

# Canadian Statistical Sciences Institute Leadership Retreat

## DRAFT REPORT

W. John Braun (University of British Columbia),  
Nancy Reid (University of Toronto), and  
Mary Thompson (University of Waterloo)

September 25, 2015 – September 27, 2015

This workshop provided a venue for leaders in statistical science in Canada, the U.S. and Europe to develop short term and long term goals for the Canadian Statistical Sciences Institute (CANSSI), and to brainstorm of future scientific directions of the Institute.

## 1 Introduction

CANSSI was established in June 2012 by the Statistical Society of Canada. The Society appointed Professor Mary Thompson of the University of Waterloo as Scientific Director, and she continued in that role until June 2015. Professor Nancy Reid of the University of Toronto, is the Director for 2015-2019.

CANSSI is a virtual institute, building research capacity across the country, and relying on the mathematical sciences institutes ( the Fields Institute, the Pacific Institute for the Mathematical Sciences ; PIMS, and the Centre de recherches mathématiques; CRM) for infrastructure and logistical support. CANSSI's NSERC funding for 2014-2019 is routed through grants to the mathematical sciences institutes.

CANSSI's current level of funding is adequate to support 8 Collaborative Research Teams, at approximately \$60,000 per year for 3 years each; several workshops and conferences; and partial support for 2-3 postdoctoral fellows. We are currently holding competitions for two CRTs; three have been running since April 1 2015, and three since April 1 2014.

The BIRS workshop provided an opportunity for focussed discussion on the future development of CANSSI, in a sequence of open sessions and breakout sessions. The following questions guided the discussion:

- What will be the role of CANSSI in the current infrastructure for mathematical sciences?
- What role will CANSSI play internationally? What international networks are most relevant to CANSSI's strategic goals?
- What will be the role of CANSSI in the current national infrastructure? What existing national entities are most relevant to CANSSI's strategic goals?
- How will CANSSI position itself for leadership in data science and in big data, both in academic and non-academic settings?
- What are the most important scientific directions to build on? Should we prioritize these, and if so, how?

## 2 Scientific Highlights

On the first evening of the program, three speakers gave concise presentations describing research projects being carried out with CANSSI's support. The first two project areas are funded through the Collaborative Research Team program, and the third area is related to a proposed Thematic Programme.

Derek Bingham (Simon Fraser University) presented the activities of his collaborative research team Earth Modelling and Computer Experiments, which has nodes at Dalhousie, UBC and Simon Fraser. Derek briefly described three ongoing research projects: ozone modelling; a deterministic tidal flow simulator for the Minas Passage in the Bay of Fundy; and calibration of glacier mass-balance models, which is complicated by the fact that glacier surfaces are non-convex regions. The team has developed sequential experimental design methodology for computational models that allows for a more efficient calibration of the model (a type of inverse problem).

John Braun (UBC) presented the activities of his collaborative research team discussed projects related to evolving marked point processes, the focus of his collaborative research team. The underlying objective of this research is to develop spatio-temporal methodology to model marked point process data aggregated over moderate to large temporal and spatial scales where the marks are possibly realizations of an interacting particle system or of another spatio-temporal process. The main application of interest is wildfire modelling where the "points" are fire ignitions, caused by lightning, for example. These points can smoulder before either dying out or growing (hence the possibility of an interacting particle process to govern the evolution). Related problems in environmental science also arise; for example, a model for storm cells is needed for a study of data from Manitoba Hydro. The problem here is to model the duration, direction and speed of the cells and larger "systems". One possibility is to model the origins of the storm cells in space-time using a modified Neyman-Scott cluster process where the underlying process is not Poisson, but possibly a log-Gaussian Cox Process. The movement of the storm cells must also be modelled, possibly non-parametrically. Partners in this research include provincial fire management agencies as well as ICLR who look to use the research to obtain improved assessments of lightning and wildfire risk in forested areas.

