



## Report on the Canadian Statistical Sciences Institute Leadership Retreat Banff International Research Station September 25 – 27, 2015

### ***Introduction***

The Canadian Statistical Sciences Leadership Retreat was held at the Banff International Research Center, September 25-27, 2015. There were twenty-six participants: the management group (Director, Deputy Director, and Associate Directors), several members of the Board of Directors, several members of the Scientific Advisory Committee, the directors or deputy directors of PIMS, Fields, CRM and AARMS, and invited members from the statistical sciences research community.

The purpose of the workshop was to brainstorm about future strategic and scientific directions and to develop short term and long term goals for CANSSI. To this end there were full group and breakout group discussions over the course of the weekend. The Appendix to this report contains the workshop program. The final session of the weekend was devoted to outlining a road map for CANSSI, with a focus on immediate, short-term, and long-term goals.

### ***Executive Summary***

We addressed six strategic areas for CANSSI: industrial collaborations, education and outreach, health and social sciences, mathematical and computational sciences, international linkages, and the Networks of Centres of Excellence program. In each of these areas we identified immediate, short-term, and longer-term activities for CANSSI to undertake. We will continue to refine these in terms of priority and timing.

CANSSI aims to take on a prominent leadership role in the area of Data Science. With this in view, CANSSI's immediate activities will be focussed on maintaining existing relationships, and establishing new relationships, with various organizations in Canada and abroad. Linkages with the mathematical sciences institutes will be deepened through joint ventures in industrial innovation, in part with Mitacs. Increased cooperation with SAMSI, NISS, and STATMOS will be sought, through working groups, for example; with Big Insight and the Alan Turing Institute; and in Canada with MILA, the new Machine Learning Laboratory at CRM. Connections with the health sciences, social sciences and computer science will be developed. CANSSI's CRTs will be encouraged to raise the profile of their activities, as an outreach mechanism for CANSSI, particularly around education. A key long term objective is that CANSSI will lead in developing a proposal to the Networks of Centres of Excellence program; it is expected that a call for letters of intent will be issued in Spring 2017.

The following sections summarize the discussions that took place around each of the strategic initiatives.

## 1. Connecting with Industry

### ***The Current Situation***

The current environment demands that the statistical community in Canada develop closer connections with industry. Although not always recognized, statisticians have a lot to offer industry, but it may be necessary to modify our approach. Traditionally it has been somewhat difficult for academic statisticians to overcome their natural reticence to engage with industry, so the number of industrial collaborations in the mathematical and statistical sciences lags those of some of the other sciences. At the same time, traditional grant funding is not keeping up with the demands for graduate student support.

The advantages of collaborations with industry are not always obvious to academics and students. Incentives for faculty or students not always clear, since the currency in academia is primarily publication of journal articles, especially in one's own discipline. Interactions with industry also require a commitment to building relationships, and this takes time.

Each of the mathematical sciences institutes has recently appointed associate directors for industrial liaison, as part of a coordinated effort, with NSERC, to increase the engagement of mathematical scientists in NSERC's research partnerships programs. Clearly statistical science has an important role in this activity. The institutes are planning to build databases of industrial partners, interested academics, and project successes.

In the US, NISS and SAMSI have an [Affiliates Program](#) for industry and universities, with six industrial affiliates, a large number of university affiliates, and seven affiliates among government agencies and national laboratories. For the most part, the Canadian mathematical sciences institutes have not directly sought out industry affiliations, but this could change with increased involvement with Mitacs.

### ***Research Partnership Opportunities***

CANSSI could serve a useful role as a pipeline between statistical science research and industry. Internships or co-ops are a great way to make progress on this front. If industry wants a stellar researcher, they also want students that they might hire later. The product for industry is the talented young person that they could then hire. A number of opportunities exist which can be exploited with minimal effort.

As pointed out above, there has been relatively little uptake of NSERC's Engage and CRD programs by mathematical and statistical scientists. Engage and CRD programs are relatively easy to apply for, and they require very little from the industrial partner. These programs could be used much more by statistical scientists; they offer a way to increase funding for research and for training of HQP.

