Paper 2653-2018

Organizational Contracts in Analytics of Things Ecosystems

Michael Goul, W. P. Carey School of Business, Arizona State University

ABSTRACT

The impending convergence of big data, analytics and IoT has been referred to as an era of the Analytics of Things (AoT). Companies working in AoT ecosystems like smart cities and inter-organizational collaborations will build business relationships through contracts related to the sharing (or not) of data and analytics resources. Data and analytics ownership issues are already challenging the legal system, and with the proliferation of AoT, these challenges will only become more pervasive. This research addresses inter-organizational data and analytics ownership contracts. Findings provide guidance for both ecosystem owners and renters of multi-tenant AoT ecosystem resources. While recent research has reported that data and analytics sharing provide a significant innovation advantage, what we know and understand about ownership contracts lags behind. This is an important research area because Chief Analytics/Data Officers will be influential stakeholders in helping their organizations navigate AoT investments.

INTRODUCTION

The fog, Internet of Things (IoT), edge computing and other related phrases have come to denote a major disruption whereby sensors and other end devices are bolstering the transition of data and analytics computing to the logical endpoints of networks. These endpoints are likely to rapidly advance in computing capability in accordance with Moore's Law. Business applications in areas like wearables, connected cars, digital signage and smart agriculture are frequently mentioned. Broader application contexts like smart electric grid, connected health and smart home can involve combinations of publicly owned and proprietary infrastructure ecosystems that fuse IoT with cloud computing, mobile phones, digital signage, etc. For example, investments in Smart Cities have resulted in parastatals - company or agency computing ecosystems owned or controlled wholly or partly by a government. For such parastatals, Smart City ecosystems can be multi-tenant in that they can 'rent out' hardware and software services in order to help society and businesses collaborate to realize solutions to urban problems. Multi-tenant ecosystems have the burden of ensuring that deployed analytics, collected and transmitted data/analytics and other related operations are all conducted under the aegis of contractual obligations. The phrase "Analytics of Things" (AoT) has come to denote a new area of inquiry related to this emerging computing context. Important research problems include:

- What will contract provisions for data and analytics ownership look like, and what will they address?
- For the parastatal that owns an ecosystem, what are the best contract terms for subscribing tenants in order to minimize ownership costs?
- For a new tenant, what is the best negotiation strategy for a contract given the nature of existing contracts between entities in the ecosystem?

The purpose of this research was to discover information about contracts for analytics and data ownership/sharing that address the research problems, above. Findings include:

- Parastatal smart city ecosystem architectures need to be designed to accommodate different data and analytics ownership/sharing contract contexts.
- Value-based contracts require a democratization of data and analytics across participating entities, and this requires more investment in the ecosystem architecture than if all relationships are highly proprietary.
- A tenant that enters early and signs data and analytics sharing contracts with later entrants are shown, on average, to own a very large percentage of the data that enters the entire ecosystem.

CONTRACT PROVISIONS AND ECOSYSTEM MODEL

In this paper, the following data and analytics contract provisions were investigated:

Fixed-price vs. Value-based contracts: In fixed-based contracts, parties agree on a unit of analysis upon which fixed service fees can be assessed. In contrast, value-based contracts refer to a division between the contracted parties of, for example, the total profit generated by all partnering stakeholders.

Data Exclusivity vs. Non-Exclusivity: When data ownership is exclusive, there is a specific entity that owns the data. Data is co-owned or open when contract provisions stipulate that data is non-exclusive.

Analytical Model Exclusivity vs. Non-Exclusivity:
An analytical model (AM) 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 AM 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 AM.

Co-mingling vs. No Co-mingling: In co-mingling, an AM 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 AM 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) in the contract. If there is no data post-use, the data solely owned by an exiting owner is deleted.

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

To investigate these contract provisions, an ecosystem model was constructed and a simulation was developed to encapsulate the model. Following is information on the ecosystem model. We consider an ecosystem as involving three major abstractions:

1) A set of Hosts: $H = \{H_1, H_2, \ldots H_p\}$

2) A set of Smart Object Hosts (or Hubs): $SOH = \{SOH_1, SOH_2, \ldots SOH_q\}$

3) A set of Smart Objects: $SO_{SOHi} = \{SO_{SOHi,1}, SO_{SOHi,2}, \ldots SO_{SOHi,m}\}$

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:

4) $S = \{S_{SOi,1}, S_{SOi,2}, \ldots S_{SOi,n}\}$

A sense can be a capability to ingest and store 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 AM that may take sense values as inputs and produce sense values as outputs. We consider an AM as being deployed by a Host through a Smart Object Host to a Smart Object. We describe an AM 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 (AMs) 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

