

Review Article

# Enhancing Government Fiscal Impact Analysis with Integrated Big Data and Cloud-Based Analytics Platforms

Vamsee Pamisetty<sup>1,\*</sup> <sup>1</sup> Middleware Architect, USA

\*Correspondence: Vamsee Pamisetty (vamseeepamisetty@gmail.com)

**Abstract:** While several application domains are exploiting the added-value of analytics over various datasets to obtain actionable insights and drive decision making, the public policy management domain has not yet taken advantage of the full potential of the aforementioned analytics and data models. To this end, in this paper authors present an overall architecture of a cloud-based environment that facilitates data retrieval and analytics, as well as policy modelling, creation and optimization. The environment enables data collection from heterogeneous sources, linking and aggregation, complemented with data cleaning and interoperability techniques. An innovative approach for analytics as a service is introduced and linked with a policy development toolkit, which is an integrated web-based environment to fulfil the requirements of the public policy ecosystem stakeholders [1]. Large information databases on various public issues exist, but their usage for public policy formulation and impact analysis has been limited so far, as no cloud-based service ecosystem exists to facilitate their efficient exploitation. With the increasing availability and importance of both public big and traditional data, the need to extract, link and utilize such information efficiently has arisen. Current data-driven web technologies and models are not aligned with the needs of this domain, and therefore, potential candidates for big data, cloud-based and service-oriented public policy analysis solutions should be investigated, piloted and demonstrated [2]. This paper presents the conceptual architecture of such an ecosystem based on the capabilities of state-of-the-art cloud and web technologies, as well as the requirements of its users.

**Keywords:** Fiscal Impact Analysis, Big Data, Cloud-Based Analytics, Government Finance, Data Integration, Predictive Modeling, Public Sector Analytics, Decision Support Systems, Real-Time Data, Financial Forecasting, Digital Transformation, Budget Planning, Cloud Computing, Data-Driven Policy, Integrated Platforms

## How to cite this paper:

Pamisetty, V. (2020). Enhancing Government Fiscal Impact Analysis with Integrated Big Data and Cloud-Based Analytics Platforms. *Journal of Artificial Intelligence and Big Data*, 1(1), 1–24.  
DOI: [10.31586/jaibd.2020.1339](https://doi.org/10.31586/jaibd.2020.1339)

**Received:** October 7, 2020**Revised:** November 25, 2020**Accepted:** December 20, 2020**Published:** December 29, 2020

**Copyright:** © 2020 by the author. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

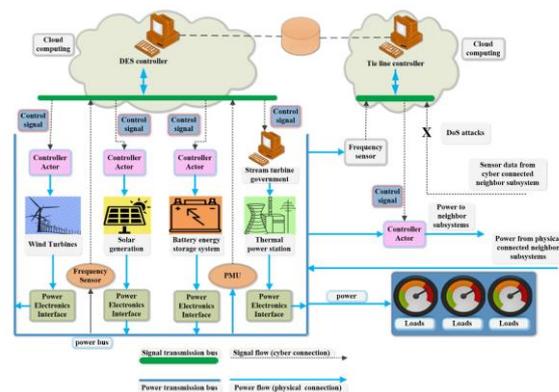
Governments rely on fiscal impact analysis to estimate the financial consequences of revenue and expenditure changes and to inform politicians, the public, and rating agencies [1]. While some governments produce comprehensive analyses detailing multi-year impacts on tax revenue, expenditure programs, and the resulting effect on net government cash flow, estimates of allocation changes in direct tax legislation tend to be less sophisticated, often being a simple assessment of changes in measured tax revenue. Nevertheless, statistical models based on "big data" can produce far more comprehensive and complex impact estimates. For example, the model demonstrated that Australia's ten separate natural resource taxes result on average in a net cash inflow each year equivalent to ten per cent of total tax revenue. They present a cloud-based application, which uses "big data" and probabilistic batch simulations to produce rapid around-the-clock automatic estimates of government cash flow changes, fiscal impact analysis was

complemented by predictions of the flow-on real gradational effects of the revenue change on deficits, the economy, and inequality.

The emergence of data-driven technologies offers significant opportunities for efficiently managing the whole public policy management domain, from monitoring the public interest to all subsequent decision-making and implementation and impact analysis phases [2]. The critical importance of timely, accurate, and consistent data in the policy development and evaluation process is highlighted, focusing on policies to address the climate crisis. The accuracy and credibility of data have become crucial in modern democracies, as a basis for good governance, appropriate public policies, and the sustainability of democracy itself and social stability. Spreading data literacy skills and knowledge through educational systems has now become a top global priority.

## 2. Background of Fiscal Impact Analysis

Large public authorities have either internal units or external companies crafting fiscal impact analyses (FIA) assessing the impact of projects or programs on the public budget. The quality of the analysis – the robustness of the analysis and the transparency of the underlying assumptions and calculations – varies significantly in regard to time, human resources, method, and data used. In regard to method, different tools have been developed by different institutions, and many tools are very well at what they do. However, there are still significant gaps in the overview tools. The tools remain separate, stand alone, and external [3]. Furthermore, many of these costs/nest disciplines have barely been derived from methods or have no accessible source in English. This means that the potential to backtrack the calculations or to change the parameters is brittle. Most public authority units for fiscal impact analysis thus face problems with post process transparency and credibility of the calculations. In particular smaller public authorities are largely at the mercy of the external companies and outsourcing services, which not only jeopardizes a coherent and interoperational methodology and comparable approach but also means a vulnerability for smaller public authorities. This often leads to a vicious circle in regard to fiscal impact analysis, especially in the smaller public authorities.



**Figure 1.** Big Data Analytics Using Cloud Computing

Fiscal impact analysis for small government units with mainly infrastructure projects on the municipal level has significantly less sophisticated methodology and tool support than program or project fiscal impact analyses on higher governmental levels. This gap in regards to methodology and tools is particularly striking with regard to countries where FIAs are an established, almost mandatory, part of the societal planning process especially for major construction projects. Research interest in on the one side methodologies and transparently available tools to assess possible fiscal impacts of planned municipal actions is considerable however the tools used are still only established in very few countries, very few tools are available, and even fewer have been documented in English [4].

## 2.1. Definition and Importance

In 2001, two researchers proposed the first initial definition of Big Data as “a collection of data sets so large and complex that it becomes difficult to process using on-hand database managing tools or traditional data processing applications.” Such a definition has become one of the most widely accepted definitions of Big Data. Since then a large number of definitions have been proposed by industry and academia and a wide range of properties have been identified such as the 5V and the 1C. The 5V consist of the Volume, the Variety, the Velocity, the Variability, and the Value. The 1C means the continuous stream of data [2]. In 2013, a high level definition of Big data was coined as “one petabyte” in storage, 100 petabytes in data set and 10 petabytes in persistence.” Since Big Data in government has different attributes than in private sector, previous researches on Big Data in business are limited in explaining the Big Data issues in public sector. In public sector, there are concerns regarding the Big Data Asset such as the lack of control tower for BDA and dispersed silo. These concerns have made the Big Data Analytics (BDA) to be considered as the new challenges for government. By developing Big Data supporting systems such as data control tower and portal with highly secured technology, government can utilise the data for public decision making.

BDA is defined as the strategies, infrastructure, and skills needed to acquire, store, analyse, and manage Big Data. A data portal is an integrated view on relevant data source which is a precondition of BDA. BDA is considered as one of the new emerging disciplines of research as well as of practical interest and its applicability in public sector has attracted researchers and practitioners. Governments in Europe Union, United States, and some Asian countries have shown an active interest in building data portals and experiments have surfaced in implementing data portals with open data. This provides a landscape in which attempts to construct data portals using similar approaches. However, little work has been published on fundamental concerns such as what to build, to build what and how to design and develop a data portal.

## 2.2. Historical Context

The ICT advances and the digitalisation of several processes is leading to the generation of vast quantities of data. These technological advances have made it possible to store, transmit and process large amounts of data more effectively. This rich data environment affects decision and policy making: cloud environments, big data and innovative data-driven approaches create opportunities for evidence-based policies. During the traditional policy cycle, data is a valuable tool for allowing policy choices to become more evidencebased and analytical. Quality checks for both input and output data are conducted and the estimates of impact on relevant variables are carefully considered [2]. Nevertheless, the public sector has some distinctive characteristics compared to other domains. At the national level, policies are often complex: (1) they tend to be multi-variate, (2) they are usually indirect, influencing the target only indirectly, through an instrument chain and (3) with a significant lag time, several years or even decades [1]. Still, modelling the costs and effects of plans and policies is ubiquitous in the public sector. Typically, in this context, modelling is of a two-step nature, starting with a macro-modelling of costs or outcomes.