Another avenue to pursue is NSERC's [Connect Grants](#) program. The three-month

connect grants provide up to \$5000 for travel, accommodation and meeting costs for academics and Canadian companies. There cannot be any prior relationship between the researchers and the companies. Larger grants are also available for larger-scale, regionally or nationally oriented events, over a one-year period.

[Mitacs](#) offers other opportunities to link researchers with industry through their Accelerate and Elevate programs. CANSSI should proceed, in concert, with the math institutes in pursuing a Memorandum of Understanding with Mitacs which could ultimately provide a streamlined review process for student researchers seeking internship opportunities at eligible companies.

### ***Other Ways to Connect***

Connecting to the media is another way for the value of statistical science in industry to be demonstrated. The role of statisticians in industrial successes needs to be more visible. This could build additional linkages between the research community and the industrial research community. Short descriptive paragraphs on past collaborations are one way of doing this.

Does CANSSI need a communications officer? Such an individual could coordinate and orchestrate various social media and conventional media announcements.

Industry is increasingly looking for individuals with data science skills, such as database familiarity, algorithmics and analytics. CANSSI could play a role in training of HQP in this area, through cooperative ventures. For example, CANSSI could develop online resources, possibly in conjunction with the mathematical sciences institutes and with industry. We expand on education initiatives in Section 2.

## **2. Education and Training**

### ***Connection with CANSSI's mandate***

Students and HQP are an important part of CANSSI's mandate to build research capacity in the Statistical Sciences. The most obvious connection between research and education/training is at the university level (undergrad, grad and PDF), where training enables HQP to become research collaborators. Training at the introductory undergraduate level can help draw students into statistics, by introducing modern statistical science, including data science, and a host of interesting and substantive scientific research areas that depend on statistics. With such emerging changes at the undergraduate level, CANSSI could play a role in undergraduate curriculum reform, including data science. CANSSI could also play a role in primary and secondary education, helping foster interest in the statistical and mathematical sciences.

### ***Undergraduate and Graduate Statistical Training***

Curriculum reform has once again become a priority on many university campuses, due in part to the advent of Data Science and Big Data. Departments are struggling to identify the core topics, the topics that can be removed, and new topics that should be added. They are also trying to identify the role of computing and how cooperation

with computer science might enhance or augment the statistics curriculum.

A number of topics that don't always receive much attention in the current statistics curriculum are data visualization, databases and SQL, and official statistics. In the health sciences knowledge of SAS is essential, but often not taught formally. Rethinking STAT 101 is another aspect of curriculum reform, keeping in mind its relevance for other sciences. What topics should be included if this is the only statistics course a student will take? Should it be a course in modelling?

The CANSSI-supported workshop on Statistics and Education held in July, 2015 established seven working groups to make progress on specific aims, and CANSSI should find ways help maintain this momentum.

### ***Primary and Secondary Education***

Raising the profile of statistics in the public school has always been a challenge, as has increasing teacher knowledge of statistics. PIMS has had some success through its summer school program for teachers.

The Education Section of the SSC runs the Census@School program; CANSSI could perhaps play a role in supporting this.

### ***Exploiting and Developing Non-Traditional Resources***

Massive Open Online Courses are being developed in several universities. The Fields Institute is planning to hold a workshop on MOOCs and Mathematics. Other resources that could be exploited in aid of statistical training and education are YouTube videos, Jupyter notebooks, and hack/reduce finishing school. Summer schools offer another more conventional approach.

### ***The Role of CANSSI***

Specific activities that CANSSI might undertake to provide leadership in this area include:

- facilitating the sharing of curriculum and its development across institutions
- sharing experience with courses run at different institutions<sup>1</sup>
- support the development of summer schools
- encourage CRTs to create lectures for use in K-12 setting
- encourage CRTs to develop researcher visits to local schools

CANSSI could play a leadership role in identifying the important elements of data science in the Canadian context; see Section 4.

