

Organizational Data and Analytics Contracting in Smart City Fog Platforms

Michael Goul
W. P. Carey School of Business
Arizona State University
Michael.Goul@asu.edu

Vineet Mishra
Ira A. Fulton Schools of
Engineering
Arizona State University
Vineet.Mishra@asu.edu

Divyesh Dnyanmothe
Ira A. Fulton Schools of
Engineering
Arizona State University
ddnyanmo@asu.edu

Abstract

Smart City infrastructures require contracts between public and private organizations collaborating in what is frequently referred to as fog computing platforms. We investigate contract provision variations from different stakeholder perspectives. Our methodology relies on complex adaptive systems theory, and we simulate different contract provision scenarios to identify patterns that emerge. The specific contract provisions we investigate in this paper are related to analytical model and data ownership paradigm variations. We find that some variations offer advantages to stakeholders that include those who participate in the smart city fog platform and those who may have ownership of smart city fog platform infrastructure.

1. Introduction

Fog computing is a highly virtualized platform that provides compute, storage and networking services between end devices and traditional cloud computing data centers. [1] The fog is envisioned to provide a number of extensions to the cloud paradigm including: 1) The fog targets the large scale distribution of data, analytics and smart devices connected through a network with rich services located at endpoints, 2) Endpoints are assumed to be a large number of geographically dispersed, mainly wireless devices that engage in real-time interactions and 3) Fog applications of analytics and big data require collectors at the edge to gather data (e.g., from mobile devices), provide semi-permanent storage facilities, deploy analytical models, and they have links to a cloud computing infrastructure where longer term storage, platform-level management and data mining can be efficiently performed.

Fog computing is a natural space for Internet of Things (IoT) applications. By bridging IoT and the cloud, fog computing takes on the tasks and roles for which cloud computing has not typically been deployed. For example, proprietary devices such as

those common to Smart City applications don't normally exist in cloud infrastructure, but niche players in IoT for Smart City applications are envisioned as serving a major role in the fog. There are a variety of definitions for the phrase 'Smart City,' but they almost always include some reference to IoT. Most definitions center on increasing urbanization, the types of problems and issues that arise in the dynamic of large population centers and the realization that such places can have serious problems in areas like traffic congestion, healthcare delivery, sustainability, deteriorating infrastructure, etc. [2]. For example, Ojo, Curry, et al. [3] find:

1. Smart City programs generally leverage physical infrastructure, information communication technologies, knowledge resources and social infrastructure to address economic regeneration, greater social cohesion, better city administration and improved infrastructure management, and
2. Smart City initiatives typically involve an interplay of technology, policy and organizational innovation that is shaped by and impact external factors like people, communities, the economy, built infrastructure, the natural environment and governance.

Several Smart City programs are ongoing in countries including Portugal, Brazil, United Arab Emirates, Korea, Japan, the USA and many others. Most are considering or have addressed Open Data as a Smart City initiative. Open Data is data that can be freely used, shared and built-on by anyone, anywhere, for any purpose. [4] For example, the Open Data Institute is centered on the mantra that, "An Open Data culture creates better cities for everyone." [5] In contrast, some view self-interest as a key reason for the importance of encouraging business ownership in many contexts including Smart Cities; business owners have a significant vested interest in ensuring a system is successful [6]. There is an important tension related to whether Open Data or a more traditional data ownership paradigm will foster more innovation in Smart Cities [e.g., 7]. Many tout Open Data as a means to stimulate entrepreneurs,

programmers, sustainability, transportation and healthcare experts, data scientists and others to identify solutions to urban issues. Open Data improves transparency, accountability and it encourages citizen participation. On the other hand, there can be issues if data is not well understood, managed improperly and/or used inappropriately in designing and deploying advanced analytics. Similarly, there are issues of security and privacy in Open Data Smart City contexts. Finally, Smart City innovations in complex verticals are seen as more likely to be motivated through proprietary structures, e.g., innovations in the traffic engineering and healthcare delivery disciplines will likely occur separately and under different ownership structures.

Many believe that a hybrid, parastatal model is applicable to managing fog computing platforms in Smart Cities. A parastatal is an entity that is owned wholly or partially by a government. Urbanization, infrastructure, and technology are aligned in smart cities, and a variety of parastatals have emerged [e.g., 8]

1. Songdo International Business District in South Korea is chartered by the municipality of Incheon, but the largest shareholder is an American commercial developer
2. The Skolkovo district of Russia is entirely government owned
3. The Delhi Mumbai Industrial Corridor includes 8 smart cities and is funded through a public-private partnership model

Because fog computing platforms for Smart Cities are and will likely continue to include the hybrid model, there is increasing need for legal instruments to foster partnerships among business, government and parastatal entities in order to achieve innovation [9]. The remainder of this paper deals with contracting issues in this business-to-business-to-government (B2B2G) context. First, we examine contract theory with an eye towards Smart City services that leverage data and analytics in deployments on fog computing platforms that are linked to cloud computing environments. We define the major issues that contracts need to address in this context. We then discuss our methodology and an example experiment. Through multiple experiments, we found important results intended to guide contract decision making from different stakeholders' perspectives. We conclude with a discussion of needed future work.