Models, representing a specific domain and reflecting existing knowledge about that domain, determine costs and effects of a policy for a limited set of conditions. These models are almost without exception based on assumptions and heuristics and they involve highly aggregate data. Typically, a second modelling, often of a statistical nature, is developed to explain costs or benefits of policies, based on analysis of existing time series data. This analysis is of a one-dimensional nature, focusing on only one or a limited number of cost variables or effects. The decision-makers in this context typically seek assurance rather than novelty in the presented output.

### 2.3. Current Trends in Fiscal Analysis

As uncovered by the literature, fiscal impact analysis entails estimating expenditures and revenues arising from a change in economic state [5]. Generally, understanding these mandated dollar values is a significant starting point for discussion in regards to fiscal analysis. Currently efforts to automate this process continue via improved assistance to governmental financial systems for modeling, estimating and reporting of aforementioned fiscal impact analysis; continued collaboration with data analysis research; and examination of alternative processes for deploying cloud-based solutions to data analyses and cost/benefit modeling of all impacts.

Research findings suggest that the pacing issue with financial reporting data release predated cloud services and affects all analysis platforms and within any format. Reporting agencies subsequently update entries within general accounting systems prior to fiscal year end close-out. Speed disparities also arise from source data entry speed differing between jurisdictions/professions; some positions analyze general entries using source data while others query the general ledger. Faster input of accounting events yields older reporting data. Merely to be able monitor the timely release of accounting events would be a major analytic advance.

Additional scope includes addressable external data, such as economic data releases that likewise provide critical context to fiscal impact analyses. Uncertainty surrounding use of external data is determined on economic growth outlook. Journals regularly compile macroeconomic data via a description of its availability and importance; access is usually through financial market platforms. The need for a cloud-based analytic service is detailed further in a companion paper; data sources suitable for this task could be examined to ensure an even playing field and intended report cadence.

#### Equation 1: Fiscal Impact Function

$$\mathcal{F}_t = \Delta R_t - \Delta E_t$$

Where:

- $\Delta R_t$  : change in revenue
- $\Delta E_t$  : change in expenditure

### 3. Big Data in Government

With Big Data and the Internet of Things, governments now have the potential access to huge amounts of data on the economic and social impact of Government policies. European economics has widely developed the theoretical foundation for these analyses, which largely remain idle in the public administration domain. This policy brief proposes to address this issue by recommending a new educational initiative in collaboration with the growing field of cloud-based big data analytics and visualization platforms for public administration. These platforms can significantly enhance a wide range of public administration activities beyond fiscal impact analyses alone. However, their usage in this domain remains minimal. There is considerable demand among government officials and public servants for applied knowledge on public administration discipline with reference to big data and cloud-based analytics, as well as the development of country-specific and domain-specific applications for fiscal impact analysis and simulation modeling of Government policy measures. There are also practical, relative easy to implement options for bringing together existing platforms with their current research and discipline-specific applications and policy brief findings with government bilateral backing and funding. This represents a unique opportunity for public administration researchers and professional associations to further enhance the dissemination, understanding, and thus the societal impact of their research. For governments and individual government services possibly no significantly easier way is available to more effective policy measures than

the wider application of these platforms, as well as strengthened collaboration with academia and public administration professional associations are important actors in this regard. Big Data has become a central topic in research and the public and private sectors alike.

There is a growing recognition that Big Data and Artificial Intelligence will have a profound transformational impact on Governments around the world. Governments are already expressing considerable interest in the potential effects of Big Data and AI on their functioning and effectiveness. It is still early days in terms of research within the fields of public policy and administration of the effects of Big Data and AI on government itself. However, there are already some great studies reporting on important individual aspects of this transformational shift in government. Big Data is thought to have a global reach and exert a profound, if often unnoticed, fundamental structural impact throughout society. Big Data and AI have a global reach and will exert a profound, if often unnoticed, fundamental impact for good and ill on governments and governance around the world. The reach of these activities is global, and so is concern about how what happens in one part of the world may affect what happens elsewhere.

### ***3.1. Overview of Big Data***

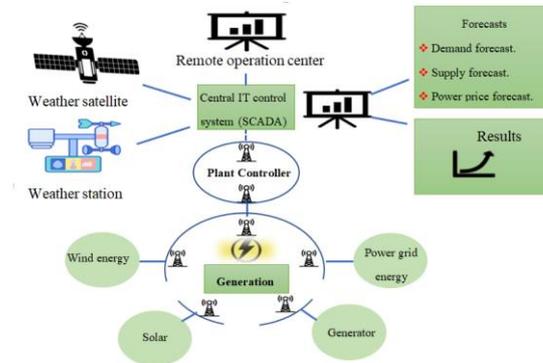
Big Data is characterized by high volume, velocity, variety, variability, and complexity in data origin and types. In both research and real-life applications, Public Sector Big Data is considered a subset of Big Data characterized by specific attributes. Some Public Sector Big Data are continuously collected such as crowdsourcing data and biometrics, and some are collected as recordings for retrospective analysis. Other data are created through interactions with governments, e.g. taxi ridership and personal expenses in case of government subsidies. Moreover, Public Sector Big Data encompass solely national scale and internationally comparative datasets. Policy management and public service delivery based on Public Sector Big Data however present significant methodological challenges concerning representativeness, interpretability, granularity, verifiability and ethics among others [1].

Increased demand for transparency and accountability in credit ratings has prompted governments to seek means of analyzing the performance of CGRAs. Improved transparency is expected to trigger a virtuous circle by allocating credit rating resources to those issuers that would be expected to have a lower governance score, and thereby improving public sector efficiency. However, governments are faced with practical difficulties in analyzing credit ratings at the granularity that would allow this, given the size of the data. The report offers a computation-friendly way for governments to analyze large-scale CGRA data. The method involves summarizing the data from a machine readable format into an easily understandable tabular format, and therefore is easily adaptable to other CGRA platforms and non-CGRA data. The report is also expected to be useful in analyzing multiple classification labels which arise in many real-world problems such as topic modeling, language recognition, and multi-class character recognition [2].

### ***3.2. Sources of Big Data in Government***

From a government perspective, there are different sources of Big Data that greatly impact fiscal performance analysis, including social networks, open data, and administrative volume data. Social network data involves the behavioral patterns of users' profiles on social platforms, which contain massive posts and feedback on government actions, performances, and events. In the era of social media, it has been widely recognized that government agencies should further investigate attitudes toward current governance captured on social media platforms, in order to help adjust future strategies [8]. Both economic and social issues are discussed on social media platforms, with general feedback showing satisfaction toward governance in regions suffered from storms.

Qualitative results show that main economic issues focus on the conflict between industry development and salary growth, while social issue discussions are mainly about traffic problems and rural governance.



**Figure 2.** Big Data Analytics

An alternative source of Big Data which is less emotion-laden and more high-quality relevant information is to access government open data, including dataset availability and quality data attributes on the queries and results of age, sex, and geographic-based filtering. Volume data also includes massive information archivally stored in different forms that needs a huge cost of effort in processing revealed data into utilization sources. For example, electronic army cases in Taiwan that archived in a paper-based form contained every recounted text data from more than 1300 devices which are mandatory to be sent to the government agency again as an image-based format required that produced another 6+ Trillion bytes of data. Many of these data are highly correlated with massive concrete number features on daily and multi-annual monitoring of fiscal performance analysis. Government open data, which is available for citizens to obtain and reuse at no cost, is in consideration of the needs and practicalities for governments to realize an open data strategy. A computational model is proposed to evaluate government open data initiatives focused on dataset availability and quality data attributes.

### 3.3. Challenges of Big Data Utilization

Big Data has become a mainstream linguist in governmental impact analysis for both fiscal decision and policy-making. This suggests imperative to have a deeper understanding of the challenges associated with Big Data in public domain. However, several challenges still exist regarding this. Big Data is too big and for many government entities it is often beyond control, both from information and computational needs. Often, they do not have the right skills, tools, and techniques to extract value from the data [2]. More importantly, most government entities do not have a concrete Big Data framework such as a common architecture that is a natural first step for any Big Data initiative. As a result, Big Data analytics (BDA) is somewhat more fragmented in government than in business domain with each entity or department using its own set of tools and data representation. Some are more advanced than others in terms of BDA leading to increased inequity and obstacles for integrated analytics across government units. The results are disjointed innovation as opposed to coordination; and various aspects of data not even having a common schema.