Funding for a Statistical Education CRT and/or Educational Initiatives is another way in which CANSSI could play a supportive role, but this would require provincial funding.

---

<sup>1</sup> The CRT in Modelling the World ran a web-based course with participants at SFU, UBC and Acadia, for example.

### 3. Health and Social Sciences

A key strength of CANSSI is the potential to develop partnerships between researchers in statistics and potential collaborators in a wide range of fields, including social sciences, health sciences and natural sciences and engineering. We especially have good contacts with a large group of researchers in biostatistics and in statistical genetics, who in turn have ongoing collaborations with scientists. We have similarly strong ties to researchers at Statistics Canada.

#### ***Connecting with Health Sciences***

Efforts to engage CIHR have not been very successful to date, in spite of the fact that CIHR does make (limited) calls for proposals in methodology; for example a program called "[Catalyst: Methods Post-Market DSE - Bayesian methods and statistical models](#)" was started in January 2015. The Collaborative Health Research Project program (NSERC and CIHR) is highly subscribed, and a few successful applications have come from statistical scientists.

Board member Michael Kramer has suggested working through provincial health agencies.

Suggestions that arose from the breakout group included:

- Identify some advocates for integration of statistical and health research. Models for this from the US and UK may be helpful.
- Consider individual memberships in CANSSI; these might facilitate or provide a forum for connecting researchers.
- CIHR established a [Partner Linkage Tool](#) as part of a call for proposals on "Healthy and Productive Work", a joint program with SSHRC. Some version of such a tool created by CANSSI would surely be welcomed by substantive researchers, if we could advertise it suitably.
- Have some targeted CANSSI-CIHR research efforts, such as CRTs, Postdoctoral Fellows, etc., or similar partnerships with provincial health research bodies. CANSSI has already committed to providing partial support for a PDF as part of the Ontario Institute for Cancer Research's Biostatistics Training Initiative, which was launched in November 2015.

#### ***Connecting with Social Sciences***

There are notable initiatives in the social sciences connected to Big Data that were described by Chad Gaffield. These are both international (See Section 5) and national. See also [this article](#) in University Affairs for a good overview.

- [The Digital Humanities Summer Institute](#) is a major training program which in 2016 will run in Victoria, BC. There are courses offered on data wrangling and visualization, and several courses on coding; some of the courses teach R, but there are no courses (this year) on statistical methods. Courses can be proposed: 2017 course proposals are due by April 1, 2016. This summer

institute is part of an international initiative.

- The Trans-Atlantic Platform is funded by the EU, and it is expected that its first call for proposals will revolve around data. Canada's partner in this is SSHRC. The President of SSHRC (Ted Hewitt) is chair of the steering committee.
- The [Digging into Data](#) challenge was a similar international initiative, involving both SSHRC and NSERC. The last call for proposals was in 2013.

Statistics Canada has demands for both undergraduate and graduate students, to work in co-op programs, on contracts, and in internships. The Mitacs model does not always work well for Statistics Canada, but Statistics Canada has internal funds that can be used for students to work on problems of interest to them. It was suggested that Statistics Canada could become an institutional member of CANSSI.

Perhaps CANSSI could, in partnership with Statistics Canada, identify challenge problems to foster collaboration. One model is [Kaggle](#), although opinions about this competition vary. Another model, perhaps more staid, is the SSC's case studies program.

CANSSI is a partner organization on a SSHRC partnership grant proposal for a National Centre for Research on Poverty and Inequality, and committed financial (\$5,000 per year) and in-kind (\$15,000 per year) support for PDFs and graduate students, working on the sub-themes measurement of poverty, and social mobility. Both of these sub-themes involve longitudinal data.

#### 4. Mathematical, Statistical and Computing Sciences

##### ***The mathematical and statistical sciences institutes***

One breakout group was charged with discussing plans for the next application to NSERC, which will be due in fall 2018. Both CANSSI and AARMS will submit separate proposals. We agreed that we would exchange ideas and draft proposals, and make the case to NSERC that a 'five-node network' is a natural evolution of the mathematical sciences portfolio.