2. Nature of B2B2G Contracting for Smart City Fog Computing Platforms

There are four paradigms for fog computing platforms designed to deliver Smart City services: 1) Enabler, 2) Neutral, 3) System Integrator and 4) Broker. Walravens and Ballon differentiate the four on the basis of whether a single entity controls the assets in the platform and whether that entity has control over the customer relationships [10]. In this context, we can consider the owner as a government or a parastatal. An enabler platform is one where the owner controls the assets but does not have control over the customers. The neutral platform is one where the owner is heavily dependent on the assets of the other actors to achieve Smart City ideals and does not control the customer relationship. The system integrator platform is one where the owner controls many of the assets, the owner establishes a relationship with end-users, and external, third-party service provider engagement is encouraged. In the broker platform, the owner is dependent on the assets of others but doesn't control the customer relationship. These paradigms can be extended by considering the data assets that are created when a Smart City fog computing platform executes, and considerable data ingestion at service delivery endpoints ensues. When data is ingested, its ownership rights can inherit properties from the paradigm under which the platform is organized. For example, in the neutral platform, the data ingested might be owned by the third-party partner who also owns the relationship with the customer. Private parties may have proprietary rights to the data collected as part of their engagement in the fog platform. Similarly, in the broker platform, the data ingested at endpoints delivering services that are under control of the owner may also be the property of a third-party, but the owner may have to negotiate and enter into a broker contract for data ownership rights since external data and infrastructure provided by a third party enables the service. In these and many other Smart City scenarios, there are complex issues of data ownership that need to be reconciled through contracts.

In addition to data ownership contract considerations, the analytical models that can be created through data mining can also be subject to negotiated ownership rights, and this process is similarly complex. For example, a predictive model that is derived from data that is co-owned may inherit ownership rights from the data co-ownership contract agreement. Consider also that a predictive model deployed to deliver a personalized service at an endpoint that, when executed, may well result in additional data ingestion. This newly ingested data can be thought of as being *influenced* by a predictive model that is co-owned, so should that subsequently

ingested data be similarly co-owned by the predictive model owners? In general, data and analytics ownership should be contracted for separately, but their ownership can be intertwined. It is important for data and analytics contracts for Smart City fog platforms to take this complexity into account when detailing contract terms. However, it not always possible to envision any and all possibilities for a Smart City initiative, so contracts aren't often all-encompassing.

Grossman and Hart and Hart and Moore define ownership as the, “residual right to control access to an asset.” [11, 12] In using the term *residual*, this definition implies that some rights can be assigned to others using what is referred to as an *incomplete contract*. The rights not assigned are residual rights in the Grossman and Hart and Hart and Moore approach. The reason that a contract might be incomplete is that its clauses, provisions, conditions and terms may not cover all actions and payments for every possible contingency that might arise. In the theory of incomplete contracts, there is an important concern called the hold-up problem. This problem occurs when parties must make non-contractible relationship-specific investments before a transaction takes place. For example, building a Smart City fog computing platform might require third-party investment in infrastructure assets before any services can be deployed and revenue can be realized. The hold-up problem manifests in the fact that the third-party underinvests because the return may be greater if there can be renegotiation at a later time – e.g., the third-party sees the situation as a multi-stage game whereby more value may be extracted by a later renegotiation. Incomplete contract theory stipulates mechanisms for mitigating the hold-up problem. A basic solution is to make a contract as complete as possible and to limit renegotiation at a later date. To this end, we have developed a Smart City services system model that aids in articulating terms for Smart City contracts. Our focus is on the data and analytics ownership issues.

3. A Smart City Multitenant System Model

A Smart City fog computing infrastructure linked to a cloud computing environment needs to be capable of integrating multiple initiatives simultaneously. This is consistent with a multitenant architectural approach. (Note: Some Smart City infrastructures may be multi-instance; our focus is on multitenant environments.) For example, there may be traffic applications and healthcare applications deployed concomitantly. Therefore, a Smart City platform owner needs to configure infrastructure for different initiatives in order

to make optimal use of resources. In this way, there is support for initiatives that come and go, there are ways to support solution refinement and redeployment, new initiatives can built on top of successful predecessors and some initiatives might provide more value than others in delivering the same service so some die while others are scaled up. The swapping in and out of instances of Smart City services needs to be efficient, and the overall management of the platform requires a global management viewpoint for the owner(s).

