# *From Data-Centric to Citizen-Centric Architecture:*
## *Architecting a Future State for Open Data in the Government of Puerto Rico*

By

**Nestor Victor Leonardo Figueroa-Rodriguez**
B.S., Industrial Engineering
Pennsylvania State University, University Park, PA

SUBMITTED TO THE SYSTEMS DESIGN AND MANAGEMENT PROGRAM
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN ENGINEERING AND MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2022



Signature of Author _____

Systems Design and Management
January 31st, 2022

Certified by Joan Rubin_____

Joan Rubin
Executive Director and Senior Lecturer
Advisor

Accepted by Warren Seering_____

Professor Warren Seering
Weber-Shaughness Professor of Mechanical Engineering in MIT's School of Engineering and a faculty co-director for the SDM program
SDM Faculty Co-Director

This page has been intentionally left blank.

This thesis is dedicated to my family -

**Lourdes, Alejandra, and Adriana**.

Also, to my **mom** and especially **dad,** who passed away while I was going through this thesis process. **Dad**, thanks for teaching me the value of education and motivating me to continue and finish even during your last days.

# *From Data-Centric to Citizen-Centric Architecture:*
## *Architecting a Future State for Open Data in the Government of Puerto Rico*

By
**Nestor V. Figueroa**

Submitted to the System Design and Management Program on January 31st, 2022, in partial fulfillment of the requirements for the degrees of Master of Science in Engineering and Management.

**Abstract**

Government institutions throughout the world have been working on transparency initiatives to build accountability with their constituents. One of these critical initiatives is *Open Data*. Open Data has as its primary objective creating transparency in Governments by liberating data sets stored in systems and databases under the custody of public agencies. Historically, this data has been challenging to find and often does not make it to the public domain. When it does, it is not easy for ordinary citizens to make it useful.

Governments have made strides to liberate these data sets. Architecturally, governments have been developing centralized systems that extract the data from many agencies to transform, store, and publish on a website for public access. However, these enterprise data architectures focus on data liberation and not on citizen value through relevant and contextual insights creation. In other words, instead of being citizen-centric, these architectures are data-centric. Since the objective is to liberate data, the full intended benefit is not realized. It creates an unintended effect: data

inequality. Only people with data skills and corporations with resources extract the value and benefit of these data sets, while ordinary citizens cannot.

Puerto Rico, a territory of the US with a population of 3.1 million, is not the exception. In 2019, the Government of Puerto Rico passed Law 121 or *Ley de Datos Abiertos*. The objective is to publish data sets generated through the interaction of citizens with government agencies. Today, there are 98 data sets published through the Puerto Rico Statistics Institute (PRSI). However, the "as-is" architecture, which complies with the current law, publishes data sets in a format usable for a few skilled citizens. It requires data exploration, analysis, visualization, and interpretation before extracting key insights and motivating follow-up actions. It is not realistic to expect this type of work from ordinary citizens.

This thesis proposes a "to-be" Open Data architecture for the Government of Puerto Rico that extends the value and benefit of any published data set. The "to-be" architecture transforms the current data formats into meaningful and actionable insights while minimizing a possible data inequality issue.

**Acknowledgments**

I am grateful to family, colleagues, classmates, professors, and advisors for their assistance, guidance, and support in completing this work. First, I am forever thankful to the SDM Executive Director, Joan Rubin. After completing my Executive MBA at MIT Sloan, with many years of experience, it is not a typical path to continue an academic career pursuing a technical master's degree. The norm is to do it backward. However, Joan believed in my mission and what this degree could do for my career.

My journey through this thesis has been unusual, although it may sound familiar to some. I started with very lofty goals to solve a complex problem until I realized it would take way beyond the time allocated for this degree. When my father's cancer came back, under the effects of anger, sadness, and despair, I redirected my efforts to contribute through this thesis to help cure his cancer. Not that easy, of course, especially when encountering a steep process to obtain data and emotionally too daunting. Then, the COVID pandemic hit the entire globe, and my father passed away months after. I found myself in complete darkness. At times, I questioned if it was worth it. However, with the support of family and friends, I got back on track. I had found a topic that I felt passionate about, in the intersection of data, systems architecture, and government transparency, directly impacting the people of Puerto Rico, where I was born and raised.

Special gratitude goes to my classmates for the endless and enriching discussions that shaped me into a better student and professional. Similarly, I would like to thank the entire SDM Program staff for understanding my unusual journey, goals,

and insatiable thirst for learning, which almost created a balancing issue of engineering and management courses. Thanks to Bill for keeping me in compliance by maneuvering a suitable course plan for me.

Thanks to my teammates at Nagnoi LLC and my business partner Miguel Colom-Mena, for supporting this opportunity to pursue my lifelong dream to come to MIT while the business continued to run seamlessly for our customers.

Finally, to my family, I cannot thank you enough for the constant encouragement and support during these years. Thanks to my sisters, nieces, and nephews for always supporting *tío*; and my parents for teaching me the value of a good education, perseverance, and hard work, while encouraging me to follow my dreams. To my wife Lourdes, there are no words to convey the sense of gratitude. Thanks for your unconditional support, especially in my absence from home, unending patience, and for taking care of our most beautiful creations: Alejandra and Adriana. My love and this thesis are for the three of you!

This page has been intentionally left blank.

# Contents

## Table of Figures

**List of Tables**

## List of Equations

*Chapter 3 Equations*

# 1 Introduction

## 1.1 Context

"*Although open data undoubtedly builds upon the fifty-year legal tradition of the right to know about the workings of one's government, open data does more than advance government accountability. Rather, it is a distinctly twenty-first-century governing practice borne out of the potential of big data to solve society's biggest problems.*"— Beth Simone Noveck

Government Open Data initiatives, particularly the development of Open Data portals, have proliferated since the mid-2000s at central and local government levels around the world. These jurisdictions make government data, primarily generated through interaction with its constituents, more open, accessible, and available for citizens, media, businesses, non-profits, and the government itself. The motivation is to use this data to promote economic growth, enable innovation, create public policy, oversee government decisions and execution, and encourage public engagement. Exploring open data, patterns discovery, and correlations are necessary to extract insights relevant to society and realize that value. As governments collect and maintain vast amounts of data, it should be treated as a public asset.

As published by the National Conference of State Legislature (NCSL) in their January 2021 article *State Open Data Laws and Policies*, jurisdictions are passing laws to make the liberation of non-sensitive government data mandatory. However, an increasing number of states are now operating from the presumption that non-sensitive government data should be made freely available in a way that makes it open,

discoverable, and usable for the public. It is undoubtedly positive and indicates the progress made.

Government data is “open” when published in formats that make it available for unrestricted, public use and reuse, without charge and intellectual property or licensing restrictions (NCLS, 2022). Such structures enable digital processing and follow documented, widely used, and publicly available standards. Open Data is machine-readable. It also includes descriptive metadata—data about data—information that describes elements such as title, author, keywords, or other characteristics of digital materials.

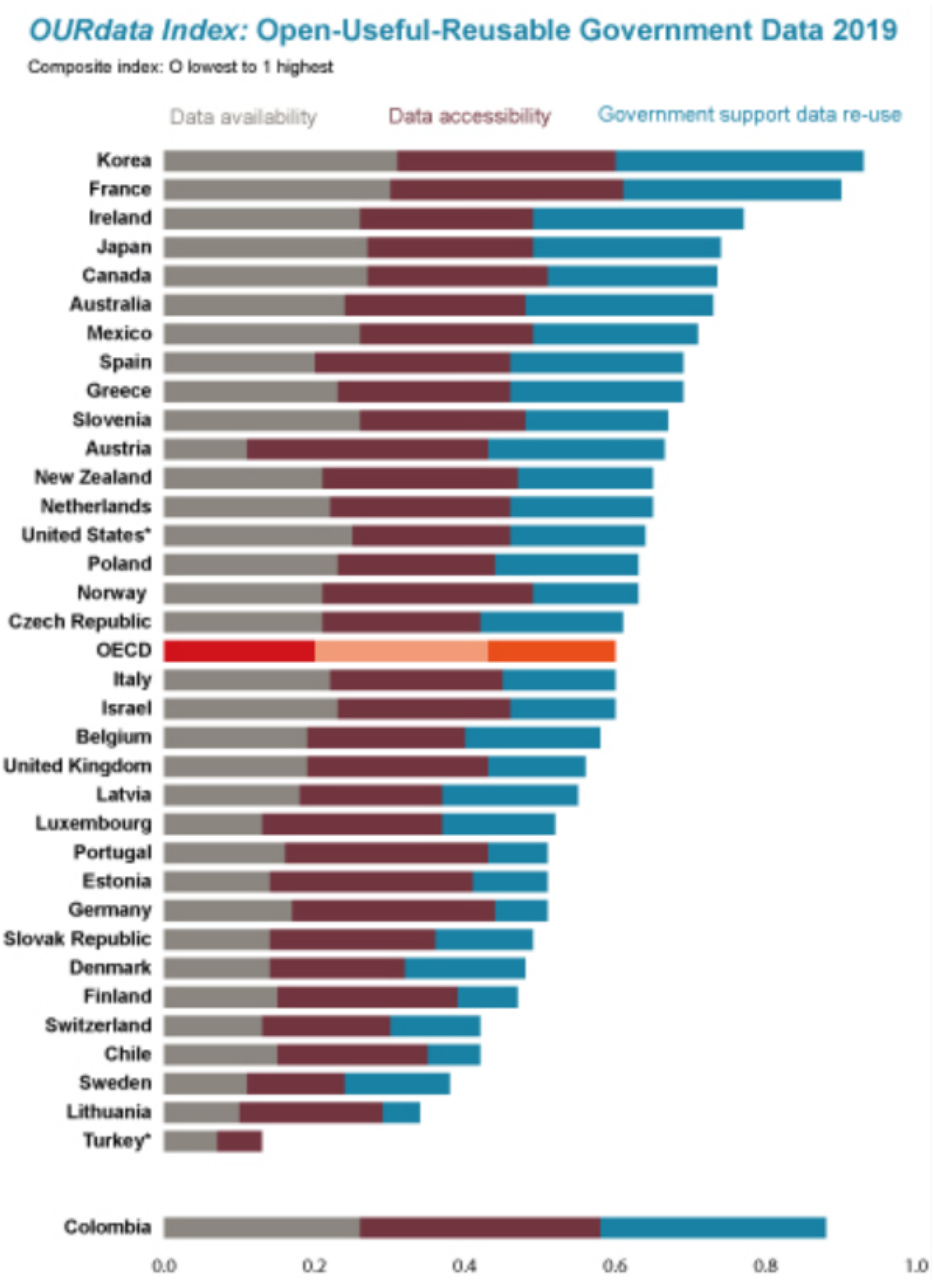
Through these Open Data initiatives, transparency helps citizens protect themselves and make informed decisions necessary for their families. For example, decisions on school selection, awareness of areas or regions with high positivity rates for a very contagious virus, crime rates in their living community, government spending, among many other use cases.

Open Data has benefits beyond what the government can envision, or simply it would be not easy to get involved. As aforementioned, the information gathered from Open Data can be used for many purposes, including creating new products, services, and businesses. The value of Open Data is demonstrated by its many uses. A good example is companies founded using open government data, such as Zillow for the house pricing, the Weather Channel for weather forecasting, and Garmin for geolocating.

**Open Data Around the World**

To obtain a general assessment of Open Data initiatives worldwide, the Organization for Economic Co-operation and Development (OECD) pays special attention to the progress of Open Data globally. This international organization works to build better policies for better lives. As its website states: "*the OECD's goal is to shape policies that foster prosperity, equality, opportunity, and well-being for all through insights to better prepare the world of tomorrow."* The OECD studied the efforts of Open Data in governments across the world.

The OECD Open Government Data (OGD) project aims to progress international efforts on OGD impact assessment. The mapping of practices across countries helps establish a knowledge base on OGD policies, strategies, and initiatives. It supports the development of a methodology to assess the impact and creation of economic, social, and good governance value through OGD initiatives. The OECD OURdata Index assesses governments' efforts to implement open data in the three critical areas - Openness, Usefulness, and Re-usability of government data. Refer to the figure below that presents the results from 2019.



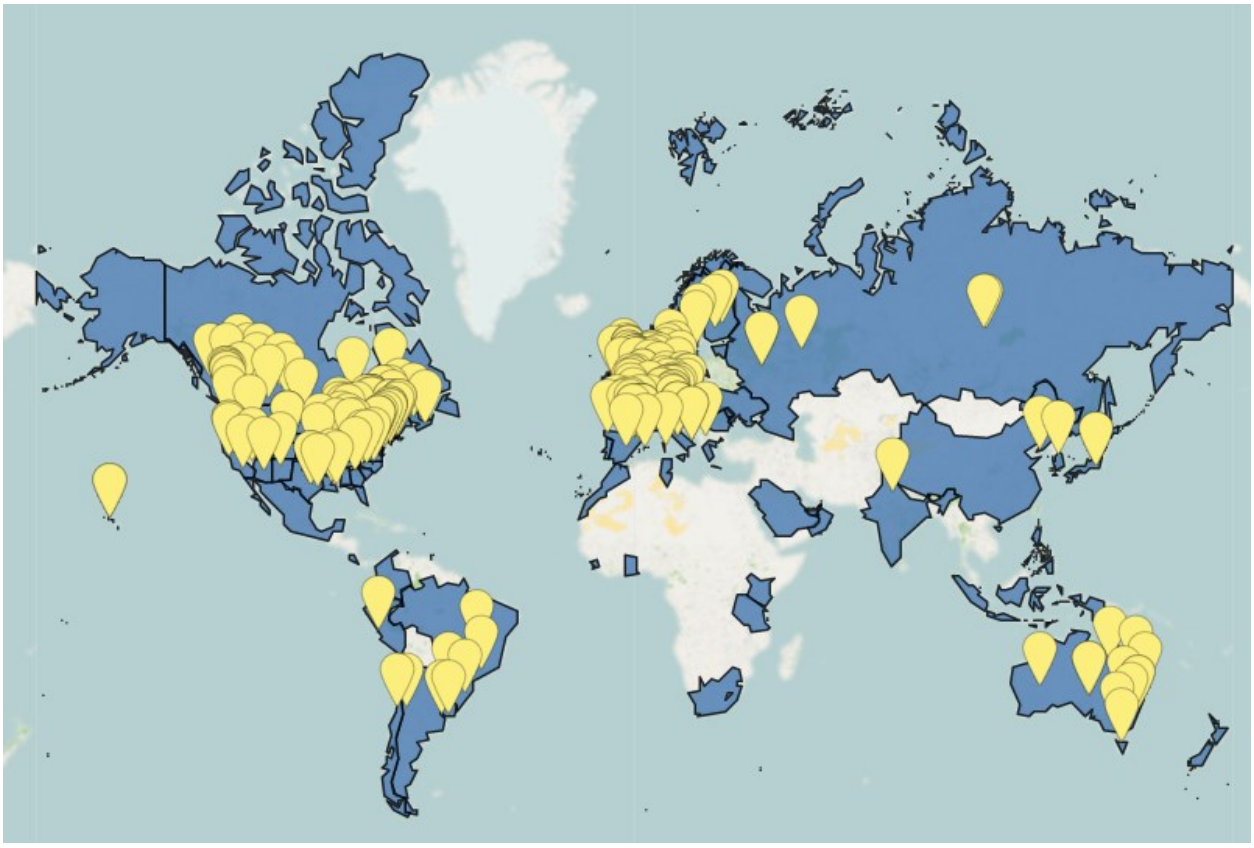

Source: OECD Government at a Glance 2019

Figure 1-1*: OECD OURdata index results*

From Figure 1-1, the efforts of Open Data focus on the availability and accessibility of the data. Given that the starting point was that the data was not even reachable and heavily guarded by governments, it is understandable that the focus of

Open Data has been on making sure that the data is liberated and transformed in a consistent machine-readable format. However, what is the point of making the data available and machine-readable if it does not translate into value to the citizens? It is a limitation that Open Data initiatives worldwide are not paying attention to yet (Krishnamurthy, 2016).

At a glance, the map below shows some of the countries with open data sites as published by the USA's federal government Data.gov. A complete list of cities, counties, and counties is available on this site.



*Source: data.gov*

Figure 1-2: Map representation of Open Data sites

Another organization that studies and promotes Open Data initiatives globally is the Global Open Data Index (GODI), a global benchmark for publication of open government data, run by the Open Knowledge Network. Their survey measures the openness of government data according to the Open Definition.

**Open Data in the United States**

In the US, forty-eight states have some form of an open data portal. The initiatives implemented by these states vary with the most robust including a wide variety of datasets, from information on public employees' salaries, population counts, and estimates, to employment and unemployment reports. In Maryland, open data portals offer population demographics by county, job growth trends, pollution reduction data, health trends, and other datasets. Recently due to the COVID-19 pandemic, many state portals have published data sets on cases, tests, hospitalizations, and deaths, which organizations like Johns Hopkins University have used to create valuable information, including the 50-state maps, charts, and data so widely viewed.

State governments accumulate data that can be used in many ways for citizens' benefit. Data availability online, such as state economics and financials, allows businesses to study the benefits of relocating to a state. The state of Colorado promotes its open data in annual public data competitions where teams compete to use public data to solve various problems. One example of a problem solved with public domain data is a platform that helps streamline water data discovery and analysis. This application helps real estate agents, developers, investors, appraisers, and consultants make better business decisions with publicly available water data.

The NCSL stated the following: “In recent years, state legislatures have taken a more active role to ensure that state government data is more accessible, transparent, and open to innovative uses that may help transform and improve citizens’ lives.” The Federal Funding Accountability and Transparency Act from 2006 require that federal financial assistance and expenditures be available through a single, searchable website. Many states shortly after that followed the example, passing Taxpayer Transparency Acts and creating centralized websites with detailed information about statewide expenditures.

One of the most noticeable initiatives in the US has been Data.gov. Data.gov is a federal government’s open data site that aims to make this institution more open and accountable. The objective of Data.gov, as it is for most Open Data initiatives, is to increase citizen participation in government, create opportunities for economic development, and inform decision-making in both the private and public sectors. The Open Government Data Act makes the Data.gov site a requirement. Federal agencies must publish their information online as open data, using standardized, machine-readable data formats with their metadata.

There are numerous states, cities, and counties that have launched open data sites today. Figure 1-3 below presents an evolving count of US states, cities, counties, and international jurisdictions with open data sites available by the end of 2021 per the data.gov site.

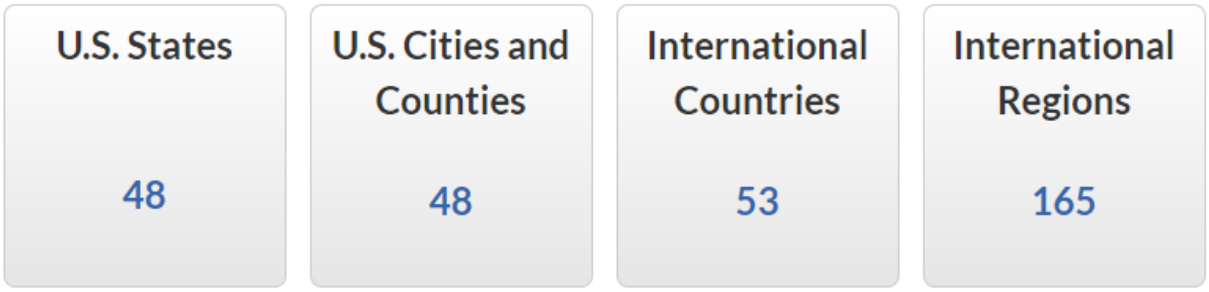
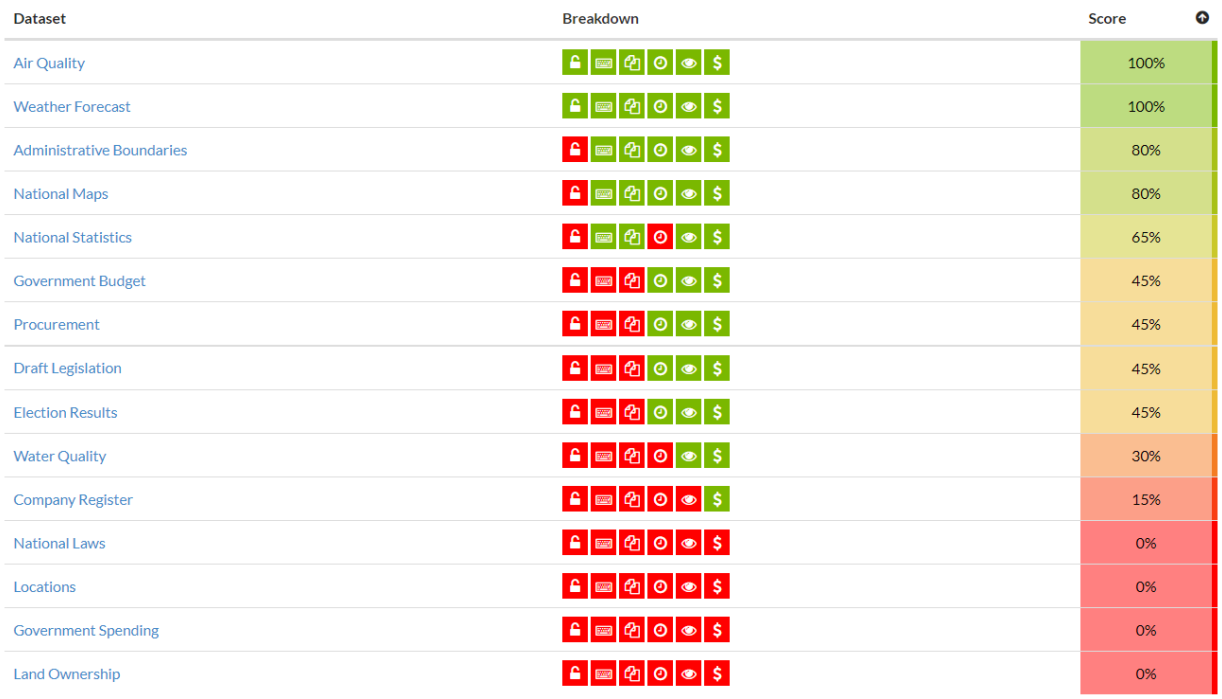



Figure 1-3: Evolving count of geographies with Open Data sites

**Open Data in Puerto Rico**

Puerto Rico has been involved in Open Data initiatives for years. Media and the public have advocated for transparency, with one of those components being the liberation of government data sets. The Global Open Data Index, in 2015, ranked Puerto Rico in the 38th position out of 94 countries with a score of 43% based on a survey. Similarly, this global index gave Puerto Rico a score of 13% in the percent of fully open datasets, as defined by the Open Definition (refer to the definition in Section 2.1 ). The figure below shows how Puerto Rico fared in the different categories.



| Dataset | Breakdown | Score |
|---|---|---|
| Air Quality | | 100% |
| Weather Forecast | | 100% |
| Administrative Boundaries | | 80% |
| National Maps | | 80% |
| National Statistics | | 65% |
| Government Budget | | 45% |
| Procurement | | 45% |
| Draft Legislation | | 45% |
| Election Results | | 45% |
| Water Quality | | 30% |
| Company Register | | 15% |
| National Laws | | 0% |
| Locations | | 0% |
| Government Spending | | 0% |
| Land Ownership | | 0% |

*Figure 1-4: Global Open Data index scores for Puerto Rico*

As of 2019, data sets were formally and legally liberated in Puerto Rico due to *Law 122 – Ley de Datos Abiertos*. Similar legislation passed in other jurisdictions to increase transparency and accountability while providing citizens with relevant data for their benefit. This Open Data Law states that the Puerto Rico Institute of Statistics (PRSI) is required to publish and disclose the list of data sets available from government agencies on its website. The Institute's mission is to prepare the policy for developing the public statistical function, coordinate the statistics production service of government entities, and require information from both the public and private sectors. The PRSI is also traditionally free of political interference and independent status from the executive branch by its organic law. This mission aligns with what is necessary to carry out the Open Data initiative for the Government of Puerto Rico.