The current mathematics and statistics envelope is constraining, in being very rigid, and long-awaited re-allocations do not seem to be occurring. The long range plan recommended a fixed wall between CTRMS (institute) funding and Discovery Grants (19%/81%), and any change in that division before 2018 seems unlikely, and inadvisable.

The leadership of PIMS, CRM, Fields, AARMS and CANSSI should start discussions with NSERC around the next competition. A modest increase in the envelope, on the order of 10%, would help to make the competition more than a shuffling of existing funds, which we all view as a worst-case scenario.

### ***The Emergence of Data Science***

It is very important to develop collaborations with those parts of computer science that comprise data science. It is possible that CANSSI, and statistics departments, will develop into data science entities over the next few years. There are a number of “Big Data” initiatives in research around the country, including a CFREF proposal from the University of Toronto.

Computer scientists tend to identify by field: machine learning/AI, computer graphics, visualization, and so on, so we will need to build these connections from the ground up. We should also form relations with other institutes for data science, both nationally and internationally. A natural place to start is with the new MILA (Machine Learning) laboratory at CRM.

CANSSI could develop a web-resource for departments and groups trying to figure out how to incorporate data science into their programs and research. We could also provide support to one or more workshops to discuss this. Although there are many data science resources on-line, we could add value by emphasizing those that make a serious effort to include statistical science.

### ***Other National Linkages***

Outreach to national groups in social sciences and health sciences is discussed in Section 3. We have reasonably strong links to environmental science through existing CRTs. Other research communities with whom we would have natural links include econometrics, actuarial science and finance, engineering, education, and business. This effort feeds into the potential for an NCE, described in Section 6.

## **5. International Outreach**

### ***American Connections***

The closest connection for CANSSI is SAMSI, and this connection must be maintained and enriched. Currently we link via the undergraduate research workshops, and by encouraging Canadian researchers to join working groups during SAMSI thematic programs. In particular it is not necessary to visit SAMSI in order to join a working group.

We could encourage more research activity by providing support to working groups that are not necessarily tied to a thematic program at SAMSI or a CRT at CANSSI. There should be potential for joint postdoctoral fellows, but the mechanisms need to be sorted out. SAMSI already shares PDFs with other institutions; perhaps we’ll need to approach NSF and NSERC to make this work.

We should also find a mechanism for CANSSI and SAMSI to have ongoing dialogue about joint workshops.

There are other networks in the US that CANSSI could develop links with, including NISS, ICERM and STATMOS; the latter is a network of researchers in environmental

sciences, and SAMSI is a node of that network. John Braun is pursuing the possibility of joint PDFs with STATMOS, and CANSSI could in principle become a node of STATMOS as well.

### ***Europe and Beyond***

The mathematical sciences have *unités mixtes* (essentially local branches) of France's CNRS, and developing links there is another way to encourage collaborations. PIMS has a relationship with the Pacific Rim Mathematical Association (PRIMA).

The Alan Turing Institute is “UK’s national institute for data science”; the Department of Statistics at Warwick (Mark Girolami) and the Department of Statistical Science at UC London (Patrick Wolfe) are closely involved.

CANSSI Board member Arnaldo Frigessi is the head of “Big Insight”, an NCE-type organization that partners with several industries in Norway. He is keen to build links with CANSSI, and to help with links to industry.

In the social sciences the Trans Atlantic Platform has partners in many countries, as described in Section 3.

## **6. Towards a Network of Centres of Excellence**

There was considerable enthusiasm for considering a proposal to the NCE Program on something like ‘big data’<sup>2</sup>. Nancy Reid will follow up with a professional consultant, who is currently working with the University of Toronto on their CFREF proposal. The following was extracted from detailed briefing notes prepared by Mary Thompson for our discussion.