We consider a Smart City service initiative as involving three major abstractions: 1) A set of Hosts: $H = \{H_1, H_2, \dots, H_p\}$, 2) A set of Smart Object Hosts: $SOH = \{SOH_1, SOH_2, \dots, SOH_q\}$, and 3) A set of Smart Objects: $SO_{SOHi} = \{SO_{SOHi,1}, SO_{SOHi,2}, \dots, SO_{SOHi,m}\}$. Smart Objects are located at endpoints where services are delivered and data is optionally digested. Smart Objects are associated with a sensor cluster owner or hub, i.e., a Smart Object Host. For each Smart Object, we assume it has a set of Senses: $S = \{S_{SOi,1}, S_{SOi,2}, \dots, S_{SOi,n}\}$. A sense can be a capability to ingest a data item (e.g., decibel level, temperature, etc.) and/or it can be an output capability (e.g., a noise, a coupon, a text message, etc.). Note that we refer to the endpoints as *smart* to imply they are capable of short-term data storage and they can execute an analytical model that may take sense values as inputs and produce sense values as outputs. We consider an analytical model (or simply an analytic) as being deployed by a Host through a Smart Object Host to a Smart Object. We describe an analytic as a function of a subset of the senses that a Smart Object is capable of, and the output of the model is also a sense that a Smart Object is capable of. In short, there can be a set of Analytical Models (AM) deployed by some Smart Object Host, SOH_j at the direction of a Host H_i or a Smart Object Host, SOH_j to some Smart Object, $SO_{SOHj,q}$, such that $AM(f(\text{Subset}(SO_{SOHj,q}, S_a, SO_{SOHj,q}, S_b, \dots, SO_{SOHj,q}, S_t))) \rightarrow (\text{Subset}(SO_{SOHj,q}, S_a, SO_{SOHj,q}, S_b, \dots, SO_{SOHj,q}, S_i))$.

A brief example can help to clarify the notations and the intent of the Smart City initiative service system model. Consider an Intel AIM Suite® video sign kiosk that can display a video while at the same time it can monitor the demographics of those who are watching the video [13]. Let's assume that the video signage kiosk is located just outside all store entrances in a particular district of a smart city. Further, assume that a district's single proprietor has purchased the video signage kiosks as a perk for stores that rent space in the district. Now, the Smart Object Host can be thought of as the district proprietor. The video sign kiosks can be thought of

as Smart Objects. The store owners may be thought of as the Hosts.

To put the service system into motion (Figure 1), consider that the video sign kiosk can display videos from suppliers to the store that help to advertise the products the store is selling. We can therefore also consider the suppliers as Hosts. When the district's stores are open and people walk around, a kiosk may be displaying a video and capturing demographics of a person who is watching intently. That kiosk has analytical models that enable examining the skull and body structure of a viewer to determine gender, age range, the viewing time and the distance from the kiosk. Further, suppose the kiosk can deploy analytical models sent to it by the stores. For example, Macy's may opt for the kiosk to send a coupon in a text message to a viewer who has watched a certain video for a certain length of time. Similarly, a supplier may have an analytical model at the kiosk that extends an up-sell offer for a viewer who showed interest in a swim suit ad that provides incentives to purchase a matching cover-up. When data is ingested at the kiosk, the analytical models use their senses to evaluate if the model is relevant, and if so, the kiosk's other senses enable it to send a text message (if permitted by the viewer). The data in the overall system includes the demographics of the viewer, the time spent watching the video, the state of how many kiosks had active viewers at a point in time in the district, the demographics of those issued texted coupons, whether a purchase was made with the coupon, etc. To further complicate matters, the store owners may be collecting data on sales to determine the most effective campaign models, so it can do data mining to build a new analytical model for deployment to relevant kiosks.

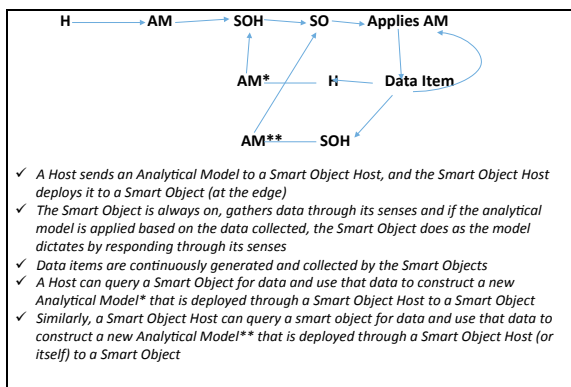


Figure 1: Sample dynamic flow

What's clear from the example above is that it can get extremely complex very quickly regarding what entities own which data and analytical models. This example is rich in the sense that there are multiple

participants, data ownership is open to contract negotiation between the parties, and the dynamics of the service system require consideration of a variety of scenarios and extended possibilities when deciding on contract terms. In fact, Federal Trade Commission Chairwoman Edith Rameriz cited Intel's AIM Suite as indicative of the privacy and security concerns for Smart City fog platforms. [14] In relation to the responsibility for privacy and security, there are many questions in terms of this scenario about ownership such as:

1. Who owns the demographic data collected at the kiosk – the district proprietor, the store, the supplier who provided the video? All of the parties?
2. How should ingested data co-ownership be determined? For example, should store owners be required to allow others to co-own data on the number of purchases (and dollar values) that resulted from the issuance of a coupon?
3. Who owns the predictive models that are constructed from data that might be co-owned?
4. Who owns the data that is collected after an analytical model is deployed that might impact the behavior of the service system as a whole?
5. If a store owner closes up, how will ownership structure be impacted?