The Puerto Rico Innovation and Technology Services (PRITS) has also been a promoter of Open Data and has established initiatives to support these efforts in Puerto Rico, many of which are in collaboration with the PRSI.

**Open Data and Data Inequality**

A research paper from Jonathan Cinnamon (University of British Columbia) in 2019 states that "the Open Data movement is a direct, if not comprehensive, attempt to challenge the data access divide." The author states that its fundamental aim is to make existing datasets (especially government data) accessible. The rationale for opening data lies in the potential impact it can make. Although opening data could reduce inequalities, the contrary is most likely to happen. With the existence of the digital divide, Open Data is an approach that could first democratize data access, usage, and knowledge to provide societal benefit by enabling decision making, transparency, appropriate governance, and economic development (Arzberger et al., 2004; European Commission, 2014; Janssen, Charalabidis, & Zuiderwijk, 2012; Sieber & Johnson, 2015; Zuiderwijk & Janssen, 2014). However, open data remains an experiment, with little evidence demonstrating that simply making existing datasets available will make an impact. Increasing evidence suggests that open data may empower private companies and the already socioeconomically advantaged in society (Gurstein, 2011; Kitchin, 2014), therefore, increasing the data inequality gap, although that is not its original intention.

## 1.2 Research Motivation

Like most jurisdictions, the focus in Puerto Rico has been on liberating data and complying with law 121. With 98 data sets released, curated, and published, Puerto Rico seems to be on the right track to transparency and in an excellent position to benefit society through these readily accessible government data. The release of data demonstrates tremendous progress, and the effort needs to continue because without publishing government data sets, there will be no Open Data.

However, the usage of these data sets by ordinary citizens is uncertain. Alfonso Quarati from the Institute of Applied Mathematics and Information Technologies Enrico Magenes at the National Research Council in Italy found that most OGD data sets are underutilized regardless of size, the software platform adopted, administrative and territorial coverage. The study included 400,000 data sets of 28 national, municipal, and international Open Government Data (OGD) portals. There is no evidence that the case of Puerto Rico is any different, as this is a worldwide trend. Puerto Rico does not collect any data about the usage and downloads of data sets published, as confirmed by the PRSI.

The publication of data needs to be accompanied by an infrastructure that can handle the data in a user-friendly way to lower the threshold for users (Janssen and Zuiderwijk, 2014). What should such an architecture look like, and how should it be designed? When analyzing the Open Data architecture of the Government of Puerto Rico through the PRSI's Open Data site, it is noticeable that the current architecture is data-centric and not citizen-centric. The design is intended to publish processed data sets in a machine-readable open format and upload those to be easily accessible to the public with their corresponding metadata. However, despite the transparency progress

and good intention, this Open Data architecture limits the value and benefit that it can provide for most citizens. As the current efforts focus on liberating as many data sets as possible and in the proper machine-readable format, an unintended barrier is created. This barrier stems from the difficulty of providing a meaningful, understandable, unbiased, and actionable interpretation of this data to work for its constituents.

People evaluate Open Data portals based on ease of use. The ease of use affects their intention to use the portal. It relates to the skills needed to excerpt and utilize the datasets (Talukder et al.). As a result, when users observe that datasets are easily usable and do not struggle to utilize them, their intention to use these datasets will increase (Zuiderwijk et al.).

Puerto Rico has an opportunity to implement an improved architecture for Open Data that also does not exacerbate the data inequality that the current architecture can potentially create. Therefore, with this context, the following questions arise:

1. What are the architecture changes and enhancements to migrate from a data-centric enterprise to citizen-centric?
2. Can a citizen-centric architecture extend the value of these liberated data sets by proposing a complementary but more usable format?
3. What are the architectural alternatives, risks, and trade-offs?
4. Can the enterprise add or change the appropriate processes, knowledge, and organization to make this happen?
5. Can the Government of Puerto Rico provide the resources and leadership to make this happen?

As the definition of “Enterprise Transformation” states (refer to Section 2.1), a change in architecture responds to radical changes in the social environment. Like in many countries, citizens are demanding more participation in government processes. In addition, when a fundamental alteration of context happens, architectures need to adapt. Puerto Rico’s Open Data enterprise and its architecture need to adapt by defining a new context to meet its citizen usability goals.

This thesis is motivated by the possibility that a citizen-centric Open Data architecture can be much more impactful for ordinary citizens versus the unrealized value of the current architecture within the Government of Puerto Rico. The impact can be on two fronts:

1. Transformed data sets into relevant citizen knowledge for action
2. Minimizing data inequality that the current architecture can potentially create

By assessing its current architecture for Open Data, identifying the areas of improvement, presenting new architecture candidates, and its implementation efforts and risks, Puerto Rico can start walking in a new path of enhancing the value of such an important asset, government data. That path can lead to an expansion of the currently limited value and put the data to work for the people of Puerto Rico.

## 1.3 Thesis Scope and Objectives

**Scope**

This work aims to assess the current architecture for Open Data in Puerto Rico and present an alternative that can extend the value of the liberated data by transforming data sets into relevant knowledge for as many citizens as possible. Relevant research already published is presented as a base to reach similar conclusions for the Open Data challenges in Puerto Rico. Publicly available virtual events and publications by the Puerto Rico Institute of Statistics, discussing current data projects and government data, are part of the input used to establish new parameters for a new architecture. Also, this work includes systems thinking, design thinking, and systems architecture methods to complement the architecture analysis.

As a proposed architecture is derived, three alternatives for its implementation are presented. This work employs aspects of the Enterprise Architecture approach of Dr. Nightingale and Dr. Rhodes from MIT (2004) to analyze these alternatives. The method is one of the few that ensures spending time developing and evaluating "could be states" given a set of desired criteria. Also, the AFE framework (Raby, Zini, 2012) used complements the Nightingale & Rhodes approach. As stated by Raby and Zini, AFE provides a more explicit and quantitative process for generating and evaluating the future state of an enterprise.

**Objectives**

Our main objective is ***"architecting the future enterprise of Open Data in the Commonwealth of Puerto Rico by using systems thinking, systems architecture, and simple management tools to maximize the use, value, and benefit of liberated data for all citizens in the island."***

The specific objectives include:

- Identify the areas of opportunities for the current Open Data initiative architecture in Puerto Rico
- To generate and select a future enterprise state and architecture for Open Data in Puerto Rico
- To discuss potential changes to Open Data Law #122 and related policy and identify areas of future research.
- Apply a combination of systems and design thinking, systems architecture, Nightingale & Rhodes, and the AFE framework (Raby, Zini) to analyze the effectiveness of Open Data architectures in the Puerto Rico Open Data enterprise.

## 1.4 Thesis Roadmap

The thesis is divided into five chapters. The goal of the chapters is to meet the objectives presented in the previous section. The aggregation of all of them attempts to respond to this work's primary purpose.

- **Chapter 1** addresses the description of the context, the research motivation, the thesis scope, objectives and organization, and the research approach used to answer the research questions.

- **Chapter 2** provides background information about terminologies, methods, and approaches employed in this work. The chapter presents the Enterprise Architecture approach developed by Nightingale & Rhodes and systems thinking and architecture methods learned in EM.411 *Foundation of Systems Design and Management* to evaluate the current architecture of Open Data in Puerto Rico.

- **Chapter 3** presents the assessment of the “as-is” architecture and derives “to-be” architecture to improve the Puerto Rico Open Data enterprise.

- **Chapter 4** presents alternative architectures that the Government of Puerto Rico could consider implementing. The AFE Framework analyzes each alternative considering effort, risk, and effectiveness.

- **Chapter 5** provides an illustrative application of a real case study.

- **Chapter 6** synthesizes the policy implications, conclusions, limitations, and areas of future work of the research.

## 1.5 Research Approach

Similarly aligned with the chapters described in Section 1.4 Thesis Roadmap, the work is divided into four different stages:

(1) assessment of the “as-is” Open Data enterprise and current architecture in Puerto Rico using existing research, enterprise assessment methods, and systems and design thinking analysis;

(2) developing a “to-be” citizen-centric architecture;

(3) generating alternatives of the “to-be” architecture and assessing those alternatives through stakeholders, risk, and effort analysis, using Nightingale & Rhodes and AFE framework approaches;

(4) the application of the recommended architecture in a case study.

1. Assessment of the “as-is” Open Data enterprise and current architecture in Puerto Rico

   The first step involved reviewing and analyzing the Puerto Rico Open Data enterprise and its current systems architecture implemented by the Government of Puerto Rico for Open Data. It involved studying relevant research with results applicable to the Puerto Rico case. Other methods used: system architecture, systems thinking combined with design thinking techniques, data gathered through public domain websites, and examining the current open data sets published by the Puerto Rico Statistics Institute.

2. Deriving a “to-be” citizen-centric architecture

The second stage involved adding new context and developing a resultant concept that addresses the needs of citizens. It includes desirable *illities* and the value that emerges when connecting disparate data sets to communicate the story hidden behind the data through context. Other activities include:

   a. reviewing literature from systems architecture, design thinking, and systems thinking learned throughout the master’s program curriculum,
   b. evaluating the Puerto Rico Open Data enterprise using Nightingale & Rhodes approach,
   c. information gathering by the many interactions with former and current Chief Information Officer (CIO) of the Government of Puerto Rico, Glorimar Ripoll and Enrique Volckers, respectively, to define the needs of citizens. Additional

data was collected from recorded events, archived online interviews, and publications by the Puerto Rico Statistics Institute,

d. presenting the analysis through mathematical abstractions of the “as-is” and “to-be” Open Data architecture.

3. Generating alternatives of the “to-be” architecture through stakeholders, risk, and effort analysis

The empirical approach uses heuristics from Matias Raby’s thesis (2012) and the AFE Framework to produce a recommended architecture alternative considering stakeholders, risks, effort, and effectiveness to create the expected value. The figures in the analysis are adapted from Raby’s thesis in 2012, where Appendix A, B, and C show the scoring methodology (not included in this work, but reference is listed in Section 7).

4. The application of the recommended architecture in a case study

The fourth phase presents a real case study putting the new architecture to work in a real scenario. The process helped adjust certain aspects of the architecture and illustrate the future states. This case study is an ongoing project within the Government of Puerto Rico called *Portal Informativo Ciudadano* (PIC), which translates to Citizen Information Portal (CIP).

# 2 Background Information

## 2.1 Common Definitions

- **Architecture** - "the fundamental design of the enterprise's strategy, organization, processes, and systems" (Glazner, 2011).” The following definitions also apply to the work presented in this thesis:
    - The structure, arrangements, or configuration of system elements and their internal relationships are necessary to satisfy constraints and requirements (Frey).
    - The arrangement of the functional elements into physical blocks. (Ulrich & Eppinger)
    - The whole consists of parts; the parts have relationships to each other; when put together, the whole has a designed purpose and fills a need (Reekie and McAdam, A Software Architecture Primer)
    - An abstract description of the entities of a system and the relationship between those entities (Crawley et al.)
    - The embodiment of *concept*, the allocation of physical/informational function to elements of form, and definition of interfaces among the elements and with the surrounding context. (Crawley)

  An architecture emerges with functionality greater than the sum of its parts. The desired outcome produces “*ilities*,” such as usability, scalability, reliability, maintainability, extensibility, among others.

- **Benefit** - worth, importance, or utility as judged by a subjective observer (the beneficiary)
- **Concept** – refers to a system vision, idea, notion, or mental image, which maps function to form. The concept is the beginning of the architecture and rationalizes the structure of the architecture (Imrich)
- **Context** – describes the interaction of a system with its environment taking into account laws, standards, and guidelines in a given design
- **Contextuality (data)** - refers to an additional data processing step to add context that transforms the data into knowledge. When the user acquires new knowledge, especially if it is meaningful, relevant, and actionable, the value and benefit of the data are realized. Contextuality is developed within the architecture (on top of the raw data) by adding visualizations, benchmarks, simple descriptions, story narratives, insights, among other forms.
- **Data Preparation** – a process to transform the data from one state to the next. It is associated with the steps for data enrichment until it is ready for consumption by the end-user.
- **Enterprise** - “one or more persons or organizations with related activities, unified operation or common control, and common business purpose" (Garner, 2009). The term “enterprise” applies to a single integrated company or collection of inter-organizational partners. With this definition, the government is also an “enterprise.” Furthermore, enterprises are activities of sub-parts of companies (Purchase, Parry, Valerdi, Nightingale, & Mills, 2011).

- **Enterprise Architecture** – it is an approach to transformation. Enterprise architecture "provides strategies/approaches to ensure time is spent developing and evaluating 'could be' states, and selecting the best alternative given a set of desired properties and criteria for the future enterprise" (Nightingale & Rhodes, 2012).
- **Enterprise Transformation** - a process within the enterprise that often involves fundamental changes to the architecture. Leading authors in the field described enterprise transformation as "a shift within a defined enterprise that is: (i) a response to radical changes in the economic, market or social environment; (ii) a fundamental alteration of context; and (iii) a step-change in performance" (Purchase et al., 2011). An Enterprise Transformation typically involves *strategic planning* and an *execution* cycle to complete it (Nightingale, Principles of Enterprise Systems, 2009). In general terms, a transformation process stimulates the need for change (at the beginning) to institutionalize (at the end).
- **Ilities** - are attributes that characterize a system's ability to respond to foreseeable and unforeseeable changes. Ilities are sometimes also known as non-functional requirements because they do not describe what a system should do but how it should be (Chin, Yau, Wah, Khiang)
- **Machine-readable** - a format in which government data can be quickly processed by a computer without human intervention while ensuring that semantic meaning is not lost (the state of Indiana, 2021)

- **Metadata** - “data about data” (Plagman, 1971). It is also information about who produced it and when among other attributes. Documents and structured datasets require metadata to be searchable and shareable.
- The Open Knowledge Foundation states that **Open** means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness). The **Open Data** definition: sets principles that define “openness” concerning data and content. Open Data can be freely used, modified, and shared by anyone for any purpose (Global Open Data Index, 2016-2017).
- **Open format** - the U.S. Government, through the Open Government Directive (http://www.whitehouse.gov/omb/assets/memoranda_2010/m10-06.pdf), defines an open format as “one that is platform-independent, machine-readable, and made available to the public without restrictions that would impede the re-use of that information.” Some examples of machine-readable formats include HTML, XML, CSV and JSON file formats.
- **Operand** – the object acted upon by a process, which may be created, destroyed, or altered. For example, “data” is the typical operand for Open Data architectures, where data is transformed from a current state to another.
- **System** - a set of interrelated elements which perform a function whose functionality is greater than the sum of the parts
- **Value** - benefit at a cost

## 2.2 Systems Thinking and Architecture Approach

The first element analyzed of the current Open Data architecture was the context used for the design. Per the definition of "Enterprise Transformation," a shift in context can generate a new architecture that transforms the enterprise, as is the case for a proposed citizen-centric architecture. Architecting "up and down" is a technique architects use to re-design an architecture as the context changes. When context changes, new needs and requirements emerge for stakeholders. The goal is to extend the value and benefits that the current architecture cannot provide.

Another component of the analysis is to depict operands and processes to generate a new function for the Open Data initiative in Puerto Rico. With the context and needs changing, the original concept changes. The new concept extends the functional architecture. Both are presented to demonstrate the differences. Two methods created as part of this work to provide additional information about the architectures are the ARI (Architecture Readiness Index) and the CCL (Correlated Contextualization Level).

Tools like Problem-Tree Analysis from Design Thinking (as shown in Figure 2-1), Design-Structure-Matrix (DSM), Object-Process Model (OPM), and OPD (Object-Process-Diagrams) complement the analysis.

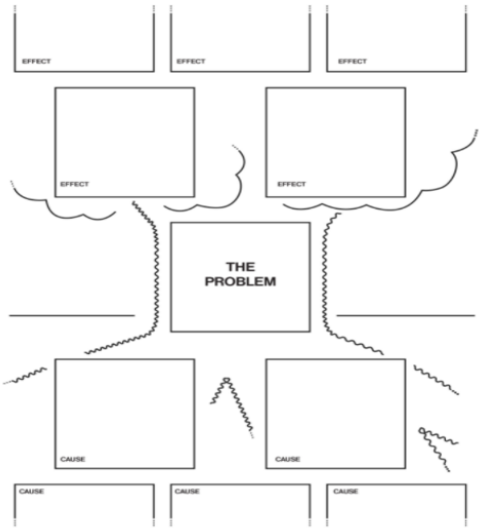



Figure 2-1: Design thinking's problem-tree analysis

## 2.3 Nightingale & Rhodes Enterprise Architecting Approach

The enterprise architecting approach developed by Dr. D. Nightingale and Dr. D. Rhodes at MIT has been evolving since 2004. It has shown to be a valuable method to support transformation initiatives in various domains, including government, healthcare, start-ups, among others. It analyzes enterprises through multiple lenses, where the Eight View Elements, as shown in the table below, is particularly useful. The approach considers evaluating alternatives when selecting an enterprise's future state.

| ELEMENT | DESCRIPTION |
|---|---|
| **Strategy** | The enterprise vision, strategic goals, business model, and enterprise-level metrics |
| **Information** | Information the enterprise requires to perform its mission and operate effectively |
| **Infrastructure** | Enterprise systems and information technology, communications technology, and physical facilities that enable enterprise performance |
| **Products** | Products the enterprise acquires, markets, develops, manufactures, and/or distributes to stakeholders |
| **Services** | Offerings derived from enterprise knowledge, skills, and competences that deliver value to stakeholders, including support of products |
| **Process** | Core, leadership, lifecycle, and enabling processes by which the enterprise creates value for its stakeholders |
| **Organization** | Culture, organizational structure, and underlying social network of the enterprise |
| **Knowledge** | Competencies, explicit and tacit knowledge, and intellectual property resident in and generated by the enterprise |

Table 2-1: Rhode's *Eight View Elements* for enterprise architecture

## 2.4 Employing the AFE Framework

The AFE Framework brings together the influence of the heuristic principles and the inputs from decision-making theory to develop a spiral framework that considers six major steps, as shown in Figure 2-2.


Figure 2-2: AFE Framework

The steps of the AFE Framework are described by Raby and Zini (2012) in their thesis as follow:

**Pre-architecting Steps:** Clarify the problem, the enterprise's current state and define the specific objectives for the future; it includes understanding the motivation for change, the enterprise landscape, the stakeholder's values, the 'as-is' enterprise, and the future holistic vision.

**Step 1, Architectures Generation:** Guides to develop several candidate architectures using a system thinking perspective. This step focuses only on generating alternatives, not on their evaluation.

**Step 2, Effectiveness Quantification:** Estimates how close each proposed architecture alternative is to what the future enterprise desires to achieve or obtain. Its evaluation considers both: future strategic competencies needed and the future values of multiple stakeholders.

**Step 3, Effort Quantification:** Compares the level of effort required to implement any of the proposed candidate architectures. It allows the consideration of trade-offs that exist among different options.

**Step 4, Risk Quantification:** Assess the level of risk associated with each of the candidate architectures. It identifies uncertainties, likelihoods, and the consequences that unexpected/ unforeseen events might have on different architectures. It provides important complementary information to decision-makers.

**Step 5, Output, Architecture Trade-offs Matrix:** This allows easy visualization of the strengths and weaknesses of the different alternatives. It helps decision-makers and architects reason about architectural decisions by showing informed trade-offs caused by the interaction of multiple elements.

The order is relevant because each step provides inputs for the next one. The model also acknowledges the importance of having an iterative process, where the evaluation of architectures using different dimensions lead inherently to improvements on the initial designs and a better final solution.

# 3 Assessing Current Open Data Architecture

## 3.1 Data Inequality and the Puerto Rico Open Data Enterprise

A study from the Global Open Data Index (GODI) titled “Creating Meaningful Open Data Through Multi-Stakeholder Dialogue” (Lämmerhirt, Rubinstein, Montiel 2017) found that public institutions should align the data with the needs of civil society groups, citizens, and other users. As mentioned in Open Knowledge International’s recent *Data and The City* report (Lämmerhirt, 2017), data infrastructures are not mere “raw” resources. They are the framework to produce and publish data. These data infrastructures are spaces for public participation, in which audiences can use data to engage with public institutions. They can lead to emerging goals focused on transparency, accountability, public participation, public service delivery, technological innovation, and economic growth. Yet, institutions are producing more information encoded in forms that prevent data publishers and public users from communicating

with one another. As the authors state, “dialogue is critical to producing relevant data that can be used by civil society.”

This report also states that data sets published, typically by governments worldwide, are not easy to understand and lack context. Still, most studies on Open Data limitations and challenges focus on accessibility and being in the proper format without considering users' data literacy and skills to explore, visualize, and interpret the liberated data sets. Also, it is not a consideration of the current Open Data architecture in the Government of Puerto Rico; it focuses on satisfying the needs of those stakeholders with the education and skills to work with the published data sets, therefore, extracting benefits for themselves and not available to most.

The lack of data accessibility and understandability within an Open Data enterprise creates data inequality. More recently, there is evidence of significant differences in the purposes that people use information technology, from activities to enhance personal or collective development, including social and economic capital, to the ability to convert this into tangible benefits (Hargittai & Walejko, 2008; Loh & Chib, 2018; Sarkar et al., 2011; van Deursen & van Dijk, 2014). These digital divides – skills/usage and benefits/outcomes – fundamentally shape patterns of inequality in a digital society, as Scheerder, van Deursen, and van Dijk (2017) established in their recent systematic review. Refer to Figure 3-1.