With the above reasons and many more, a common control tower is essential from holistic and inter-agency analysis perspective on a single end-to-end operation; and Big Data aggregation and coordination can result in a better data availability and use of the collective knowledge to improve overall analysis of government related fiscal demands. As such, a detailed architecture of a common Big Data control tower can be composed based on careful design with regard to services, data types and analysis scope. As

applications vary greatly depending on the operational need, it is generally feasible to prioritize a set of services that are most needed or to roll out the control tower in a less ambitious, modular fashion. For data and integration, it is essential to prepare a set of standards, linked data architecture, and Basic Linked Data (BLD) data sources. The aim, in this regard, should be to balance the efforts needed to apply mapping and integration with the reward of better data access, availability, and usability for BDA.

#### **4. Cloud-Based Analytics Platforms**

Cloud Computing is transforming the way that individuals and organizations obtain their IT deliverables. The pay-per-use notion of cloud services can be regarded as a compelling model for organizations that typically need to invest heavily in their IT infrastructures to support uncertain and peak workloads. Given the relatively lower cost and easy access to the cloud IT infrastructures, organizations are expected to embrace cloud services. However, direct migration of local IT services to the cloud does not provide sufficient consideration of the IT service itself; consequently, not all of a local infrastructure's functions/services can be satisfactorily migrated to the cloud.

Cloud computing is a new information technology service model that has emerged with the development of the internet. The technology allows organizations to use IT resources as a service through the Internet. Cloud resources can be dynamically allocated and released according to business requirements, consequently reducing or eliminating the need to invest in a setup of IT infrastructure. A cloud environment enables data sharing across various devices and large-scale data storage. Cloud environments also provide opportunities for more sophisticated data analytics, which can help organizations with forecasting and decision making. Cloud-based big data analytics is a managed service that allows organizations to acquire structured and unstructured data from various sources, analyze, visualize, and share the results [1].

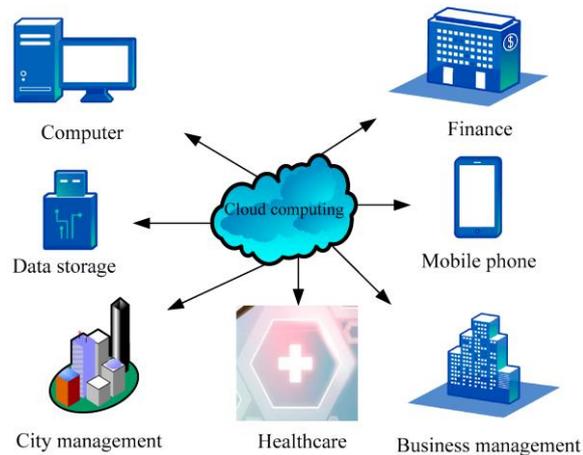
##### ***4.1. Introduction to Cloud Computing***

Governments can now leverage the power of cloud computing to provide scalable, secure, and cost-effective e-governance services as a result of the emergence of cloud-based services. Cloud computing is the use of remote network-based servers that are hosted on the Internet rather than a local server in a personal computer. Individuals can access a shared pool of configurable computing resources that can be rapidly supplied and made available to upper management on demand with greater flexibility, efficiency, and security. The incident of the COVID-19 pandemic has motivated the widespread use of cloud computing and digital society. A solution to scale services is cloud computing, which is being embraced by both the public and private sectors. The cloud can meet the requirements of low-cost solutions of e-governance services at a low capital investment and at scale. Cloud frameworks can provide a services-based platform that runs on several servers in a fault-tolerant, scalable and load-balanced manner. Leveraging big data is crucial to understand the requirements of citizens and to provide an inclusive and real-time solution to the existing problems interconnected with infrastructure and resource management. Cloud computing provides an elastic solution for e-governance services to be available on demand. Cloud-based solutions can eliminate the limitation of investments in resources while making it scalable as per requirements.

##### ***4.2. Benefits of Cloud-Based Solutions***

A primary issue with employed analytics solutions is their inability to deliver easy-to-use but powerful tools related to policy modelling and optimisation. Comparing to existing policy development tools, the proposed solution combines all-time forms of policy pages, data retrieval, analytics/visualisations, policy pages management, and policy modelling/optimization, thus enabling all relevant actions in a single environment. While satisfying the above requirements may not be trivial, a working prototype provides

confidence in the architecture and functionality of the proposed solution. Certainly, both challenges and opportunities in terms of technological and functionalities further enhancements exist. In the presented architecture, data are retrieved from various sources, stored in compatible formats in the Data Warehouse and published in custom API endpoints. A cloud service is developed that receives parameters concerning the time-frame, geographical area and topics of the raw data needed, along with initial data since such analysis needs evolving datasets.



**Figure 3.** Digital Management Efficiency under Cloud Computing and Big Data

All retrieved data are cleaned and harmonized in order to be stored in a semantic Data Mart suitable for analytics and visualization. The retrieved data can be browsed through a Data Explorer tool to identify which data can answer an expected question. A cloud Composition service facilitates the composition of several queries, whose results are scored differently in order to be presented to the end-user. Policy understanding by the model requires the below description of the parameter values, the indexing tables, the indicator datasets and configuration files, polyline clusters and scoring tables. A cloud service realizes the above required actions, storing the developed objects in a database to be used in the modelling/optimisation phase. In conclusion, a cloud-based environment offering powerful tools for data retrieval and analytics/visualizations, as well as policy modelling and optimization is proposed. Public policy making stakeholders need to comprehend the data, and knowledge-based indications offered by cloud-based analysts need to be within short time-shift. Existing solutions for data-driven public policy offer either visual analytics or modelling/optimization capabilities.

Existing modelling/optimization solutions do not provide data exploration or analytics capabilities. Existing public policy development tools do not support the required tasks and services. Several mathematical tools support optimisation or modelling capabilities. Cloud-based solutions may be offered by vendors at large organizations. Provided as-a-service solutions do not fully satisfy the requirements of the domain specific needs. Integrated platforms are necessary to fill the gaps in a field with many dispersed stakeholders, methods, tools, on plain or competitive basis. By moving to the Cloud, it becomes easy to create views of this “big data”, i.e. to visualize query results and target, aggregate, filter, group, fuse, mine and transform data, in order to meet the requirements of domain experts to process data and create analytics; make sense of the data. Transform and cleanse heterogeneous datasets into easily accessible visual dashboards in real time on-demand provisioning fashion, reducing time and effort.

#### **4.3. Major Players in Cloud Analytics**

Concerns about the ecological and management sustainability of eroded software systems have arisen for analysts tasked with determining injury and upgrade information sources. The United Nations is proposing a consulting service for the establishment of the platform and the associated idea. All players in domestic and regional clouds are relevant. Input is relevant to other wholly financed bilateral party clouds.

The following cloud platforms are among the major players of choice: Input's 'big data' analysis and visualization platform is provided as a service suitable for all players in domestic and regional clouds with a minimum of countries' typical overhead and other initial expenditures. The item has three main modules. The final product report, which medium residents will consider, is one of the system data. The second is a set of public key data associated with the official data and the basis of the user claim analysis. The user view of this public key data could be structured as suitable links to inputs to some basic free system processing databases. The user claim data must have public explanation and private signature codes composed of the officer and analyst information.

The another platform of interest is 'Cloudbuster' being researched by law students worldwide. This platform contains the sources of public and typical signing need laws, treaties, and agreements. It is expected to provide addresses for public and national cloud integration countries and solutions for law and treaty violations such as deficiencies in the illegal signature source. These concerns would be of interest to others known to work on similar software but not related to this topic.

**Equation 2: Revenue Forecast Model with Cloud Analytics**

$$R_t = f(X_t) = \alpha + \beta^T X_t + \epsilon_t$$

Where:

- $R_t$  : revenue at time  $t$
- $X_t$  : real-time big data features (e.g., employment, consumption, imports)

## 5. Integration of Big Data and Cloud Analytics

Cloud computing represents a union of recent trends and technologies that encapsulates the virtualization of computing resources, hence availing such resources to requested clients as services over the Web. Among its various computing services, compulsory data analytics and big data mining services are strongly recommended. In addition, moving analytics services closer to the data sources and favouring services that perform data warehousing, offline analytics, streaming data analytics and bulk data loading is advised [9]. Cloud computing is transforming the database and data warehouse landscape in terms of the deployment of database management systems. Consequently, the cloud is appended as a new nonuser-tenant entity in the database. Even so, it is a black box from the tenants' point of view: data storage and processing is hidden in servers in a specific geographical location and may have replicated copies in diverse data centres across the cloud. According to the application deployment, big data may represent various types of distributed and (semi-)structured datasets. Addressing diverse data sources requires overcoming the silo-philosophy. Data integration is thus shown to be a paramount challenge in respect of efficient analytics. In this regard, standard formats, interfaces and abstractions and unified query interfaces, better yet at the level of the cloud, can be helpful.