Natural extensions to the system model include the situation where more than one Host is contracting with a single Smart Object Host, Smart Objects are co-owned, owned data by a party who exits should be deleted (or is allowed to persist if co-owned), etc. In the following section, the issues related to contract terms are summarized with respect to the Smart City multitenant system model and the example scenario above.

4. Data and Analytics Ownership Contract Provision Options

Contract provisions addressing ownership and other related characteristics can potentially influence the design and operations of Smart City fog platforms. In the following, we discuss six contract provision options. Several options pertain to contract provisions involving data and analytical model ownership.

Fixed-price vs. Value-based contracts: In fixed-based contracts, the parties agree on some unit of analysis upon which service fees will be assessed. For example, the messaging between H and SOH may be billed on a per-message basis or on the basis of the time duration of a contract. In contrast, value-based contracts refer to a division between the parties

of, for example, the total profit generated by partnering stakeholders. For example, if there is a profit from a partnership of H and SOH stakeholders in a Smart City fog platform, then contract provisions will specify how the profit is to be divided between stakeholders.

Data Exclusivity vs. Non-Exclusivity: When data ownership is exclusive, there is a specific entity that owns the data. For example, a contract may stipulate that an SOH owns all the data generated by the SOs connected to it. Data is co-owned or open when contract provisions stipulate that data is non-exclusive.

Analytical Model Exclusivity vs. Non-Exclusivity: An analytical model is owned by the entity responsible for its creation if an exclusivity provision is included in a contract. In addition, only the creator of the analytic is allowed to modify it. When an AM is non-exclusive, then it is co-owned or open. Any of the co-owners can modify the analytic.

Co-mingling vs. No Co-mingling: In co-mingling, an analytical model created or derived through data that is co-owned is thereby co-owned by the same entities who co-own the data. If there is no co-mingling, the entity that creates the model owns the model exclusively.

Data post-use vs. No data post-use: If data post-use is a contract provision, then if a data owner leaves a partnership governed by a contract, the data that entity owns becomes the property of the remaining owner(s). If there is no data post-use contract clause, the data solely owned by an exiting owner is deleted.

Analytical model post-use vs. No analytical model post-use: If an analytical model post-use provision is included in a contract, then an analytical model owned/co-owned by a exiting partner remains the property of the remaining owner(s). If there is no analytical model post-use, then an analytical model solely owned by an exiting owner is deleted.

Note that the contract provisions above may all exist at the same time in Smart City fog platform contracts. In the course of the life of such a fog platform, contracts will change between parties, parties themselves may come and go, and data and analytical models may constantly evolve.

5. Methodology

We view an instance of a multitenant smart city fog platform with a set of participating organizations and a platform owner(s) as a complex system. Such a system exhibits dynamism in the sense that new participants may join over time, some participants leave, participants can enter into contracts with each other and then dynamically change the provisions of those contracts, and new IoT hardware and software can be deployed and/or removed. All of these changes can

occur quickly, simultaneously and in isolation of other changes that may be taking place in the platform. For these types of systems, a research methodology relying on complex adaptive system theory has been found relevant [e.g., 15, 16]. Given the dynamics of smart city fog platforms, it follows that there is no single governing equation or rule that controls the system. Instead, there are many interacting parts with little central control. There is an unpredictable future trajectory of a multitenant fog computing platform given the nonlinear variety of changes and evolutions it may go through. The fog platform changes and adapts in accordance with changes that occur from one snapshot to the next. It is akin to a moving target with continuously changing equilibrium states.

Using a complex adaptive systems research methodology typically requires construction of a system simulator that allows the researcher to conduct experiments by manipulating control variables, observing behavior and modifying variables to derive new conclusions. While it may be straightforward to define the rules of behavior for participants in a complex system, it is not obvious what patterns might emerge when individual participants follow those rules over time. Thus as a complex system adapts, patterns may or may not emerge that provide evidence of the holistic behavior of the system.

In the case of data and analytics ownership contracts, a complex adaptive systems methodology will necessarily involve experimentation, observation of emerging patterns, variable manipulation and additional experimentation. The following section describes an example experiment.

6. Example Experiment

Consistent with the complex adaptive systems methodology, we constructed a simulator that elaborates the fog platform abstractions of Section 3's Smart City system model. The simulator supports experimental observance of fog platform executions by facilitating manipulations of a large number of variables. It also supports the establishment of different sets of dependent and independent variables. For example, the simulator supports varying the topologies created for a specific Smart City initiative abstraction. Topology can be described in terms of the Hosts (H), Smart Object Hosts (SOH) and Smart Objects (SO). A particular topology might start with a stochastically generated number (within a given distribution) of H, SOH and SO entities. Each topology forms a graph with edges connecting

different H entities with SOH entities to indicate there is a contract for engaging in a Smart City initiative. SO entities are connected to SOH entities to indicate an ownership relationship, i.e., an SOH owns an SO. A topology usually has a larger number of smart objects (SO) than it has smart object hosts (SOH). For example, there may be many stores (following the Intel AIM Suite scenario, above) – or many Hs. There may be one district proprietor (SOH) who purchased all of the kiosks (SOs) to place throughout the district. Since there are kiosks (SOs) at each store in the district, the topology is a graph connecting H to SOH and SOH to each of the SOs. Note that if the number of SOs is larger than the number of SOHs, then the topology indicates that there are many owners for each SO.

