A Big Data Architecture for Chile

Diego Arenas
da60@st-andrews.ac.uk
The University of St Andrews

Motivation:

A real-time planning and analysis architecture is proposed on this poster. The design makes use of modern big data technologies and intends to connect and explore synergies between private and public systems already implemented in the country.

The aim for this distributed data architecture is to held the chilean data and information, accountable for the citizens and transparent from the government. It encourages the use of Open Data and Open Source software. We need to distinguish between rapid changing data and slow changing data. For example born rate can be considered a slow changing data, while the demand of public transportation can be considered as a high changing data. We proposed an architecture to support both types of data.

Design

A distributed architecture has no central authority to control all the system. Each component is meant to be reliant, self managed, scalable, and fault tolerant. The communication between components is through the implementation of APIs, which is the piece that should be implemented in order to create this data architecture.

The exchanged data should be non identifiable by default and the exceptions must comply with the privacy and security standards. The requests and use of the platform should be public to anyone scrutinise its utilisation.

The microservice architecture, shown in Figure 1, allow the systems to interact without interference. Any system can be off at any time with zero impact to the other systems.
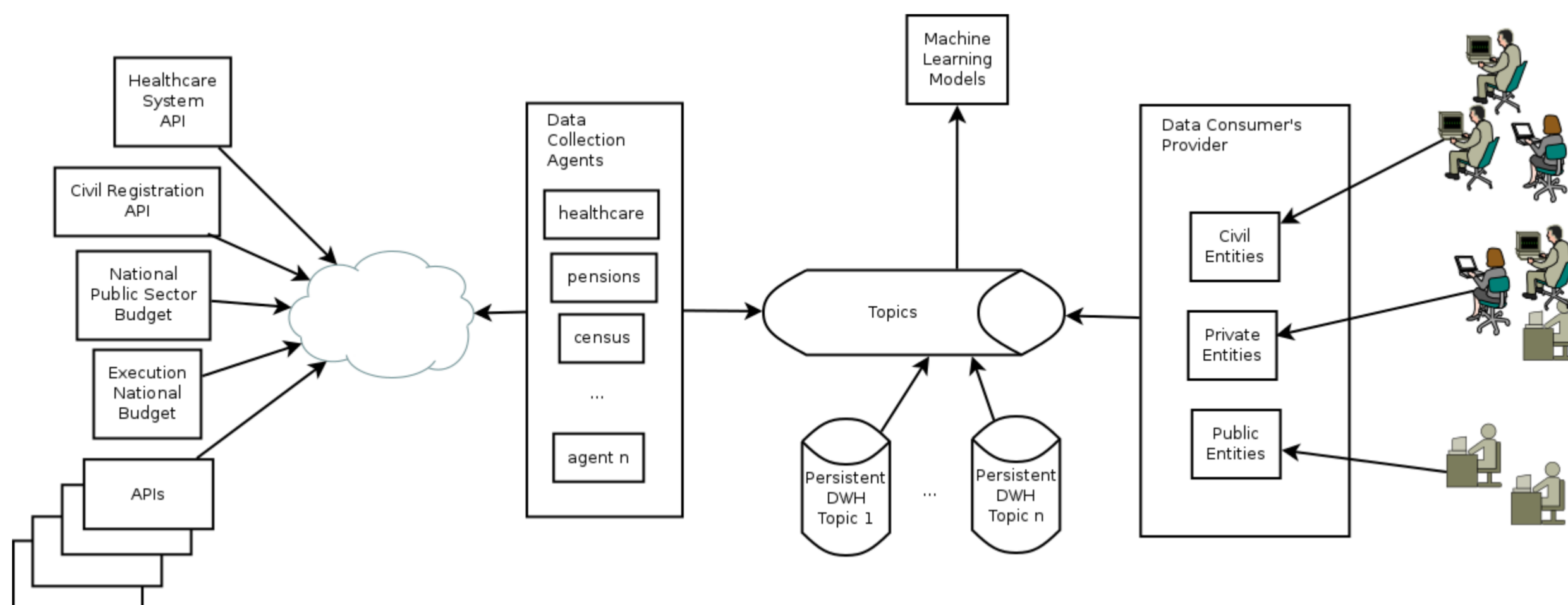

Figure 1: Microservice streaming data architecture for national systems.

