

Open Government Data Ecosystems: A Closed-Loop Perspective

Mahdi M. Najafabadi

University at Albany

mahdi@alumni.albany.edu

Luis F. Luna-Reyes

University at Albany

lluna-reyes@albany.edu

Abstract

Open data initiatives have opened new alternatives in creating benefits for the public through secondary use governmental data. From some perspectives, benefits will come from the development of innovative applications using the data, and from the new business models enabled by these applications. From other perspectives, open data applications offer an opportunity for increased citizen participation, improved transparency and accountability. Although the number of published governmental datasets has increased in many countries, producing the expected benefits – and even measuring them – has proven difficult. Creating the expected benefits depends on the development of an ‘ecosystem’ of government actors and private stakeholders that enables multiple forms of interactions and value creation. We propose that modeling and simulation of this open data ecosystem can expand our understanding of its enablers and barriers, leading to improvements in policy making and ultimate outcome of open data initiatives.

1. Introduction

Opening government information is becoming a public policy adopted by governments around the world. Most democratic societies recognize the right to access, use, and reuse information produced by the State [28] – with some exceptions for situations in which data disclosure is in conflict with another social value such as individual privacy or national security [25]. In fact, considering the structure and settings of modern societies, some scholars suggest that having access to public information is no longer a privilege but a human right [16].

Although opening public information is not a new phenomenon, it has been recently revitalized through the creation of Open Government Data programs (OGD or ‘open data’ for short), as well as emerging technological and social trends. Recent technological advancements, for example, have created the opportunity for storing and sharing large amounts of data in open and re-usable formats. On the other hand, social trends on open data have been promoted at least in part by the US Federal Government Initiative, expressed in president Obama’s memorandum on his

first day in office in order to have a transparent, participatory, and collaborative government [18]. This initial initiative has been followed by state-level and city-level initiatives of open data to unleash an innovation potential for economic development [19]. The main idea behind these US policies is to allow entrepreneurs to create value from government data by developing new businesses around the data assets in a crowdsourced manner. At the international level, the trend can be observed in the Open Government Partnership (OGP), which started in 2011 with 8 participating countries, and contains 69 countries as of mid-2016 (see <http://www.opengovpartnership.org>).

Although there are few very successful examples of value creation and innovation through open data [22], the current status of most open data efforts and the information about the applications of available datasets indicate that creating value through open government data is not an easy task. In recent years, the number of government datasets that have been made available to the public has grown rapidly, but the number of applications, their use, and consequently the value realized by the society has grown in a much slower pace. The United States’ official open data portal data.gov, for example, features over 183,000 datasets, but less than 100 applications using these datasets, as of June 2016. This gap between the number of datasets released and the number of applications signals some barriers in unleashing the full potential of opening data.

Some scholars have focused on the risks, challenges, and impediments on open data effectiveness, and have suggested that it is not sufficient to only make the data freely available and accessible for the public [25,26]. Recently a self-sustaining closed-loop perspective has been emerging in this context, which suggests the notion of an open data ecosystem. In this perspective, a good open data ecosystem will allow effective interaction between actors in the value-creation chain starting from the data sources in the government, all the way down to the end-users, and there is also a relation back to the government which closes the feedback loop [17,8,27,9,6,10].

This paper aims to contribute to this line of thinking by employing computer simulation techniques, particularly system dynamics, to study the structure and characteristics of an open government data ecosystem.

Our goal is to better understand what are the key factors and processes needed to promote an effective and sustainable open data ecosystem. We also aim to gain insights and identify both success and failure modes for open data initiatives.

The paper is organized in six sections including this introduction. The second section constitutes a review of the literature on open data and open data ecosystems. Section three introduces our selected method and initial data. Section four introduces a preliminary simulation model that includes main elements of an open data ecosystem. In section five, we present three simulation scenarios that can be used to explore and understand this preliminary model. The last section of the paper includes a brief discussion of the simulation results and also our future paths of research and model development.

2. Literature Review

Traditionally, citizens' access to governmental data has taken place through a pull process. Data was initially created and recorded as a result of government operations, following the logic of internal processes and internal information use. As expected, data in government systems is designed to serve government internal operations or provision of services. Citizens usually needed to ask for specific information through a formal request, and government offices might (or might not) send data to the citizen that asked for it. Current directions in open data initiatives are more about embedding the technical requirements for openness in the ongoing processes of recording and maintaining the data. This allows the governments to push the datasets into the public sphere near to real time, and make them available for general public's creative applications. Consequently, citizens may take advantage of the information stored by the government in a faster pace, more straightforward, and at a lower cost – i.e. making the government data 'open by default' [26]. However, data in government systems is still prepared to satisfy the needs and uses of internal operations and provision of services. Most frequently, this new trend of maintaining 'born-open' data lacks data curation for new uses, and thus limits the potential value of the data, and might even involve a possibility for conflicting views on new uses [10].