Other variables include the number of scripts that are executed in a single simulator run, and there is a variable for managing the total number of simulator runs for a particular experiment. Actions in a script might include adding or deleting hosts (Hs), adding or deleting smart object hosts (SOHs), adding or deleting smart objects (SOs), adding or deleting analytical models (AMs), etc. Scripts are simulations of events that occur in the Smart City service system, and they are serially executed primitive operations within the simulator. Scripts have actions that are stochastically determined. In the experiment described below, many of the independent variables are manipulated while two dependent variables are observed: in-fog communication cost and fog-to-cloud communication cost. Table One lists the variables included in the current version of the simulator.

Topology (Number of H, SOH an SO entities and the contract/ownership links between them)
In-fog messaging and data access costs
Fog-to-Cloud messaging and data access costs
Number of scripts executed per simulator run
Count of topology simulations for a given topology in a run
Number of Hs added or deleted in a run (add and delete scripts are randomly generated according to a given distribution)
Number of SOHs added or deleted in a run (add or delete scripts are randomly generated in a run according to a given distribution)
Number of SOs added or deleted in a run (add or delete scripts are randomly generated in a run according to a given distribution)
Number of Analytical Models (AMs) added, deleted or updated in run
Contract Type: Fixed (flat fees for a level of service) vs. Value (parties share in value that the relationship generates)

Table One: Simulator variables

The main benefit of the simulator is that it can be configured to examine a particular contract type. For example, selecting from the provision options in Section 4, above yields a contract instance. At this point, the simulator is constructed such that all of the dyads in a particular topology agree to the same contract (e.g., all H and SOH entities in a topology

agree to a common set of contract provisions). In other words, all H to SOH contract provisions are fixed for a run. A run means the simulator will: REPEAT 1) Generate a topology, 2) REPEAT 3) Generate a script, 4) Serially execute the script and accumulate the in-fog communication costs and the fog-to-cloud communication costs (currently counted as a cost of 1 per link followed in the topology), UNTIL the required number of scripts are generated, UNTIL the number of topologies generated are robust enough to draw general conclusions from the run.

In-fog communication costs are those related to sending messages from SOH to SO (and vice-versa) to reflect the actions required in executing a script. In contrast, H to SOH communications are from the cloud (where the datacenters belonging to each H are presumed to exist) to SOHs and from SOHs (in the fog) to Hs (in the cloud), i.e., cloud-to-fog and fog-to-cloud – we refer to both cases as fog-to-cloud) to reflect the actions required in executing a script. We modeled the databases stored at Hs, SOHs and SOs – and the database model remains consistent across cases investigated in the following experiment.

In this experiment, we considered two specific sets of contract provisions consistent with the discussion in Section 1, above, regarding tensions between open and more proprietary Smart City services’ fog computing platform initiatives. We assume the context is one where all parties are entered into the same fixed-based contracts for each case examined using the simulator.

Case 1: The first case we investigate assumes analytical model exclusivity, open data, no co-mingling, no analytical model post-use. Basically, this context is one where the ideals of open data proponents enable entrepreneurs to create new services (via proprietary analytical models) that will generate revenue for innovators, but where all data remains open - yet upon exit, an entrepreneurial partner can remove owned analytical models from the fog computing platform.

Case 2: The second case we investigate assumes analytical model non-exclusivity, open data, co-mingling, and analytical model post-use. In this scenario, there is an even more open environment assumption since analytical models have shared ownership and all data is open.

To examine differences using the simulator, we established distributions for generating random topologies. Basically, we used the same distributions for Hs and SOHs, but we increased the size of the SO distributions to reflect that there will likely be many sensors associated with a Smart City initiative:

H/SOH Distributions

Low: 5 to 50
Medium: 51 to 200
High: 201 to 500

SO Distributions

Low: 10 to 100
Medium: 101 to 500
High: 501 to 1000

We refer to the topologies generated in the following by using a three letter acronym associated with abbreviations of Low, Medium and High. For example, the abbreviation LLL implies a topology where the number of Hs is low, the number of SOHs is low and the number of SOs is low. In this way, we ensured we could observe that a simulator run produced sufficient observations for each topology for a varying number of scripts. We varied the number of scripts in a similar low, medium and high manner. The low number of scripts is 20, the medium is 50 and the high is 100. The intent is to use the simulator to examine a robust mix of scripts as well as a robust set of topologies.

For Case 1 with 100 scripts, Table 2 shows the results of the simulator. Note that in Case 1, there are considerably more in-fog costs than fog-to-cloud costs for all topologies. This finding was consistent across the 20 and 50 script size runs. Table 3 shows information about the variables we sought to keep consistent across all runs and script executions in order to isolate the in-fog and fog-to-cloud cost variations only. The small standard deviations of the means across runs are indicative that these variables were kept within suitable range.