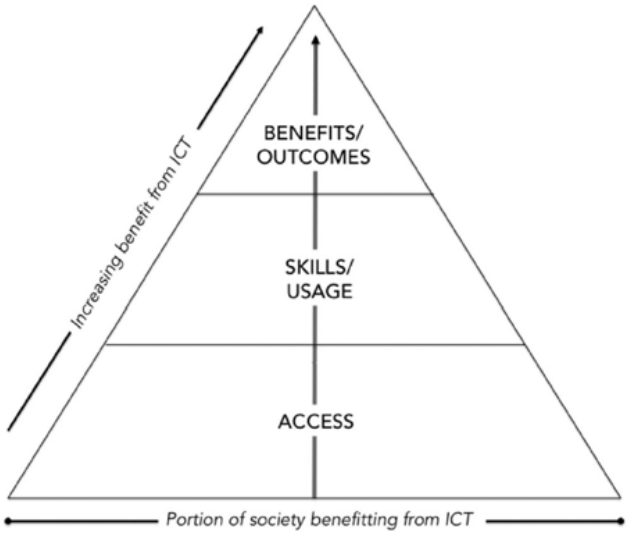



Figure 3-1: Levels of benefits for a portion of society

Puerto Rico is not the exception. Digital divide leads to data inequality. Without the proper skills, citizens cannot benefit from Open Data initiatives. The proposed architecture should use data independent of the user's skills. Otherwise, the benefit will be a few, exacerbating the data inequality gap.

## 3.2 An Eight Steps Architecture Analysis

This assessment of the current Open Data enterprise in Puerto Rico addresses the following elements:

1. State of the Open Data enterprise
2. Stakeholders in the Open Data enterprise
3. Validity and relevance of context
4. Causes and effects of current limitation
5. Generation of a new concept that addresses the limitation
6. A systems architecture that responds to the concept generated

7. The value pathway for the “to-be” architecture
8. A mathematical abstraction of function for the “as-is” versus “to-be” architectures

### 3.2.1 The State of the Open Data enterprise in Puerto Rico

Using Rhode’s Eight View Elements for Enterprise Architecture to assess the Open Data enterprise in Puerto Rico, the following limitations were found:

- *Process*: the focus is on complying with the law to liberate data
- *Strategy*: the strategy fails as Open Data initiatives unintendedly are not designed as a service for all citizens
- *Product*/*Services*: too centric on data and not on servicing the ordinary citizen
- *Organization*: legislature and the executive care mostly about compliance with the law versus the value to its citizens
- *Information*: it is not presented as information but as raw data
- *Knowledge*: presumes that citizens have the skill to process this data to make it valuable and actionable

The current Open Data enterprise focuses on liberating data sets and not on how the data can be used or benefit the citizens of Puerto Rico. Figure 3-2 shows a print screen of the format of a published data set, *Residential Technology Assessment 2016*, taken from the Puerto Rico Statistics Insititute site. The data set has the corresponding metadata and shows a raw, machine-readable format that requires data skills to work with this set and convert it into usable relevant knowledge.

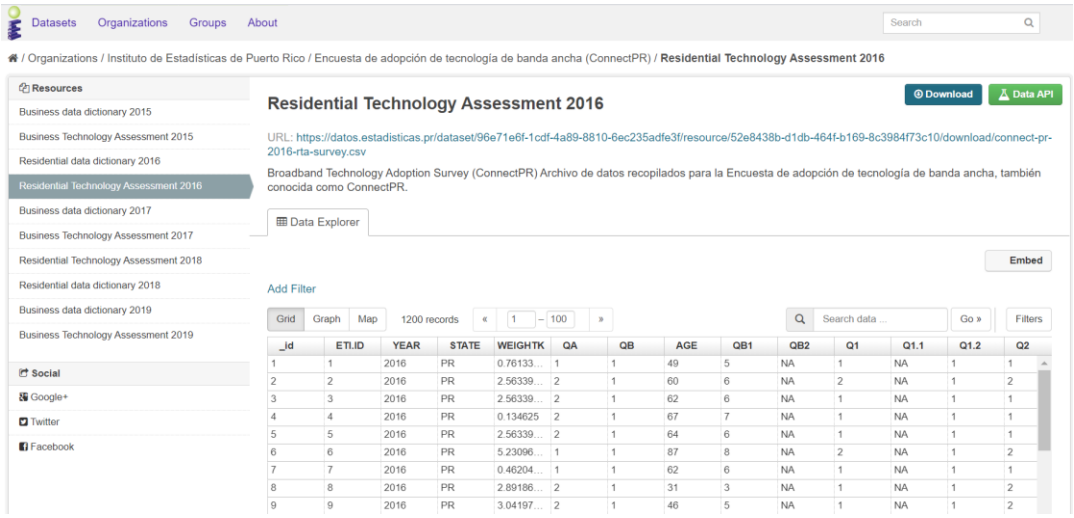



Figure 3-2: Example of the format from a current data set published

Figure 3-3 shows the structure of the enterprise view elements. The elements impact each other within a connected system like government, specifically for the Open Data enterprise.

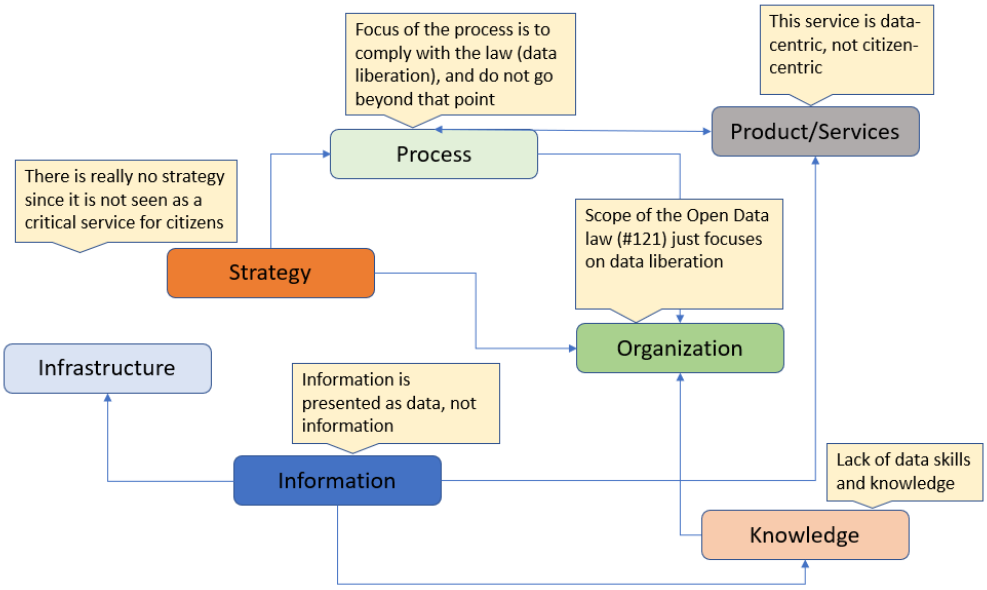



Figure 3-3: Enterprise view elements for Open Data in Puerto Rico

Four of these elements have the most impact: strategy, organization, processes, and knowledge. Figure 3-4 shows specific gaps to address within those elements to

continue improving the Open Data enterprise in Puerto Rico. For example, as part of a focused strategy, the government should not abandon its mission to serve its constituents in every dimension of public service. A culture of *service through data* must be part of the mission.

The Puerto Rico Statistics Institute, as the custodian of Open Data in Puerto Rico at this time, has made good progress in improving some of those gaps from Figure 3-4. However, the highlighted sub-elements within strategy, process, and knowledge make the most impact in increasing citizen participation. Government employees involved in managing data within the Open Data architecture should have the skills to know how to present data in an actionable format. Those should be a priority to improve the Open Data enterprise.

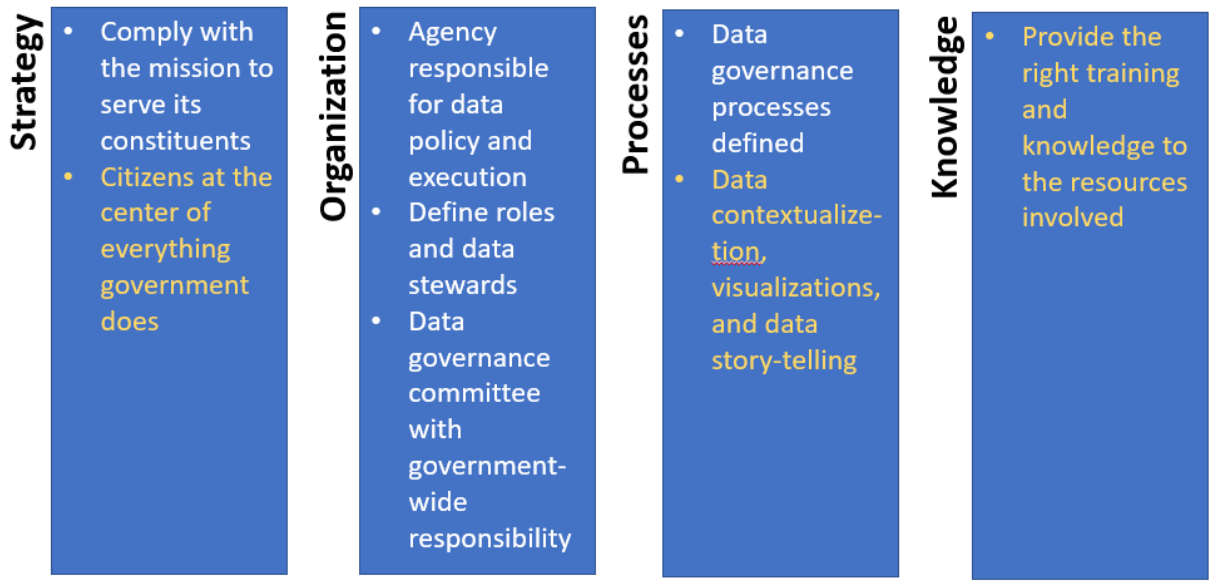



Figure 3-4: Gaps in Puerto Rico's Open Data

The Open Data Enterprise in Puerto Rico consists of data stored in technology systems and disparate databases within the different government agencies. Agencies

use code, databases, and other technical infrastructure instruments to create open data sets. Those agencies act as sources to the Open Data system architecture. Figure 3-5 presents a system decomposition view of the Puerto Rico Government Systems that impact Open Data. As shown, government agencies use three main instruments, code, databases, and basic technical infrastructure (cloud services, security protocols, etc.) to build the data sets sent to the core Open Data systems architecture.

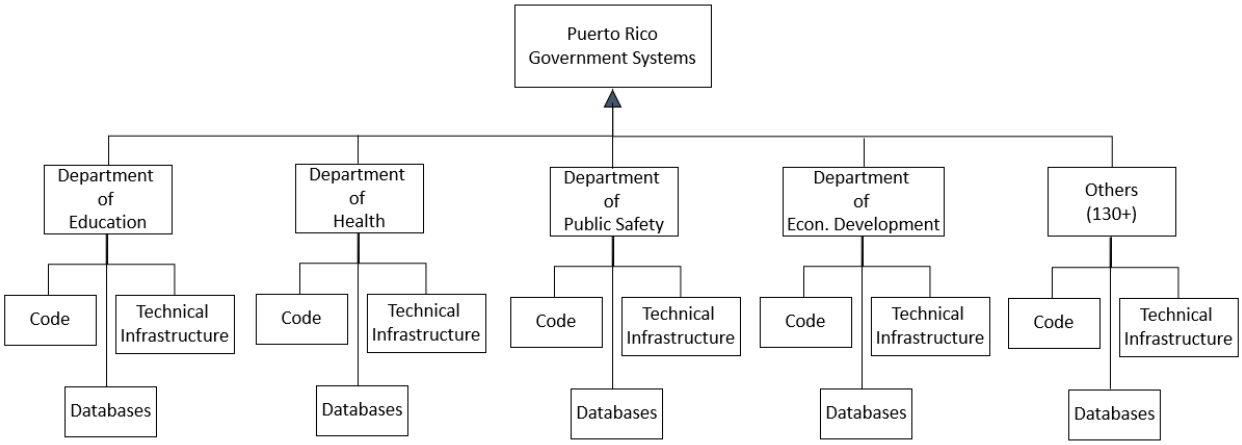



Figure 3-5: Puerto Rico systems decomposition view

The current functional architecture is presented in Figure 3-6. This figure shows the internal operands (data) and processes for the Puerto Rico Open Data Architecture. At present, it uses three main instruments for data transformation until it is ready for publishing within the context of how those data sets are being utilized today, most to comply with the law.

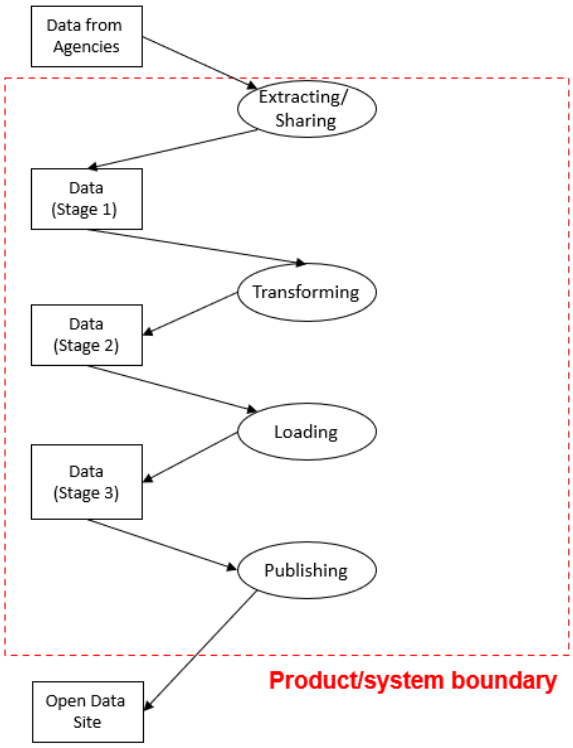



Figure 3-6: Current functional architecture

The following systems architecture structure, Figure 3-7, depicts the current Open Data enterprise in Puerto Rico. It shows a focus on ingesting and preparing data sets for publishing. Those data sets are categorized by topic/theme or source agency within the data access site object (user interface). There are processes to update those data sets as they become available. The Puerto Rico Statistics Institute currently publishes 98 data sets through the official Open Data site for Puerto Rico.

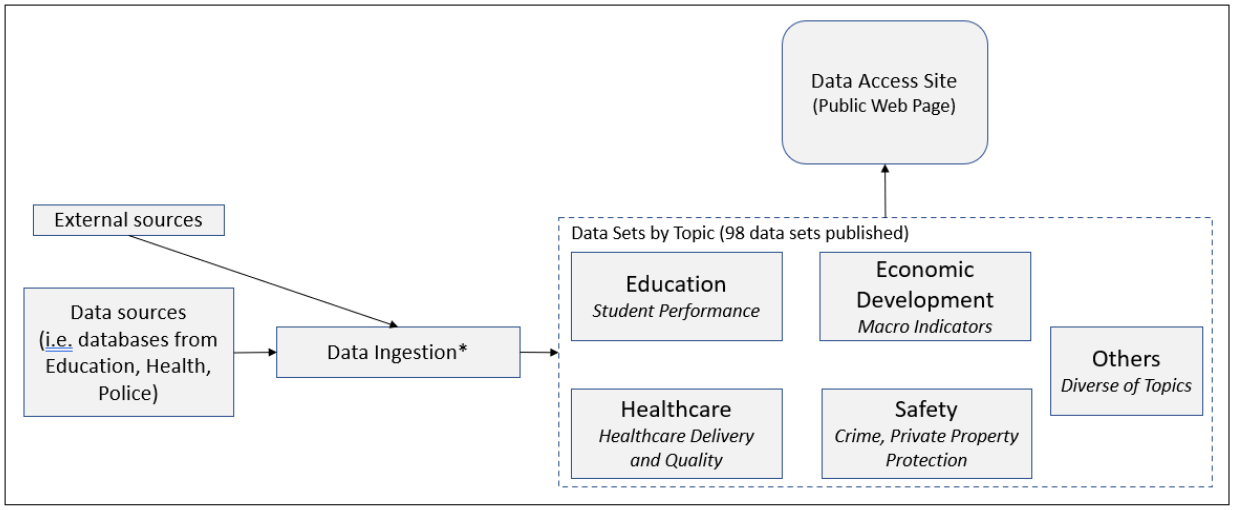



Figure 3-7: Current systems architecture

### 3.2.2 Stakeholders for Open Data in Puerto Rico

One of Vivek Kundra's 10 Principles for Improving Federal Transparency, presented to the House Committee on Oversight and Government Reform on July 14th, 2011, states the importance of providing equal access and incorporating user feedback. Kundra expands on this principle by adding, "this provides a common view of data to **all stakeholders** to foster collaboration, incorporate user feedback to help identify high-value, meaningful data sets, and set priorities to continuously drive and improve future planning and processes." Identifying those users or stakeholders is a critical step when evaluating architectures.

For the Open Data initiative in the Government of Puerto Rico, there are six main stakeholders, as shown in Figure 3-8. Four stakeholders can benefit the most: *citizens, media, academia, and entrepreneurs*. The media has historically been using public domain data to make the government accountable. Media enterprises hire data journalists to extract the benefit from this data. Similarly, academia and entrepreneurs

are stakeholders with the sophistication to benefit from these data sets in their current state. Academia uses this data for research, and entrepreneurs typically create or enhance their startups with public domain data that can help monetize their product or offering. Therefore, for media, academia, and entrepreneurs, the performance of the current enterprise works well due to the technical and data skills they possess in their organizations.

As highlighted in Figure 3-8, ordinary citizens are the most important stakeholder that this Open Data enterprise should serve but are the most overlooked from how Open Data publishes these data sets.

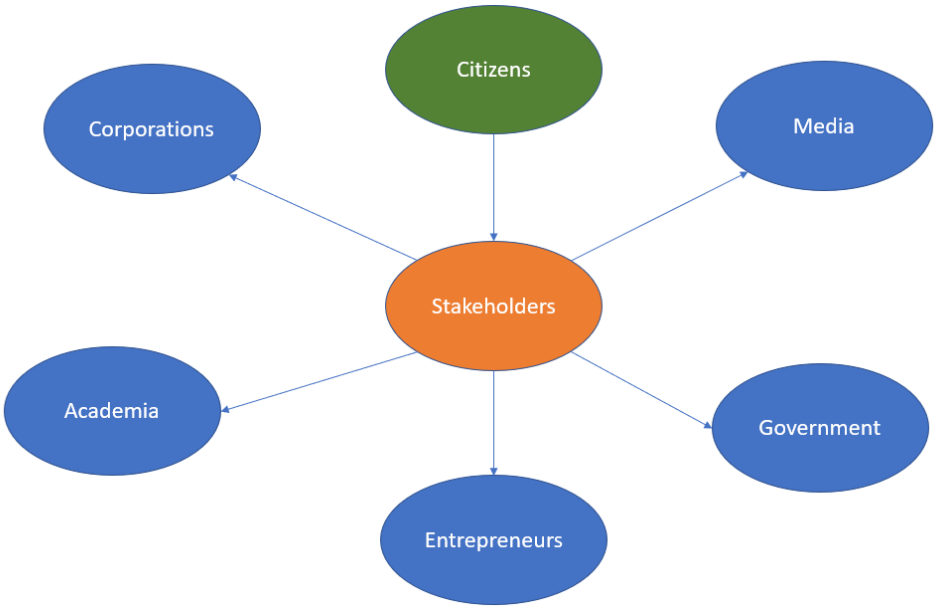



Figure 3-8: Stakeholder identification

The interests presented in Table 3-1 are typically not met (Ubaldi, 2013). These interests are consistent with those discussed with the former and current Chief Information Officer (CIO) of Puerto Rico.

The following table shows the interest for these stakeholders:

| Stakeholder | Main Interest |
|---|---|
| Citizens | To easily understand the meaning of these data sets without having the technical skills and be able to make decisions to benefit their families and living community (decisions such as hospital selection based on mortality rate); open participation and collaboration with government (Ubaldi, 2013) |
| Media | Hold the government accountable |
| Government | Measure internal performance |
| Entrepreneurs | Monetize their product and services with public domain data |
| Academia | Research, collaboration with government, teaching, influence policy |
| Corporations | Influence policy, enhance their product and services, customer segmentation, market research |

Table 3-1: Stakeholders main interests

Figure 3-9 shows the current performance of the enterprise versus its importance for the stakeholders. Specifically for citizens, this performance represents an obstacle to unlocking the value of these data sets that the government of Puerto Rico publishes. Citizens have the best improvement opportunity out of all stakeholders. The government itself has a lot to gain with the improvement potential, as it can help

measure their performance to get better. The other stakeholders have fewer improvement opportunities as the current architecture provides value given their organizations' data skills.