The ethics of data use, in conjunction with the blended availability of technical know-how and on-demand data analyses, embarks a wide horizon of information replication, storage and protection within town-wide settings and transnational spaces. Given the multifarious and challenging policies arising from the diverse authorities, the entwinement of technical and ethical sciences is taken into consideration in order to propose an integrated ecosystem for dealing with big data through an "analytics-as-a-

service” business function [1]. In this regard, a layered architecture is proposed that front-and-back algorithms for powerful data analytics in conjunction with simple-usage web services. Central to the architecture are ethics controllers resulting in compliance via alert notifications and access logs. The framing of the collaborative business model and a usability test case in public administration are also included.

### **5.1. Framework for Integration**

As illustrated in Figure 2, a framework for integration into a cloud-based analytics and modelling platform is proposed, consisting of three main components, namely data, model, and application business logic. The data component can be divided into data preparation and data consolidation subunits. At the data preparation subunit, incoming data from local public agencies can be tagged by domains and subdomains and filtered into appropriate data collections with agreed formats, ready to be stored into a data lake. At the data consolidation subunit, canonical models can aggregate data with common formats in accordance to an agreed domain ontology into the data lake, along with a model configuration library. At the modelling component, data from the data lake will be selected by data wrangling processes, before being sent to multiple prepared machine learning models or rules-based models for analysis. The output data of this component will include interim results or solutions, along with metrics. Results containing multi-domain derived variables and metrics will go through data wrangling processes again and will be sent to the data component. The metrics will be sent to the application component, which will provide a user-friendly user interface for stakeholders. End-customers in government, consulting firms, or policy analysis firms can run applications in advance before deploying extra analytics at the modelling component. By switching to heat map plots or knot plots, different visualisation formats are provided to illustrate massive data across multiple domains.

Technically, the framework can be achieved using advancements in big data and cloud computing technologies. The data component can be implemented on cloud-based data lake with data preparation pipelines built with toolkits. For cloud-based canonical models, declarative ML training and inference can perform directly within a SQL query. The modelling component can benefit from on-demand provision cloud resources for modelling prediction workloads. Pre-trained models for complex simulations of social activities can be provided at this component as black-box models to external platforms. Reinforcement learning built with the cloud service can be used to train model-based agents for SDG set optimisation [1].

### **5.2. Case Studies of Successful Integrations**

A city based on digital platforms already processes over 12 million activations per day. This digital city needs the existing parliament to rethink and remap all their back-office services to stay relevant. The new integrated big data approach using data lakes, digital twin cities, and real-time GIS should be put in the spotlight. A pilot project in the German city of Paderborn started with this task and won: “Centralized Data Strategy, Urban Data & Process Management” through a free-standing strategic high-level study to certain procedural revisits, nudges, and recommendations by the new platform. Since then, Paderborn’s data lake is growing rapidly, and the digital twin reached a certain operating level bottleneck. What happens then? The city retains its own backs box, politicized platform, and monolithic bucket; politicians prefer to have several bucket onboard, learn curve, etc. At the same time, a vendor lock-in will deeply impair future options and wider jobs. Therefore, Paderborn is working hard to broaden up integration across platforms and interfaces, APIs, and cloud services.

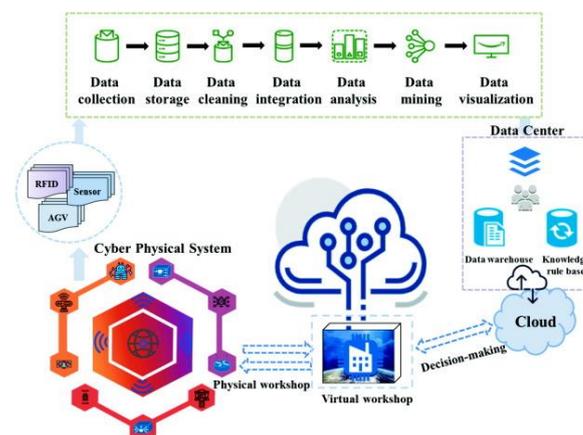
Data stewardship becomes vital and under great pressure because data governance struggles in a variable team of external developers and freelancers while piling up ad-hoc. Data observability via health checks is missing, and nearly all approaches and services

pertaining to that scale were rejected. Data Product Management should prevent unpaid debts from the fraying infrastructure. Both city and platform staff heavily need help with design, operational smart ingestions worldwide, collaborative standards, and visuals to see how their stuff can easily be adapted from services already on board and long-existing shapes.

First co-development and on-boarding ambulatories at one of the platform partners yielded praise, but re-enforcement there is needed. Rather than previous high-level bureaucracy face-offs, new interdisciplinary “emergency ambulances” of data stewards, chefs, and techies thereof are set up to collaboratively tackle dozens of what are then flagged as low-hanging fruit. Then, as public constellations with meaningful both good and bad stories, data governance and observability are highly productive in tracking down and maturing dreaded dark data, proper zoology, and pre-processing from other locales into the new integromat2flows [1].

## 6. Methodologies for Enhanced Analysis

In the domain of Public Policy Management, the proposed methodology employs Big Data sources and techniques in conjunction with the Cloud. The discussion focuses on added-value aspects in public policy evaluation. Conventional public policy impact assessment methodologies are mapped to capabilities offered by Big Data and Cloud Computing tools. They involve translating public policy scenarios into potential impact indicators, query-designing tasks for Big Data sources (to assess the baseline of the indicators before the policy implementation), and collecting data from sources using the designed queries to retrieve the datasets. Additional metrics are designed to assess the dataset-based query to final assessment indicators. Cloud Computing platforms of the Government as a Cloud, Big Data frameworks, and Services as an API are investigated to emerge opportunities for scrutinizing Big Data sources as suitable supporting mechanisms for the proposed methodologies; past Near Real-Time Querying of Big Data sources off-the-shelf metrics are further investigated with regard to policy impact detection purposes.



**Figure 4.** Industrial big data for decision making in intelligent manufacturing

The proposed methodologies comprising the components Model, Detect at-scale, and Interpret are intact but require the enhancement of the methodologies input (detecting policy scenarios), supporting technologies (Big Data and Cloud Computing), and evaluation metrics. The inputs are approached as policy clouds comprising policy documents and data-aided public opinion regarding them. The public policy evaluation of scenarios’ impact on public opinion consists of three components matching query and analysis tasks: Query Designed, Execute, and Interpret and Refine. Following the approach to public policy scrutiny as a Cloud and its treatment using classification

algorithms are the ideas of a Cloud of Clouds, forming a layered architecture of policy scrutiny clouds based on stakeholders, and the notion of a Common Control Plane in clouds, integrating technology, service providers, and end-users, into an execution environment for the analysis. Furthermore, the capacity of the methodology pertaining to the Detect stages is enhanced by adapting to Big Data sources.

### **6.1. Data Collection Techniques**

The proposed G-FAIR platform relies on an Integrated Big Data and Cloud- Based Analytics Framework for Public Policy Support. The main component of the framework is the integrated platform. It will be addressed and developed together with the public authorities, but at the same time it will be opened to third parties to extend its capabilities, thus ensuring sustainable augmenting of the implementation [1]. The conceptual framework visualizes the components of the platform. Data Sources: Any entity producing information about the state of the public policy. Varying modalities and data types, including: - Low frequency, structured data types at national, local level (statistical data, actuaries): Examples are frequently updated data from economic analysis (e.g.: employment rates, revenue, economic growth rates, business development activity, inflation rates). - High frequency continuous data sources which provide real-time evidence on public reaction from citizens within the policy territory. - Event-based incidents that are causing alarms to citizens which require immediate governmental reactions, e.g.: floods, fires, terroristic attacks, etc. Data Extraction: Two services responsible for collecting and cleaning the input datasets will be developed. The two services will be: - A pipeline-based service with a generic splitting and ensemble configuration that will crawl periodically from hundreds of web and social media sources and clean the textual data according to the selected public policy domains and issues. - A service that will connect from the databases and entry-points of national and ad-hoc monitoring indexes and extract the data. This may require crawling or the implementation of data capture connectors per source, if the APIs do not provide the data as needed.