Topology	In-Fog Costs	Fog-to-Cloud Costs	Count
LLL	1,418,901,602.00	1,322,350.38	293
LLM	1,538,680,994.00	1,432,207.60	266
LLH	1,614,220,422.00	1,502,766.21	225
LML	1,470,203,498.00	1,368,115.69	236
LMM	1,582,192,066.00	1,472,875.60	276
LMH	1,613,721,673.00	1,502,277.50	233
LHL	1,638,695,953.00	1,544,229.20	245
LHM	1,620,924,447.00	1,507,126.10	272
LHH	1,569,527,453.00	1,460,256.30	190
MLL	1,578,840,869.00	1,469,341.75	258
MLM	1,641,345,498.00	1,526,573.50	249
MLH	1,583,851,551.00	1,472,606.90	210
MML	1,680,872,818.00	1,563,652.80	229
MMM	1,668,962,464.00	1,552,893.90	234
MMH	1,486,440,631.00	1,383,571.07	173
MHL	1,546,653,065.00	1,440,613.29	259
MHM	1,506,906,948.00	1,401,363.67	242
MHH	1,564,982,110.00	1,456,073.47	194
HLL	1,577,668,984.00	1,468,729.05	265
HLM	1,582,113,627.00	1,472,909.88	275
HLH	1,636,453,909.00	1,522,136.13	199
HML	1,572,133,471.00	1,463,513.17	265
HMM	1,701,662,490.00	1,582,938.81	241
HMH	1,533,345,334.00	1,427,737.19	182
HHL	1,574,900,090.00	1,765,919.21	235
HHM	1,578,362,780.00	1,468,635.34	234
HHH	1,726,634,953.00	1,603,335.39	158

Table 2: Case 1 - 100 scripts – dependent variables

CASE 1	20 Scripts		50 Scripts		100 Scripts	
	Mean Of Mean	SD Of Mean	Mean Of Mean	SD Of Mean	Mean Of Mean	SD Of Mean
Avg #H added	9,795	313	17,572	626	49,666	2,159
Avg #H deleted	9,779	317	17,565	629	49,961	2,170
Avg #SOH added	9,808	324	17,514	625	49,758	2,172
Avg #SOH deleted	19,519	631	34,820	1,244	98,398	4,280
Avg #SO deleted	9,762	307	17,403	622	50,048	2,174
Avg AM added	3,224	104	5,815	209	16,694	723
Avg AM deleted	3,217	91	5,812	210	16,694	723

Table 3: Case 1 – 100 scripts – control variables

For Case 2 with 100 scripts, the results of the simulator are shown in Table 4. Compared to case 1, the Fog-to-Cloud costs are substantially higher, but the in-fog costs are fairly consistent. This appears to be a surprising result since Case 2 is presumably a more open Smart City context than Case 1 – analytical models are non-exclusive, data is open, there is no Post-use limitation. Why would the fog to cloud costs be so substantially higher for a more open environment when contrasted with a more proprietary, closed environment? Note that the reason isn't due to some variation in the control variables for Case 2 as shown in Table 6. The standard deviations are quite low.

Topology	In-Fog Costs	Fog-to-Cloud Costs	Count
LLL	1,435,689,941.53	61,961,562.09	303
LLM	1,515,670,584.67	65,384,928.09	278
LLH	1,481,367,365.68	63,936,550.59	235
LML	1,480,691,085.18	63,884,041.67	256
LMM	1,579,861,442.77	68,132,964.47	249
LMH	1,494,748,210.50	64,503,463.77	242
LHL	1,575,744,830.25	67,943,358.97	280
LHM	1,477,016,530.93	63,716,149.85	260
LHH	1,472,672,706.10	63,557,916.19	185
MLL	1,447,297,831.08	62,473,617.45	264
MLM	1,508,190,762.60	65,002,924.60	258
MLH	1,482,376,442.15	63,962,473.70	232
MML	1,448,178,701.00	62,496,678.75	237
MMM	1,451,499,536.20	62,636,967.48	241
MMH	1,634,016,700.78	70,418,365.52	201
MHL	1,555,942,572.18	67,080,782.60	224
MHM	1,631,792,201.34	70,381,041.20	209
MHH	1,491,970,865.30	64,354,923.21	200
HLL	1,548,963,157.15	66,825,366.92	253
HLM	1,522,050,702.91	65,625,290.48	227
HLH	1,369,411,272.57	59,118,090.50	174
HML	1,449,491,552.14	62,542,104.78	230
HMM	1,414,640,491.44	61,042,665.31	232
HMH	1,538,028,657.29	66,301,070.94	178
HHL	1,531,807,089.06	66,090,009.35	234
HHM	1,477,779,161.17	63,756,352.70	220
HHH	1,564,282,284.76	67,447,879.74	176