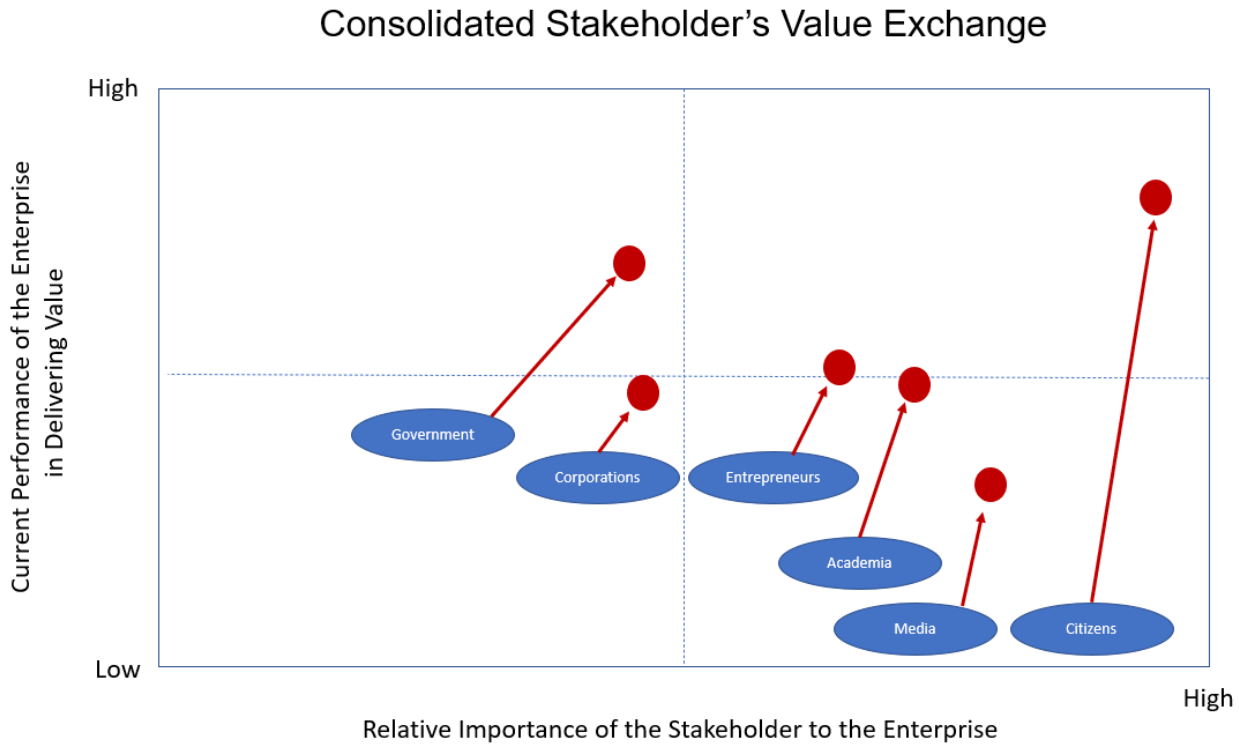



Figure 3-9: Consolidated stakeholder's value exchange

### 3.2.3 Validity of context

The current Open Data enterprise in Puerto Rico was designed based on compliance with the Puerto Rico Open Data law #122 of 2019. It is a valid context but limited. It falls short in meeting the citizens' needs, as presented in Table 3-1. Complying with the Open Data law does not translate into benefits for citizens. As context informs the system's requirements, design, and function, a new context should be defined to place citizens at the center of the Open Data enterprise in Puerto Rico.

Expanding a context beyond compliance with the law and the mere liberation of data sets becomes vital to achieving the desired results. It is not about eliminating what

has been done up to this point or even changing it completely. It is about having new aspirations and setting new goals for Open Data in Puerto Rico. A new goal can be: “to make open data usable for most citizens without requiring to have the technical skills to extract the meaning and insights out this data.” Citizens need information that can translate into knowledge, not raw data. Complying with the law and liberation of data is still very important. Without data available in the first place, there is no Open Data. However, a context focused on the needs of the citizens inevitably derives a new concept for architecture. A new concept extends the functionality of the current enterprise architecture (refer to section 3.5), which then produces a “to-be” Open Data systems architecture that should meet the new goal for Puerto Rico Open Data.

### 3.2.4 Causes and effects

A problem-tree analysis from design thinking complements the assessment of the current architecture, as presented below in Figure 3-10. This framework is helpful because it focuses on maximizing the experience for the end-users. The main problem is at the center of the diagram and shows the cascading causes and effects on multiple levels.

From this analysis, the problem is the lack of usability for citizens of the published data sets, as stated in Section 3.2.2. The problem causes are (from Figure 3-10 as depicted in a parent-child relationship):

- *a focus on data liberation* –the focus is to comply with the law
- *incomplete vision* – it is not based on citizens needs
- *not understanding the audience* – there is no awareness of the skills required to extract value and benefit from this open data

- *limited law (law 121)* – there is no provision in the law to make sure that citizens can benefit from the liberation of this data
- *wrong perception of transparency* – transparency is not just about publishing data; people need to understand what those 98 liberated data sets in the Puerto Rico Open Data architecture mean

From the effects side, the following were identified (from Figure 3-10 as depicted in a parent-child relationship):

- *data inequality* – opens the gap between the ones that can benefit from this data versus those that do not have the skills
- *infrequent access to the published data sets* – many citizens are not accessing these data sets; therefore, the impact is questionable
- *not serving its purpose of accountability and transparency* – with little use, citizens are unaware, which limits accountability and transparency
- *only a few can benefit from the liberated data* – it is the proper architecture for the ones that can work with the data in the current format
- *costly effort without seeing the benefit* – this requires an investment from taxpayers' money for the development and maintenance of this technology

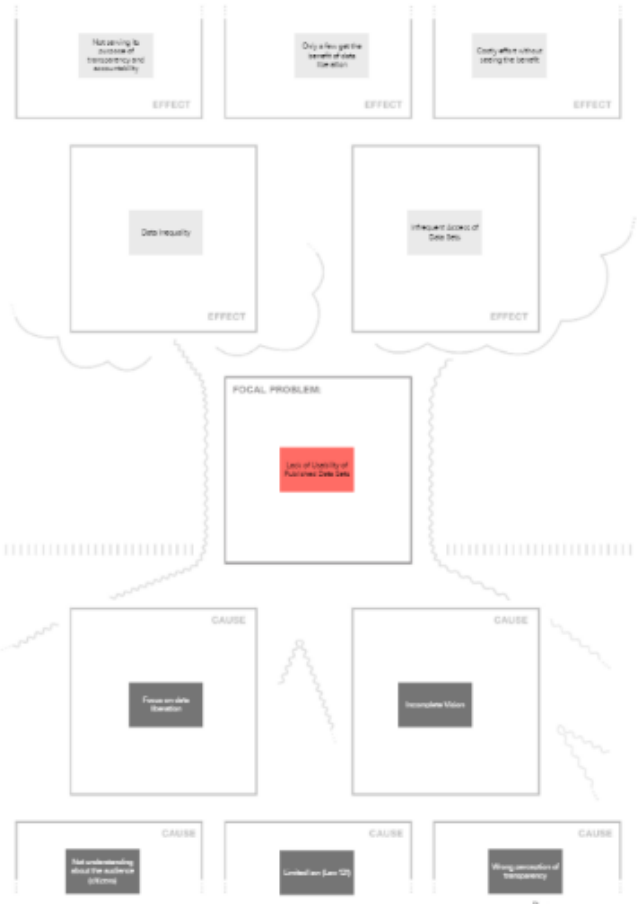



Figure 3-10: Problem-Tree Analysis

### 3.2.5 Generation of a new concept that addresses the limitation

A redefined context can lead to additional requirements and a concept driven by newly set goals to comply with the needs of citizens. The concept in which the Puerto Rico Open Data architecture was designed has limitations in scope, is more data-driven and is less focused on citizens. In these situations, architects should architect “up and down” to confirm needs and goals and validate the current concepts against those needs. Figure 3-11 depicts the activity of architecting “up and down.” It shows how “needs” impact “goals,” “goals” connect with “concepts,” and “concepts” lead to new, modified, or completely redesigned “architecture.”

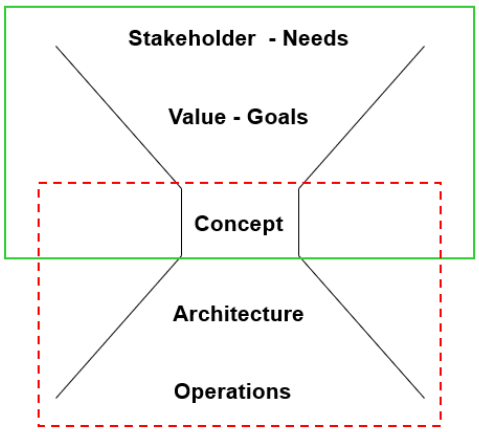



Figure 3-11: Architecting “up or down” (Crawley)

A new concept should be generated for the Puerto Rico Open Data architecture and must comply with the citizens’ needs presented in Table 3-1. Complementary to the current concept of the architecture, it needs to include additional processes to transform the data further. Beyond the data preparation, *contextualizing* is the transformation process required to take the raw data and convert it into a meaningful and easy-to-understand format.

Figure 3-12 presents this concept with two additional objects/instruments marked within the established boundary. These instruments or objects are the code required to transform the data (operand) and the resource that creates the code. These are additional instruments added to the existing “preparing” process. Data preparation is shown in this figure as the process that adds value by changing the data state.

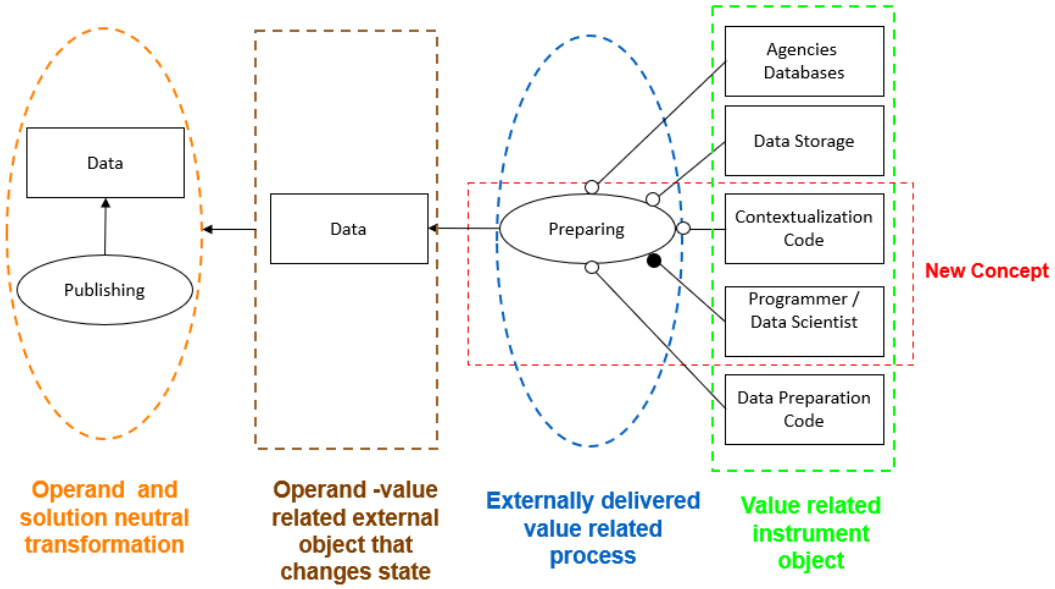



Figure 3-12: Boundary of the new concept within Open Data for Puerto Rico

However, the data preparation process is multi-function, as shown in Table 3-2. Those internal functions add new value to "publishing." The table also shows the form and function of the concept. *Form* is the code developed by the data scientist that uses the data storage infrastructure. *Function* is the contextualization that creates value and benefits the end-users (citizens). This combination presents a new concept within the Open Data architecture: *contextualizing the data with code produced (i.e., by a data scientist) within the data preparation process for general audience clarity and usability*.

| Function: | Internal Function: | Form | |
|---|---|---|---|
| Publishing | Extracting | Extraction Code | |
| | Transforming | Transformation Code | |
| | Loading | Loading Code | |
| | Contextualizing | Contextualization Code<br>Data Storage<br>Data Scientist | Concept Added |

Table 3-2: Multi-function concept

Another way to depict the selected concept is in Figure 3-13. The specific form object (context code developed) uses the generic object form (the data storage infrastructure) to add contextualization to the operand (data). Although the solution-neutral process ("publishing") continues to be the architecture's primary function, it adds a new format for publishing the data. It is the format that should enhance the value for the benefit of the citizens.

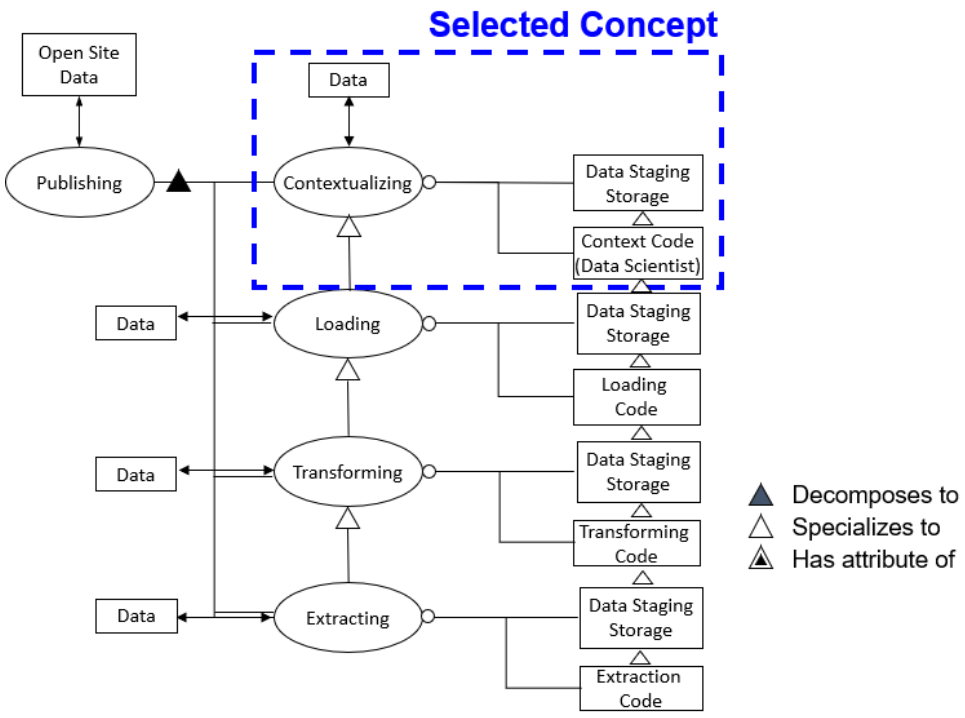



Figure 3-13: Selected concept with the expanded processes

### 3.2.6 A systems architecture that responds to the concept generated

From the context and concept generated, the "to-be" architecture needs to have the following attributes directly linked to its function. Those attributes or "ilities" expected are:

- *reliability* – data needs to be consistent and timely

- *accessibility* – always accessible and updated based on the corresponding latency of the data elements included within each set
- *intelligibility* – consistently in a machine-readable format
- *contextuality*/*usability* – data needs to be useful and transformed into knowledge leveraging visualizations, context, narratives, and data story-telling that derive actionable insights

From the *Foro Virtual de Estadísticas y Tecnología* held on May 7th, 2021, presented by the Puerto Rico Statistics Institute, Puerto Rico has been performing at acceptable levels in terms of reliability, accessibility, and intelligibility. However, contextuality/usability are attributes that do not exist in the current architecture. Table 3-3 presents a comparison of "ilities" of "as-is" architecture with what a "to-be" or proposed architecture should have. The "+" sign means that the "ility" is present in the corresponding architecture, and "None" means it is not present.

| | "As-Is" Architecture | "To-Be" Architecture |
|---|---|---|
| Reliability | + | + |
| Accesability (openness) | + | + |
| Intelligibility | + | + |
| Contextuality / Usability | None | + |

Table 3-3: "Ilities" comparison

A comparison between the "as-is" or current (left) and a potential "to-be" functional architecture (right) is presented in Figure 3-14. This figure shows the internal operands (data) and processes for the Puerto Rico Open Data Architecture. Specifically shown are the newly added contextualization process and stage 4 of data transformation.

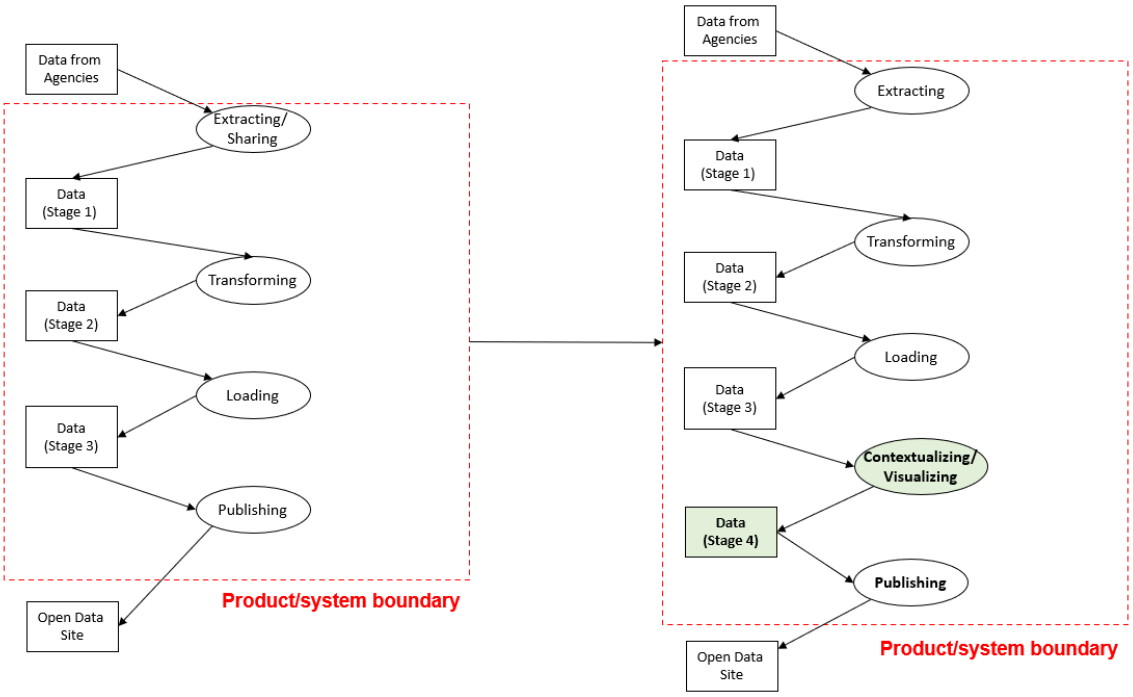



Figure 3-14: Functional architecture comparison

### 3.2.7 The value pathway for the "to-be" architecture

The architecture delivers value when the externally delivered operand (data) changes its state through the action of the processes (processing = data contextualization) enabled by the instrument (code and storage), as shown in Figure 3.15. The value state changes from existing to desired when additional data processing (for easy understanding and interpretation) is completed.

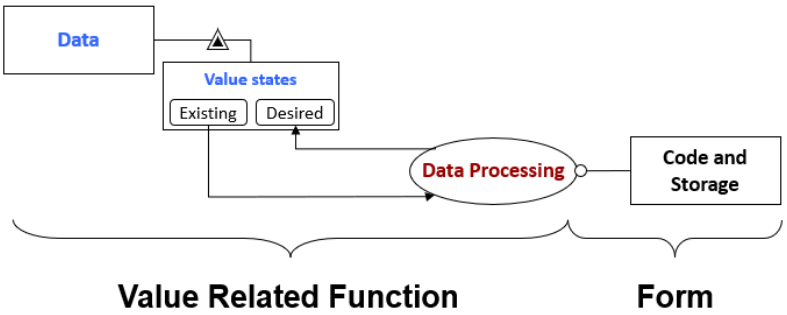



Figure 3-15: Value function and form for architecture comparison

In conclusion, data contextualization produces value and benefits to Open Data. Contextualization can take many forms. Narratives, storytelling, graphical insights, comparisons, and connecting correlated data sets are standard techniques for contextualization. Figure 3-16 presents the evolution to guide a future Open Data architecture design centered around a data format that any individual can consume. Notice the extra layer at the center. This extra layer enables building a citizen-centric open data enterprise versus data-centric as it exists today in Puerto Rico.

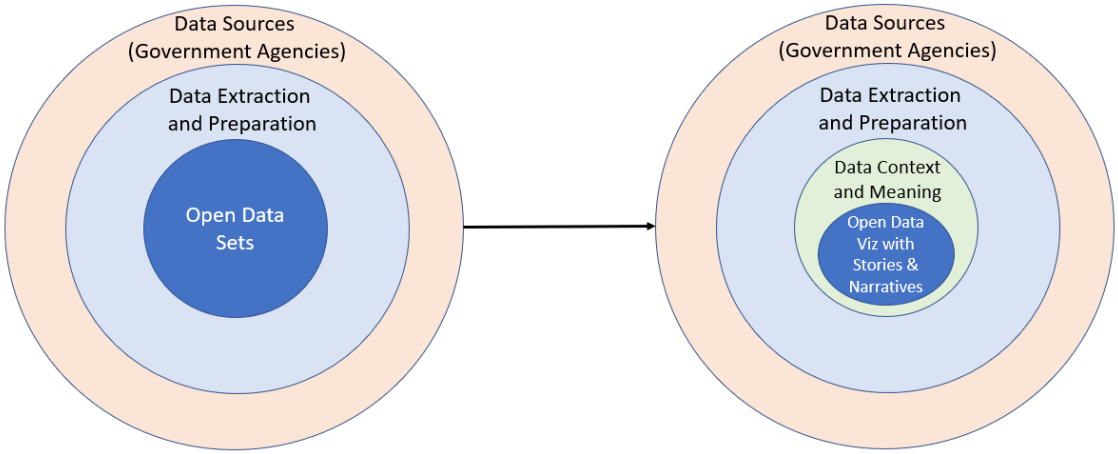



Figure 3-16: Framework for value creation in Open Data

Figure 3-17 shows the resulting systems architecture with contextualization added through data storyboards and visualizations step before publishing within the Open Data site. Also, notice how data sets are interconnected, representing possible correlations, which helps build insights when applying systems thinking into the architecture.

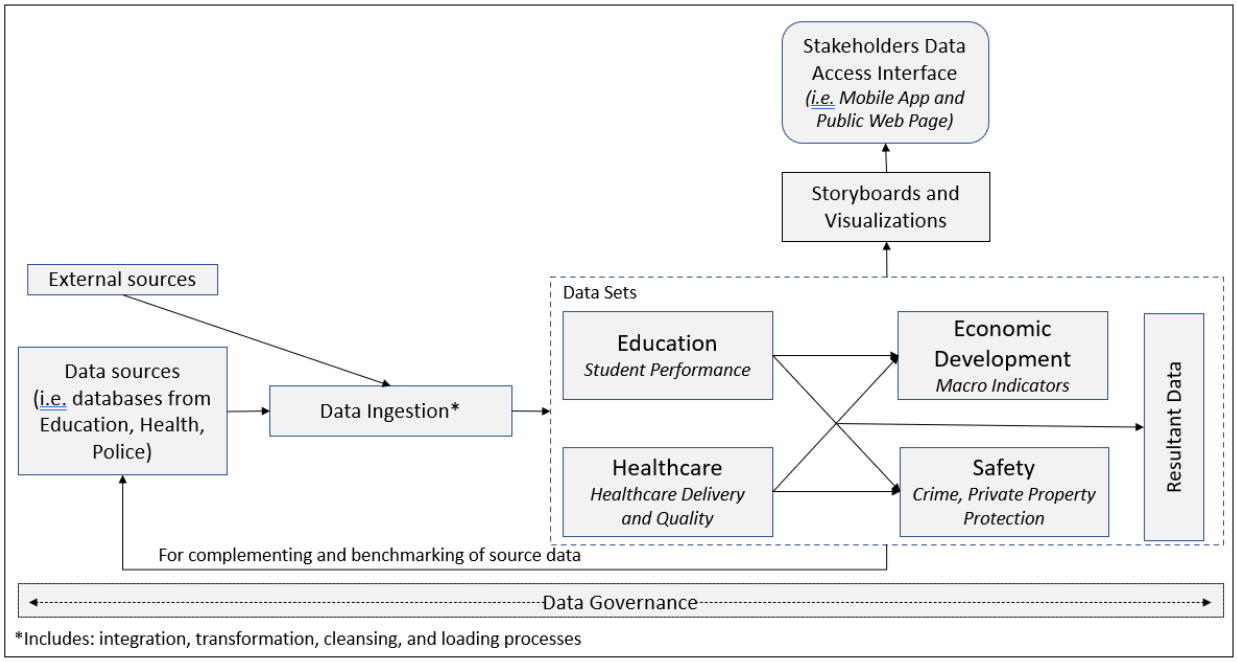



Figure 3-17: “To-be” Open Data architecture structure

### 3.2.8 A mathematical abstraction for the “as-is” versus “to-be” architectures

Open Data architectures have focused on having the processes and technical infrastructure to publish data. The two key attributes that Open Data has focused on providing within its architecture are:

- *Accessibility* of the data set refers to data released and delivered to the architecture for access in the public domain
- *Intelligibility* of the data set refers to applying data engineering processes to prepare the data and produce a consistent and machine-readable format to view, explore, and download.

