1) Event Description

Closing Conference: Statistical and Computational Analytics for Big Data
June 12-13, 2015
Dalhousie University
http://www.fields.utoronto.ca/programs/scientific/14-15/bigdata/closingconf/index.html

Organizing committee:
Hugh Chipman, Acadia University
Stan Matwin, Dalhousie University
Nancy Reid, University of Toronto

Sponsors: Fields Institute (through thematic program), CANSSI, AARMS,

Invited speakers:
Hugh Chipman, Acadia
Jean-François Plante, HEC
Lisa Lix, Manitoba
Stephanie Shipp, Virginia Tech
Stan Matwin, Dalhousie
Evangelos Milios, Dalhousie
Andrew Rau-Chaplin, Dalhousie
Roseanne Minghim, Dalhousie
Rob Beiko, Dalhousie
Roger Grosse, Toronto
Einat Gil, Toronto

Panel discussants: Stan Matwin, Hugh Chipman, Jean-François Plante, Nancy Reid, Einat Gil

Short description
This conference was part of the six-month thematic program on "Statistical Inference, Learning, and Models for Big Data". The thematic program ran from January to June, 2015, and was organized by the Fields Institute, with a number of allied activities at PIMS, CRM and AARMS.

This conference was presented jointly with the Institute for Big Data Analytics (IBDA) at Dalhousie University and sponsored by the Canadian Statistical Sciences Institute (CANSSI) and AARMS. The conference gave an overview of highlights of the thematic program, and showcased research at IBDA on text mining, high-performance computing, visualization, bioinformatics, and privacy.

The conference consisted of 1.5 days of research presentations. Coffee breaks, lunch and a mixer on the evening of June 12 provided ample opportunity for networking and discussion. The conference program is appended at the end of this report.

2) The number of participants, broken down into two categories: regional, and national/international.

The closing conference had 71 participants (26 faculty, 2 government, 8 postdoc, 35 students). 16 students received partial support from the conference for travel ($500/person maximum). 68 participants were from Canada, 1 USA, 1 Brazil, 1 Germany. Of those from Canada, 40 were from Atlantic Canada.

3) Details of any resulting publications (conference proceedings, etc).

A manuscript is in preparation summarizing the thematic program has been drafted and is currently under review.:

"Statistical Inference, Learning and Models in Big Data" by Beate Franke, Jean-François Plante, Ribana Roscher, Annie Lee, Cathal Smyth, Armin Hatefi, Fuqi Chen, Einat Gil, Alex Schwing, Alessandro Selvitella, Michael M. Hoffman, Roger Grosse, Dieter Hendricks, Nancy Reid
http://arxiv.org/abs/1509.02900

The pdfs of all presentations were posted on the CANSSI website at
http://www.crm.umontreal.ca/CANSSI/news-events/big-data-thematic-program-closing-workshop-talks/

4) Scientific Highlights

This workshop was organized jointly with the Institute for Big Data Analytics (IBDA) at Dalhousie University. Talks presented by Chipman, Gil, Grosse, Plante, Lix and Shipp were over-views of the research presented at selected workshops during the thematic program on Statistical Inference, Learning and Models for Big Data. The other talks presented research at IBDA on text mining, high-performance computing, visualization, bioinformatics, and privacy.

The event was well-attended by regional researchers and by statisticians from across Canada who were visiting Halifax for the annual meeting of the Statistical Society of Canada.

As with the other events in the Big Data thematic program, the diversity of fields involved in Big Data research was evident at this conference. At the heart of nearly all the presentations was a substantive data science application. For example, Stephanie Shipp gave an interesting application involving data on first responder calls by fire departments, and the optimal allocation of services. Although the applications were quite varied, a number of common themes emerged. In this context, issues such as privacy of individuals represented in databases, feasibility of computation, including high-performance computing, combination of different databases (e.g., in health policy research), including diverse data sources (geolocation, text, transactions, social media) and data visualization were discussed.

In a panel discussion on education in data science participants Chipman, Gil, Reid, Matwin and Plante addressed a series of questions about the 'ideal' data science program:

- how traditional university degrees in statistics or computer science need to be modified to provide students with sufficient skills to work with big data
- the importance of soft skills such as data exploration, collaborative projects and communication in data science
- education of primary and secondary-level students in data science
- cross-disciplinarity and partnering with subject-matter experts in data science applications
- "data strategists" vs. "data scientists"
- Making room in a university curriculum for new topics: what should we be teaching less of?

