Big Data

SCIENCE AND
INDUSTRY
ENDOWMENT
FUND

SIEF supports Big-Data Knowledge Discovery: Machine Learning meets Natural Science.

the **challenge**

In every research area the size of data is exploding faster than current tools allow scientists to interpret it. Many future scientific insights are likely to be buried ever more deeply, and subtlety, in ever larger amounts of experimental data. However, this kind of data-centric research is a game changer for science, so there is a need for development of new techniques for handling, deciphering, analysing and making sense of this vast quantum of valuable data that is being collected.

the **response**

The Big Data Knowledge Discover Project is developing new methods and tools for applying machine learning in the natural sciences, using three case studies. The case studies involve: improving the understanding of the motions of tectonic plates in geology; modelling how different plant traits allow plants to effectively compete for resources in forest ecosystem; and the development of tools to explore and understand complex laser systems, looking for new laser applications.

the **collaboration**

This research initiative brings together leaders in Information and Communications from National ICT Australia Limited (NICTA) and SIRCA limited with world class discipline leaders in the data-intensive sciences of Geo Sciences, Life Sciences and Physical Sciences (Macquarie University and University of Sydney).

projected **impact**

The tools and techniques to create, manage and analyse massive amounts of scientific data have only recently been released. By developing breakthrough tools, the Project is enabling new ways to interpret and draw insight from these massively big data sets.

- Establish "Big Data" best practices in a range of natural science disciplines ensuring a consistent framework is used

- Open access to the technologies allows for translation of application into other fields

- Enable machine learning on large geophysical datasets combined with the plate tectonic history of the earth, to drive discovery of resources that are buried under cover, leading to greater economic value from as yet undiscovered deposits.

- Understanding drivers of forest diversity enables forest structure and biodiversity to be predicted for the first time, from first principles, allowing better and more sustainable management of the earth's forests.

- By using the 'virtual' techniques of machine learning and data driven experimentation, discovery of new functionality in laser systems is possible without the need to undertake complex and exhaustive investigations. This will lead to the fast-tracking of industrially useful applications for potential use in areas such as communications and manufacturing.