Today, the Open Data architecture in Puerto Rico is a function of accessibility and intelligibility. Refer to Equation 3-1.

$$OD_{arch} = f(A_s, I_s)$$

Equation 3-1: Open Data architecture as a function of A and I

Where,

$OD_{arch}$ = represents open data architecture

$A_s$ = accessibility of the data sets *s* contained within the architecture

$I_s$ = intelligibility of the data sets *s* contained within the architecture

As explained in sections before, the limitation presented in this architecture is the absence of context applied to the data sets. When considering the citizens as the audience, it is crucial to provide data usability for the people. Otherwise, it is an architecture that only provides access and format or accessibility, A, and intelligibility, respectively. Adding context to the data generates insights and helps formulate an interpretation. It expands the user base to many more citizens in Puerto Rico. A citizen-centric architecture that meets the needs of the ordinary citizen has accessibility, intelligibility, and contextuality. To review the definition of *contextuality* presented in Section 2-1, it is as follow:

- *Contextuality* refers to an additional data processing step to add context that transforms the data into knowledge. When the user acquires new knowledge, especially if it is meaningful, relevant, and actionable, the value and benefit of the data are realized. Contextuality is developed within the architecture (on top of the raw data) by adding visualizations, benchmarks, simple descriptions, narratives to create stories, highlighting insights, among other forms.

We then take Equation 3-1 and expand to add this third attribute of contextuality. Equation 3-2 presents an improved architecture as a function of accessibility A, intelligibility I, and contextuality C.

$$\mathbf{OD_{Arch}' = f(A_s, I_s, C_s)}$$

Equation 3-2: Improved Open Data architecture as a function of A, I, and C

Where,

**$A_s$** = accessibility of the data sets *s* contained within the architecture

**$I_s$** = intelligibility of the data sets *s* contained within the architecture

**$C_s$** = contextuality added to the data sets *s* contained within the architecture

The mathematical representation of the "to be" architecture is presented below as *ODArch'* (prime notation to denote improvement from current architecture). The model contains only multiplicative parameters, in binary format (0, 1), to represent accessibility, intelligibility, and contextuality for each data set *i*. A factor is multiplicative when it has a global effect across the overall system (Boehm, Valerdi, Lane, & Brown, 2005). Its general form is:

$$ODArch' = \sum_{i=1}^{n} (AIC)_i$$

Equation 3-3: Architecture with factors A, I, and C for each data set i

Where,

**ODArch'** = Open Data "to be" architecture

$\mathbf{A_i}$ = accessibility of the data set $i_{th}$, represented as a binary variable, where 1 means that the data set is accessible, and 0 when it is not

$\mathbf{I_i}$ = intelligibility of the data set $i_{th}$, represented as a binary variable, where 1 means that the data set is intelligible, and 0 when it is not

$\mathbf{C_i}$ = contextuality of the data set $i_{th}$, represented as a binary variable, where 1 means that the data set is transformed to add context and interpretation, and 0 when it is not

**n** = number of data sets published

This simple equation can help track that the architecture has considered the required functionality to serve the citizen's needs best. Table 3-4 presents four scenarios:

1. A, I, and C is present yielding AIC equals 1
2. A and I is present, but no contextuality is added, yielding AIC equals 0
3. A is present, but no intelligibility added (the data set is not machine-readable or with quality issues), yielding AIC equals 0
4. A is not present; therefore, no accessibility to the data set (with no access, there is no I or C), yielding AIC equals 0

Table 3-4 also presents an assessment of the situation for each scenario and a recommended action.

| Data Set (i) | Accessibility (A) | Intelligibility (I) | Contextuality (C) | Product (AIC) | Assessment | Recommended Action |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | The goal for each data set in the architecture is to have A, I, and C. | No action. |
| 2 | 1 | 1 | 0 | 0 | This is the case in the current architecture, when there is no contextuality C. | Add contextuality |
| 3 | 1 | 0 | 0 | 0 | This represents a flaw in the architecture with intelligibility. Cannot get to C without I. | Work on data set format |
| 4 | 0 | 0 | 0 | 0 | Data set is not accessible yet. Potential reasons: not delivered by the government agency, or not published yet. | Enforce the law, or complete data preparation process |

Table 3-4: Scenarios for $ODArch'$

The Architecture Readiness Index (ARI), presented below in Equation 3-4, extends $ODArch'$ to create an index that can help quantify the readiness of the architecture to serve its citizens best from the perspective of having considered factors A, I, and C. This equation takes the result of $ODArch'$ and divides it by the number of data sets *n*.

$$ARI = \frac{\sum_{i=1}^{n}(AIC)_i}{n}$$ which equivalent to $$ARI = \frac{(ODArch')}{n}$$

Equation 3-4: Architecture Readiness Index (ARI)

Let's run a scenario to see how this index works and learn the meaning of its output. Table 3-5 presents the architecture of 10 data sets, or n=10. For each data set, the product of (A)(I)(C) is calculated, which is a binary result. The selected scenarios present four cases where data sets have all functionality and the rest where either A, I, or C is missing as represented by a 0 score.

| Data Set (i) | Accessibility (A) | Intelligibility (I) | Contextuality (C) | Product (AIC) |
|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 0 | 0 | 0 |
| 3 | 1 | 1 | 0 | 0 |
| 4 | 1 | 0 | 0 | 0 |
| 5 | 1 | 1 | 1 | 1 |
| 6 | 1 | 1 | 1 | 1 |
| 7 | 1 | 1 | 0 | 0 |
| 8 | 1 | 1 | 1 | 1 |
| 9 | 1 | 0 | 0 | 0 |
| 10 | 1 | 1 | 0 | 0 |
| | | | *Total (ODArch')* | *4* |
| | | | ***ARI*** | ***0.4*** |

Table 3-5: Scenario using ARI

The ARI of the sample architecture in Table 3-5 yields a 0.4 score or that only four data sets met the needs of citizens. With a maximum score of 1, we can conclude that this architecture has a readiness score of 40%. An index score of 1 or 100% should be the goal for most Open Data architectures. No architecture should have an index score less than 1 unless there are valid reasons. The point is that every data set, once accessible and in the proper format, should be contextualized to make it meaningful.

The representation of correlations among data sets from multiple government agencies is a way to capture a system of systems behavior (system thinking element of the architecture). The services and policies established by different government agencies impact each other. Table 3-6 presents how crucial data about education, safety, public health, and economic development can be correlated and contextualized as such. An education system failure can be the result of mediocre public health policy. The lack of a good education hurts the economic development of a community. It is harder for students who drop out of school to find a job. Eventually, it can lead to crime.

Lastly, unemployment can be correlated to high crime rate regions (Raphael & Winter-Ebmer 2001).

| | Education | Safety | Health | Economic Development |
|---|---|---|---|---|
| Education | X | C | C | C |
| Safety | C | X | | C |
| Health | C | | X | |
| Economic Development | C | C | | X |

Table 3-6: DSM for contextualization in "to-be" Open Data architecture

These correlations help in the interpretation and are part of contextuality. As a system of systems with interconnectivity among agencies, there are multiple behaviors within the data that can be relevant. For example, a hike in the crime rate at a municipality can result from a depressed economy in the region. That is the result of a correlation built between the crime data and economic development. The objective is not to prove causation, as correlation is not causality, but valuable correlations make the data more usable through context.

Equation 3-5 *ODArch"* (double prime notation to denote the added improvement from the prime architecture) adds the correlation factor. This factor accounts for the possible correlations between variables within each data set through a correlation coefficient *r*.

$$ODArch'' = \sum_{i=1}^{n} \sum_{j=1}^{n} (AIC)_{ij} r_{ij}$$

Equation 3-5: Architecture with factors A, I, C, and r for each data set *i* and *j*

Where,

$\boldsymbol{ODArch''}$ = improved Open Data architecture from prime (ODArch') version

$\mathbf{A_{ij}}$ = accessibility of data sets $i_{th}$ and $j_{th}$, both represented as a binary variable, where 1 means that the data set is accessible, and 0 when it is not

$\mathbf{I_{ij}}$ = intelligibility of the data sets $i_{th}$ and $j_{th}$, both represented as a binary variable, where 1 means that the data set is intelligible and 0 when it is not

$\mathbf{C_{ij}}$ = contextuality of the data sets $i_{th}$ and $j_{th}$, both represented as a binary variable, where 1 means that the data set is transformed to add context and interpretation, and 0 when it is not

$\mathbf{r_{ij}}$ = coefficient of correlation between applicable variables within each of the data sets $i_{th}$ and $j_{th}$. This value range is between -1.0 and 1.0. A correlation of -1.0 shows a perfect negative correlation, while a correlation of 1.0 shows a perfect positive correlation. A correlation close to 0 indicates no linear relationship between two variables or anything in between from entirely negative to a perfect positive correlation. **NOTE:** Perfect correlations, in this case, can occur when calculating r for the same variables within the same data set files (i.e., data set $i_{th}$ = data set $j_{th}$). These correlations should be excluded or deducted from the total. This adjustment is not taken into account in Equation 3-5.

**n** = number of data sets published

We are assuming that correlation *r* within $ODArch''$ applies when Contextuality *C* equals 1. When *C*=1, the data is published with context. In other words, correlations

happen as part of a contextuality analysis to add more meaning to the data. If there is no Accessibility *A*, Contextuality *C* does not occur. Similarly, if there is no Intelligibility I, there is no Contextuality *C* either. From Table 3-5, only when the AIC = 1, correlation *r* applies.

Correlations within the architecture are an additional dimension of contextuality. In $ODArch''$ correlation *r* is presented separately from contextuality as correlations do not always exist between data sets. Meaningful correlations do not exist for most data sets. Moreover, the objective of $ODArch''$ is to mathematically represent correlations in the relationship between data sets in the architecture. This calculation needs meaning by creating a comparison through a ratio or index.

To add meaning to $ODArch''$, divide by $ODArch'$. It creates a ratio or a percentage of the overall correlation of data sets within the architecture. The resultant provides a directional sense of the level of correlation in the architecture and does not judge if it is good or bad. Of course, and if properly executed, more correlations are synonyms of better contextualization, which means more meaningful communications to the public. The relationship in Equation 3-6 is the *correlated contextualization level* (CCL). NOTE: The equation takes the absolute value of the correlation coefficient *r* (versus using $r^2$) to eliminate the negative correlations and enable a relationship with the denominator without significant complexity.

$$CCL = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (AIC)_{ij} |r_{ij}|}{ODArch'}$$

Equation 3-6: Measures the correlated contextualization level within the architecture

To give an example, refer to Table 3-7. As a hypothetical example, this table shows an architecture with n=10 or ten datasets. Then, it offers some correlations between specific data sets, represented with the coefficient *r* in the intersection cell of data set i and j (green cell). These data sets are the ones that in Table 3-5 showed contextualization or C = 1 (data sets 1, 5, 6, and 8). The actual correlation coefficient r for the applicable intersection in this table is not as important, just that there is a good and well-thought-out correlation derived.

| Data Set (i, j) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | r | | | r | | |
| 2 | | | | | | | | | | |
| 3 | | | | | | | | | | |
| 4 | | | | | | | | | | |
| 5 | r | | | | | r | | | | |
| 6 | r | | | | | | | r | | |
| 7 | | | | | | | | | | |
| 8 | r | | | | | r | | | | |
| 9 | | | | | | | | | | |
| 10 | | | | | | | | | | |

Table 3-7: Correlation relationship between specified data sets

Table 3-8 runs a scenario for Equation 3-6 with the same architecture where n = 10. Out of the 10 data sets, there is a correlation for 4. Therefore, n = 4 for the eligible data sets, instead of n = 10. The other data sets do not qualify for correlation as contextualization or C = 0.

| Data Set (i) | Accessibility (A) | Intelligibility (I) | Contextuality (C) | Product (AIC) | r | \|r\| |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 0.7 | 0.7 |
| 2 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 1 | 1 | 0 | 0 | 0 | 0 |
| 4 | 1 | 0 | 0 | 0 | | 0 |
| 5 | 1 | 1 | 1 | 1 | -0.8 | 0.8 |
| 6 | 1 | 1 | 1 | 1 | 0.75 | 0.75 |
| 7 | 1 | 1 | 0 | 0 | 0 | 0 |
| 8 | 1 | 1 | 1 | 1 | 0.9 | 0.9 |
| 9 | 1 | 0 | 0 | 0 | 0 | 0 |
| 10 | 1 | 1 | 0 | 0 | 0 | 0 |
| | | | | *Total (ODArch')* | | *3.15* |
| | | | | **CCL** | | **78.8%** |

Table 3-8: Hypothetical example for CCL

The result shows a correlated contextualization level (CCL) of ~79% (3.15 / 4.00). NOTE: The correlation coefficient numbers presented are hypothetical numbers to demonstrate the example. Four out of the ten data sets in the architecture have a noticeable correlation. These correlations cannot be forced to increase the CCL. As stated before, close to zero correlations among data sets do not make the architecture less effective. However, if those are present, architects need to make every effort to present them through the contextualization mechanism and method chosen.

# 4 Generating and Analyzing Alternatives for the "To-Be" Architecture

Using the suggested architecture presented in Figure 3-17 and based on how to serve the stakeholders' needs best, three alternatives address the identified gaps as part of this work.

The architecture in Figure 4-1, or *Alternative A*, leverages the resources and organization of the government. The particularity is that this alternative is to continue running, maintaining, and hosting by the Government of Puerto Rico. Within the government, the Puerto Rico Statistics Institute should continue to be the organization to take on this task because of the autonomy as established by its organic law. It is essential to maintain the independence of special interests from the government. The data assets need to stay with the highest level of integrity. These efforts should also be in collaboration with the Puerto Rico Innovation and Technology Services (PRITS).

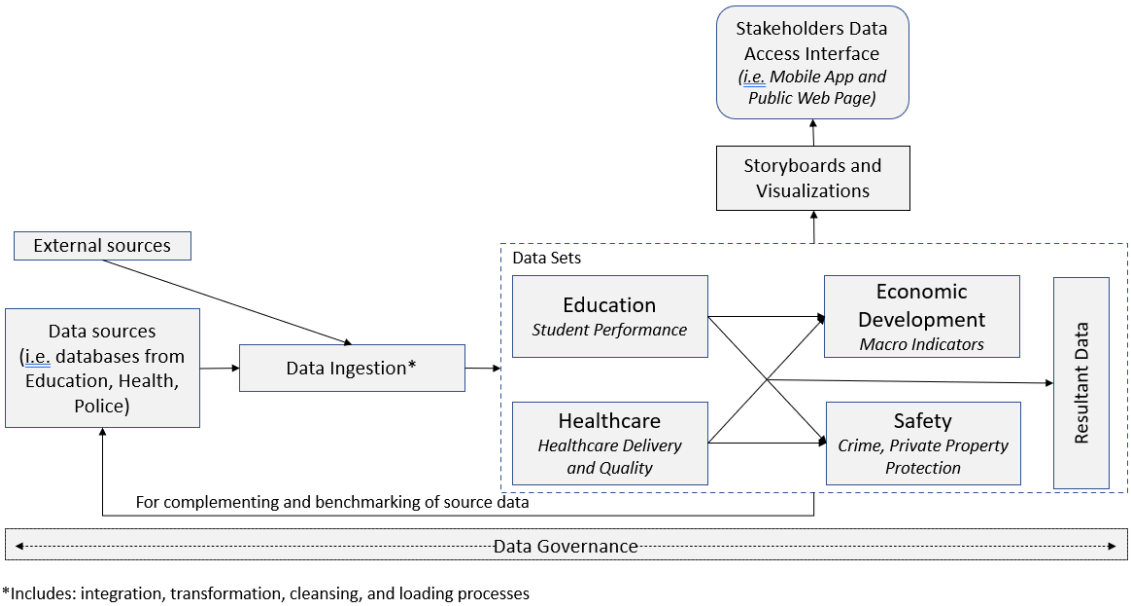



Figure 4-1: Alternative A managed by the Government of Puerto Rico

Alternative B, shown in Figure 4-2, differs in which the demarked components are executed by an external organization, typically a not-for-profit group representing multiple sectors (multi-sectorial organization or MSO). The government would continue to obtain the data sets from the corresponding government agencies. The data sets would be delivered in a pre-agreed frequency and data transfer mechanism (i.e., secured FTP, among other options). The MSO would be responsible for executing all the required data tasks to publish the results in the required format and specifications of the "to-be" architecture presented in Section 3.2.7.

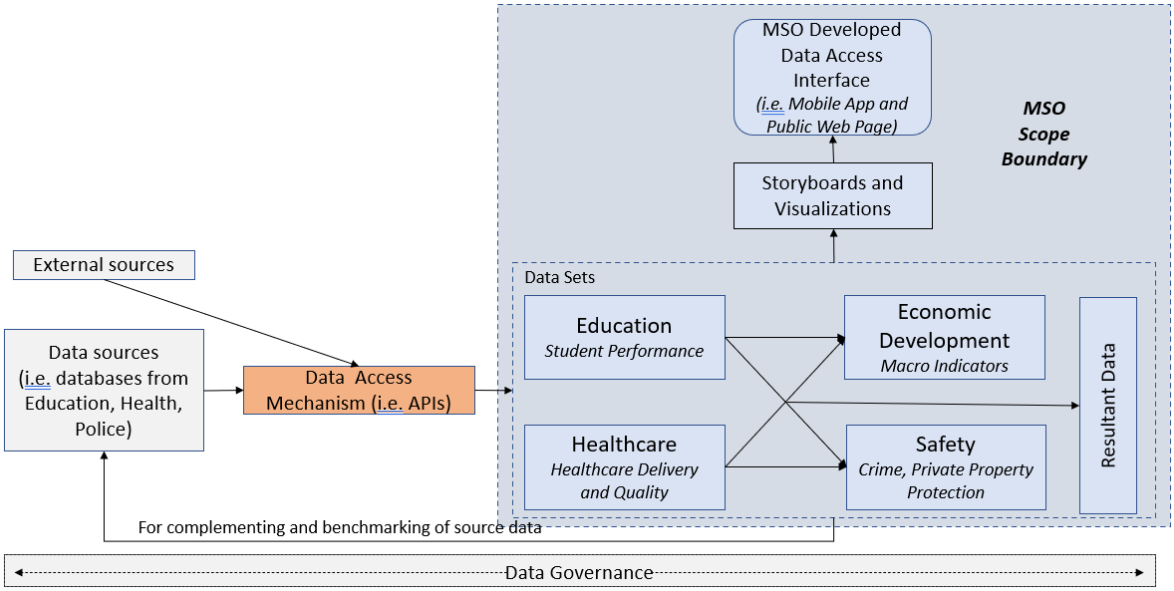



Figure 4-2: Alternative B managed by a Multi-Sectorial Organization (MSO)

Alternative C presented in Figure 4-3 gives the responsibility of managing the Open Data architecture to the citizens. It means any citizen would have access to the raw data sets responsible for deriving meaning, adding contextualization, proper visualizations, and narratives to make the data useful for the general audience. The

government will not have much control of the output. Still, it will continue to be responsible for obtaining the data sets from government agencies, executing data preparation tasks, uploading in a data storage infrastructure, and providing visualization and communication tools. Citizens would perform all the required data tasks required to publish within the specifications and contextualized format of the "to-be" architecture presented in Section 3.2.7.

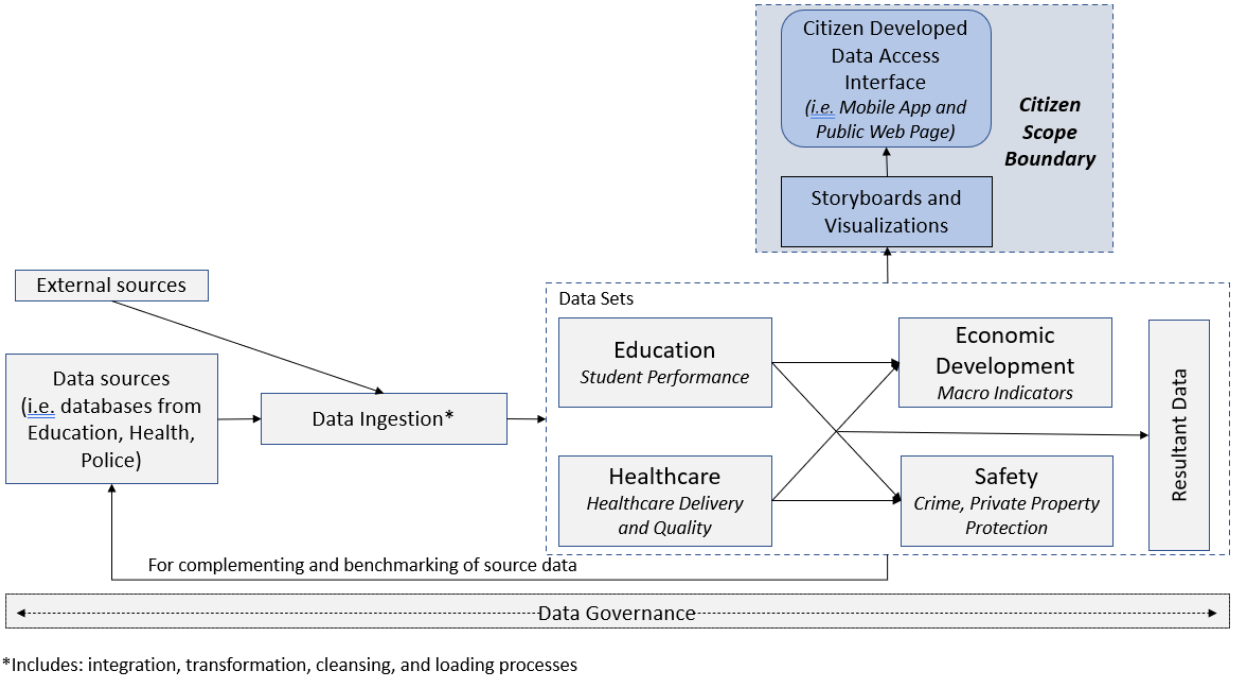



Figure 4-3: Alternative C managed by citizens

## 4.1 Analyzing alternatives for the "to-be" architecture

In the form of "ilities," competencies are required to ensure that the alternative architectures produce the desired results. For the "envisioned future state," and as strategic competencies, Figure 4-4 states that data must be accessible and relevant within a proper infrastructure that enables Open Data requirements. "Ilities" such as

usability, availability, and reliability contribute to compliance with the strategic competencies needed to build the future state.