5) $AM(f(Subset(SO_{SOHj,q},S_a, SO_{SOHj,q},S_{b, \ldots}\ SO_{SOHj,q},S_t)) \rightarrow (Subset(SO_{SOHj,q},S_a, SO_{SOHj,q},S_{b, \ldots}\ SO_{SOHj,q},S_t))$
A topology for an instance of an ecosystem therefore has a set of interconnected Hosts, Smart Object Hosts and Smart Objects connected to Smart Object Hosts. Smart Object Hosts and Hosts can store data permanently (e.g., in a local database, in the cloud, etc.). Data can also be stored in a sharable area (e.g., shared space in the cloud). A topology is also characterized by connections between Hosts and Smart Object Hosts to indicate business relationships representing manifestations of the existence of contracts. From basic constructs, the abstractions above and the contract provisions discussed, complex topologies can be realized in the ecosystem model. For example, if a college football stadium has sensors for decibels, humidity, video feeds, etc. (Smart Objects) stationed around a stadium, and clusters of those sensors are connected to hubs where data is collected (Smart Object Hosts), then analytical models may be sent by the hub to the decibel and video sensors (Smart Object Hosts to the Smart Objects) in order, for example, to display the loudest cheering section on the jumbo screen. Companies may purchase rights to leverage the stadium infrastructure to advertise and sell goods, e.g., a soda company (Host) collects humidity and sensor data to be used in AMs to offer a deal at the concession stand for a soda special. A topology for game day will feature different company participants, different analytical models that may be deployed, etc. Contracts between the company and other companies as well as the stadium owner will dictate such characteristics as whether profit at the concession stands is to be split between participants or it is solely to the benefit of the concession stand owner. On game day, there may be many tenants renting infrastructure to deploy AMs for their business purposes. At the end of the game, the tenants leave, and the infrastructure goes dormant until the next event.

Figure 1 is a depiction of the simulation used to investigate different contract provisions using the ecosystem model. The simulation generates random topologies for a contract configuration and events that are dispatched to be executed against the topology. Global variables tabulate coordination costs (messages to enact a contract by sending and storing data in distributed databases as designated by those contracts) and communication costs such as when an event requires a topology change through H, SOH or SO deletion, addition, AM revision, etc. A variety of scripts are executed against the topology with each script characterized by a set of pseudo-random global functions. Experiments were conducted under different topologies, e.g., H/SOH distributions were Low (5 to 50), Medium (51 5o 100) and High (201 to 500). The number of SO distributions varied from Low (10 to 100), Medium (101 to 500) and High (501 to 1000). An instance of a topology can therefore be categorized as LLL or a Low number of Hs, a Low number of SOHs and a Low number of SOs. All combinations of distributions were tested in experiments that investigated a fixed set of contract provisions. Thousands of experiments were conducted to determine findings.

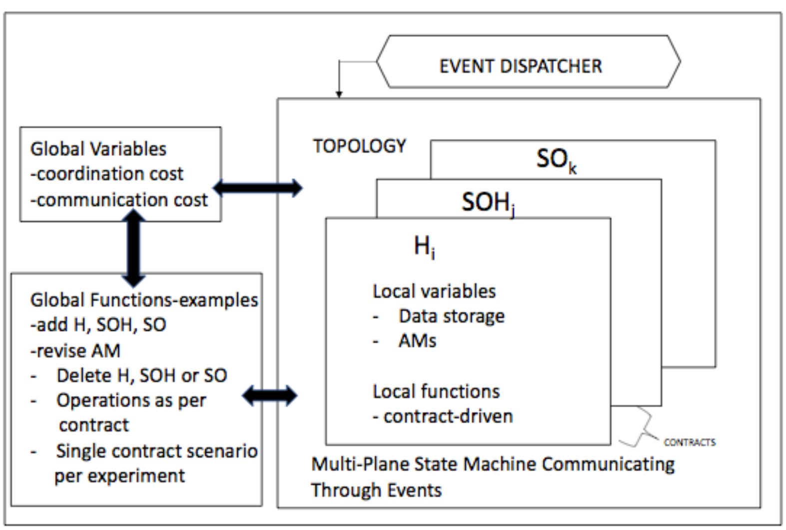

Figure 1. Depiction of Simulation

DISCUSSION AND FINDINGS

Thousands of experiments were conducted for different ecosystem contract arrangements. Each contract arrangement was executed against scripts of different complexities (i.e., 20, 50 and 100 events in pseudo-random scripts that added/deleted hosts/smart object hosts/smart objects, analytical models, data, updated models, etc. As above, a variety of topologies were subjected to the scripts in a contract arrangement to assess that contract arrangement and to enable it to be compared with another contract arrangements. Figure 2 shows the result of one contract arrangement and exemplifies the summary data collected for the variety of topologies under consideration. The communication and coordination costs were calculated to then be compared with other simulation runs in order to draw conclusions through comparing contract arrangements. The 'count' variable indicates the number of times a particular topology was generated as part of a simulation run.

