The Alan Turing Institute Programme on Data Centric Engineering
Response to the National Infrastructure Commission's Call for Evidence

Background

The Alan Turing Institute is the national institute for data science, with a mission to make great leaps in data science research to change the world for the better. The programme on data-centric engineering (DCE) will develop critical data analytic capabilities to address the challenges in improving the performance and resilience in engineering systems and national interdependent infrastructure nexus. The evidence presented in this document will be based on the Turing Program for DCE will focus on 3 grand challenge areas of:

1. **Resilience:** Resilient and Robust Infrastructures,
2. **Monitoring:** Monitoring Safety of Complex Engineering Systems,
3. **Design:** Data Driven Engineering Design under Uncertainty.

Key Technologies

The key technologies that have been identified by the Turing Institute revolve around developing and employing state-of-the-art data analytic methods to better understand and predict the behaviour of engineering systems and infrastructures. ***Measures for Functional and Connected Resilience:*** the Turing Institute is developing signal processing and network science approaches to better understand both functional resilience of individual infrastructures, as well as connected infrastructures. Examples of functional resilience include critical health monitoring of multi-scale infrastructures (i.e., from large-scale bridges to small-scale wind-turbine power generators) using advanced dynamic threshold and motif detection techniques. Examples of connected resilience include understanding modelling national interdependent infrastructure networks (i.e., water-supply network and multi-modal transport network hubs) and developing graph theoretic bounds on both hard-protective resilience and soft-dynamic resilience investment strategies. ***Learning from Historical Data:*** the Turing Institute is developing new advanced statistical modelling to learn from past events to predict the outcome of future scenarios, such as disruptions. One example is understanding how signal failures and accidents in the London Underground can cause passengers to select new multi-modal transport choices and the cascade effects it may cause. By using advanced machine learning techniques to predict the outcome of different perturbations, public safety and economic efficiency can be improved by developing appropriate communication strategies to inform citizens. ***Data-Driven Operations:*** In terms of driving the design and operations of existing and new infrastructures, the Turing Institute is developing combined structured and unstructured data analytic tools to understand both consumer demand and experience at an unprecedented spatial-temporal accuracy. By combining social media, business, geo-spatial, and operational data, complex engineering systems can be designed, deployed, and dynamically driven to meet the rapidly changing patterns in consumer demand. Examples of ongoing research include working with National Grid to predict demand cycles, and with telecommunications vendors to design a new generation of flexible communications systems, and with water supply operators to deploy cyber-physical systems to optimise consumer driven supply.

Market, Commercialization, and Investment

The data-driven monitoring and analytics research are embedded in current and future smart cyber physical systems. These smart systems are at the forefront of the competition between utility companies. In the data-driven smart water management business alone, more than $20bn of investment from 2016 to 2025 is set aside for investment in smart meters, data analytics and data management. Data analytics can deliver lower energy expenditure (20-40%) and lower OPEX (30%) through dynamic supply reconfiguration and deliver superior supply to rapidly changing demand patterns. This translates to significant savings in the $12bn US and $11bn EU water sectors. In the $5bn data-driven telecommunications market, social data analytics is increasingly used to drive small cell deployment and self-organisation-network functionalities. Together, the data analytics techniques developed at the Turing Institute both serve to empower decentralised local decision autonomy and improve the efficiency of centralised control systems. The system-of-systems approach to modelling the resilience of connected infrastructures can identify the cascade impact of different environmental and terrorism risks. The outputs are directly applicable to a wide range of commercial stakeholders and government level policy makers, with the goal of forming a credible scientific basis for investing in resilient networked infrastructures and nexus hubs.

Overcoming Barriers

Barriers to data-driven engineering have multiple dimensions. Technologically, the Turing Institute is actively building the hardware and expertise capacity to support a national effort to facilitate the sharing of data and verification of algorithms. Politically, it is important to have common agreement in recognising the importance of data in addressing societal challenges across local authorities and develop common understanding in best practice in collaboration and implementing solutions. Standardization of data, analytical methods, and sensor systems are also important to improve data analytic and solution development efficiency.

The Alan Turing Institute Programme for Data Centric Engineering would be pleased to provide further detail of any of the issues raised above, either in writing or by way of oral evidence. This response was initiated by Professor Mark Girolami (mgirolami@turing.ac.uk) and coordinated by Dr Weisi Guo with additional contributions from Dr Din-Houn Lau, Dr Ricardo Silva, and Professor Julie McCann.