Erica Moodie (McGill University) discussed causal inference and genetics, motivating the Short Thematic Program that will take place next summer at CRM, with funding from both CRM and CANSSI. Causal inference lies at the foundations of scientific discovery, because it asks the question, "Why"? Within the fields of statistics and epidemiology, the discovery that randomized controlled trials can eliminate the biases that confound observational studies is a notable and important achievement. However there are many contexts in which randomization is impractical, unethical or prohibitively expensive; this is particularly so in genetics. Statistical causal inference provides a methodology to identify causal relationships in genetics when randomization, at least in its conventional formulation, is not feasible. The scale of current data and technology available to scientists also poses new challenges to data analysis and experimental design, as the number of possible hypotheses and experiments increases rapidly. There is a need for new statistical methods for causal inference which can help identify the experiments most likely to show an effect, aiding in protocol design, and in effective use of newly acquired data, which may have been obtained at great expense. The new methodology must scale both computationally and inferentially.

## 3 Brainstorming Sessions

Much of the meeting time was spent in "Blue-Sky" discussions and in break-out group discussions in order to identify potentially fruitful directions for CANSSI.

In the initial discussion, the question was raised as to whether CANSSI should be a physical institute or whether it should continue as a virtual institute. A possible model, though not the only one, is that of PIMS, which has a relatively de-centralized structure.

Connections with industry were considered by one of the breakout groups. Several ideas were generated there, including using CANSSI as infrastructure to showcase what statisticians can offer to industry and to facilitate linkages, and student and faculty exchanges, with industry. Mitacs internships are an important part of this, and CANSSI should actively pursue NSERC's Industrial Workshop program. It is important to keep records on these activities, including keeping a database of students who have been supported this way. Is there potential for CANSSI to expand its membership program to corporate and other non-academic

members?

Another group focused on education and training, all the way from K-12 through postdoctoral mentoring. It was suggested that CANSSI could play a more active role in supporting the CensusAtSchool program which was recently taken over from Statistics Canada by the Statistical Society of Canada. New ideas about how to present introductory statistics to undergraduates were discussed, including the importance of moving quickly into Data Science, partnering with computer scientists. Many online resources are available, and it was noted that a well-developed MOOC on big data with tech add-ons would be extremely popular.

Connecting to the Health and Social Sciences was the theme of the third breakout group. Discussion focused on the role of CANSSI as promoting and facilitating collaboration with health and social sciences on important problems requiring new data science, and alerting researchers to targeted funding opportunities. CANSSI could advocate for more funding for methodology research. The Collaborative Health Research Project program (NSERC and CIHR) is well-subscribed and an expansion would be welcome. SSHRC offers an interesting potential opportunity for CANSSI. Its Trans- Atlantic Platform is an international and cross-disciplinary effort which has funding from Canada, the US and the UK: the funds flow from the agencies to researchers in their countries. The first call for proposals was about Digging into Data (joint with NSERC and CFI).

International connections were discussed in another breakout session. Here, suggestions for international joint postdoctoral fellowships and exchange programs for PhDs were suggested as possibilities for CANSSI to pursue. SAMSI is one of many institutes to collaborate with in this way. Working groups, again with SAMSI, but also with STATMOS, offer another avenue for CANSSI to facilitate international connections for Canadian statisticians.

Other possible international connections include CNRS (France), PRIMA (Pacific Rim), the new Alan Turing Institute (UK) and Eurandom (The Netherlands) should also be considered.

One of the breakout groups discussed the possibility of a proposal in Data Science, to the Network of Centres of Excellence, in collaboration with PIMS, CRM and Fields. This idea received an enthusiastic response when brought back to the main forum. The LOI for this program will be due in 2017.

## 4 Outcome of the Meeting

The meeting concluded with a session in which the road forward was mapped out, with identification of short, medium and long term goals and activities. Ongoing dialogue with the mathematics institutes and with SAMSI for joint workshops was identified as a top priority, including the active participation of CANSSI in SAMSI programs.

The future looks bright for CANSSI and for statistics and data science in Canada. For this to be fully realized, however, it will be important for CANSSI to maintain connections with young statistical investigators in Canada and abroad through networking, to help them to build maintain connections with industry andwith international institutes, and to foster mentoring opportunities. This next generation of investigators will ultimately determine the destiny of CANSSI.