2.1. What is Open Data

At least in theory, opening government data can unleash capacities for generating public value by crowdsourcing the process of value generation to the whole community of citizens [7]. In this view, the number of open data applications, number of users of

the applications, and the value generated by the applications and perceived by the users as well as the general public – should follow the trend in which governmental datasets are made open. Open data ecosystem perspectives are trying to capture the big picture in which all the actors in this context are seen as endogenous entities, and are interacting with each other in a closed loop.

This perspective can effectively illustrate the types of public value that may be generated by open data in the government sector. A number of abstract models have been suggested that depict the relationship between several entities in the open data ecosystem. These conceptual models show how opening data by government may have an effect on the rest of the entities in the whole ecosystem, resulting in several types learning back on the government sector [6]. These models contribute to the understanding of the interactions among different elements in the open data ecosystem as well as their possible consequences. However, conceptual models usually provide less understanding in terms of tracking back consequences of policy choices and their effects on the various entities and also on the ecosystem as a whole. This is where a simulation model can generate more insights about the long-term results and possible side-effects of policy choices and help discover some causes of what is going on in an open data ecosystem.

2.2. Open Data Benefits

Conceptually, open data can be used by anyone in the society for creating value. However, the actual value created by open data may not all be immediately recognized, as opening data as a free resource frequently requires an intermediary to turn that informational asset into something valuable for end-users – mostly, ordinary people in the society. Scholars have found that opening data can create social and economic benefits directly via one or more of the categories below [1]:

1. Releasing public and commercial value, by creating an environment in which the needs of end-users can be identified and addressed in a crowdsourced manner;
2. Accountability of government agencies and public officials as a result of transparency; and
3. Participatory government, which fosters the soul of democracy by giving people a mechanism to take their part in public decision making processes, encouraging people's engagement as they realize how impactful they can be.

The different types of benefits above are generated from different dynamics in the open data ecosystem, thus might be generated and realized in different

timeframes. Also as a side-note, although governments may require more effort at the beginning of their open data initiative, learning processes will lead to more efficient procedures for data governance in governmental agencies as it becomes a routine [21]. As we described before, there may also be a tendency toward embedding the technical and social characteristics in the original data to facilitate the process of making data publicly available. Consequently, crowdsourced use of open data can lead to new insights for government policies in a plethora of issues, enhancing policies toward better results, and creating a greater buy-in for governmental policies. There will also be a learning curve for the society. As more open data applications are used, the people will better know how to take advantage of it [13].

2.3. Challenges and risks

Although there are huge potentials in opening government data, there are also several barriers and impediments in effectiveness of open data initiatives in governments [25,12,4]. Scholars have also studied these factors in addition to some risks that might be generated by opening government data, which should also be considered for improving the effectiveness of open data initiatives. In a recent review of the literature, Zuiderwijk et al. [26] suggest three vast categories for all impediments they have found:

1. Data access impediments, comprised of data availability, access issues and findability problems;
2. Data use impediments, comprised of usability issues, understandability difficulties, quality impairments, linking and combining data challenges, comparability and compatibility issues, and metadata incompleteness or irrelevance; and
3. Data deposition impediments, comprised of difficulties in interaction with data provider, and challenges in opening and uploading the datasets.

There are also some risks associated with open data efforts in government. In another research project, Zuiderwijk and Janssen [25] conducted a number of interviews with field experts to identify key risks associated to open data efforts. In their research, they identified general risks, as well as risks related to the implementation and management of open data efforts (see Table 1).

There are also other situations in which some actors in the government sector might resist opening data, especially in cases where opening data can threaten their power position. The latter situation can mostly be seen in not-so-democratic governments, where the freedom of press is even at risk, and where in the extreme

situations, the only operational information sources are those owned and/or operated by the government.

Table 1. Risks associated to open data efforts in various levels [25].

Open data efforts	<ol style="list-style-type: none"> 1. Risk of violating legislation by opening data 2. Difficulties with data ownership 3. Privacy can be violated unintentionally 4. Published data can be biased 5. Misinterpretation and misuse 6. Negative consequences of transparency 7. Open data may have negative consequences for the government 8. Decisions made on poor information quality 9. Timeliness: embargo period prohibits the publication of recent data
Implementation of open data efforts	<ol style="list-style-type: none"> 10. Opening data as an afterthought 11. Little attention for public value and solving societal problems 12. Unclear responsibility and accountability 13. Not citizens but others profit from open data
Management of open data efforts	<ol style="list-style-type: none"> 14. Wasting resources to publish invaluable data 15. No priority given to data publication 16. Limited information about data publication policy published

2.4. Open data ecosystem

The main idea of the ecosystem perspective is that the interoperable entities that create the open data environment are forming a closed-loop system. Thus, instead of a one-way perspective to the open data, as the society and the government communicate, the benefits for both government and society can be leveraged [8]. This ecosystem is comprised by the data that is made available to the public, the government agencies that made it available, and also captures a feedback from society in response to opening governmental data [8]. As it is shown in Figure 1, the chain of benefits corresponding to the flow of data in this ecosystem starts from the datasets already residing in governmental databases, feeds into the applications developed by governmental or non-governmental developers, and transforms into benefits when the applications are used by the end-users in society. In the ecosystem perspective, the benefits generated in society reinforces open data initiatives for more value creation as a delayed and secondary effect [17,8,27,9,6,10].