Table 4: Case 1 - 100 scripts – dependent variables

CASE 2	20 Scripts		50 Scripts		100 Scripts	
	Mean Of Mean	SD Of Mean	Mean Of Mean	SD Of Mean	Mean Of Mean	SD Of Mean
Avg #H added	8,866	386	22,470	941	45,527	1,448
Avg #H deleted	9,008	386	22,567	946	45,025	1,438
Avg #SOH added	8,790	381	22,402	942	45,086	1,448
Avg #SOH deleted	17,697	763	44,646	1,877	89,071	2,861
Avg #SO added	8,888	388	22,506	949	45,122	1,448
Avg #SO deleted	8,798	385	22,703	945	45,302	1,452
Avg AM added	2,963	128	7,531	316	15,049	480
Avg AM deleted	2,963	128	7,531	316	15,049	480

Table 5: Case 1 – 100 scripts – control variables

We find that the reason Case 2 requires more fog-to-cloud communication is because that in order to have non-exclusive analytical models, there is a need to communicate all model changes to Hosts. In other words, the open environment in this experiment leads to a high cost for fog-to-cloud communication. Democratization of analytics and models comes at a cost.

7. Overall Findings

Table 6 summarizes findings from thousands of experiments that we conducted. These can be viewed as starting points for elaborating more sophisticated decision models by applying actual smart city infrastructure cost estimates and other financial parameters. For findings I - IX, the contract type

assumption for the finding is given. Next, the contract’s provision and the perspective relevant to the finding are stated. Timing refers to when a participant may opt to enter into the smart city dynamic, and partnering refers to the nature of the relationships among stakeholders. Finally, findings are shown relevant to the given scenario.

Row I refers to the experiment discussed in section 6 of this paper. Here, the contract was fixed, data and analytics are open, and the perspective is from the vantage point of the smart city fog platform owner. Timing doesn’t matter in this case, and the partnering relationship is one of open sharing. The important finding here is that the platform owner will need an infrastructure with a larger capacity, all other things being equal, in contrast to an architecture where there contract provisions are proprietary. Democratization of analytics and data comes at a cost to the infrastructure owner.

In Row II, taken from a participant’s perspective, if the partner is an early entrant into the smart city infrastructure, and that participant sustains for a relatively long period of time, then under the data post-use paradigm, that participant can anticipate co-owning a large percentage of the fog platforms ingested data during that time period. In fact, the average across thousands of simulation runs was that the participant would co-own 73% of that ingested data.

Row III is similar to the case of Row II, however, here we found that an early entrant in a data post-use contracting environment that partners with a another early entrant that operates in a complementary area can together co-own even more data than in the scenario of Row II. In Row IV, we consider the scenario where there are no data post-use contract provisions. In this scenario, a participant should not exchange any value for a post-use contract with another participant because the data ingested into the whole system will likely be fairly quickly flushed. The scenario of Row V. is similar – it indicates that trying to negotiate for data post-use in a fog platform instance where no other parties are sharing data post-use is not a good option.

In scenario VI, we consider a late entrant who may target a long-term, early entrant with a data sharing agreement, and we find the late entrant can indeed benefit from that relationship. In this scenario, if the late entrant picks the correct partner, the averages we observed were that 50% of the fog platforms ingested data may well be shared.

Scenario VII addresses the context where participants have exited in an environment characterized by data non-exclusivity and analytic post-use. The fog platform owner will require some

sort of garbage collection mechanism that executes on a regular basis in order to avoid unnecessary infrastructure costs.

In Row VIII, where there are analytics co-mingling and data non-exclusivity, predictive model updates can become an issue. We observed highly bi-modal distributions in communication and coordination costs for predictive model updates. In instances where update latency is significant, it will be very important to have large communication pipes.

Finally, in Row IX, we found throughout our experiments that early performance results were not reliable; the simulation almost always took significant time to stabilize. This means that fog platform owner should not rely too much on early system performance results in taking measures to improve the infrastructure performance.

8. Conclusion

The purpose of this research was to investigate organizational data and analytics contracting in smart city fog platforms. We identified and considered different contract provisions including analytical model and data co-mingling/no co-mingling, post-use and no post-use and exclusivity/non-exclusivity. We also consider fee-based and value-based contract types. Using complex adaptive systems methods and extensive experimentation, we identified emergent patterns that can guide stakeholders negotiating preferred contract positions. Stakeholders include both smart city fog platform owners and participants in these multitenant platforms. Our future work requires improving the simulator so that multiple contract types and provisions can be mixed in hybrid contexts rather than be held consistent for each experiment. In addition, we plan to consider additional contract provisions and more real-life scenarios. Finally, we plan to leverage our simulation results to construct analytical models that include financial/cost/revenue detail.

9. References

[1] Bonomi, Flavio, et al. "Fog computing and its role in the internet of things." *Proceedings of the first edition of the MCC workshop on Mobile cloud computing*. ACM, 2012.