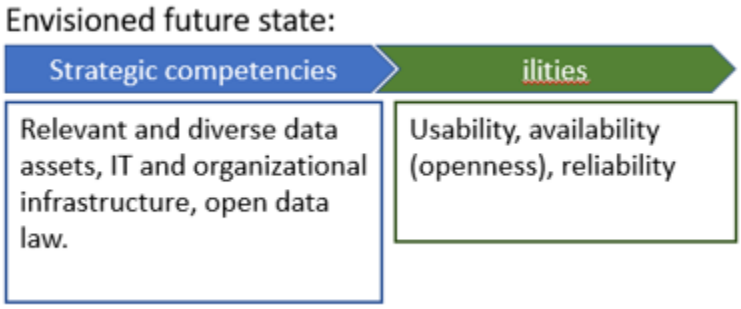



Figure 4-4: Envisioned future state

For the stakeholder's future values, in Figure 4-5 below, it is essential to present the data to enable decision-making. Data needs to be updated, cleaned, and shown usefully.

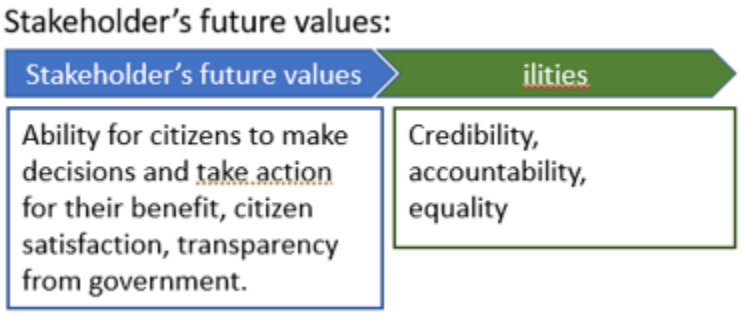



Figure 4-5: Stakeholder's future value

Each of the "ilities" identified is crucial for the success of the selected alternative.

- *Availability* – the data needs to be available and accessible in a timely fashion
- *Reliability* – the data needs to be consistent in its format and free of errors as much as possible

- *Usability* – the data needs to be published in a format that is meaningful, easy to understand, and useful for most citizens
- *Credibility* – the insights presented need to reflect the truth of what the data says and not be manipulated for convenience
- *Accountability* – enable participation and collaboration to monitor public policy and execution by government
- *Equality* – the architecture should not create the conditions to displace citizens from disadvantaged educational and socio-economic backgrounds that do not have the skills to understand what Open Data produces

Each "ility" can emerge successfully. Figure 4-6 presents how it can process the data reliably once the architecture complies with its availability condition. From there, a design with usability is possible. If the usability design is truthful, reflecting what the data yields, maximizing integrity, and minimizing biases, it earns credibility with citizens. With credibility, accountability is enforced, and ultimately, the architecture can minimize data inequality.

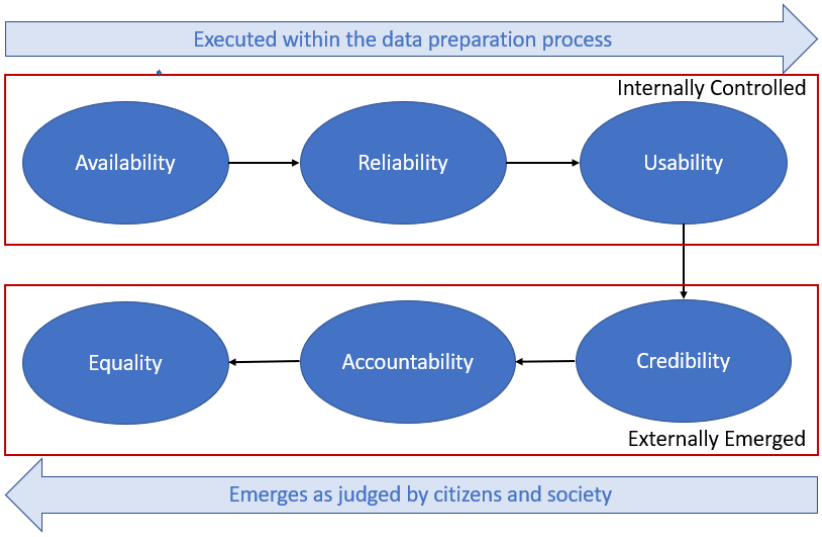


Figure 4-6: Interdependencies of “ilities”

Table 4-1 gives more weight to equality and usability in the effectiveness quantification analysis. Credibility is the second *ility* most weighted for effectiveness. Alternative A shows to be the most effective, followed by Alternative B and then C. It means that for this dimension, giving the full responsibility of the architecture within the Government of Puerto Rico, particularly within the Puerto Rico Statistics Institute, and in collaboration with PRITS, produces the highest effectiveness score. The hypothesis is that the architecture is very fluid from the process perspective. In other words, the end-to-end process, from obtaining the data to publishing it, relies on the responsibility of the same institution or team. It provides more control in the process and more cohesive execution.

| Criteria | Weight | Total Weight | Alternative A | Alternative B | Alternative C |
|---|---|---|---|---|---|
| **Credibility** | **15%** | | 0.69 | 0.45 | 0.36 |
| Does the proposed architecture improve the credibility with government? | 60% | 9.0% | 5 | 3 | 2 |
| Can it keep the data sets updated (updatability) as agreed per policy? | 40% | 6.0% | 4 | 3 | 3 |
| **Equality (data)** | **15%** | | 0.6 | 0.75 | 0.54 |
| Does the proposed architecture reduces the data inequality? | 60% | 9.0% | 4 | 5 | 4 |
| Can it be contextualized to make it relevant to most citizens and communities? | 40% | 6.0% | 4 | 5 | 3 |
| **Accountability** | **10%** | | 0.38 | 0.42 | 0.36 |
| Is it best to make the government accountable? | 60% | 6.0% | 3 | 5 | 4 |
| Can it create a performance management system to have government metrics? | 40% | 4.0% | 5 | 3 | 3 |
| **Usability** | **25%** | | 1 | 1.1 | 0.75 |
| Can the proposed architecture promote the best experience for its users? | 60% | 15.0% | 4 | 4 | 3 |
| Can it make it relevant to its users? | 40% | 10.0% | 4 | 5 | 3 |
| **Availability (data openness)** | **10%** | | 0.5 | 0.3 | 0.3 |
| Can the proposed architecture be in the best position to influence data openness? | 60% | 6.0% | 5 | 3 | 3 |
| Can it make fast and agile the publishing of new data sets? | 40% | 4.0% | 5 | 3 | 3 |
| **Reliability** | **10%** | | 0.4 | 0.46 | 0.3 |
| Does it promote good governance processes to increase reliability? | 60% | 6.0% | 4 | 5 | 3 |
| Does the proposed architecture provides too much freedom that can affect the true meaning of the data? | 40% | 4.0% | 4 | 4 | 3 |
| **Total Effectiveness Score** | | | **3.57** | **3.48** | **2.61** |
| **Ranking** | | | **1** | **2** | **3** |

Table 4-1: Effectiveness quantification

The effort required is another element analyzed for these alternate architectures. With a score of 1.35, the effort is less with Alternative A. The government has possession of the data, controls its governance, accountability over the independent owners, the infrastructure in place, and the resources to carry on this initiative. Refer to Table 4-2 to see the analysis of effort for Alternative A.

***Alternative A: "Driven by Government"***

| Effort Driver | Sub-Factors | Very Low | Low | Nominal | High | Very High |
|---|---|---|---|---|---|---|
| **People's Drivers** | **Score** | | | **2.12** | | |
| *Leadership Support* | Motivation | 1.86 | 1.37 | 1.0 | 0.78 | 0.61 |
| | Data culture | | | | | |
| | Familiarity/trust architecture | | | | | |
| *Citizen Acceptance* | Trust in the data and the purpose | 1.70 | 1.30 | 1.0 | 0.82 | 0.66 |
| | Understand the value of the data | | | | | |
| | Involvement through use and feedback | | | | | |
| *Staff/Contract Capacity* | Current internal competency | 1.42 | 1.19 | 1.0 | 0.88 | 0.76 |
| | Attract talent | | | | | |
| | Qualified external vendor | | | | | |
| **Complexity Drivers** | **Score** | | | **1.00** | | |
| *# Stakeholders Involved* | Number of stakeholders | 0.66 | 0.82 | 1.0 | 1.30 | 1.70 |
| | Diversity of stakeholders and their interest | | | | | |
| *Level of Change* | Level of consensus | 0.66 | 0.82 | 1.0 | 1.30 | 1.70 |
| | Policy/regulation/legal changes | | | | | |
| *Architecture Complexity* | Interdependencies | 0.66 | 0.82 | 1.0 | 1.30 | 1.70 |
| | Coordination | | | | | |
| | Requirements ambiguity | | | | | |
| **Operational Drivers** | **Score** | | | **0.927** | | |
| *Schedule Contraints* | Time to implement | 0.66 | 0.82 | 1.0 | 1.22 | 1.49 |
| | Intermediate goals | | | | | |
| *People Involvement* | Knowledge gaps | 0.76 | 0.88 | 1.0 | 1.19 | 1.42 |
| | Support for the solution | | | | | |
| *Infrastructure Investment* | Cloud and technology investment | 0.76 | 0.88 | 1.0 | 1.19 | 1.42 |
| **Technical Drivers** | **Score** | | | **0.686** | | |
| *Technology Maturity* | Tools and methodologies maturity | 1.28 | 1.13 | 1.0 | 0.88 | 0.78 |
| *Integration Complexity* | Diverse data systems | 0.78 | 0.88 | 1.0 | 1.13 | 1.28 |
| | Data readily available for integration at the source | | | | | |
| | Data governance policy and processes | | | | | |
| **Total Score** | | | | **1.35** | | |

Table 4-2: Alternative A effort analysis

With a score of 2.14, the effort is more impactful with Alternative B. This organization will still depend on the government to provide the data, provide infrastructure and resources, including data expertise and money. It will not be easy for this non-for-profit to raise private money for this effort. In addition, it will need to hire or contract expertise to do the technical work. Refer to Table 4-3 to see the analysis of effort for Alternative B.

***Alternative B: "Driven by MSO"***

| Effort Driver | Sub-Factors | Very Low | Low | Nominal | High | Very High |
|---|---|---|---|---|---|---|
| **People's Drivers** | **Score** | | | **0.82** | | |
| *Leadership Support* | Motivation | 1.86 | 1.37 | 1.0 | 0.78 | 0.61 |
| | Data culture | | | | | |
| | Familiarity/trust architecture | | | | | |
| *Citizen Acceptance* | Trust in the data and the purpose | 1.70 | 1.30 | 1.0 | 0.82 | 0.66 |
| | Understand the value of the data | | | | | |
| | Involvement through use and feedback | | | | | |
| *Staff/Contract Capacity* | Current internal competency | 1.42 | 1.19 | 1.0 | 0.88 | 0.76 |
| | Attract talent | | | | | |
| | Qualified external vendor | | | | | |
| **Complexity Drivers** | **Score** | | | **2.21** | | |
| *# Stakeholders Involved* | Number of stakeholders | 0.66 | 0.82 | 1.0 | 1.30 | 1.70 |
| | Diversity of stakeholders and their interest | | | | | |
| *Level of Change* | Level of consensus | 0.66 | 0.82 | 1.0 | 1.30 | 1.70 |
| | Policy/regulation/legal changes | | | | | |
| *Architecture Complexity* | Interdependencies | 0.66 | 0.82 | 1.0 | 1.30 | 1.70 |
| | Coordination | | | | | |
| | Requirements ambiguity | | | | | |
| **Operational Drivers** | **Score** | | | **1.047** | | |
| *Schedule Contraints* | Time to implement | 0.66 | 0.82 | 1.0 | 1.22 | 1.49 |
| | Intermediate goals | | | | | |
| *People Involvement* | Knowledge gaps | 0.76 | 0.88 | 1.0 | 1.19 | 1.42 |
| | Support for the solution | | | | | |
| *Infrastructure Investment* | Cloud and technology investment | 0.76 | 0.88 | 1.0 | 1.19 | 1.42 |
| **Technical Drivers** | **Score** | | | **1.130** | | |
| *Technology Maturity* | Tools and methodologies maturity | 1.28 | 1.13 | 1.0 | 0.88 | 0.78 |
| *Integration Complexity* | Diverse data systems | 0.78 | 0.88 | 1.0 | 1.13 | 1.28 |
| | Data readily available for integration at the source | | | | | |
| | Data governance policy and processes | | | | | |
| **Total Score** | | | | **2.14** | | |

Table 4-3: Alternative B effort analysis

With a score of 5.82, the effort is more impactful with Alternative C. This organization will still depend on the government to provide the data. However, it will be difficult to articulate a reliable and consistent governance structure for this architecture. Credibility typically is affected in this model, as its output is not moderated or curated. These are barriers to overcome; however, it takes time. Refer to Table 4-4 to see the analysis of effort for Alternative C.

***Alternative C: "Driven by Citizen"***

| Effort Driver | Sub-Factors | Very Low | Low | Nominal | High | Very High |
|---|---|---|---|---|---|---|
| **People's Drivers** | **Score** | | | **1.28** | | |
| *Leadership Support* | Motivation | 1.86 | 1.37 | 1.0 | 0.78 | 0.61 |
| | Data culture | | | | | |
| | Familiarity/trust architecture | | | | | |
| *Citizen Acceptance* | Trust in the data and the purpose | 1.70 | 1.30 | 1.0 | 0.82 | 0.66 |
| | Understand the value of the data | | | | | |
| | Involvement through use and feedback | | | | | |
| *Staff/Contract Capacity* | Current internal competency | 1.42 | 1.19 | 1.0 | 0.88 | 0.76 |
| | Attract talent | | | | | |
| | Qualified external vendor | | | | | |
| **Complexity Drivers** | **Score** | | | **2.21** | | |
| *# Stakeholders Involved* | Number of stakeholders | 0.66 | 0.82 | 1.0 | 1.30 | 1.70 |
| | Diversity of stakeholders and their interest | | | | | |
| *Level of Change* | Level of consensus | 0.66 | 0.82 | 1.0 | 1.30 | 1.70 |
| | Policy/regulation/legal changes | | | | | |
| *Architecture Complexity* | Interdependencies | 0.66 | 0.82 | 1.0 | 1.30 | 1.70 |
| | Coordination | | | | | |
| | Requirements ambiguity | | | | | |
| **Operational Drivers** | **Score** | | | **2.062** | | |
| *Schedule Contraints* | Time to implement | 0.66 | 0.82 | 1.0 | 1.22 | 1.49 |
| | Intermediate goals | | | | | |
| *People Involvement* | Knowledge gaps | 0.76 | 0.88 | 1.0 | 1.19 | 1.42 |
| | Support for the solution | | | | | |
| *Infrastructure Investment* | Cloud and technology investment | 0.76 | 0.88 | 1.0 | 1.19 | 1.42 |
| **Technical Drivers** | **Score** | | | **0.994** | | |
| *Technology Maturity* | Tools and methodologies maturity | 1.28 | 1.13 | 1.0 | 0.88 | 0.78 |
| *Integration Complexity* | Diverse data systems | 0.78 | 0.88 | 1.0 | 1.13 | 1.28 |
| | Data readily available for integration at the source | | | | | |
| | Data governance policy and processes | | | | | |
| **Total Score** | | | | **5.82** | | |

Table 4-4: Alternative C effort analysis

Table 4-5 presents the final quantification of the effort for each of the alternate architectures. As the ranking shows, Alternative A gives the lowest score for effort, followed by Alternative B and then C.

| | Architecture A | Architecture B | Architecture C |
|---|---|---|---|
| Total Effort Score | 1.35 | 2.14 | 5.82 |
| Ranking | 1 | 2 | 3 |

Table 4-5: Effort quantification

As part of the analysis of architectures, some uncertainties need to be addressed, their likelihood, and their impact. These uncertainties are internal or external. Internal uncertainties can be controlled or minimized more easily versus external uncertainties outside the control of the "architect" or group responsible. Refer to Table 4-6 for a qualitative analysis of uncertainties, which are then quantified in Table 4-7.

| Uncertainty | Architecture A | | Architecture B | | Architecture C | |
|---|---|---|---|---|---|---|
| | Likelihood | Impact | Likelihood | Impact | Likelihood | Impact |
| **Internal** | | | | | | |
| Lack of data liberation | Unlikely | Severe | Unlikely | Severe | Unlikely | Severe |
| Bad quality of data | Medium | Severe | Medium | Severe | Medium | Severe |
| Inadequate infrastructure and resources | Unlikely | Negative | Likely | Negative | Likely | Negative |
| Improper management of data (governance, selection, interpretation) | Unlikely | Negative | Unlikely | Negative | Likely | Negative |
| **External** | | | | | | |
| Changes or elimination of open data law to limit transparency | Unlikely | Catastrophic | Unlikely | Catastrophic | Unlikely | Catastrophic |
| Loss of public interest | Unlikely | Severe | | Severe | | Severe |
| Media miscoverage | Medium | Negative | Unlikely | Negative | Unlikely | Negative |
| Lack of awareness of its existance | Medium | Negative | Unlikely | Negative | Unlikely | Negative |

Table 4-6: Likelihood versus Impact

Table 4-7 below shows that all three alternatives yielded a similar risk profile when quantifying the uncertainties. Alternative C presents a slightly higher risk due to governance issues (internal). Alternative A and B show two "high" denominations, but also those risks were concentrated in fewer elements than Alternative C.

| Uncertainty | Architecture A | | Architecture B | | Architecture C | |
|---|---|---|---|---|---|---|
| | Risk | Score | Risk | Score | Risk | Score |
| **Internal** | | | | | | |
| Lack of data liberation | Moderate | 1 | Moderate | 1 | Moderate | 1 |
| Bad quality of data | Moderate | 1 | Moderate | 1 | Moderate | 1 |
| Inadequate infrastructure and resources | Moderate | 1 | High | 3 | Moderate | 1 |
| Improper management of data (governance, selection, interpretation) | Low | 0 | Low | 0 | High | 3 |
| **External** | | | | | | |
| Changes or elimination of open data law to limit transparency | Low | 0 | Low | 0 | Low | 0 |
| Loss of public interest | Low | 0 | Low | 0 | Moderate | 1 |
| Media miscoverage | Moderate | 1 | Low | 0 | Low | 0 |
| Lack of awareness of its existance | High | 3 | Moderate | 1 | Moderate | 1 |
| **Quantification** | **7** | | **6** | | **8** | |
| **Risk Profile** | **Medium** | | **Medium** | | **High** | |

Table 4-7: Risk quantification

Figure 4-7 shows the results when taking the architecture alternatives through the heuristics, method components, and graphing the results in three dimensions. Alternative A presents the least effort and highest effectiveness and is very close to the Alternative B results. On the contrary, Alternative C presents the highest risk, effort, and the lowest effectiveness.

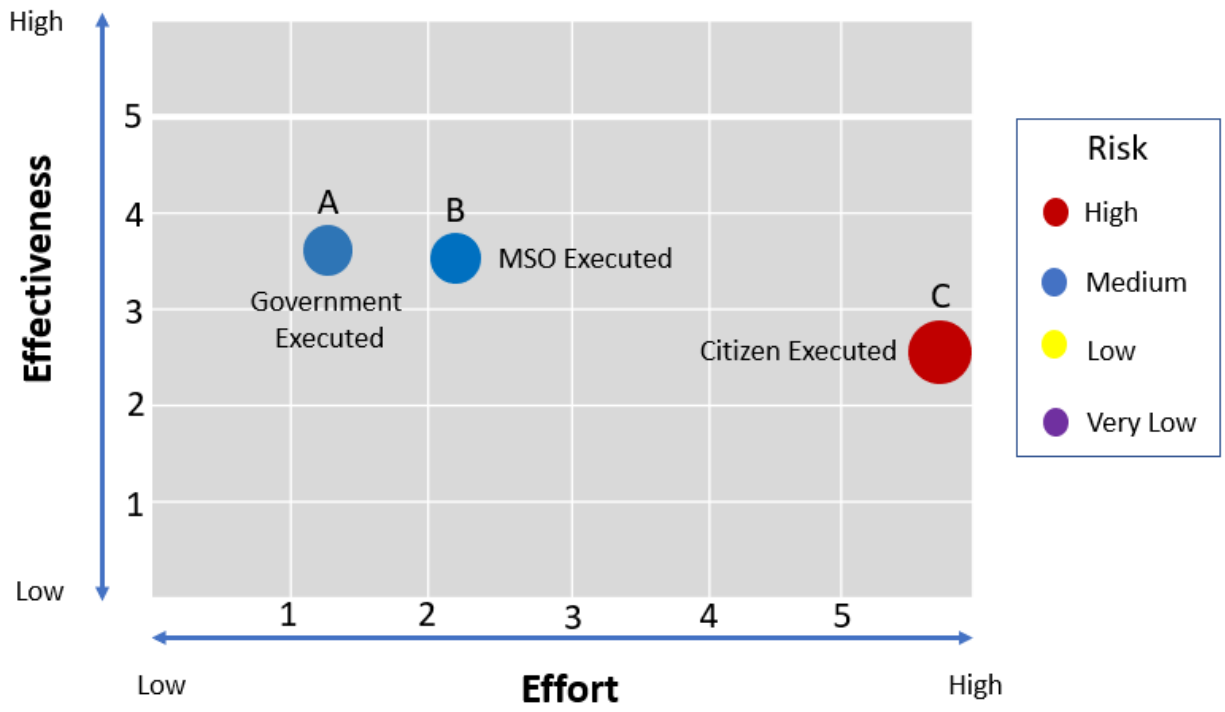



Figure 4-7: Trade-off matrix

## 4.2 General Analysis

The matrix facilitated the analysis of the different alternatives for the Government of Puerto Rico. The architecture alternatives were similar in risk. However, Alternative C 'Citizen Executed' was quickly discarded as a feasible alternative because of its low score in effectiveness and implementation difficulties. Regarding the other two options, there are trade-offs. Alternatives A and B appeared not dramatically different in terms of effectiveness, but B requires more effort than A. Moreover, Alternative A is easier to implement.