NCE is a suite of programs that “mobilize Canada’s best research, development and entrepreneurial expertise and focus it on specific issues and strategic areas”. It is a joint initiative of NSERC, SSHRC, CIHR, Industry Canada and Health Canada.

An NCE network “is a virtual collaboration model connecting researchers, highly qualified personnel, administrators, manager, and directors, across public, academic, private, and not-for-profit sectors and who geographically span the country”. It supports research and development activities.

The program of networks has five sectors: Health and Life Sciences (25 networks); Information and Communication (5 networks); Environment (5 networks); Natural Resources (4 networks); Manufacturing and Engineering (2 networks). There is also a cross-sectoral category (5 networks).

The funding periods vary, as well as the amounts, but \$2M - \$3M per year for 5 years is not unusual. Some have been renewed once or twice, for a maximum of 15 years. The

---

<sup>2</sup> So much so that Jim Colliander bought the domain name “insightscanada”.

average award per network life is \$54M, the average number of researchers per network per year is approximately 60, the average number of institutions per network per year is approximately 30, and the average HQP FTE per network per year is approximately 270.

The original Mitacs (Mathematics of Information Technology and Complex Systems), which became Mprime, was funded from 1999 -- 2014. Its research themes included information processing, biomedical and health, communication, networks and security, environment and natural resources, and risk and finance. It might be described as cross-sectoral today.

The next NCE program competition will be in 2019, and a call for letters of intent is expected in **Spring 2017**. In the 2015 competition four new networks were approved and one was renewed. 83 eligible LOIs for new networks had been received, of which 10 were invited to full application, and 4 were successful.

The following rules apply

- The networks are to be pan-Canadian. Among their activities they are to “create functional, multi-regional and interdisciplinary” research teams.
- The area of proposed research and researchers must cover the mandates of at least two of the three federal granting agencies.
- Support of organizations is required. Although the program does not require matching funds, NCE networks must demonstrate relevance and collaboration with their stakeholders by leveraging resources from non-NCE sources.
- The proposed governance and management structures should be sound, with a team that has the necessary strengths to deliver on the objectives of the network. With five-year funding cycles, it is important for a network to be able to rapidly ramp-up its activities to ensure that it can meet its targets within a five year time frame.
- Before receiving funds, networks and centres must be incorporated as not-for-profit entities.
- A network must have a host organization, where the network’s/centre’s administrative centre is located. Typically this is a university.
- Typically, the personnel include a Scientific Director, a Network Manager, a Financial Officer, a Business Development and Partnership Officer, and a Communications Manager.

As an example, the Marine Environmental Observation Prediction Response Network (MEOPAR) has research opportunities like our CRTs, but are to be multi-sectoral and to cross granting agencies. Partner types are federal and provincial government, industry, other, and there are many of them. The Research Management Committee includes two members of our community.



# Banff International Research Station

for Mathematical Innovation and Discovery

**CANSSI Retreat**  
**September 25 – 27, 2015**

## SCHEDULE

### FRIDAY

- 16:00** Check-in begins (Front Desk – Professional Development Centre - open 24 hours)
- 18:00 – 19.30** **Dinner**
- 19.30 – 21.00** – Introductions  
– Overview of CANSSI (Mary Thompson)  
– Scientific Presentations (Derek Bingham, John Braun, Erica Moodie)  
– Goals for Saturday (Nancy Reid)

### SATURDAY

- 7:00-9:00** Breakfast
- 9:00 – 10.30** Full Group – blue sky brainstorming session
- 10.30 – 11.00** Coffee Break, TCPL
- 11.00 – 12.00** Breakout I
1. Connecting with Industry
  2. Training and Education
  3. Connecting with Health and Social Sciences
- 12.00 – 13.30** Lunch
- 13.30 – 15.00** Breakout II
4. Mathematical Sciences, 2019 and beyond
  5. International Outreach
  6. National Outreach and NCE Program
- 15.00 – 15.30** Coffee Break
- 15.30 – 17.00** Full Group – reports from break-out groups