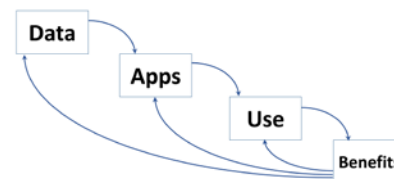


Figure 1. Abstract layers of open data ecosystem and the feedback mechanisms.

Consequently, as a result of this feedback mechanism, once the benefits of opening data are perceived by the society, government will be pushed by the ecosystem to open more data. Therefore, the availability of data becomes more of a “core expectation” in the society, and the government is pushed to increase the capacity (i.e. the efficiency and effectiveness) of policies and procedures to open data.

Some conceptual models have been suggested for an open data ecosystem. Helbig et al. [10] studied two specific applications in New York City and the City of Edmonton. Dawes et al. [6] studied the cases of New York City and the City of St. Petersburg. These cases show the importance of all elements in Figure 1 in the creation of value. The quality of both data and meta-data, for example, are key elements in facilitating the development of applications. In fact, one of the developers interviewed by Helbig and her colleagues identified these factors as key in his motivation to developing a mobile app to access information in road work in the city of Edmonton in Canada. On the other hand, opening restaurant inspection data in New York City, without curating the data for new uses to help citizens, generated more conflict than value. Many negotiations among restaurant associations, government agencies and consumer associations had to happen before agreeing on a data format that may be useful for all participants in the ecosystem to obtain value.

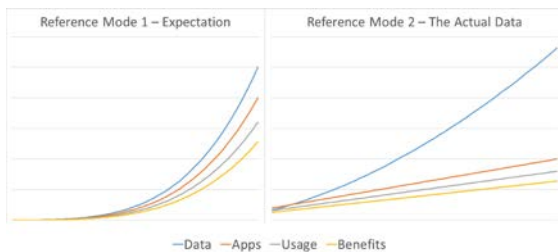


Figure 2. Suggested reference modes for open data ecosystem layers.

Although all reinforcing processes suggested in the literature point to an exponential growth in both the number of datasets and open data application use, as we described previously, it appears that in the studied cases, applications and their use have not followed the trend by which the datasets had been made open (see Figure 2), and thus neither has perceived benefits of open data. In this research, we are looking for insights about the reasons of these patterns of behavior.

3. Methods and data

In our ecosystem perspective to open data, we intend to analyze a closed-loop environment in which several variables are interacting over time and the relation

between several variables suggest several feedback effects. We are building over the work of Helbig and her colleagues [10], who proposed a conceptual stock-and-flow map to represent relationship in the ecosystem on the basis of two case studies. Their proposed map used the grammars of system dynamics, and we are proposing then to continue with this same approach to understand the dynamics of opening data.

3.1. System dynamics as the preferred approach

System Dynamics (SD) is a research method that has strengths in depicting the internal structure of complex closed-loop systems in terms of stocks, flows and feedback loops [20,24]. The method is appropriate to study socio-technical systems such as open data ecosystems given that main elements in conceptualizing dynamic systems represent main capabilities (stocks), activities (flows) and recursive interactions representing both synergies and policy resistance (feedback loops) [2,15]. Moreover, system dynamics has been successfully used to understand digital government phenomena [3,5,14], including the dynamics of open government [23]. Causal relationships codify through mathematical formulations main assumptions about the nature of such relationships, allowing the researchers to better understand the likely effects of their assumptions. Experimenting with small conceptual models has been recognized as a valuable tool in the process of developing public management theories and policies. That is to say, simulation experiments can generate insights about effective policies and some consequences that one might not have considered in the first place. In general, System Dynamics enable us to “experience the long-term side effects of decisions, speed learning, develop our understanding of complex systems, and design strategies for greater success” [24].

3.2. Open Data in the City of Edmonton

Our model builds upon a case study in which the open data ecosystem comprised of governmental agencies, the application developers, and users in the City of Edmonton were studied [10]. In this case, the story presents an opportunity seeking behavior. That is, activities in the case are the result of groups of actors seeking opportunities to advance their own interests.

The City of Edmonton administration started an open data initiative in in early 2009, with the commitment of increasing transparency. This institutional commitment motivated action from the Office of the Chief Information Officer (OCIO) as well as from the agencies owning most of the city’s information assets. The OCIO staff’s actions included

identifying datasets that were candidates for opening to the public and including those sets in an online data catalog (<http://data.edmonton.ca>).

OCIO staff was looking for datasets that were both easy to be made available because of its technical characteristics, but also for committed agencies willing to collaborate with them in the whole technical process. One of those agencies, the Department of Transportation, had a number of datasets that were potential candidates for inclusion in the catalog. Among all candidate datasets, Helbig and her colleagues describe the road construction data as an ideal candidate “because it was a relatively small dataset, simple in structure and easy to render in a more accessible visual form using their new platform, in other words, low hanging fruit.” On the other side, DOT also had a clear idea of the value of opening this dataset for the unit. Moreover, road construction data had also potential value to city residents, who experience travel delays, disruptions to local businesses, noise and other environmental impacts in neighborhoods, as a result of the construction projects. Additionally, construction projects had also impact on activities of other agencies on charge of other services such as water, sewer, and other utilities.

