Big Data for Evaluation

Nicholas S. Vonortas[1], *Ozcan Saritas*[2]

[1] Institute of International Science and Technology Policy and Department of Economics, George Washington University, USA; São Paulo Excellence Chair, Department of Science and Technology Policy, Institute of Geosciences, University of Campinas, Brazil

[2] Institute for Statistical Studies and Economics of Knowledge, National Research University Higher School of Economics, Russian Federation

Demand is increasing for evidence that can establish the effectiveness and success of public policy and programs, as demonstrated by the passage of the Evidence-Based Policy Commission Act in 2016 (the Ryan-Murray Act). Several years ago, the late John H. Marburger III, science advisor to President George W. Bush, issued a call for new methods for making science, technology and innovation policy more evidence-based, resulting in the establishment of NSF's Science of Science and Innovation Policy (SciSIP) program. However, rapid social, technological, economic and political transformations are in motion around the world, leading to the increased complexity and unpredictability of short and long term futures. This poses particular challenges for developing evidence of the effects and outcomes of STI policies, as conditions shift more quickly than the traditional cycles of data collection. Current approaches to formulating prospective policy options and evaluation utilize quantitative and qualitative methods including future-oriented technology analyses, roadmaps, options theory, econometric analysis, bibliometrics, data envelopment analysis, surveys, expert/advisory panels and interviews.

What is new and exciting nowadays, however, is the revolutionary opportunity introduced by Big Data Analysis (BDA), which can enable a more rational, rigorous and systematic approach for informing policies even under conditions of uncertainty. BDA has been increasingly recognized by public and private research institutions, policy and strategy makers, universities, as well as corporations, as a potentially very useful tool for identifying opportunities and addressing grand challenges in applying innovation to socio-economic development.

BDA enables more dynamic and adaptive STI policies covering the whole policy cycle from agenda setting through explorative research (ex-ante long term, foresight), design and formulation (ex-ante short term), implementation (monitoring) and impact assessment (ex-post, feed-forward). BDA approaches could influence the STI policy by developing customized sources of data and procedures for analysis to be used at each stage. Large amounts of data from a wide variety of sources can be selected and scoped in line with policy requirements. These range from published materials like scientific publications, patents, media, and social networks; to speech, sensor, satellite and visual data, which are generated on a continuous basis across the world. Useful intelligence can be extracted through bibliometric, semantic and syntactic analysis, and then transformed through data processing and visualization to become important inputs into overall policy and strategy formulation.

Thus, BDA offers a discourse or set of methods to inform STI policy, rather than aiming to directly affect the eventual goals of the policy. BDA helps to manage 'information overload' by handling large amounts of data and information so as to reduce the effort involved in obtaining useful input for policy making and evaluation. Besides this systematization function, BDA provides outlier detection and treatment; neural and social network analyses; business process and risk analyses; predictive models, simulations and scenarios. Thus, from the very first stage of agenda setting, BDA helps to detect key trends and patterns of activity, which may signal socio-

economic and STI developments; to identify "weak signals" of what is yet to emerge; and to provide early warning of "wild card" developments for more effective and proactive STI policy.

Furthermore, BDA provides largely unbiased and inclusive input allowing hearing everyone's voice, from academic and scientific communities to policy and industry as well as citizens at all levels of governance from macro (global, national, and regional) levels to micro (local) levels. Besides potentially empowering participation in policy and strategy making, BDA creates opportunities for monitoring and creating opportunities for networking, partnership and cooperation in and between different stakeholder groups through research, business and other collaborations. Input from BDA can help to measure impacts of STI policies as well as to develop indicators for their monitoring and evaluation. Continuously developing visualization techniques enables the integration and communication of large amounts of data into STI policy processes in fast and efficient ways.

The proposed session, titled "Big Data for Evaluation", will explore the advancements in Big Data Analytics and link those advancements to the changing nature of STI policy formulation, implementation and evaluation. In an era where 'data' is arguably becoming one of the main sources of competitiveness, the session will highlight the variety of concepts and methods for using BDA for Evaluation. It will also point out areas of particular concern for the foreseeable future – e.g., sensitivity to noisy data, ill-formed conceptual treatments, mob emotions.

The session will include a mix of leading experts from academia, think tanks and industry, with considerable expertise and experience in using BDA to support decision-making in R&D-intensive agencies, companies, and institutions.