Big Data and Economic Development

Measuring the impact of economic development activities is critical, as demonstrating impacts guides continuous improvement in the field, as well as assisting in securing resources to further growth at the local or regional scale. Economic development has many stakeholders, including elected members of council, ratepayers, governments at all levels and economic development professionals interested to learn from the success of others. There is a pressing need to develop robust metrics on the impact of economic development programs and activities. These measures are needed in order to:

- justify the expenditure of scarce public sector resources by local governments, or indeed any tier of government;
- provide an evidence base on progress towards economic development goals;
- help local government agents to better allocate funds across economic development plans by monitoring and evaluating their impacts and ROI (return on investment);
- assist in the management of economic development funds by helping local governments to monitor their performance and impacts and take appropriate corrective actions;
- assist economic development practitioners in selecting the most appropriate and effective strategies in order to achieve their economic development goals; and,
- build momentum in economic development efforts by creating knowledge of local success that can be shared with businesses, the community and other key stakeholders.

Until this point in time, measuring the success of economic development activities has been challenging because of the difficulty – and high cost – associated with using conventional methods to assemble meaningful measures that reflect economic development outcomes. Output measures are readily assembled, but outcome measures – that speak directly to the impact of economic development efforts – are often difficult to develop and implement (Beer, Haughton & Maude 2003; Turok 1989).

Contemporary approaches to economic development measures suffer from several shortcomings:

- Official data is often only available irregularly, or at time periods that do not match the needs of the organization. Moreover, the scale of the data analysis is often challenging as key data may only be available at a very broad scale. This outcome reflects both privacy concerns and the limitations of the data collection process.
- There are several business data sets within the private sector, but these data sets are often complex and do not necessarily cover the types of enterprises – Small to Medium Enterprises (SMEs), business start-ups, emerging industries – of interest to economic development agencies.
- Primary data collection (i.e. the gathering of data for the specific purpose of evaluating economic development activities) is one solution, but attribution is an unavoidable problem (Turok 1989). Businesses often benefit from economic development actions but either overlook who provided that assistance, or are unaware of the boost they received. Moreover, the surveying of businesses or consumers can be expensive and often results in output measures, rather than outcome measures.

New technologies and data sources can shed light on the additional value created by economic development programs and activities, and programs and activities which have economic development outcomes. New data sources are increasingly becoming available globally, nationally and locally. Often referred to as 'big data', these alternative information sources reflect the new reality of a digital life, and can include data from various sources (i.e. platforms, devices and technologies) such as mobile phones, social media, on-line booking websites or applications, transactions and smart sensors installed by governments or the private sector.

The monitoring and capturing of big data is not an easy process, and it is also a non-stop process. This is because big data is changing and created on a continuous basis, as well as characterized by seven complex factors (Gunther et al. 2017): (1) *volume* refers to the large size of datasets generated through the fast development of technologies and applications like Internet of Things (IoT), Artificial Intelligence (AI) and user generated content (UGC) from online social media platforms; (2) *variety* represents the diversity of data sources and multimedia formats, e.g. text, videos, photos, likes, comments, website statistics, log data; (3) *variability* related with the data whose meaning can vary significantly in context; (4) *velocity* refers to high speed of data generation through the growth of interconnected devices generating enormous amounts of data in real-time, so that big data changes and is created with the speed of light; (5) *veracity* represents the data reliability, as data is worthless if it's not accurate; (6) *visualization* science of visual representation of data and information in a way that is easily readable, accessible and understandable. Big data is recognized as a key source for creating (7) *value*, since it is considered that it contributes to more efficient and effective operations like optimal price setting, optimizing supply chain, minimizing errors and improving customer satisfaction. M-Brain (2016) identified another characteristic of big data that makes its management difficult, (8) *virility*, which means that big data is quickly spread around through viral online practices and social networks and by being spread and shared it is being augmented, enriched, changed which again makes its monitoring, capturing and understanding quite difficult and laborious. Figure 1 illustrates each of these eight points.