One goal of the closing workshop was to 'introduce' researchers from the IBDA to the statistical sciences community, and vice versa. The second goal was to give a high level overview of the thematic program to interested researchers. We feel that both goals were successfully met with the workshop. The panel discussion was very helpful for the academic statisticians who participated, as many are involved in their institutions in introducing programs or courses in data science.

5) A full income and expenditure account.
See attached spreadsheet

6) Any other notable information - and any photos if you have them.
No other information

Appendix: Scientific Program

Friday June 12

9:00 - 9:10	Opening Remarks
9:10 - 9:50	**Hugh Chipman**, Acadia University, *Statistical and computational challenges in networks and cybersecurity*
9:50 - 10:30	**Jean-Francois Plante**, HEC Montréal, *Challenges, Tools and Examples for Big Data Inference*
10:30 - 11:00	Coffee
11:00 - 11:40	**Lisa Lix**, University of Manitoba, *How Big Data and Causal Inference Work Together in Health Policy*
11:40 - 12:20	**Stephanie Shipp**, Virginia Tech, *Policy meets Social and Decision Informatics*
12:20 - 1:30	Lunch
1:30 - 2:10	**Stan Matwin**, Dalhousie University, *Big Data meets Big Water: Mining Ocean Vessel Trajectory Data*
2:10 - 2:50	**Evangelos Milios**, Dalhousie University, *Exploiting Semantic Analysis of Documents for the Domain User*
2:50 - 3:30	**Andrew Rau-Chaplin**, Dalhousie University, *Scaling up to Big Data: Algorithmic Engineering + HPC*
3:30 - 4:00	Coffee
4:00 - 4:40	**Rosane Minghim**, Dalhousie University/University of São Paulo, *Multidimensional Projections and Tree-based Techniques for Visualization and Mining*
4:40 - 5:20	**Rob Beiko**, Dalhousie University, *Microbial genomics for rapid investigation of infectious disease*
5:30 - 7:00	**Reception**

Saturday June 13

9:00 - 10:20	Panel discussions on training for data science at the undergraduate and graduate levels.
10:20 - 10:40	Coffee
10:40 - 11:20	**Roger Grosse**, University of Toronto, *Highlights from the deep learning workshop*
11:20 - 12:00	**Einat Gil**, University of Toronto, *Learning about Big Data among Secondary School Students in a technology-supported collaborative learning environment*

Statistical and Computational Anlytics for Big Data
JUNE 12-13, 2015

BUDGET

Income

CANSSI	$6,000.00
AARMS	$4,500.00
Fields	$4,235.42
Advance Registration	$2,370.00
On-site registration	$650.00
total	**$17,755.42**

Expenditure

poster printing	$535.24	
poster postage	$94.45	
speaker travel	$5,667.45	
student travel	$7,988.39	
catering	$2,209.95	
reception: bar services	$188.84	
Administrator	$1,000.00	
Processing fee (Fields)	$71.10	
total	**$17,755.42**	$0.00

Travel Details
Students

Elnaz Bigdeli	$496.38
Nkumbuludzi Ndwapi	$500.00
Xiaomei Yu	$500.00
Paula Murray	$494.59
Muz Mallo	$500.00
Harris Quach	$497.98
Gong Zhang	$497.98
Arman Didandeh	$500.00
Abdollah Safari	$500.00
Farhad Shokoohi	$500.00
Ribana Roscher	$500.00
Omidali Aghababaei Jazi	$500.00
MOHAMMAD NIKOU SEFAT	$500.00
Chris Gravel	$500.00
Fuqi Chen	$501.46
Ya Yu	$500.00
Total	**$7,988.39**

Invited Speakers

Jean-Francois Plante		$1,262.26
Lisa Lix		$919.93
Stephanie Shipp		$1,434.70
Roger Grosse		$1,157.31
Einat gil		$893.25
	Total	**$5,667.45**

On-site registration:

Stephen Brooks	Dal	$100.00
Arnold Mitnitski	Dal	$100.00
Reuben Pereira	U. of Toronto	$50.00
Yanqing Yi	Memorial	$100.00
Danyiel Silver	Acadia	$100.00
M. Shameen Iqbal	Acadia	$50.00
Geoff Mason	Acadia	$50.00
Michael Foster	Dal	$100.00
	Total	**$650.00**