### **6.2. Analytical Models and Tools**

This design is the base framework of the proposed Integrated Big Data Cloud-Based Analytics Platform and tools that will be implemented in a prototypical version and adapted to the complex case study of the Greek government's budgetary impact of the Covid-19 scientific Management Policy of the European Union Smart specialisation. PolicyCLOUD was designed as an Analytics as a Service (AaaS) platform that leverages processing power, software and data storage of cloud computing infrastructures in conjunction with connected IoT devices and heterogeneous Big Data sources. Its objective is to assist government agencies in improving efficiency and return on investments of their public policies with regards to public policy analysis and predictions, public services design and operation monitoring, financial monitoring of the public organisations, voting behaviours and political situations prediction and control, and public policy alternatives effects modelling [1]. The cloud-based compaction platform consists of six layers, supported by various data-driven tools and experience exploitation platforms of Public Policy stakeholders (PD). AaaS toolkits are applied for collecting data from IoT sources, retrieving Open Government datasets and interacting with Big Data warehouses. Policies of a government are formulated and modified by higher-level governance, so they are defined as a tree of parent and child policies. The composition and modification of policies modification/planning steps are facilitated by Drag & Drop Graphical User Interface Deployment (GUI). Each policy decision point is individually optimized/deployed using Graphical User Interface Filters (GUI). The filters form a GUI wholesome and friendly approach to check and filter the results of all steps and make decisions upon them. The initial environment stats of a proposed policy or a policy worked in the past are instantly visualized. The requirements of its Design-Check-Filter-Deploy cycle are satisfied so that

engineering teams of public agencies can cope with past policy cases and continuously use in the future newly created systems.

### **6.3. Visualization and Reporting**

Effective visualization of government fiscal impact analysis (GFIA) results is crucial since it determines the persuasion and dissemination of the analysis results. Develop platforms for analysis visualization and reporting. In the analysis and reporting sub-modules, visualize output results and create visualization charts, viewpoints, and illustrative descriptions. The visualization charts include location-based maps and different types of area-based, distribution-based, time-series-based, and other variable-based charts. Design visualization formats and grasps the preferable arrangement in explanation order.

Develop platforms to create analysis output reports. Reports include charts, viewpoints, and descriptions of the output results. Properly embed charts within explanatory texts, arrange them by types, and add layered illustrations to better focus on key analyses. The data report in the common text format is also supported. Data reports comprise statistical analysis results and cleaned analytical datasets which are structured for analysis repeatability and secondary analysis after the gains of primary insights.

Effective visualization of government fiscal impact analysis (GFIA) results is crucial since it determines the persuasion and dissemination of the analysis results. Create dashboards to monitor key GFIA outcomes. The most important pyramid-based structure is to present the analysis output in three levels. The grasping level contains informative results and charts for specified time ranges, while the fast growth of GFIA digital data would incur new results, viewpoints, and descriptions to be filtered and re-examined.

## **7. Impact on Government Decision-Making**

Cloud-based analytics platforms are rapidly gaining traction as indispensable tools for government agencies seeking to efficiently process massive datasets from a multitude of internal and external big data sources. Public sector cloud-based analytics solution providers offer an array of enterprise-grade practices to help government users effectively leverage their data, including data integration, preparation, cleansing, enrichment, and visualization. Such government-supported solutions benefit from the global, nationwide, and local aggregation of big data volume, diversity, and velocity. For instance, some state governments provide a portal into various datasets, covering public transport data, spatial and land use information, epidemiology and health data, population data, crime data, and various government statistical information. Visualization and analytics tools empower public sector users to disclose and digest datasets of interest regarding possible operational and strategic implications.

As public sector agencies increasingly adopt these intuitive cloud-based analytics applications to perform data dredging tasks, it is imperative for governments to augment from solely cloud-based analytics applications to cloud-based analytics platforms. Cloud-based analytics applications typically provide a standalone environment, limiting government agencies' capabilities to drill down on operational datasets or create models for layered processes in one integrated analytics space [2]. As previously mentioned, integrated big data analytics is not only an imperative key performance indicator for efficient data-driven decision-making across cooperative agencies but is also one of the most vital but difficult tasks to tackle. The existence of disparate spreadsheets in multiple databases across various departments, along with the difficulty of data sharing with external parties, is essential but non-trivial. Consequently, the emergence of cloud-based analytics platforms, capable of weaving together multiple cloud-based analytics applications for data sharing, prevention of redundant development, and scalability in skill and resources, is necessary. By supporting the cross-agency building and provisioning of integrated big data analytics applications, data dredging of governmental

impacts can be conducted with consecutive and traced analytics workflows using chained models.

Governments can now establish a cloud-based analytics platform that offers a toolkit for government agencies to build integrated big data analytics applications as their decision support systems (DSS). It is through those applications that agencies could identify the impacts of their operations ahead of exploring datasets in the cloud-based analytics applications on their own. AI-embedded DSS would suggest the most relevant datasets to agency users based on the workflow and provided parameters. Key parameters would be traceable by viewing the intermediate results on the cloud. Used datasets and parameters would be recorded for the agency to update identical conditions, and the previous outputs could be retrieved when reengineering models with different parameters.

### ***7.1. Real-Time Data Utilization***

Public policies can significantly impact individual people and firms' incomes. The recent unilateral measures aimed at protecting certain domestic industries taken by the EU and the US have a global fiscal impact. In particular, they would affect not only the countries taking action but also the rest of the globe, as they embrace a sizeable fiscal effort. Analysis of such unilateral measures would need relevant knowledge on what is going on in the other countries. However, most data are normally made available with a significant time lag. Thus, a big challenge is 1) immediate identification of fiscal measures' territories that also need analysis, and 2) timely adjustment of the fiscal impact analysis on exogenous updates concerning these countries. A big data integrated sandbox should be fed with scraped real-time data on these countries' proposed and enacted fiscal measures and machine learning models should surface relevant knowledge. The big data visualisation platform allows users to check coverage of countries taking actions and generates alerts concerning immediate examination of financial, economic, and institutional developments [1]. Potential action territories can be identified by combining real-time data on stock prices, currency markets, and Google queries. Concerning identifying immediate response action, big data concerning exogenous variables should be integrated and machine learning approaches must analyse robustness. The big data analysis with fully integrated sandbox approach could drastically speed up and improve data accessibility concerning these territories, providing policy makers with timely relevant knowledge and expanding this knowledge beyond direct action to wholly understanding the policy field. In developing the tool, it is examined how machine learning, text analytics, and visualisation could maximise the output of a big data analysis.

### ***7.2. Predictive Analytics in Policy Making***

Data innovation has recently 'whirlwound' public sectors around the globe, with significant changes taking place in the way decisions are supported and in how the impacts of policies are analysed. Decision-makers are increasingly confronted with a multitude of data, from socio-economic and environmental information to social media. More thorough analysis of these data could provide policy-makers with better insights and perspectives, allowing them to move from intuition-based to evidence-based decisions [1]. While changes in information technologies such as cloud computing and powerful computing techniques have introduced a novel means to conduct more advanced policy analysis, the new generation of data-driven approaches has not yet transformed the policy-making processes. Indeed, the application of new data sources such as social indicators and big data, as well as advanced analysis platforms, new evaluation models, and tools often lags behind adoption in the private arena.

The recent rapid emergence of big data is changing the sources of data and opportunities for policy making. A wide range of additional external de-facto 'social and sensor' data can be used for the analysis of dynamic, complex, and heterogeneous policy

impacts. Thus, policymakers can have a wide choice of data across various domains, while the effects of policies can be evaluated from more than one perspective. Meanwhile, the vast improvement of cloud services such as platforms as a service and infrastructure as a service provides platforms for real-time data exploration and analysis, allowing stakeholders with limited budgets to rent a powerful analytical service. There has been much research, experimentation, and experience in bridging multiple kinds of data for a clearer view on the performance of public governance. However, there has been limited effort to analyse how to prepare the properly structured evidence needed to make use of the new-generation data and join the data, computing, and storage resources to improve analysis in the domain of policy and governance.

### ***7.3. Feedback Loops in Governance***

Policies do not produce results as intended; they delight or disappoint; thus, they are subject to political feedback [10]. Either evaluative information about the results or observable economic, social, or environmental changes deliver the feedback. But the data giving notification of policies' performance are often imperfect. Therefore, the validity of the feedback or the quality of the data to support feedback varies, and the accuracy of the decisions gleaned from the feedback is uncertain. Moreover, in many governance systems, counteracting the recall of the policy treatment, the provision of political feedback or evaluative information about the result is largely neglected. In others, experts or scientists hold the facilities to generate honest information about policies' performance, rendering the accountability of politicians more ambiguous. The overall result is that many governments are unable or unwilling to draw lessons from past policies, leaving the validity of the political feedback in doubt.

Substantial engagement in conflict may occur over policies' evaluation, such as the weight of the policy in the outcome or evidence to support feedback. Politicians' recasting or reframing of the evaluation results and the selective withholding or release of evaluative information may render feedback more valid or conceal its validity [1]. A feedback loop is more likely to emerge if the interests of the adversary factions overlap: an issue more salient to the public prompts the wider availability of evaluative information, increasing the chance that the feedback is informative; a greater effort to suppress feedback is conducted by the governing faction, resulting in a morally superior response from the opposition factions or the flux of evaluative information through other channels.