OCIO was in charge of the initial application development, and they proposed the development of an interactive map showing main construction projects. Making data available and updated, in this specific case, does not impose a large workload to DOT because data is mostly static. The decision regarding construction projects is made once a year during the city budget process. It is uncommon to add or cancel projects after the initial plan. However, start and end times for each project are somewhat more dynamic. This specific dataset offers additional opportunities because there was no need to change business process or to modify the data in an important way before it was made available. OCIO staff, however, commented that sometimes much more work and adaptation was needed.

The city website presented construction project information in a single interactive Google-type map interface, including a description of the construction types. Local news media reported on the new application increasing rapid growth in public access. Additionally, a local application developer decided to create a mobile app for smart phones and similar devices to access the map interface. He saw the data as useful for the population and decided to make a contribution to his new home city, not as a commercial venture. In an interview, the local developer reported that the high quality of the data made very easy for him to create the mobile app. The data was available in a standard format, with all metadata needed to support application development.

Helbig and her colleagues [10] report a short lifecycle on this application, with a fast increase in data views and downloads, increasing to a monthly maximum of 1200, declining gradually to less than 250 views per month four months later. OCIO and DOT staff explained the increase and reduction on the basis of the number of construction projects in the city. Among the main benefits for the DOT, staff identified a reduction in costs and public complaints about travel disruptions. OCIO staff identified an increase in their reputation as the main benefit, as well as their learning and capability building for future projects. For citizens, the main source of value was better information to plan their commuting and traveling in the city.

4. The model

In this section of the paper, we introduce a preliminary simulation model. We started the modeling process by creating a high-level causal model of a generic open data ecosystem. We then formulated the relationships and added some details to the structures to allow model formulation through adding actual mathematical formulas for relationships.

4.1. Conceptual model description

Our conceptual model is comprised of three interacting sectors: the government sector (includes agencies in charge of opening data), the developers sector (represents the middle tier which facilitates the use of openly available data for the citizens), and the society (includes citizens who take advantage of the applications and realize benefits from open data).

The government sector is a key determinant in the open data ecosystem. The top layer of Figure 3 depicts the dynamics in the government sector. Just like in the case of Edmonton, we have assumed that there will be a capacity for the government sector in opening data that allows processing and opening datasets to the public.

Thus, the open data capacity (the top stock represented by a rectangle), sets the rate of which the government opens its datasets (“Publishing Govt. Datasets” – the valve in between “Govt. Datasets” and “Published Govt. Datasets”). This can be thought of as a representation of the governmental resources allocated for its open data program – in our case, this is the OCIO, DOT, and other agencies’ staff dedicated or assigned to work on opening datasets to the public.

Our current conceptualization assumes that resources are limited, and that resources allocated to OGD projects programs mostly depend on institutional

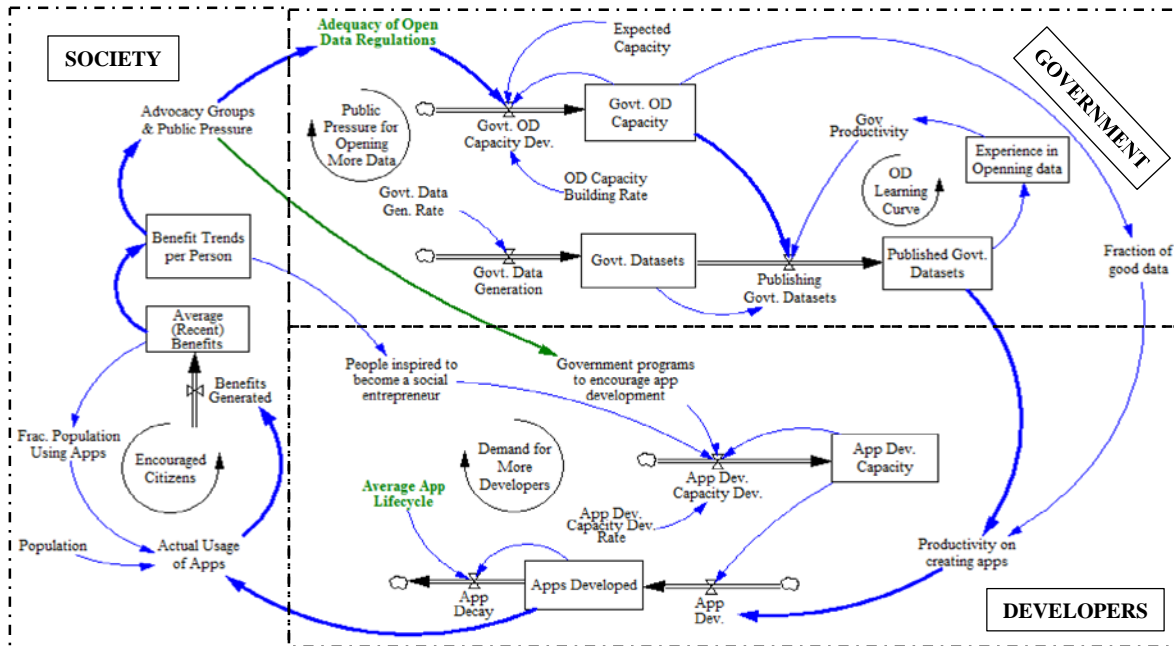


Figure 3. OGD ecosystem's main sectors and feedback processes.