# Banff International Research Station

for Mathematical Innovation and Discovery

## SUNDAY

- 7:00-9:00** Breakfast
- 9:00—10.30** Road-map forward
- 10.30 – 11.00** Coffee
- 11.00 – 13.30** Management group de-briefing

Group 1. Will Welch\*, Alejandro Adem, Derek Bingham, Don Fraser, Luc Vinet, Nell Sedransk, Paul McNicholas, Richard Smith

Group 2. Hugh Chipman\*, Alexandre Leblanc, Christian Robert, Jeannette Janssen, Jim Colliander, Liqun Wang, Matheus Grasselli, Yogendra Chaubey

Group 3. Erica Moodie\*, Chad Gaffield, Jack Kalbfleisch, Mary Thompson, Paul Gustafson, Shelley Bull, Wesley Yung

Group 4. Nancy Reid\*, Alejandro Adem, Alexandre Leblanc, Jeannette Janssen, Jim Colliander, Luc Vinet, Matheus Grasselli, Will Welch

Group 5. John Braun\*, Chad Gaffield, Christian Robert, Derek Bingham, Jack Kalbfleisch, Nell Sedransk, Richard Smith, Yogendra Chaubey

Group 6. Mary Thompson\*, Erica Moodie, Hugh Chipman, Liqun Wang, Paul Gustafson, Paul McNicholas, Shelley Bull, Wesley Yung, Don Fraser



# Banff International Research Station

for Mathematical Innovation and Discovery

## **BREAKOUT GROUPS – Morning**

*The questions below are simply to get things started – there may well be other/better questions to be discussed. Be creative!*

1. Connecting with Industry – Will Welch
  - Opportunities: Mitacs, NSERC Programs, ... ?
  - How does NSERC/Mitacs MOU change the landscape?
  - What do we do on Big Data?
  - What should CANSSI be doing to build and strengthen industry connections?
  - What strategies do we need for initiating discussions?
  - What ideas from the blue-sky morning session are most relevant for your topic?
  - ???
  
2. Education and Training – Hugh Chipman
  - Summer schools?
  - Distance courses in CRTs and outside?
  - New activities: High school? Undergrad? Grad?
  - What should the relationship be with SSC and CMS educational initiatives?
  - Feedback from math institutes on their educational activities?
  - What ideas from the blue-sky morning session are most relevant for your topic?
  - ???
  
3. Connecting to Health and Social Sciences – Erica Moodie
  - Joint programs?
  - Strategies for initiating discussions?
  - For example, graduates with policy experience + data science, or with biology + computational skills, very much in demand; presumably this is true for many areas of social science and health?
  - Should CANSSI respond to this; if yes, how?
  - What ideas from the blue-sky morning session are most relevant for your topic?
  - ???



# Banff International Research Station

for Mathematical Innovation and Discovery

## **BREAKOUT GROUPS – Afternoon**

*The questions below are simply to get things started – there may well be other/better questions to be discussed. Be creative!*

4. Mathematical Sciences, 2019 and beyond – Nancy Reid
  - Preparing for the CTRMS 2019 competition --
  - Discussions with NSERC?
  - Ideal relationship?
  - National network?
  - Expanding the envelope?
  - What ideas from the blue-sky morning session are most relevant for your topic?
  - ???
  
5. International Connections – John Braun
  - What do we want to gain from these?
  - How to decide on importance?
  - How to ensure the connections are more than ‘name only’?
  - What opportunities exist now?
  - What ideas from the blue-sky morning session are most relevant for your topic?
  - ???
  
6. National Outreach and the NCE Program – Mary Thompson
  - Linking with communities: computer science, industrial engineering, environmental science, applied math, ... ?
  - Bioinformatics, biostatistics, finance, actuarial science, ... ?
  - Building bridges and managing the scope: order of importance?
  - Strengths, weaknesses, opportunities?
  - What ideas from the blue-sky morning session are most relevant for your topic?
  - ???