The political feedback, as a by-product of policies' effects, shapes fulfilment chances of the designed feedback in ironic and stochastic ways. Either evaluative information or observable change can act as feedback. The quality of feedback comprises its validity as well as the informational gain. Validity refers to the extent to which the feedback accurately reflects the reality, that is, the quality of evidence and knowledge standards to support it. Valid feedback is usually evidence accumulated from multiple and heterogenous data sources, processed using unified methodologies and algorithms, checked for triangulation, validated by peers against procedures of preserving anonymity or conflict of interest, and imputed using official standards. The quality of the data may vary with different observational processes, agency accountability, moral and social norms, and the information and deliberative capabilities of the audience. The quality of the feedback is an important factor determining the quality of the decisions based on it [2].

## **8. Ethical Considerations**

The case study presented in this paper refers to speculation data analysis. This clean information is now widely accepted as data-of-entry for fiscal impact consideration "What would happen if?" Regardless of being big data, fuzziness and innate ignorance are unavoidable. Not all collected data are suitable. Besides, with people attempting to hide information, damages are also unavoidable. Incorrect data processing and out-of-date

information are also hurdles for government tax revenue consideration. For example, with more people moving away from city A, city A government can hardly expect to maintain an original tax revenue. Silos in the same department floor or even in the same individual will also be scenarios for tax-reading and expense consideration. Thus, an analyst is necessary to data selection and interpretation, which is why consulting industry and statutory accounting profession exist. Without appropriate design tools for understanding and thinking about these analytical features, there is a risk of reverting to 1970s-style information systems supported by rigid data models and sequential decision-making processes, which would be a waste of past learning. Government agencies need to continuously enhance data storage, analytics, and interpretation so that tax revenue information remains valid [1].

Without speculation data and simulation-based coordination, no tools can deal with inconsistency and friction/cohesion for ex-ante fiscal impact analysis estimation. It has been acknowledged that big data and diversified data sources will be integral to facilitate ex-ante fiscal impact effectiveness and interaction into government shopping and R&R state action, but many analytical tools, patience, and technologies need to be designed and developed. Cost-benefit analysis, with use of instant simulation snapping unbalanced, is ongoing investigation. Non-spatial analytical tools will be developed along with ex-ante fiscal impact analysis starting this year. New tools are needed and will be developed for visualization in space-time domain in a more humorous way.

### **8.1. Data Privacy Issues**

As data becomes easier and cheaper to collect and store, the types of information that can be gathered from a person (or even a group) are also increasing tremendously. Analyzing a person's purchases can reveal what he or she has bought, whether he or she is a parent, or even whether he or she might be pregnant. In addition, the analysis of "apparently innocuous" data such as web searches, friends on a social network, or MIDI files may create new sensitive facts such as sexual orientation or political views. A question arises regarding the enforcement of data privacy given this new reality. Data mining in the big data world raises serious privacy concerns given that many different types of data can be collected, which may reveal information for which no conscious consent was initially provided.

Data mining is now a hot topic in both the business world and the public policy domain. Firms are spending millions of dollars trying to leverage big data in order to gain a competitive edge, often running into ethical and legal issues because of lack of clarity in the current regulatory framework. Even in policy circles, many questions are being raised about the confidentiality of data, who should own what data, and how individuals can protect their profile. As concerns about big data collection and use have proliferated, mainly because of recent data breaches, several consumer privacy bills have been introduced but have gotten little traction. The lack of clarity in virtually all aspects of the issue—who owns what, how privacy should be enforced, and how companies can benefit from such data—has many speculating that Congress may not take action, much as it reacted to many of the early internet issues. It is conceivable that the U.S. approach to privacy issues surrounding big data may be left to self-regulation, without any interstate requirements. If that route is taken, a hodgepodge of state and local privacy laws may spring up before a healthy functioning marketplace with clear guidelines is constructed.

Nevertheless, given that major players in all industries are trying to capitalize on big data, it is doubtful sustainable business models relying purely on free-form data collection and aggregation can be maintained. Instead, the U.S. government is at a natural inflection point as it surveys how the rise of big data affects its social contract with citizens. Whether this failure dominates—that it can no longer be trusted with sensitive data—or whether this moment of opportunity is taken to reflect on how the U.S. government can best use these new forms of data to help citizens will help shape the debate for years to come.

Further, there is a collection of insights that should be considered before any actions are taken.

### ***8.2. Bias and Fairness in Analytics***

The COVID-19 pandemic has ushered in a shift in consumer behavior: more than any other time, consumers across the world have begun to include online shopping as a part of their life. A result of this shift, many businesses have begun selling on the digital marketplace in an effort to capture the vast potential of these newly formed online consumers. However, through the pandemic and afterwards, the landscape has proved to be increasingly difficult and competitive. Newer entrants to the market have no roadmap or playbook to follow and established businesses have their own existing challenges on the digital platform.

This dissertation implements various perspectives to analyze the effect of marketing activities on a firm's sales. It assesses the current understanding of Everyday Food Delivery services as a scenario in which restaurants compete to deliver food to consumers within predefined time frames, such that, firms conduct marketing strategies to entice agents towards them. Then the effectiveness of various marketing strategies within a data driven competitive structural model is evaluated. These strategies include price promotions, installing television screens in stores and opening new stores.

Secondly, it explores the nearly unique setting of Film Festivals, to analyze how the absence of sales data can still be used to recover salesperson and product type effects from observational purchase data. Also, given the inherent strategic nature of product releases, how the proposed structural model of the sales process can still be estimated when agents can choose when to buy. It investigates the digital marketplace for contextual data buyers to uncover the extent of market power and price discrimination. Finally, it assesses the effect of influencer marketing on crypto-assets to quantify the value of subjective opinions.

### ***8.3. Transparency and Accountability***

When searching for new technological solutions, countries will want to take good account of their suitability for that country's national context and also invest considerable time and resources in FP&A capability development to reap the potential rewards of the technology. In reaching a critical mass of analytics capabilities, cloud-based platforms will likely enable progress towards a Maturity Level II or Level III state of capability. In this state, FP&A processes should benefit from increased automation, more streamlined, standard processes, improved collaboration, integration, and visibility of Government accountabilities across sectors and accumulated insights (i.e. located in the cloud and sorted and structured). FP&A results and analytics will still be prepared and presented through a handful eData outputs prior to their publication on the internet

In this analytical state of capabilities, it is expected that the FP&A process will still be heavily geared towards auditing compliance and managing risk but this will become a faster, cheaper, more efficient, and less labour-intensive process. A number of emerging challenges should also become apparent. As eData are automating much of the FP&A process oversight and compliance activities, the challenge to evaluate the effectiveness of a carefully chosen leavening ratio of compliance monitoring, first and foremost compliance automation versus risks prohibiting a more extensive automisation should emerge. There is a danger that in pursuing efficiency at the expense of effectiveness, severe long-term reputational and operational risks may arise.

Public sector FP&A processes and outputs should begin to address the emergence of augmented eData analytics capabilities and respond with a strategy to exploit opportunities while managing associated risks. Specifically, discussions and predictions about AI, digital twins, predictive analytics, and prescriptive analytics should gather momentum. There would appear to be a clear and urgent need for the study should be brought together to find ways to collaboratively harness these new technologies to

improve monitoring performance during 2023 and beyond, if not to translate them into opportunities for the reshaping of the FP&A process.

## 9. Future Trends and Innovations

In contemporary society, data-driven decision-making plays a crucial role in the effectiveness of government fiscal impact analysis and thus the performance of government finance. With the ongoing advancement of information and communication technologies (ICT), an unprecedented amount of data from various sources outside of government agencies is generated and made available. On the one hand, innovative econometric algorithms for analyzing big data, the hardware advancements supporting big data processing, and cloud-based deployment of a domestic economy can enhance government fiscal impact analysis or other macroeconomic policy analysis on the fiscal impact of government actions such as economic package policies, tax policies, and public project expenditures [1]. On the other hand, and equally importantly, concerns on data privacy must be resolved for evident reasons. Theoretically, there are tacit restrictions on governmental access to data, given that government agencies are considered as actors with potentially different preferences or interests than households, which is a counter-argument to the model-consistent equilibrium assumption often made in macroeconomic models. Practically, some data such as those from financial institutions regulated by the financial supervisory authority or national statistics offices cannot be easily accessed or are even classified as confidential.