## 4.3 Recommendation

Based on the multidimensional analysis, Alternative A is the most viable. There are several reasons, including:

- owners of the data (closer to the data)
- government has direct oversight over the data custodians throughout the agencies
- easier to hold accountable than citizens or other external organizations with less visibility
- the government already has the infrastructure in place
- proper funding due to compliance with law 121

The “Government Executed” alternative (A) resulted more effectively with the least effort for the reasons mentioned above. Risks are similar for alternatives A and B. Alternative C resulted in higher risk, and it presents governance concerns as the output is difficult to control and validate.

# 5 Case Study: Citizen Information Portal of the Government of Puerto Rico

In this section, the "to-be" architecture is analyzed through a real-life project. This project is the Citizen Information Portal (CIP), an Open Data initiative of the Government of Puerto Rico (datos.pr.gov). The CIP has a public portal managed by the Puerto Rico Innovative and Technology Services (PRITS). It is not the same as the official Open Data site in Puerto Rico carried on by the Puerto Rico Statistics Institute (PRSI). The CIP's design originates from the context and concept presented in Section 3.

Figure 5-1 shows CIP's systems architecture. Component A, shown in boundary marked A, represents correlations between data sets. Component B, shown in boundary marked B, represents the contextualization of the data.

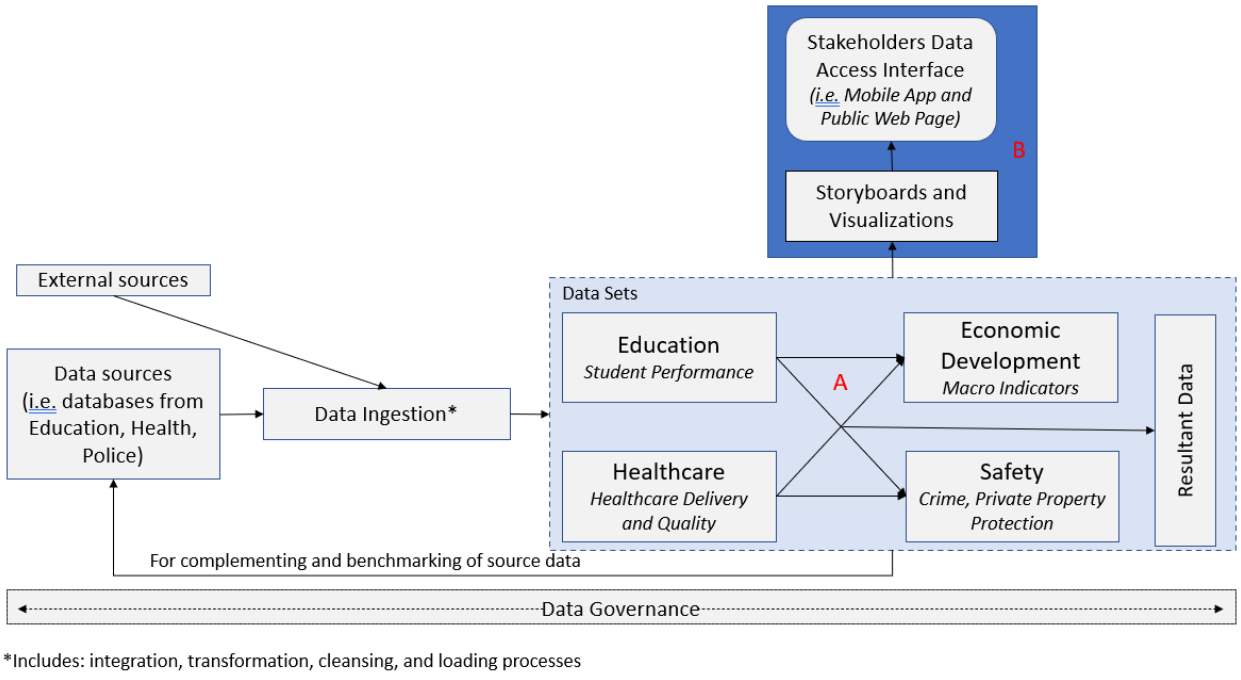



Figure 5-1: Citizen Information Portal (CIP)'s architecture

This section presents differences based on the following measures, which demonstrate the incremental value versus what exists today:

- *Visuals* – provide knowledge through easy-to-understand visuals
- *Use case* – this is to illustrate the benefit with a real example
- *Portal usage statistics* – although the Puerto Rico Statistics Department do not track this, as confirmed by email on January 4th, 2021, there are higher numbers for the new architecture (refer to Figure 5-10)
- *Contextualization* – this is to demonstrate a correlated contextualized example

*Visuals*

PIC uses information cards as a format to publish data. This data is presented through visuals that contain context, straightforward descriptions, and a vital section for key insights ("Puntos Claves" in Spanish). This section of "Puntos Claves" presents the essential points that any user or citizen needs to learn or understand about the data set. Figures 5-2 and 5-3 show the components of the information card designed for all users.

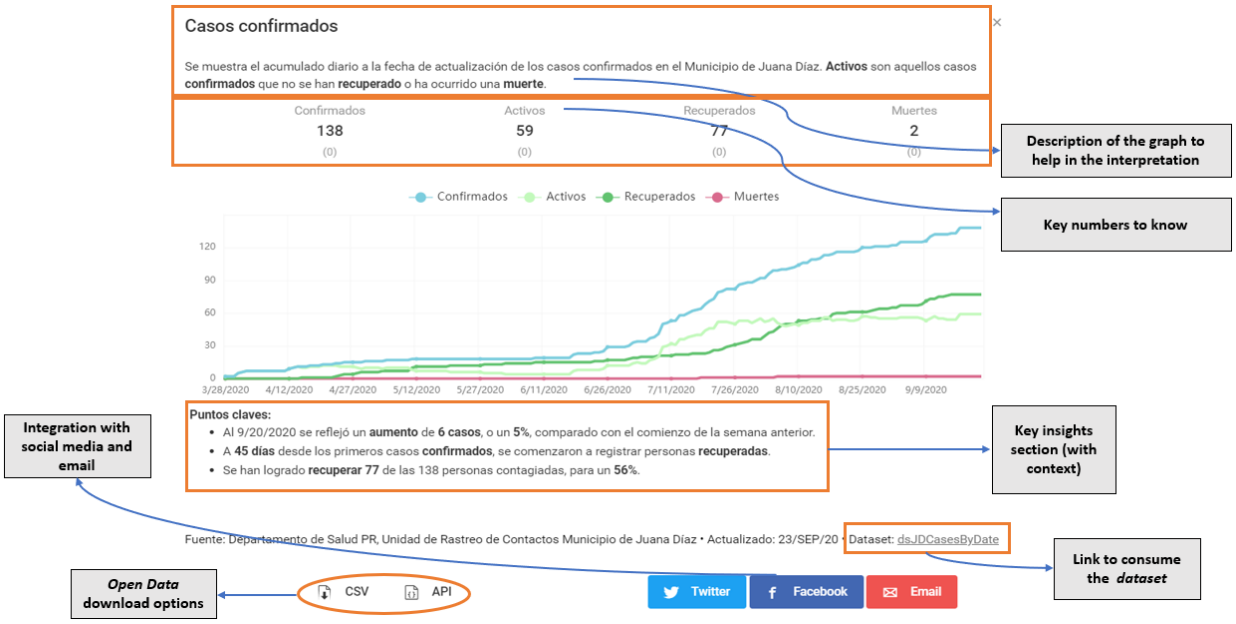



Figure 5-2: CIP sample information card

Other components of the information card, as shown, are the graph description, highlight of summarized numbers, integration with social media, link to the data set, and data download options.

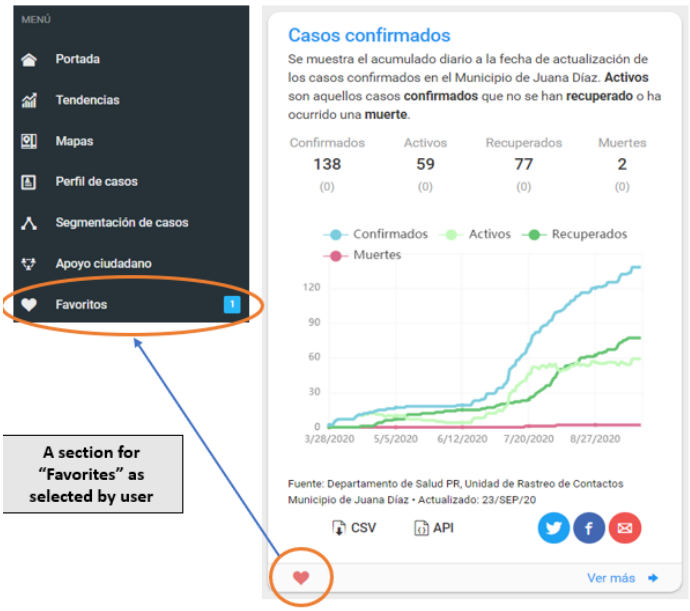



Figure 5-3: CIP's sample information card with a section for favorite cards

The visuals at the CIP are organized by topic. Figure 5-4 shows examples of these visuals by topic.


Figure 5-4: CIP visualization examples by topic

As Figure 3-2 from Section 3.2.1 shows, the current format for publishing data sets in the Open Data site of Puerto Rico is in tabular raw form. However, the CIP was designed with a concept and context that helped make Open Data for Puerto Rico more meaningful for users.

To provide examples of the insights that CIP presents today, refer to Figures 5-5 and 5-6. Figure 5-5 shows data about family affairs. As denoted in the “Puntos Claves” section, infants and children between the ages of 0 (months) to 4 years old are the group with the highest percentage of cases without an order from the court to receive

food benefits.  With this information, citizens can claim the Department of Family Affairs or receive an explanation of why the most vulnerable segment of the population is not receiving the benefit that they have the right to obtain.

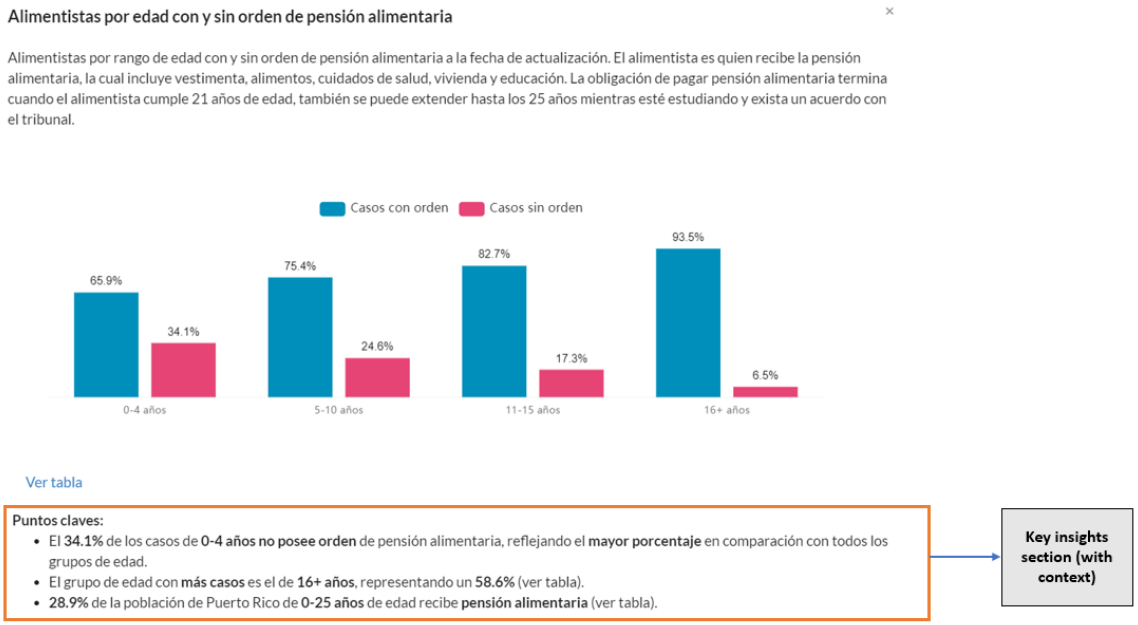



Figure 5-5: CIP information card with insights about family affairs

Similarly, Figure 5-6 shows how students from the public system in Puerto Rico decrease their proficiency dramatically in subjects, especially in math, once they reach sixth grade.

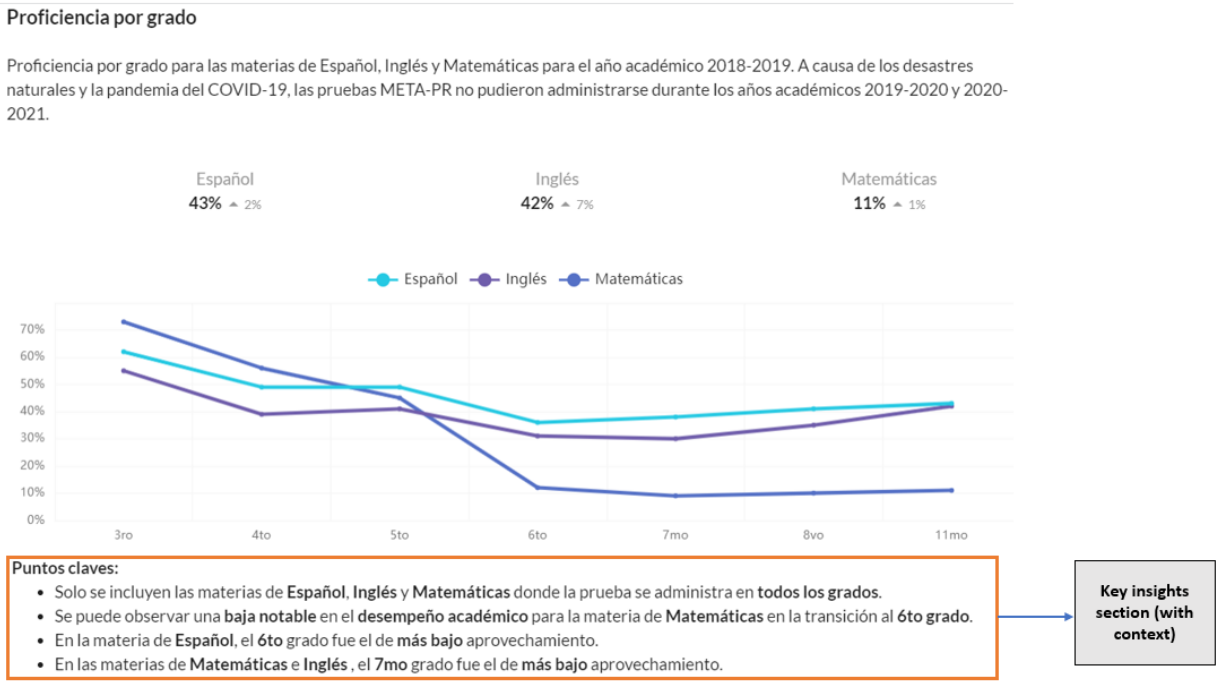



Figure 5-6: CIP information card with insights about education

With this information, parents, academics, and other stakeholders can start a conversation with the Puerto Rico Department of Education to investigate why this happens and what preventive actions should be implemented. Moreover, parents can investigate further to learn if this trend is consistent in their children's community schools.

*Use case*

The objective of presenting this use case is to demonstrate the difference between the current Puerto Rico Open Data site that publishes raw data and tabular format versus the CIP site that shows data with context and insights. Key insights or "Puntos Claves," descriptions, and contextualized visuals help citizens acquire knowledge to make the best decision for their benefit. As aforementioned, the spirit of this comparison is not to judge one site as being better than the other. Both were

designed under different circumstances, objectives, and contexts. However, as this thesis has proved, the Puerto Rico Open Data strategy needs to rethink the concept and move from a data-centric architecture to a citizens-centric.

This use case is related to COVID as it currently is a very relevant topic for most people. Citizens are looking for a balance between protecting themselves and their families and living with the proper precautions.

Let's present the case: *"a fully vaccinated 55-year-old lady of name Jenny living in the municipality of Ponce, Puerto Rico, with an approximate population of 125,000, would like to find a place for lunch. Acknowledging that the positivity rate of COVID is on the rise, she is looking to minimize risk. She knows the benefits of vaccination and would like to find a restaurant in a region with the highest vaccination rate possible or with the lowest cases registered. Jenny decides to go to the public health section of the Open Data site to learn more about the COVID situation. She finds a link to a file in CSV format, as shown in Figure 5-7. However, Jenny does not know how to download or open the information on her phone. Even if she accomplishes that, Jenny does not have the skills or the time to do the analysis to extract the relevant insights she needs.*

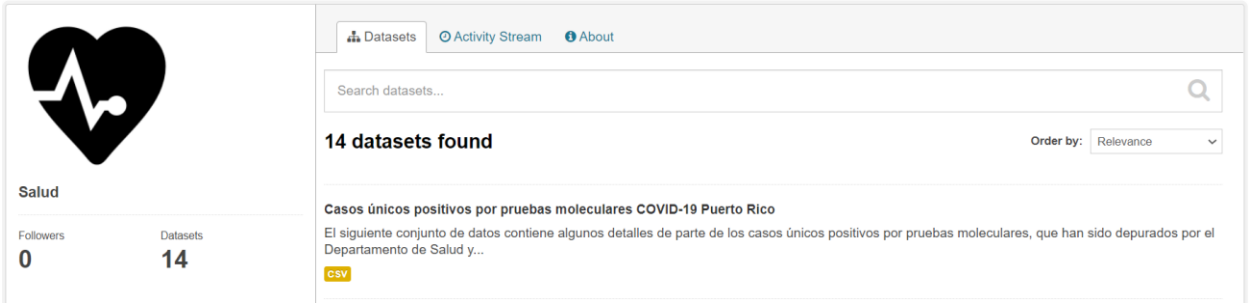



Figure 5-7: COVID cases data set in CSV format

*Then, Jenny finds out about the Citizen Information Portal or CIP and enters the site. The site rendered very well on her mobile phone and went to the health section. There she saw a map of Puerto Rico depicting the vaccination rates. Intuitively, she deduces the intensity of the color but decides to read the description and the insights section. Residing in Ponce, she chooses to touch on the demarked area of "Ponce" within the map, as shown in Figure 5-8. She quickly learns that the vaccination rate is 74%. Then, she decides to examine the rates beyond Ponce and notices that the vaccination rate of the municipality of Villalba is 89%, as presented in Figure 5-9. This number is significantly higher, and she decides to find a place for lunch in the municipality of Villalba. As this municipality is only 25 minutes away, Jenny decided to go to Villalba."*

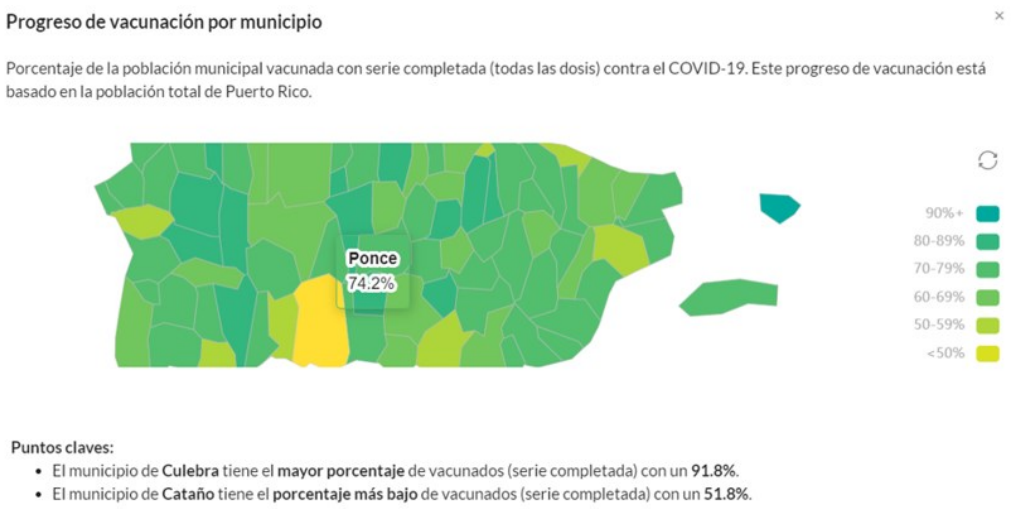



Figure 5-8: Map showing the vaccination rate for Ponce

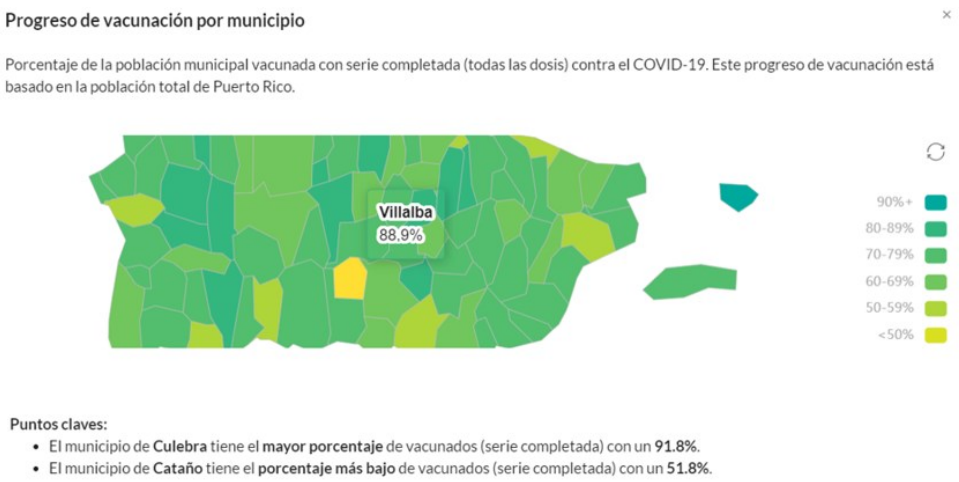



Figure 5-9: PR map showing the vaccination rate for Villalba

*Portal usage statistics*

As mentioned in Section 1-2, the PRSI does not track usage or download statistics of the 98 published data sets through the Puerto Rico Open Data site. The only statistic is the number of *followers* for each data set visible through the Open Data site. Although this is not representative of the data set usage, it is a proxy for interest in the data set. The number of followers is consistently zero in all data sets.