- [2] Chourabi, Hafedh, et al. "Understanding smart cities: An integrative framework." *System Science (HICSS), 2012 45th Hawaii International Conference on*. IEEE, 2012.
- [3] Ojo, Adegboyega, Edward Curry, and Fatemeh Ahmadi Zeleti. "A Tale of Open Data Innovations in Five Smart Cities." *System Sciences (HICSS), 2015 48th Hawaii International Conference on*. IEEE, 2015.
- [4] James, L. "Defining Open Data," Retrieved 6/10/2016 from <https://blog.okfn.org/2013/10/03/defining-open-data/>
- [5] Open Data Institute, "A smart city is an open city," Retrieved 6/10/2016 from <https://theodi.org/smart-cities>
- [6] Van Alstyne, M., Brynjolfsson, E., Madnick, S. Why not one big database? Principles for data ownership (1995) *Decision Support Systems*, 15 (4), pp. 267-284.
- [7] Barns, S., "Mine your data: open data, digital strategies and entrepreneurial governance by code," *Urban Geography*, 37:4, 554-571.
- [8] Khanna, P. "The rise of hybrid governance," McKinsey & Company, Retrieved 6/10/2016 from <http://www.mckinsey.com/industries/public-sector/our-insights/the-rise-of-hybrid-governance>
- [9] Trung, H.L. and S. Dustdar, "Principles for Engineering IoT Cloud Systems," *IEEE Cloud Computing*, March/April 2015, 68-76.
- [10] Walravens, N. and Ballon, P., "The City as a Platform: Exploring the Potential Role(s) of the City in Mobile Service Provision through a Mobile Service Platform Typology," (2011) 10th International Conference on Mobile Business, 60-67.
- [11] Grossman, Sanford J.; Hart, Oliver D. (1986). "The costs and benefits of ownership: A theory of vertical and lateral integration". *Journal of Political Economy* **94**: 691–719.
- [12] Hart, Oliver; Moore, John (1999). "Foundations of Incomplete Contracts". *The Review of Economic Studies* **66** (1): 115–138.
- [13] Intel Aim Suite Analytics, Retrieved on 6/10/2016 from <https://aimsuite.intel.com>.
- [14] Neiger, C. "The Government Reveals Its 3 Biggest Concerns About The Internet of Things, Retrieved on June 6, 2010 from <http://www.fool.com/investing/general/2015/01/10/the-government-reveals-its-3-biggest-concerns-abou.aspx>
- [15] Nordin, F., A. Ravald and J. Mohr, "Capabilities for Managing High-Technology Business Networks," Proceedings of the 31st Annual Conference on Industrial Marketing and Purchasing, Kolding Denmark, August 25-29, 2015.
- [16] Albino, V., L. Fraccascia and I. Giannoccaro, "Exploring the Role of Contracts to Support the Emergence of Self-Organized Industrial Symbiosis Networks: An Agent-Based Simulation Study, *Journal of Clean Production*, 2015.

	<i>Contract Type</i>	<i>Contract's Provisions</i>	<i>Perspective</i>	<i>Timing</i>	<i>Partnering</i>	<i>Findings</i>
I.	Fixed	Data and Analytics Non-Exclusivity, Post-Use	Smart City Platform Owner	All	All Partners Share in Ownership	Open architectures have higher infrastructure capacity costs than more proprietary architectures
II.	Simple	Data Post-Use	Participant	Early Entrant, Sustaining Stakeholder	All Partners enter contracts with post-use provisions	An early entrant will co-own a large amount of the total ingested data
III.	Simple	Data Post-Use	Participant	Early Entrant with a Complementary Partner	Post-Use Contract with Complementary Partner	The two stakeholders will co-own an even larger amount of the total ingested data in comparison to I., above
IV.	Simple	No Data and Analytics Post-Use	Participant	Doesn't Matter	No Data and Analytics Post-Use	Best in Volatile Situations where Data and Analytics will be fairly Quickly Flushed Anyway
V.	Simple	Data and Analytics Post-use	Participant	Doesn't Matter	One Party Wants Data and Analytics Post-Use, but No Others Do	Being a hold-out for Data and Analytics Post-Use Provides Little Advantage Unless Contracting With all Partners Post-use
VI.	Simple or Value	Data and Analytics Sharing with Early Entrant	Participant	Late Entrant	Partnering with an Early, Sustaining Stakeholder that has Used Non-Exclusive Data and Analytics with Co-Mingling	A late entrant can gain advantage by partnering with an early entrant that was openly sharing data and analytics
VII.	Simple or Value	Data Non-Exclusivity and Analytics Post-Use in Contract Mix	Smart City Platform Owner	Later	Case When Partners Have Exited	The ecosystem owner must have a garbage collection capability for data and analytics that are no longer owned by any active partner
VIII.	Any	Analytics Co-Mingling, Data Non-exclusivity	Smart City Platform Owner	Any	Case When Most Partners Co-Mingle Analytics	There can be bi-modal volatility in adaptation costs for predictive model updates; ensure there are big pipes and low latency for communication peaks when necessary
IX.	Any	Any	Smart City Platform Owner	Any	Any	Infrastructure owners should not rely on early performance measures to gauge long term ecosystem performance

Table 6: Summary of Findings from Multiple Experiments