frameworks – i.e. regulations or administrative initiatives like in the Edmonton case, as well as appropriate data management routines to support born-open data or otherwise catalyze opening data down the road. In the long run, such institutional frameworks may be modified as a result of a change in public pressure advocating for open data – they might get better or worse. The idea here is that procedures should be kept up to date to remain effective, and that updating data and processes require resources in an ongoing basis. In an equilibria situation, the need to update the procedures and the rate of which they are kept up to date are equal, and thus the effect remains constant. However, once the motivation for opening data exceeds or fall below that level, it can affect the governments capacity of opening data. We have used a goal-gap structure to simulate raises in the capacity over time according to the capacity-building rate, up to the time it reaches a predefined desired – cap – level.

Given its relationship to institutional frameworks and its effects on opening data, government capability is also a main determinant of the data quality in the opened datasets – represented as “Fraction of good data,” which represents the number of datasets that, like the dataset on read construction in Edmonton, have the quality and value to be open to the public. This variable will in turn have a high impact on the quality of developed applications, assuming that high quality data will lead to high quality applications (through “Productivity” of the applications, in the Developers sector, at the bottom of Figure 3). Just like the developer in Edmonton

commented, the quality of the data made easy for him to develop a mobile app. In this preliminary model, we have decided to leave out of the model boundaries the details of data creation, and assumed that the government is generating new – openable – datasets on a flat rate. We have also considered a learning-curve effect in opening governmental data, just like OCIO staff reported in the Edmonton case. In this way, the more data the government releases, the more it becomes productive in releasing data and thus, more data can be released. However, as it is suggested in the literature, that effect is mostly visible only in the very initial stages of the process, which is reflected on our preliminary model.

In the Developers sector, once again we have used a capacity stock with the same philosophy. It represents the availability of open data application developers. This capacity is one of the determinants of the application development rate. Some of this development capacity resides inside government, like OCIO in the Edmonton case. Some other capacity comes from local developers that get inspired to develop apps just like it was also illustrated in the case.

Just like in the case, people can get inspired to become open data application developers if there are expected benefits associated with the application (on average), and also in case the benefits of using open data applications is realized in the society (the loop in part in the Developers section in Figure 3 shows this chain of consequences). Thus, in case people realize there are still more benefits to be generated by more applications,

they are motivated to become application developers. Government initiatives will also accelerate this process. We have shown this relation as a green arrow in Figure 3.

On the other hand, as discussed earlier, “Productivity on creating apps” is influenced by the quality of datasets. That is to say, it is easier for a developer or any user in general to make sense of data that is carefully cleaned and that includes all necessary metadata to make sense out of its primary context. We have also assumed that there will be a decay rate for the applications since most of these applications require maintenance to remain functional.

Finally, the left side of Figure 3 illustrates a very abstract level of use and realization of benefits by society. This structure contains a reinforcing feedback loop indicating that the more benefits perceived in the society from using open data applications, the more fraction of people will become users of such applications. “Actual Usage of Apps” is represented by the number of downloads and in addition to the “Population” and “Frac. Population Using Apps,” there is another feedback into it from the “Developers” sector: Of course, potential use is constrained by the number of active applications available to the users. Although we recognize that this the number of downloads is a limited view of benefits from open data, we decided to use this metric in this preliminary version of the model.

The largest feedback effect goes from the society back to the government, through advocacy groups and public pressure. Public pressure from advocacy groups will have a feedback effect on the resources that the government has allocated both for encouraging open data application development, and opening datasets at agencies’ level. This pressure is positive in cases such as in Edmonton, where the local media promoted the use of open data increasing both the impact of initial datasets and the motivation of government agencies to continue opening data. In Figure 3, this full feedback process is presented with bold blue arrows.

4.2. Model formulation

Formulating a conceptual system dynamics model requires adding more details – i.e. more variables – and making some assumptions about the range of meaningful possible values for some inputs. The next step would be assigning mathematical formulas to relate different stock, rate, and auxiliary variable, and set initial values of the model. We used the qualitative data from the Edmonton case to formulate the relations in the formal model, and we run extensive sensitivity tests to build confidence in our current preliminary formulation. Following common system dynamics practices, we also used some common formulations from project-like

models and molecules [11]. For some part of the society sector, for example, we are using the “Trend” molecule as the basis of connecting benefits generated from the number of downloads, and the “Perceived Benefits” by the society. The “Benefit Trends per Person” variable will be the inspirational motive for people to realize if they are willing to become application developers in the open data context. We used the trending structure because the perceived benefits in the society does not encourage people forever. Excitement from realized benefits will be damped through time, and unless there are new benefits generated, the society will no longer be interested in the application in cycles like the ones suggested in the Edmonton case. The Vensim implementation of the model is available through email from the authors.