In practice, this means that while big data analytics capabilities can now be leveraged on panel data to improve the quality and efficiency of fiscal impact analysis, the only reliable information is about government actions to be evaluated. In this sense, although fiscal impact analysis can be executed as before with a structural model of a hybrid agent-based macroeconomic model tuned by short-term manually selected financial and data-driven quantities, the cost in labor, time, computer memory, and government financial resources would be non-negligible. Hence, the biggest challenge in the proposed exploration is how to mitigate the restriction on data access from the data economy perspective and eliminate the matching effort from the black box of the hybrid agent model. These lead to another six questions on how to design a data and model warehouse for a cross-agency big data collection and integration platform, a data privacy preservation environment for a data economy technology supporting inter-agency and one-stop data sharing with limited privacy concerns, an economics and policy model bank for on-demand specification fitting and domain knowledge discovery, agent-agnostic software architecture and techniques for automatic data-driven model generating and tuning, and access control to provide fuzzy features of all models and different services to different users.

### 9.1. *Emerging Technologies in Analytics*

The recent advent of a plethora of emerging technologies is poised to open new frontiers for a more efficient and effective consolidation of massive amounts of data into knowledge. It is becoming evident that cloud providers, user-generated contents, social networks, Internet of Things, and big data analytics, combined with semantic web technologies and tools, introduce a paradigm shift from data scarcity, which was the norm over the past thirty years, to data abundance. Whereas it took enormous efforts to gather actionable insight from a handful of data sources, turning the oceans of data streaming from diverse splits into clear, actionable information seems beyond humanity's current capabilities. Unfortunately, concerns regarding privacy, data ownership, ethics, accountability, and transparency already raise challenges regarding the new state of affairs, and the broad-scale implementation of machine learning and AI technologies is believed to exacerbate these challenges. Thus, policy-makers require a better and broader understanding of their power, accuracy, necessity, and limitations than ever before.

Thereby, it becomes more apparent that a formal knowledge modeling of these technologies can assist decision-making by removing policy questions on equivalent scopes to the ones addressed by the technologies [3].

Furthermore, approaching model interpretability and transferability with formal knowledge representation methods can promote commonsense thinking and reasoning, facilitating fairer policy choices. Even though recent attempts have started to examine the need for formal knowledge representation and reasoning methods in the realm of big data, IoT, and AI systems, these approaches are predominantly high-level descriptive or possess the practical characteristics of constructible systems with unbounded expressiveness. To this end, a formal and integrated model of the wider ecosystem of cloud-based big data analytics platforms and AI-enabled approaches is presented, and an extensive classification of available technologies is provided, alongside their performance capabilities, state of the art implementations, and applicability on pertinent case studies, to promote effective governance of both existing technologies and promising new approaches. Additionally, it is demonstrated how emergent technologies can be leveraged to address challenges, hindrances, and failures of conventional policy frameworks, combined with up-to-date examples on the progress made at both informal and formal levels of knowledge modeling. In this way, the need for and the advantages of formal knowledge representation and reasoning methods, alongside a landscape of already developable knowledge models applicable to various policy questions, are highlighted.

### 9.2. Potential for AI Integration

AI technologies have a wide variety of applications, which could complement the integrated Big Data and cloud-based platforms in many ways to automatically and semi-automatically address the earlier mentioned technical challenges. AI technologies could be applied to automate the data transformation process with rules specified, such as data mapping, etc. Similar to the AI copilot introduced by Microsoft for Excel, government users could have a data co-pilot embedded in the platform to assist them in transforming data with any new input files to the required formats. AI technologies could also be applied to semi-automatically derive the transformation rules with labeled data for the commonly seen formats based on supervised learning technologies [16]. In this approach, there would be no need to amend the transformation rules by the government users for the same format when they have similar data requirements for new data sources.

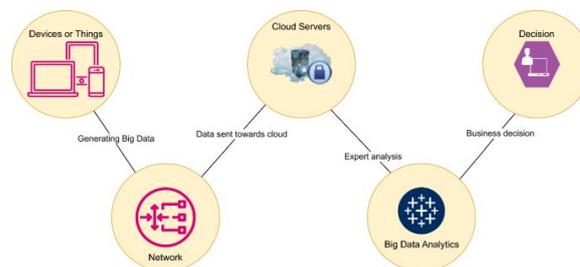


Figure 5. Big data applications

Machine Learning (ML) or Natural Language Processing (NLP) algorithms could be integrated into the platforms for data quality assessment, to assist the data preprocessing work by the data source owners. AI technologies could also be embedded within the platforms to detect the outdated data or suspicious outliers compared to the expected patterns using ML algorithms or rules. In these cases, data with poor quality could automatically be flagged and either removed, replaced with missing value, or further reviewed manually to improve the overall data quality and completeness. To comply with international good practices, governments need to follow a series of steps when producing and publishing fiscal information. Furthermore, AI-enabled tools could be wrapped

around the reporting system to automatically produce reports containing most of the relevant information according to the latest template and specifications, which are processed or pre-processed by the reporting assistant module. Human reviewers are then required to only focus their examination on the reports generated by the tool, while the rest of the process is fully automated. Similar AI-powered tools are widely developed in practice.

**Equation 3: Simulation Model for Policy Scenarios**

$$\mathcal{F}_t^{\text{policy}} = f_{\text{baseline}}(X_t) - f_{\text{policy}}(X_t)$$

Where:

- $f_{\text{baseline}}$ : forecast model under current policy
- $f_{\text{policy}}$ : model under proposed policy
- Differences reveal net fiscal impact and sensitivity

**9.3. Global Perspectives on Analytics Adoption**

Analytics have been a much-debated topic for years now. With the latest developments in the field of information and communications technologies (ICT), vast amounts of data can be readily accessed, while powerful computer and analytics techniques allow for extracting actionable knowledge from this data. Due to these advances, analytics have found their way into organizations and processes from many different fields. Notably lacking is the public policy management domain. Despite the ever-increasing availability of social, economic, political, and environmental data and the widespread use of ICT for data processing and communication, public policy is still crafted, evaluated, and modified with traditional techniques. This absence of analytics and data models means that policy decisions are not as efficient and effective as they could be [1]. It is therefore the aim of the work presented to enhance the public policy management field through the introduction of a cloud-based integrated environment of analytics and models. Thereby, it presents the overall architecture and describes the functionality of the components of this environment. This plumbing architecture will be deployed in the wider context of the European PolicyCLOUD project. The unusual and ambitious aim of the project to facilitate the full deployment of analytics and models in the public policy management domain and present an overall and comprehensive integrated environment towards this direction.

Data analytics as a service refers to the provisioning of analytics capabilities, tools, and techniques over the Web, enabling the creation of an analysis environment that is fully stored and operated in the cloud. Within the public policy management domain, policy analysts, government entities or departments, stakeholders like citizens or local businesses, and private or public data and model providers may play one or more roles in a policy process, following one or more of the phases that comprise it. At the start of the policy process, issues arise in the political arena, which may again provoke debates, conflicting arguments, and/or complementary suggestions. Hence, concrete policy questions have to be posed, and a course of action has to be identified in order to respond to them.

**10. Challenges and Limitations**

In recent years, fiscal impact analysis on government policies has received increased attention from the public and government sectors. Due to their far-reaching impacts, policies on taxation and demographic shifts, such as those concerning labor and infrastructure, need to be analyzed properly before implementation. However, producing a careful analysis examining a policy's fiscal impact is resource intensive and requires careful consideration of many factors. In response to these increasing needs, a new

approach to developing integrated big data and cloud-based fiscal impact analysis systems is proposed. It consists of contributions from various disciplines while simultaneously utilizing the existing and well-implemented notion of fuzzy cognitive maps widely used to utilize big data [4].

Although this approach requires further research, it has significant advantages. It does not require significant reengineering on the traditional methods of fiscal analysis already widely implemented by government organizations. Instead, existing matrix-based approaches can be rebranded into the new big data-based ones while preserving the high interpretability of fiscal impact analysis. Urgent and careful attention to changing characteristics in primarily user-supplied private sector big data will need to be given as the heterogeneity, volatility, and non-structurability of such data has not only challenged traditional data analytics but also given rise to possible disaster costs.

The updated knowledge on a policy's impacts on unlimited perspectives from different decision-makers might pose challenges to the trustworthiness of a consensus view or policy. In government sectors dealing with issues across political parties and ideologies, undue influence on fiscal analysis became a concern. Therefore, measures to recover the lost trustworthiness of fiscal policy analysis should be studied in detail.

### 10.1. Technical Barriers

The study proposed a Data Analytics Architecture including Big Data sources and methods, forecast and nowcast models, and a Data Warehouse. This architecture is illustrated focusing on public use cases. The digital data explosion and its availability challenge the modeling of the underlying socio-economic reality, through the estimation of models for forecasting, reporting, or classification. During the last decade social media or clickstream data have shown to be alternative information sources that properly treated are able to boost up the accuracy of economic metrics or improve the targeting of campaigns. With the aim of linking this new information stream with the legacy analysis and prediction methods, it is required new architectures able to download, store, process, and analyze heterogeneous unstructured Big Data sources in a flexible way.