Topology	Communication	Coordination	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

Figure 2 Example Execution of the Simulation

We have eight findings from our simulations to report. Finding One refers to the context where the contract is simple, data and analytics are openly shared, and the perspective is from the vantage point of the ecosystem owner. The important finding here is that the owner needs an infrastructure with a larger capacity, all other things being equal, in contrast to an architecture where contract provisions are proprietary. Democratization (complete sharing) of analytics and data comes at a cost to the ecosystem owner. That additional cost may need to be passed on to tenants.

Finding Two is relevant from a participating stakeholder's perspective (e.g., a tenant). If the tenant is an early entrant into an ecosystem, and that participant sustains for a relatively long period of time, then (under carefully managed data post-use contracts) that participant can anticipate co-owning a large percentage of the ecosystem's ingested data during that time period.

Finding Three is similar to Finding Two except that an early entrant in an ecosystem with data post-use partners with another early entrant that operates in a complementary business area. Together they can co-own even more of the ecosystem's ingested data than in the scenario of Finding One. Finding Four focuses on 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 ecosystem will likely be fairly quickly flushed. The scenario of Finding Five is similar: it indicates that trying to negotiate for data post-use in an ecosystem where no other parties are sharing data post-use is not a good option.

In Finding Six, the focus is on a late entrant who may target a long-term, early entrant for a data sharing agreement. The simulation study shows that the late entrant can indeed benefit from that relationship. Finding Seven addresses the situation where participants have exited an AoT ecosystem characterized by data non-exclusivity and analytics post-use. The ecosystem owner will require some sort of garbage collection mechanism that executes on a regular basis in order to avoid unnecessary infrastructure costs associated with data and analytical model storage/maintenance.

Finding Eight focuses on an ecosystem where there are predictive analytics co-mingling and data non-exclusivity. Here, analytical model updates can become an issue. The research resulted in highly bi-modal distributions for communication and coordination costs for analytical model updates. In instances where update latency is significant, it will be very important to have large communication pipes between ecosystem participants in this contract scenario.

CONCLUSION

The emergence of AoT combines complex analytics and IoT highlighting the need for new research on contractual instruments that will shape data and analytics sharing in order to foster inter-organizational innovation. The notion of multi-tenant ecosystems can serve as a lens that focuses on the business aspects of complex AoT. Contracts will be the glue between stakeholders in AoT ecosystems.

This report's guidance on contracts in AoT contexts considers different ecosystem perspectives, e.g., the owner of the ecosystem as well as participants. Considerations address data and analytics sharing, the nature of ecosystem investments, and different contract strategies based on timing.

The Chief Analytics/Data Officer's (CA/DO) role in AoT deployment is crucial beginning with the early stages of design and continuing through data and analytics sharing contract negotiation. After contracts have been decided, CA/DOs should be actively involved to ensure that contract provisions are adhered to. As discussed above, AoT contracts and ecosystem investments will be linked as certain contract provisions imply the need for certain infrastructure capacities and networking capabilities.

More research is needed to provide guidance to CA/DOs regarding contract negotiation issues surrounding data and analytics ownership. This research provides first steps. Since industry is still in the initial stages of analytics and IoT convergence, there is some lead-time. That said, companies who are early movers are already stipulating contract provisions for prospective partnership models for AoT. Companies such as Axon (formerly Taser) have introduced body cameras for police officers, thereby including cloud providers and judicial systems. All Traffic Solutions has advanced sensors and wireless communications to usher in a new era of traffic management, thereby including cloud providers and hardware vendors. These early movers will be watched closely; some may fail and others could succeed beyond expectations. With business and competitive shifts underlying these ventures, AoT is likely to be a major disrupter for a long time to come.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Michael Goul
W. P. Carey School of Business, Arizona State University

Michael.Goul@asu.edu
https://wpcarey.asu.edu/people/profile/67283