5. Simulation results

In this section of the paper, we show preliminary model experiments. The simulation runs introduced here are only examples of more extensive sensitivity testing done with the model. We choose to have as the base scenario for this model a city with a brand new open data program (see Figure 4). In this city, government starts opening datasets at time zero, i.e. the beginning of the project. In this base run, it is assumed that, just like the case of Edmonton and OCIO staff, governments allocate some resources to prepare and open datasets as a way to start the project. In this base city, OG regulations are not perfect, and are mostly a set of promises from the central administration to become more transparent, as it is in many cases. We capture in the model the adequacy of these regulations as a dimensionless constant ranging from 0 to 1 (0.5 in the base case). As it is shown in the graph, society perceives a lot of benefits from the new policy at the start of the project, but quickly stop perceiving such benefits. This behavior results from the fact that the perception of benefit is formulated in the model depending on the rate of app downloads. This changes quickly at the beginning, but later becomes slower. In a sense, the formulation captures the sense of novelty that is experienced with apps in the Internet, and also is trying to capture the reduction of interest in the Road Construction application in the Edmonton case. As a result of this initial interest, government starts to use more resources in opening data, but gradually reduce this resource allocation to the basic resources required by law.

Public interest also motivates app developers to become interested in the development of new applications. However, our current model captures a non-sustainable ecosystem, where potential developers become disillusioned and stop using their time. It is

important to mention that our current model is not assuming any specific business model, and developer interest is just a function of perceived benefits. Finally, “Apps Developed” follows a pattern similar to the pattern of app developers, reflecting the initial interest in creating applications and later applications getting old and obsolete because of lack of maintenance.

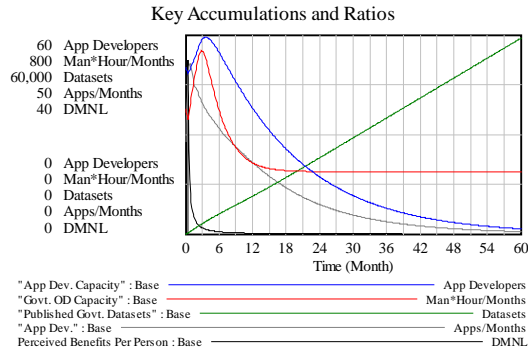


Figure 4. Base run for the OGD Ecosystem.

In order to show some of the potential behaviors that can be generated by our preliminary model, we decided to show three additional scenarios (see Table 2). The scenario “Better regulations” represents a city that starts with a much better framework to open data, and instead of using a value of 0.5 for the Adequacy of Open Data Regulations, this scenario starts with a value of 0.9. Our second scenario is looking for a case where the applications developed are more relevant to the public and have a “Longer life cycle.” In this scenario, the application decay rate is doubled from 24 to 48 months.

Table 2. Different scenarios.

	Base	Better regula.	Longer lifecycle	Govt. support
Adequacy of open data regulations	0.50	0.90	0.50	0.50
Applications lifecycle (decay rate)	24 months	24 months	48 months	24 months
Government programs to encourage app dev.	No govt. support	No govt. support	No govt. support	Support for 36 months

Finally, the last scenario involves the use of government resources to develop applications. In this scenario, government supports application development for 36 months starting in month 6. In any city, this support may take many different forms. It can represent direct use of resources such as OCIO in the city of Edmonton to application contests and other direct support to developers using government funds.

Figures 5 through 8 show comparative runs for most variables introduced in the case scenario. We are not

showing the comparative graph for the number of open datasets, because this graph is exactly the same for all scenarios. That is to say the city starts with no open datasets, and gradually opens them to finish with about 60,000 open datasets after a period of 5 years. The main reason for this pattern of the results on open datasets is associated with the behaviors in figures 5 and 6.

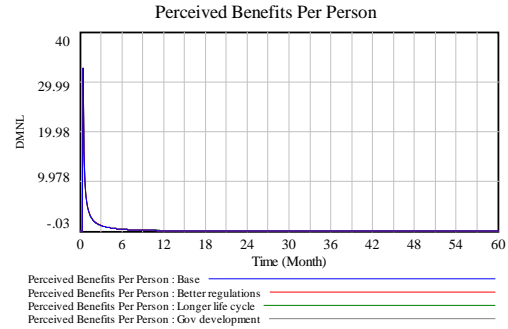


Figure 5. Perceived benefits across scenarios.

As it is shown in Figure 5, society is quickly receptive about the benefits of all scenarios, but also loses interest in the potential benefits quite fast. In this sense, initial interest for community members pushes government to allocate some additional resources in the early months of the initiative, but the pressure fades away with the lack of results in the form of benefits, allowing government to go back to a minimum use of resources (see Figure 6). These resources are the same for all runs with the only exception of the case where the city has better OGD regulations, which promotes the allocation of more resources

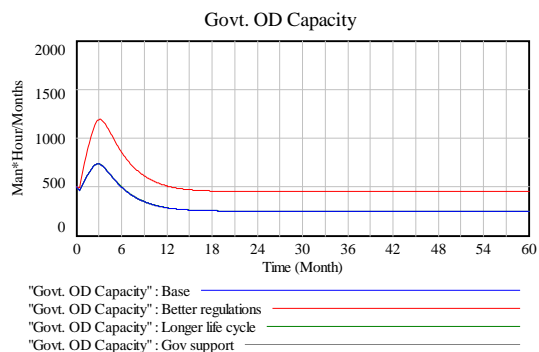


Figure 6. Govt. capacity for OGD across scenarios.

