University of Aberdeen
Peter Edwards
05/2018

University Partner:
University of Aberdeen

Academic Supervisor - Name and Email Address:
Peter Edwards - p.edwards@abdn.ac.uk

Suggested Project Title:
Realising a Flexible Quality Framework for Managing Data Assets

Suggested Project Summary:
Data provenance is recognised as one of the key enablers of a trusted information infrastructure - as it exposes the history of data and any subsequent manipulation/modifications. Consider the case of a faulty IoT device that reports incomplete or erroneous data - that has the potential to result in real physical consequences. Only by understanding the context of the device and the data it produces can we (or automated systems) determine if the data are reliable and actionable. While veracity is often listed as one of the seven Vs of Big Data, along with volume, velocity, variety, variability, visualization and value - it is perhaps the most important and least studied. This project will explore the following questions:

- How can data veracity measures (metrics) be encoded and enacted within a data ecosystem?
- How can data provenance be used to support new forms of veracity checking and anomaly detection?
- How can data policies be framed to reason about data veracity, and recommend appropriate decision-making actions?

Any organisation with creates and manages data (across a range of industrial and public sectors) should have an interest in data veracity. Potential outcomes from the proposed research include: a language for characterising data veracity (metrics); novel mechanisms for context-aware anomaly detection; a flexible policy framework able to analyse and recommend actions to be taken relating to data assets.

Collaboration Sought for the Project:
Any organisation that produces and curates large volumes of data; could be in the public sector, where open data publishing has created a real challenge around data quality. The nature of the project requires real-world use cases.

Benefit to the Industry Sponsor:
Access to expertise on data provenance, data veracity, intelligent policy frameworks/automated decision making. The project offers a partner an opportunity to better manage its data assets, with greater assurance of quality.

Published or Private?:
Yes