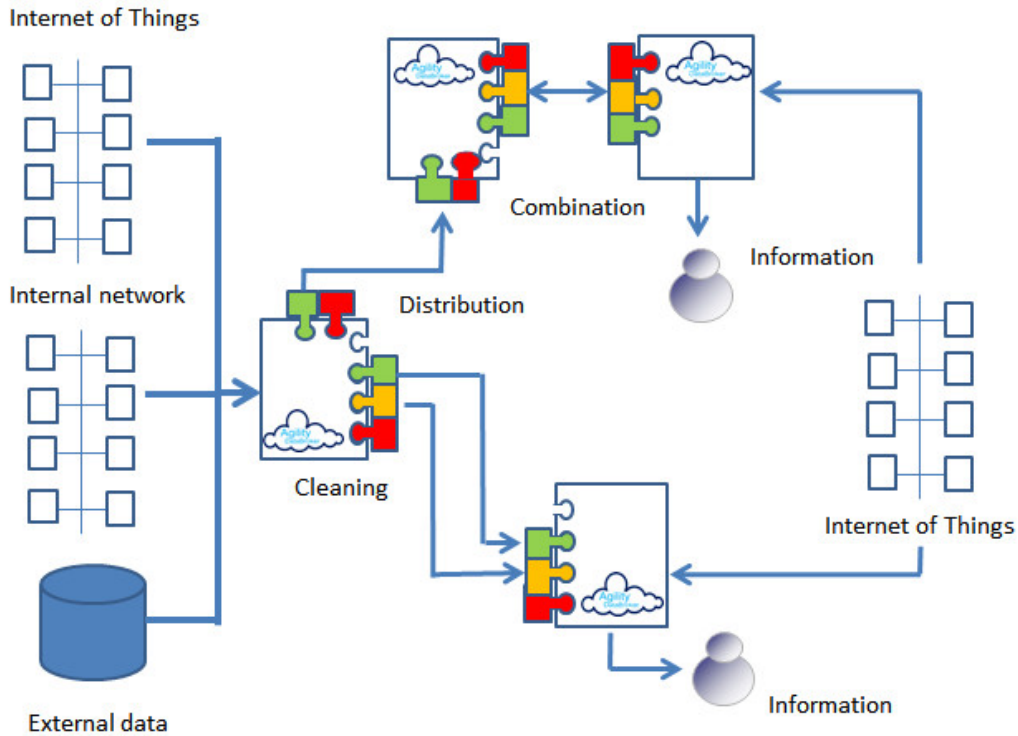
To run efficiently the Smart City requires access to multiple sources of clean data. We believe that Agility DataBroker can help facilitate that.

Summary

The Smart City will rely heavily on the Internet of Things to provide data. The plethora of data generated needs to be cleaned as close to source as possible and the relevant datasets automatically distributed to the authorised recipients who can consolidate that data with data from other data sources to produce the information they require.

Agility DataBroker from Arjuna Technologies Ltd. performs that role while coping with the politics of data. Service agreements between the data owners and data beneficiaries are implemented by policies within the software and ensure that all federated parties benefit accordingly.

www.arjuna.com

About the author

Arjuna Technologies Limited is a world leading innovator in distributed computing, and has an international reputation for delivering mission-critical products to global software vendors. Products developed by the company over the last decade are currently utilised by many thousands of businesses. Today Arjuna is applying its twenty years of research and development experience to enable the next major development in IT - the Federated Cloud.

Arjuna sells and supports Agility DataBroker directly and through its partners out of its offices in the North East of England.

Contact us at info@arjuna.com

Arjuna Technologies Ltd.
The Core
Science Central
Bath Lane
Newcastle-upon-Tyne
NE4 5TF
United Kingdom

Tel. +44 (0)191 243 0676
Email: info@arjuna.com
Web: www.arjuna.com