Figures 7 and 8 include comparative runs for the number of developers interested in using OGD to develop applications, and the number of applications themselves. As it is shown in Figure 7, none of our scenarios imply a sustainable ecosystem. Developers’ interest follows a social perception of benefit. The lack of any business model in our current preliminary model also limits the sustainability of the system.

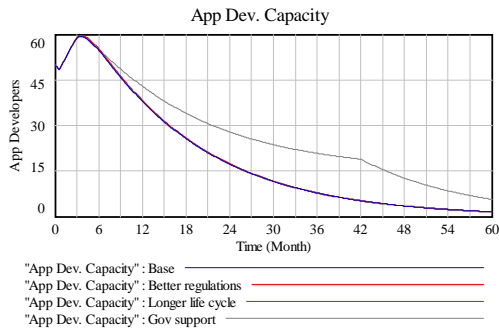


Figure 7. Application developers across scenarios.

Apps developed (Figure 8) grow fast in the first year to reach a maximum and then slowly become obsolete. The number of apps stops growing because of the lack of developers in the last months of any simulation. Developing capacity is not enough to maintain applications in any of these scenarios.

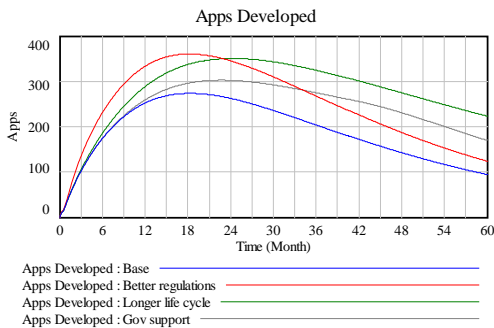


Figure 8. Apps developed across scenarios.

6. Discussion and final comments

In this paper, we introduced a preliminary model of an open government data ecosystem that builds on previous efforts of conceptualizing such an ecosystem [10]. Our preliminary results are consistent with current observed experiences that show the difficulties of building a sustainable OGD ecosystem. In a sense, our preliminary model replicates the same problems of current open data systems as they are reported in the literature. That is to say, the model represents a group of government officers that work with enthusiasm to prepare and publish datasets. In turn, dataset use is mediated by developers that prepare applications to be used by citizens as passive users [17]. In this way, although including some feedback effects, our preliminary model fails to include the necessary collaborations needed to build an OGD ecosystem [6,8,10,27]. At its minimum, an OGD ecosystem requires basic feedback mechanisms that represent collaboration among different stakeholders involved in the ecosystem. Harrison and her colleagues discuss the

need for policy makers to have the goal of explicitly and purposefully building ecosystems [8]. Otherwise, just like in the scenarios presented in this paper, an initial excitement may quickly fade in some developers, applications, some benefits for the society, and many datasets posted in government portals.

Contrary to our initial expectation, the current system represented in the model is not sustainable even in the scenario where government provides resources for application development. Although the effect of these additional resources are reflected in the second larger number of applications in the long run, longer life cycles is the scenario that yields more applications. We believe that this result supports the argument introduced in the previous paragraph. That is to say, applications that result from conversations among several stakeholders in the ecosystem will most likely be the ones with a longer time of utility and interest in the society.

A key reason for the lack of sustainability of the ecosystem in the simulation model lies in the current formulation of perceived benefits, which assumes that society demands more of an “augmented” service over time as core services. Thus, high perceived benefit from initial applications will not be sustained over time unless developers in the model keep up with a fast enough development and continuous improvement or maintenance. Although we believe that the basic assumption behind our current formulation is robust, the model requires additional refinements and exploration of alternative formulations in this specific area. Moreover, we also believe that our current definition of benefits is quite simplistic, and better definitions of benefit needs to be built into the model. Such definitions require the understanding that “benefit” constitutes a multi-dimensional concept that implies different sources of value for different actors in the ecosystem.

The ecosystem perspective to OGD is capable of connecting different actors along the value-chain starting from the datasets residing on the government sector, up to the end users in the society, and back into the government as a feedback effect [6,17]. This closed-loop system has the potential to enhance our understanding of the impacts of different policy choices and environmental assumptions on the realization of benefits. Our ongoing research shows our initial conceptual efforts towards the development of a simulation model that can generate more insight into these aspects.