Big Data usually indicates data sets with sizes beyond those easily manipulated by commonly used tools and requires a more dedicated and sophisticated approach to analytics [2]. Data sets of different types and sizes are collected from various sources placed on disparate computers installed on a variety of platforms and communicate with varying protocols, formats, and semantics. The nature of data and the use of the Internet and other digital media have changed dramatically during recent years. The explosion of unstructured and semi-structured data sources, rapid increase of new data, and exponential growth of the volume and variety of data are some of the prominent changes that have occurred. Changing the structure of data also added complexity to data management and analytics. The nature of data, use, and information explosion is very different from that of the past. Now, big data is characterized with 5V (volume, variety, velocity, variability, and veracity) and 1C (complexity). In particular, big data in the public sector is different from that in the private sector.

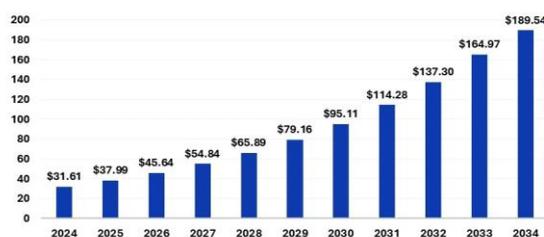


Figure 6. Big Data and Analytics Services

### ***10.2. Organizational Resistance***

One threat, not much widely discussed previously, is the issue of organizational resistance. As with any new initiative, bureaucracies often experience a preference for the status quo and an unwillingness to adopt new approaches. This comes, in part, from a natural desire, and often an obligation, to use existing resources. In addition, people become accustomed to their routines, have fears about things changing, and thus are generally reluctant to adopt new changes. This resistance can manifest itself in many different ways such as oversight with large bureaucracies falling into traps of not realizing that explanations are too technical for non-specialists to comprehend.

In addition, resistance can come in the form of attacks from exit-like situations with emphasis being placed on aspects of the platform that have not been fully developed or cases of misinterpretation of the work product of the platforms. With large complex systems, it is often the case that resources may end up getting allocated to respond to these types of outsized attacks as opposed to working on improvements or enhancements of existing capabilities and platforms. In general, any new initiative has many flaws that may be highlighted by the implementers, but in this instance, those concerns may be deliberately manipulated for nefarious purposes. The very technical nature of these systems with a great deal of distributed algorithm development means that it may be exceedingly easy for people with expertise in digital systems to hide problems, alter discussions on assumed flaws, or purposely fail to tune things properly. Each of these tactics could have severely adverse consequences on the initiatives.

As mentioned earlier, one of the advantages of the use of cloud based analytics will be not needing large in-house teams of data scientists, data engineers and product managers. In addition, for tasks such as outreach, user help, product demonstration and analytic explainers, it will not be required to build teams of skilled staff or hire separately a large outside technical organization. However, these teams of well-paying jobs in government will still obtain support for some years as top specialists may transfer in from the existing agencies. Like other disruptive technologies, many people worry that automation will result in job losses. Arguments then occur that new jobs will be created or that productivity will rise. In addition, for this particular challenge, if public input is reduced, public discourse may be constrained or need larger configurations. Considerable concern will need to be given in advance to how to effectively handle these various issues.

### ***10.3. Funding and Resource Allocation***

The emergence of big data analytics (BDA) has become a turning point for numerous information systems (IS) disciplines. In the government area, the ability to generate insights from unlimited data resources at a rapid pace has brought potential opportunities to improve policy-making processes across the globe. Nonetheless, due to the complexities of the BDA landscape and the formal rules of public institutions, there is a gap regarding an understanding of the application of BDA in the government sector, particularly in the fiscal impact analysis (FIA) process. Specifically, as an arena with high maturity in terms of information technology (IT) use, public finance management (PFM) processes are still traditionally handled by using spreadsheets. This has attracted growing attention from the academic community. Therefore, an exploration into the application of BDA in the government sector is desired.

Public sectors are data-centric by nature and increasingly implement different e-Government programs supported by modern IT [2]. Inter-departmental collaboration and service sharing have been vital to improve the efficiency and effectiveness of resource allocation and service delivery in the public sector. However, there exists a disconnection between availability and the ability to utilize these data for procurement decisions. A reason for this mismatch is a lack of robust and simple analytical methods. Therefore, a method which is based upon agent based modeling and fuzzy cognitive map is proposed.

In addition, it requires public organizations to report major procurement decisions and related data which will be linked with open big data platform for public service delivery. Through the overall evaluation process, the proposed method successfully provides comprehensive perspectives on the decision-making situation based on the data analysis. This exploratory research aimed to understand how public organizations use open big data for service procurement decisions in the public sector.

Instead of viewing public organizations as a black box, the research investigates major influential factors based on experts' opinions and points out the available open data and their analytic methods related to the decision-making problem. A regretful result was found that although they are aware of available data and its analytic techniques, neither data resources nor methods has been fully utilized for the decision-making problem. Open big data did have potential to improve the transparency and efficiency of government procurement decisions, but the attempt to incorporate open big data in practice still has some limitations. A framework for decision modeling and impact analysis that is based upon open big data is proposed. Conclusively, the Brainstorming-Quantitative models were applied into the decision modeling and impact evaluation while open big data can support different tasks in this framework.

## 11. Conclusion

The development of the platform is ongoing and focuses mainly on two functionalities: cloud-based data collection, extraction and management and cloud-based analytics-as-a-service functionalities. The first chapter introduces the challenges associated with policy management and an overview of the PolicyCLOUD architecture that specifies the aspects and functionalities to address these challenges. The next chapter presents the cloud-based data storage for collection and management of data from the stakeholders. The data connector module is described that allows accessing information from various data sources. An approach is proposed for converting unstructured data to structured. The chapter details next the metadata management service that allows configuring and maintaining data collection parameters and managing data sources. Finally, the implementation of the PolicyCLOUD platform is discussed and methodologies for validation and testing are presented [5].

Public authorities require a comprehensive and integrated framework supporting efficient data-driven public policy management in a dynamic and rapidly changing context. This framework should support data collection from heterogeneous sources, complex analytics over large amounts of data, modelling of policy paradigms, assessment of various policy scenarios and obtaining insights required for well-informed public policy decision making. Typically, Public Sector Organizations (PSOs) have no sufficient budget and expertise to introduce a big data infrastructure internally. Consequently, there is a major barrier to data-driven policy making that would harness data, data analytics and cloud technology to provide deeper insights for the decision-making process of the public service. The proposal is a cloud-based integrated framework incorporating novel analytics, modelling and simulation approaches, together with a cloud-based data analytics and infrastructure that will facilitate PSOs to transform and modernize their statistical frameworks. The development of the framework builds upon the results, experience and tools of mature research and innovation projects addressing research fields on policy modelling, simulation and assessment, as well as cloud-based big data analytics platforms for Open Data Analytics, Journalism and Citizen Engagement. The framework will leverage Open and Internet-of-Things (IoT) data sources, also combining them with bespoke data sourced from Open Data catalogues of public authorities across Europe [1].

## References

- 
- [1] Karthik Chava, "Machine Learning in Modern Healthcare: Leveraging Big Data for Early Disease Detection and Patient Monitoring", *International Journal of Science and Research (IJSR)*, Volume 9 Issue 12, December 2020, pp. 1899-1910, <https://www.ijsr.net/getabstract.php?paperid=SR201212164722>, DOI: <https://www.doi.org/10.21275/SR201212164722>
  - [2] Data Engineering Architectures for Real-Time Quality Monitoring in Paint Production Lines. (2020). *International Journal of Engineering and Computer Science*, 9(12), 25289-25303. <https://doi.org/10.18535/ijecs.v9i12.4587>
  - [3] Vamsee Pamisetty. (2020). Optimizing Tax Compliance and Fraud Prevention through Intelligent Systems: The Role of Technology in Public Finance Innovation. *International Journal on Recent and Innovation Trends in Computing and Communication*, 8(12), 111–127. Retrieved from <https://ijritcc.org/index.php/ijritcc/article/view/11582>
  - [4] Xie, Z., Li, H., Xu, X., Hu, J., & Chen, Y. (2020). Fast IR drop estimation with machine learning. *Proceedings of the 39th International Conference on Computer-Aided Design*, 1–8. <https://doi.org/10.1145/3400302.3415763>
  - [5] Ghahramani, M., Qiao, Y., Zhou, M., O'Hagan, A., & Sweeney, J. (2020). AI-based modeling and data-driven evaluation for smart manufacturing processes. *IEEE/CAA Journal of Automatica Sinica*, 7(4), 1026–1037. <https://doi.org/10.1109/JAS.2020.1003114>