Systems

The idea is to provide data access to public, private, and civil parties. The multiple data sources have been collected for years and the intention to spark innovation, improve public policies based on data at local and country level.

Systems like the Census, demographic and housing data. Civil registration data sources. The national budget for the public sector and its execution. Tax data from organisations. Financial data such as debts in the banking system. Utilization and demand of healthcare and pensions. The educational system in terms of demand, results. The banking and online payment systems can be added as well. Data from demand and generation of energy data along the country.

New data sources can be added in the future making this architecture extendible. For example information from tourism, and international trades could be included at any time.
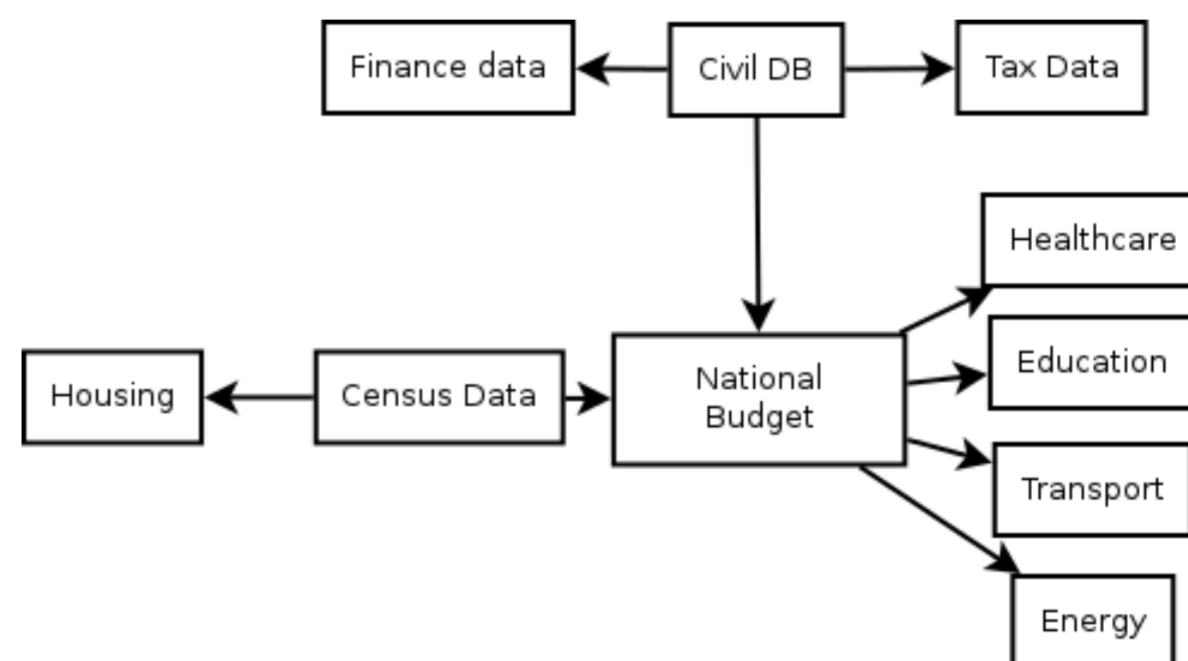

Figure 2: Relationships between data sources.

Conclusion

The current status of Big Data technologies make possible to integrate multiple information systems for the benefit of the citizens and development of the country.

The system can be extended with machine learning and artificial intelligence algorithms for better planning, forecasting, ealry information needs detection.

This architecture can be implemented by any government or organisations with the data available.