The preliminary model presented in this paper, as any other model, constitutes a simplified representation of current understanding of a phenomenon. We captured some key variables of an OGD ecosystem with the purpose of generating insights about the situation. Our next steps consist of further refinements of our formal model of an OGD ecosystem. Another line of study also

includes adding a business perspective to this model, and also trying to capture the way in which different stakeholders interact regarding their interests in the OGD. Our current experiments suggest that these refinements require additional empirical explorations to better understand the processes involved in opening data, business models for OGD Ecosystems, and different measures of benefit from the point of view of different participants in the ecosystem. This suggests a more comprehensive case-based research to feed into the modeling efforts and let us build a more generalizable understanding of how an open data ecosystem can work in different governmental settings.

7. References

- [1] Attard, J., Orlandi, F., Scerri, S., and Auer, S. A systematic review of open government data initiatives. *Government Information Quarterly* 32, 4 (2015), 399–418.
- [2] Black, L., Carlile, P., and Repping, N. Dynamic Analysis of Cross-Boundary Behaviors Emerging from Similar Organizations. (2001).
- [3] Black, L.J., Cresswell, A.M., Luna, L.F., et al. A dynamic theory of collaboration: a structural approach to facilitating intergovernmental use of information technology. *IEEE* (2003), 12 pp.
- [4] Conradie, P. and Choenni, S. On the barriers for local government releasing open data. *Government Information Quarterly* 31, (2014), S10–S17.
- [5] Cresswell, A.M. Evolution of a Dynamic Theory of Collaboration: Modeling Intergovernmental Use of Information Technology. In *35th Hawaiian International Conference on System Sciences*. 2002.
- [6] Dawes, S.S., Vidiasova, L., and Parkhimovich, O. Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, (2016).
- [7] European Commission. *Digital Agenda: Turning government data into gold*. 2012.
- [8] Harrison, T.M., Pardo, T.A., and Cook, M. Creating Open Government Ecosystems: A Research and Development Agenda. *Future Internet* 4, 4 (2012), 900–928.
- [9] Heimstädt, M., Saunderson, F., and Heath, T. From Toddler to Teen: Growth of an Open Data Ecosystem, A Longitudinal Analysis of Open Data Developments in the UK. *eJournal of eDemocracy and Open Government* 6, 2 (2014), 123–135.
- [10] Helbig, N., Cresswell, A.M., Burke, B.G., and Luna-Reyes, L.F. The Dynamics of Opening Government Data. 2012. <http://www.ctg.albany.edu/publications/reports/opendata>.
- [11] Hines, J. Molecules of Structure: Building Blocks for System Dynamics Models. *Systems Thinking World*, 2005. <http://www.systemswiki.org/images/a/a8/Molecule.pdf>.
- [12] Janssen, M., Charalabidis, Y., and Zuiderwijk, A. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management* 29, 4 (2012), 258–268.
- [13] Lee, M., Almirall, E., and Wareham, J. Open data and civic apps: first-generation failures, second-generation improvements. *Communications of the ACM* 59, 1 (2015), 82–89.
- [14] Luna-Reyes, L.F. and Gil-Garcia, J.R. Using institutional theory and dynamic simulation to understand complex e-Government phenomena. *Government Information Quarterly* 28, 3 (2011), 329–345.
- [15] Luna-Reyes, L.F. and Gil-Garcia, J.R. Digital government transformation and internet portals: The co-evolution of technology, organizations, and institutions. *Government Information Quarterly* 31, 4 (2014), 545–555.
- [16] Mathiesen, K. Access to Information as a Human Right. *SSRN Electronic Journal*, (2008).
- [17] Pollock, R. Building the (Open) Data Ecosystem. 2011. <http://blog.okfn.org/2011/03/31/>.
- [18] President Barack Obama. Transparency and Open Government. 2009. http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment.
- [19] President Barack Obama. Executive Order -- Making Open and Machine Readable the New Default for Government Information. *The White House*, 2013. <http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government->.
- [20] Richardson, G.P. and Pugh, A.L. *Introduction to system dynamics modeling with DYNAMO*. MIT Press, Cambridge, MA, 1981.
- [21] Rouder, J.N. The What, Why, and How of Born-Open Data. *Behavioral Research Methods*, (In Press).
- [22] Sayogo, D.S. and Pardo, T.A. Understanding smart data disclosure policy success: the case of Green Button. *ACM Press* (2013), 72.
- [23] Scholl, H.J. and Luna-Reyes, L.F. Uncovering Dynamics of Open Government, Transparency, Participation, and Collaboration. *IEEE* (2011), 1–11.
- [24] Sterman, J. *Business dynamics: systems thinking and modeling for a complex world*. Irwin/McGraw-Hill, Boston, 2000.
- [25] Zuiderwijk, A. and Janssen, M. The negative effects of open government data - investigating the dark side of open data. *ACM Press* (2014), 147–152.
- [26] Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., and Alibaks, R.S. Socio-technical Impediments of Open Data. *Electronic Journal of e-Government* 10, 2 (2012), 156–172.
- [27] Zuiderwijk, A., Janssen, M., and Davis, C. Innovation with open data: Essential elements of open data ecosystems. *Information Polity*, 1,2 (2014), 17–33.
- [28] Access to Information Laws: Overview and Statutory Goals. *Good Law and Practice*, 2012. <http://www.right2info.org/access-to-information-laws/access-to-information-laws>.