Figure 5-10 presents usage statistics of the CIP. These are numbers only from a soft launch since April 2022. As of January 2022, no official public announcement has been made about this site. Still, the numbers are in the thousands, and indeed, can improve dramatically. As the government promotes and educates about the CIP, its visitors' numbers and usage is expected to rise significantly.

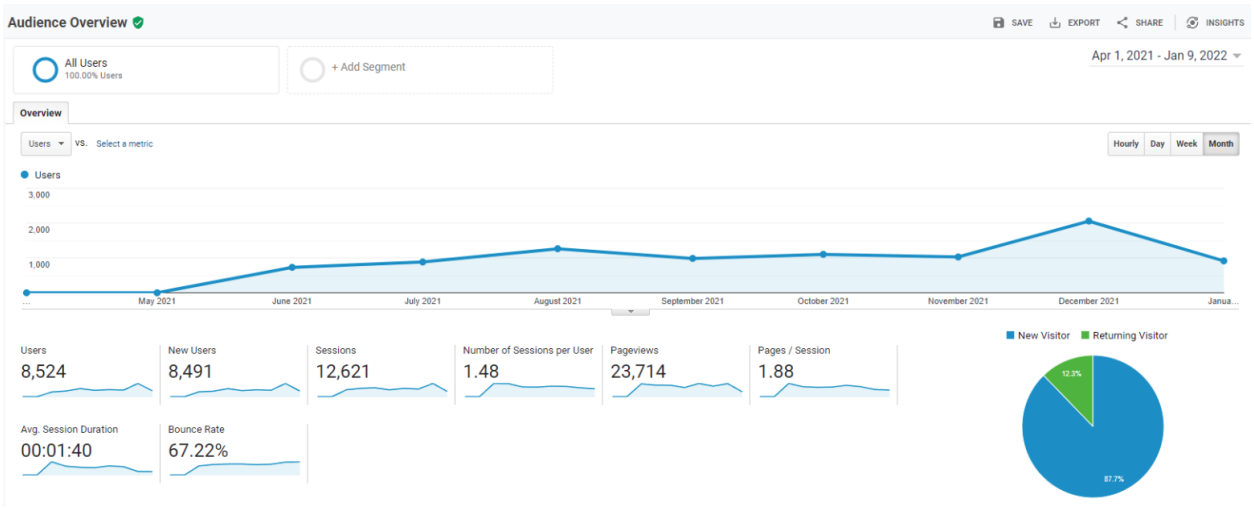



Figure 5-10: CIP usage numbers from April 2021 to January 9th, 2022

*Contextualization*

Open Data sites typically present data as a list of data sets with corresponding metadata. An example of this is shown in Figure 5-11, taken from the PRSI site.

Conjuntos de datos

| Inventario | Enlaces |
|---|---|
| Encuesta de alfabetización | Encuesta de alfabetización |
| Canasta de bienes y servicios | Canasta de bienes y servicios 2000-2001 |
| Cartografía censal / Topologically Integrated Geographic Encoding and Referencing (TIGER) | Cartografía censal / Topologically Integrated Geographic Encoding and Referencing (TIGER) |
| Censo decenal de población y vivienda | Censo decenal de población y vivienda |
| Comercio externo de bienes de Puerto Rico | Exportaciones y envíos de bienes desde Puerto Rico , Importaciones y envíos de bienes hacia Puerto Rico |
| Conjunto de datos comunes sobre las escuelas públicas | Conjunto de datos comunes sobre las escuelas públicas |
| Educación postsecundaria | Educación postsecundaria |
| Empleo Asalariado No Agrícola | Empleo Asalariado No Agrícola (CES) |
| Empleo Asalariado No Agrícola por industria | Empleo Asalariado No Agrícola por industria |

Figure 5-11: Data sets list from the PRSI Open Data site

As these data sets are published independently, there is no context added that could create a valuable data story to be communicated. As a system of systems, the government captures data that shows this interdependency consistent with the behavior

in real life. As many studies have shown, education impacts economic development, safety, and the health state of any country. That is why it is essential to find these patterns and correlations in the data. Combining visualizations add context that helps in the interpretation. Figure 5-12 shows how incorporating these data sets from separate topics adds good context.

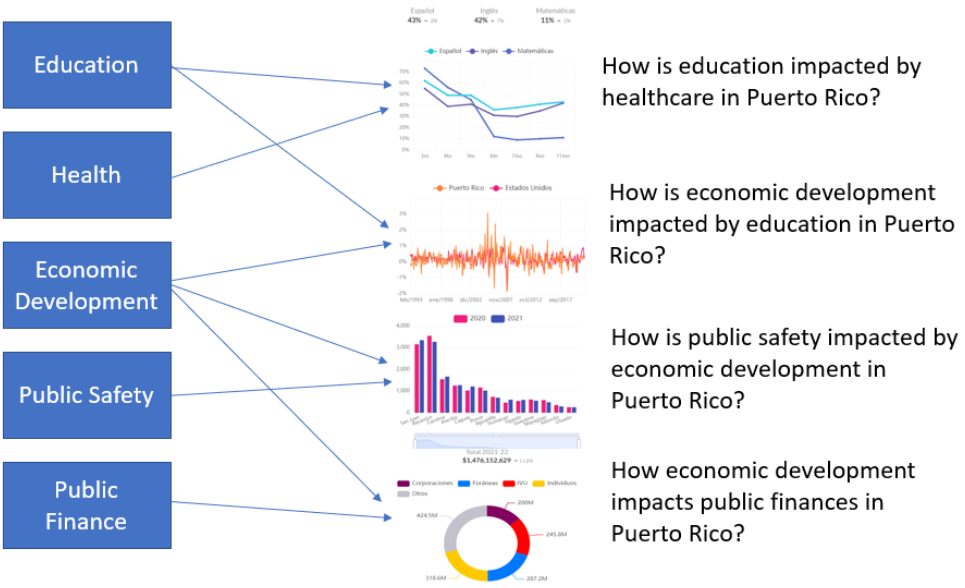



Figure 5-12: Visualization examples of the “to-be” architecture with correlations

When combining data sets from separate government agencies that do not communicate with each other, interesting correlations can emerge. The example from Figure 5-13 takes data from the Mental Services Administration within the Puerto Rico Department of Health and the Puerto Rico Forensic Sciences Institute. Mental Services has a suicide prevention hotline to provide urgent assistance to distressed citizens. The figure shows that in the months when the hotline is the busiest, fewer suicides are registered in Forensic Science Institute. Although this sounds logical, it is a correlation with no proven causality. However, context helps citizens judge the effectiveness of this

type of service for citizens. That is a powerful proposition to show Open Data with context.

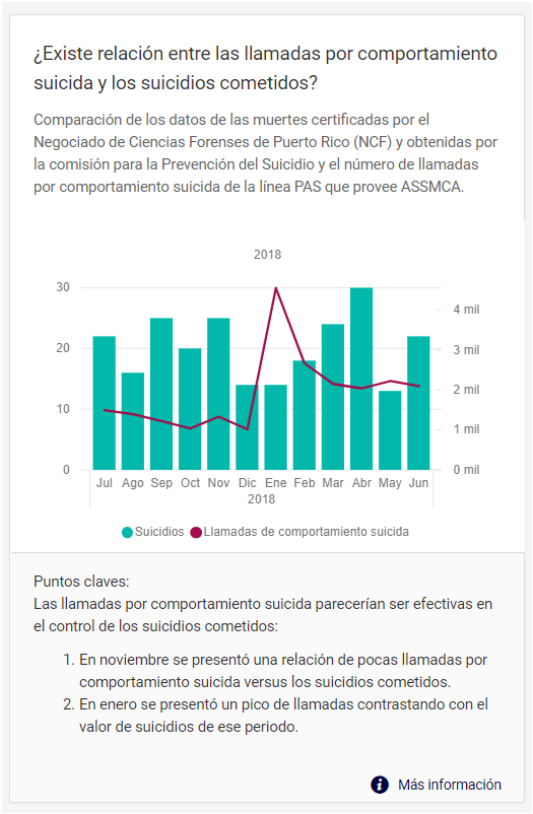



Figure 5-13: Correlation between data sets from different government agencies

This example also shows a systems perspective with data more connected. This inter-connectivity between data elements is communicated through a visualization, including a simple narrative and relevant insights, which adds value to any Open Data initiative, and certainly Puerto Rico.

# 6 Policy Implications, Conclusion, and Future Work

As discussed in Chapter 1*,* Open Data in many countries, including Puerto Rico, aims to liberalize data. However, Open Data policy needs to expand to make this public data actionable for citizens. This thesis proposes an architecture that considers how data is presented to citizens, adding context and a format that is easy to understand without needing data wrangling skills. Open Data in Puerto Rico needs a transformation to embrace a new architecture to meet this objective. Government leaders can follow the structured process presented in this thesis to validate the recommendation given and reasoning in architecting the future of Open Data in Puerto Rico, assessing the tradeoffs that exist among alternatives. It raises questions like, what are the architecture changes and enhancements to migrate from a data-centric enterprise to citizen-centric? Can the enterprise add or change the appropriate processes, knowledge, and organization to make this happen? Can the Government of Puerto Rico provide the resources and leadership to make this happen?

The work presented in this thesis used a mixed-method research approach. It included the following techniques and methods: system architecture, system thinking, design thinking, Nightingale & Rhodes, AFE framework, external research, public domain data, and interviews to generate the core proposition – *a new architecture for the future enterprise of Open Data in Puerto Rico.*

To summarize the results, this chapter includes a discussion on the policy implications of this research, general conclusions, limitations, and insights regarding potential areas of follow-on research.

## 6.1 Policy Implications

The policy implications of this work span across two major areas: (1) its effect on open data policies, not only for the Government of Puerto Rico but possibly within governments in other parts of the world; (2) its role in minimizing data inequality by envisioning a future open data architecture that considers the benefits for a larger citizen audience versus focusing its efforts in just liberating government data.

First, one of the important policy implications of this work is its effect on amending Open Data laws, such as law 121 in Puerto Rico, on mandating beyond liberating data that can be easily understandable for the ordinary citizen. As shown, this requires exploring the data and communicating responsibly using visual and written language that can prompt action by citizens. The key is not to leave most of the population blind. Policy needs to change to add meaning and unbiased interpretations of the published data.

A second policy implication of this work is minimizing data inequality. Only data analysis skilled individuals can make sense of the liberated data and take actions accordingly for their benefit, leaving ordinary citizens behind. It creates a limitation in the true democratization of the data. If the data is generated from taxpayers' interactions with the government, it is the government's responsibility to put the data to work for its constituents. Otherwise, and although not the intention, the potential consequence is a gap in data access. This perpetuates a situation where privileged citizens take advantage of this data over others that have equal rights.

Governments, specifically the Government of Puerto Rico, can assess its current Open Data policy to minimize data inequality by providing an enterprise architecture that focuses on delivering value for most citizens.

## 6.2 Conclusion

This thesis proposed an architecture for the future enterprise of Open Data in Puerto Rico. The primary objective is to generate an enterprise transformation by recommending an architecture that meets the needs of the citizens. The process used for this work showed a method to assess the current architecture, derive a “to-be” architecture, identify implementation alternatives, and make a recommendation. Then, a real-life project demonstrated the difference between a data-centric architecture and a citizen-centric architecture. The former focuses on data liberation without adding additional value to the data. In contrast, the latter has a citizen-centric design that provides an easy-to-understand data view. Why liberate data if it is not going to serve its citizens?

The recommended architecture alternative focuses on key “ilities” that are not typically considered in these Open Data initiatives. These key “ilities” go beyond transparency and openness to add usability and actionability for most citizens. This paradigm change gives another dimension of transparency to the government of Puerto Rico and probably to others around the world. Both former and current CIOs of Puerto Rico recognize the need to extend the value of the existing architecture to a broader group of citizens. The value is realized by taking the data sets and providing meaning with different techniques from data journalism, such as descriptions, visualizations, and actionable insights. It makes the data usable and understandable while closing the inequality gap.

A system thinking and architecture design tools, along with a quantitative evaluation, led to selecting the enterprise's future architecture. The application of the

AFE Framework to guide the architecture alternatives selection process proved to be a valuable tool to support the selection of the future state.

Finally, the recommended architecture was presented through a case study that demonstrated the added value to citizens when considering context to deliver a usable output. It can be the next generation of Open Data policies, laws, and regulations in Puerto Rico and worldwide.

## 6.3 Research Limitations and Future Work

This section identifies the most significant limitations of this research and outlines areas of future research.

a) The case study research is limited in its representativeness. The recommended architecture was tested and adjusted using the CIP project within the Government of Puerto Rico. Applying the changes to other Open Data architecture transformation projects is needed to validate the conclusions. One of the areas of future research is to analyze the feedback given by practitioners to identify improvement opportunities to Open Data enterprises around the world.
b) Although requested, and after several attempts, the Puerto Rico Statistics Institute did not provide data on the number of downloads and usage of the 98 data sets published through its public website. Although it is a misrepresentation, a possible proxy for this measure is the number of “followers” for each data set available from the Open Data site. Applying external research already conducted on Open Data to the case of Puerto Rico proved to be helpful, mainly because of the large sample used to reach those conclusions.

c) Another limitation is that the selected/recommended architecture, presented through the case study in Chapter 5, has not been announced to the public in Puerto Rico. Preliminary data, collected during the CIP pilot and soft launch period, has been used to explain and forecast better outcomes of the recommended architecture. Currently, there is a process to continuously collect usage data of the CIP. It is essential to constantly gauge the usage results to confirm the conclusions of this work.

## 6.4 Future Work

An essential component of the recommended architecture is extracting the meaning of and interpretation of the liberated data sets. The best value potential of the architecture is also the most vulnerable. Therefore, to complement the recommended architecture, future work includes determining the best practices to extract meaning from data sets. A global governing body can create standards for this level of interpretation and data representation. Where to draw the line of an unbiased versus biased interpretation? How to make sure that governments in power do not anchor citizens on a specific view or provide them with the wrong understanding? It is a powerful tool for influence, and it needs to be audited as necessary. Data interpretations for Open Data should align with standards and regulations that the governing body, either a policy-based institution or a think tank, defines considering best practices already in place.

# 7 Bibliography

A. Zuiderwijk, M. Janssen, Y.K. Dwivedi (2015): “Acceptance and use predictors of open data technologies: drawing upon the unified theory of acceptance and use of technology” Govern. Inf. Q., 32 (4) (2015), pp. 429-440

A repository of Federal Enterprise Data Resources. Data.gov, https://resources.data.gov/. Accessed November 2021

Britney Pay. “Digitizing Taxonomy: Introduction Dewey to Plagman” eFileCabinet 2016, www.efilecabinet.com/digitizing-taxonomy-from-dewey-to-plagman-2/

Chin, Yau, Wah, Khiang (2013-2014): “Framework for Managing System-Of-Systems Ilities” DSTA Horizons, www.dsta.gov.sg/docs/default-source/dsta-about/framework_for_managing_system_of_systems_ilities.pdf?sfvrsn=2

“Complete list of open data sites as of 2019” September 18th, 2019, https://s3.amazonaws.com/bsp-ocsit-prod-east-appdata/datagov/wordpress/2019/09/opendatasites91819.xls. Accessed October 2021

Custer, S., Masaki, T., Sethi, T., Latourell, R., Rice, Z., & Parks, B.C. (2016). “Governance Data: Who Uses It and Why?” AidData at William & Mary and the Governance Data Alliance.

Dr. Cliff Whitcomb. “Sea Connector Family and Seabase Architecture.” Systems Engineering & System Architecture Presentation to Naval Postgraduate School October 21, 2004, https://nps.edu/documents/103424733/107333295/Sea+Connectors.pdf/36649a53-5566-4fe5-9284-21f83ee49009

Figueroa, Nestor. “Re: AYUDA: Datos para tesis.” Received by Jacobo Orenstein-Cardona, 4 Jan. 2022.

“Foro Virtual de Estadísticas y Tecnología.” *YouTube,* uploaded by Instituto de Estadísticas de Puerto Rico, May 7th 2021, www.youtube.com/watch?v=uKhl7l_k1Yw

Steven Raphael, Rudolf Winter-Ebmer. “Identifying the Effect of Unemployment on Crime.” The Journal of Law and Economics, Volume 44, Number 1, April 2001), pp. 259-283.

Instituto de Estadísticas de Puerto Rico. General statistical publications, 2021, https://estadisticas.pr/en. Accessed October 2021.

Instituto de Estadísticas de Puerto Rico. Open data sets, 2021, https://datos.estadisticas.pr/. Accessed October 2021.

Jonathan Cinnamon (2019): Data inequalities and why they matter for development, Information Technology for Development, DOI: 10.1080/02681102.2019.1650244

Jonathan Gray, Danny Lämmerhirt (2017): “Data and The City How Can Public Data Infrastructures Change Lives in Urban Regions.” Accessible at: https://blog.okfn.org/files/2017/02/DataandtheCity.pdf

Justia US Law. “2020 Indiana Code Title 5. State and Local Administration Article 14. Public Records and Public Meetings Chapter 3.3. Government Data 5-14-3.3-6. Machine Readable" justia.com, https://law.justia.com/codes/indiana/2020/title-5/article-14/chapter-3-3/section-5-14-3-3-6/

Krishnamurthy, Rashmi. “Liberating data for public value: The case of Data.gov.” International Journal of Information Management. Volume 36, Issue 4, March 21, 2016, Pages 668-672.

LaFortune & Ubaldi. “2017 OURData Index: Methodology and Results” OECD, 2017, https://www.oecd-ilibrary.org/governance/oecd-2017-ourdata-index_2807d3c8-en

Lämmerhirt, Rubinstein, Montiel. “The State of Open Government Data in 2017” OKFN.org, https://blog.okfn.org/files/2017/06/FinalreportTheStateofOpenGovernmentDatain2017.pdf. Accessed November 2021

LexJuris Puerto Rico. “Ley de Datos Abiertos del Gobierno de Puerto Rico.” Lexjuris.com, www.lexjuris.com/lexlex/Leyes2019/lexl2019122.htm. Accessed November 2021

Md Shamim Talukder, Liang Shen, Md Farid Hossain Talukder, Yukun Bao “Determinants of user acceptance and use of open government data (OGD): An empirical investigation in Bangladesh” Technology in Society, Volume 56, 2019, Pages 147-156

MIT OCW Site. Systems Architecture at MIT, IAP 2007, https://ocw.mit.edu/courses/engineering-systems-division/esd-34-system-architecture-january-iap-2007/lecture-notes/lec1.pdf. Accessed December. 2021.

National Conference of State Legislatures. “State Open Data Laws and Policies” NCSL.org, https://www.ncsl.org/research/telecommunications-and-information-technology/state-open-data-laws-and-policies.aspx

Olivier de Weck. “Establishing a Common Language and Set of Methods for Systems Design and Architecture” 2002 Doctoral Seminar Engineering Systems Division at

MIT, https://dspace.mit.edu/bitstream/handle/1721.1/58743/esd-84-fall-2002/contents/calendar/GSSem6_Oli_de_Weck_on_Systems_Architecture.pdf

Open Data Catalog of Available Data Sets. Data.gov, https://catalog.data.gov/dataset. Accessed October 2021

Open Knowledge Foundation. “Data and the City: New report on how public data is fostering civic engagement in urban regions” Blog.okfn.org, https://blog.okfn.org/2017/02/09/data-and-the-city-new-report-on-how-public-data-is-fostering-civic-engagement-in-urban-regions/. Accessed November 2021

Open Knowledge Foundation. “Global Open Data Index for Puerto Rico” OKFN.org, https://index.okfn.org/place/pr/. Accessed November 2021

Open Knowledge Foundation. “Methodology” OKFN.org, https://index.okfn.org/methodology. Accessed November 2021

Open Knowledge Foundation. “Open Definition” Opendefinition.org, www.opendefinition.org. Accessed November 2021

Organization for Economic Co-operation and Development (OECD). “Open Government Data” OECD.org, https://www.oecd.org/gov/digital-government/open-government-data.htm. Accessed November 2021.

O’Reilly Online Site. Part 2 Analysis of System Architecture, “What is Architecture” https://www.oreilly.com/library/view/system-architecture-strategy/9780136462989/xhtml/fileP70004959190000000000000000A18.xhtml. Accessed December 2021.

Rivera Perez, Emilsson & Ubaldi. “2019 OURdata Index: Policy Paper” OECD, 2020, https://www.oecd.org/gov/digital-government/policy-paper-ourdata-index-2019.htm

Systems Design and Management. MIT, Fall 2018 (September), “EM411: Foundations of Systems Design and Management” Systems Architecture Module by Prof. Ed Crawley. Accessed December 2021.

T2Informatik GmbH. “System Context.” T2Informatik.de, https://t2informatik.de/en/smartpedia/system-context/

Ubaldi, B. “Towards Empirical Analysis of OGD Initiatives” OECD-ilibrary.org, 2013, https://www.oecd-ilibrary.org/governance/open-government-data_5k46bj4f03s7-en

Ubaldi, B. "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives," OECD Working Papers on Public Governance, No. 22, OECD 2013 Publishing, Paris, https://doi.org/10.1787/5k46bj4f03s7-en

US Congress. “H.R. 4174 - Foundations for Evidence-Based Policymaking Act of 2018 (115th Congress 2017-2018)” Congress.gov, https://www.congress.gov/bill/115th-congress/house-bill/4174/text#toc-H8E449FBAEFA34E45A6F1F20EFB13ED95 Accessed November 2021

USA Facts. “Global Open Data Index for Puerto Rico” Usafacts.org, www.usafacts.org. Accessed November 2021

U.S. General Services Administration. “Open Government.” Data.gov, https://www.data.gov/open-gov/. Accessed November 2021.

V. Venkatesh, et al. (2003): “User acceptance of information technology: toward a unified view” MIS Q., 27 (3) (2003), pp. 425-478

V. Weerakkody, et al. (2017): “Open data and its usability: an empirical view from the Citizen's perspective” Inf. Syst. Front, 19 (2) (2017), pp. 285-300

W3C. “Data on the Web Best Practices” w3.org, www.w3.org/TR/dwbp/#metadata