

IPS 232: Data Science and ISI

Invited Paper Session at the ISI World Statistics Congress

Virtual, July 13, 2021

Agenda

Part 1: Readout from the ISI Data Science Working Group (DS WG)

Speakers: Daniel Jeske, Jürgen Symanzik, and Paulo Canas Rodrigues

Part 2: Audience Participation Discussion:

“Practical Challenges Around Data Science Applications”

Led by Elisabetta Carfagna



Part 1: ISI Data Science Working Group Readout

1. Charge from ISI President John Bailer to develop recommendations for connecting and highlighting Data Science at the ISI.
2. WG was convened in Fall 2019 and co-chaired by:
 - Daniel Jeske, University of California, Riverside, ISBIS President-Elect
 - Jürgen Symanzik, Utah State University, IASC President
3. 21 WG members including representatives from each of the seven ISI Associations, representatives from five ISI Special Interest Groups (SIGs), three members from the ISI Council, and four regional ISI members (see Appendix).

4. WG Timeline

- October 2019 – October 2020: Co-chairs and ISI President John Bailer organize membership and logistics
- WG Meeting Dates (all meetings virtual and recorded):
 - 11/30/2020
 - 1/29/2021
 - 2/12/2021
 - 2/26/2021
 - 4/30/2021 (Final meeting, WG is retired)
- Between the last two meetings a letter went out to ISI Association Presidents and SIG Chairs requesting feedback and reactions to recommendations from the WG in five areas that we will share here today.
- Recommendations regarding:
 - Branding
 - Publications
 - Capacity Building
 - Data Warehousing
 - Global Outreach

Recommendations Regarding **Branding**

- Do not integrate “Data Science” into the name of the International Statistical Institute.
- Update the mission statement to include Data Science in the verbiage.
- Hold off on starting a new Association of ISI that is centered around Data Science.
- Form a special interest group on Data Science within ISI and encourage broad membership with representation from synergistic association.

Recommendations Regarding **Publications**

- ISI journals should review their aims & scopes to make sure Data Science is discussed as a topic area of interest.
- Encourage special issues and/or new sections on Data Science within the ISI journals.
- Encourage emphasis on applications and data analysis to distinguish the focus from other methodological journals (including JDSSV).

Recommendations Regarding Capacity Building

- Contribute to training up the world in Data Science by placing focused emphasis on basic skills rather than advanced skills, paying attention to diverse backgrounds and varying access to software packages.
- Create Youtube channels to freely disseminate training materials

Recommendations Regarding Data Warehousing

- Establish a committee with ISI Societal memberships that creates and maintains an index to websites around the world that have rich data sets for both research and educational objectives.
- Create strategic partnerships, for example with United Nations, to align their interests in data sets with this service effort.

Recommendations Regarding Global Outreach

- Develop a Data Science mentoring service to help practitioners in countries with limited mentoring capability.
- Create strategic partnerships that allow ISI to amplify their efforts through participation and through the weight of having ISI along with them.

Initial Feedback from Associations and SIGs

1. ISBIS

- More active BLOG with “Data Science Alert” section.
- Introduce Data Science Section in the ISBIS journal, Applied Stochastic Models in Business and Industry.
- Two new tracks to webinar series:
 - Track 1: Training Up The World in Data Science (Youtube Channel)
Example: “Introduction to Using Pivot Tables in Microsoft Excel”
<https://www.youtube.com/watch?v=Lrydpi8REpw&t=9s>
 - Track 2: Advanced Data Science
- Continue/Expand the new series “Play with Real Data”.
Example: One event on December 31st, 2020 presented a problem and then January 30th 2021 discussed solutions.

Initial Feedback from Associations and SIGs

2. IASC

- Hold a leading data science competition by ISI (similar to the ACM KDD Cup).
- Monthly webinars by leading scientists would be great. This does require persistence over a longer period.

Initial Feedback from Associations and SIGs

3. TIES (1 of 3)

- Which of the recommendations (Branding, Publications, Capacity Building, Data Warehouse, Data Science Without Borders) are of interest to your Association/SIG?

- TIES has interest in the following recommendations:
 - Capacity Building: Two members from the TIES liaison and outreach committee planned to put together a workshop on this topic.
 - Publications: Special issue on Data Science in TIES-Environmetrics.
 - Data Warehouse: Index to websites with rich datasets.
 - Data Science Without Borders: Possibly a mentoring service.
 - Branding: Maybe one for the TIES management to consider.

Initial Feedback from Associations and SIGs

3. TIES (2 of 3)

- What would the end goals of your Association/SIG be that would address the recommendations of interest?

The general aim is to implement initial steps and propose a roadmap to promote Data Science through TIES which we can feedback to ISI and action. It also aims to promote a broader TIES participation including volunteers to be involved in particular activities.

In relation to data sharing the aim is:

- Give TIES community better indication of what is out there.
- Quality of data sharing websites – how accessible are they?
- A database of funding opportunities in area.
- List of data science centres.

Initial Feedback from Associations and SIGs

3. TIES (3 of 3)

- What would the initial steps be as you start moving toward those goals?
 - Workshop planning as part of the TIES 2021 conference program. Starting May 2021.
 - Start planning for a special issue in Environmetrics on Data Science in Environmental Applications. End of April 2021.
 - Call for participation/action among TIES members via Newsletter. Early June 2021.

- Who are the people in your group who would be interested in connecting directly with our working group?
 - At least three members of TIES expressed their interest.

Initial Feedback from Associations and SIGs

4. Astrostatistics Special Interest Group

- Interested in Data Warehouse idea.
- Interested in short course on tools (and problems) which are relevant to astrostatistics.

5. Sports Statistics Special Interest Group

- Interested in dissemination of data science methods in special issues of ISI journals.
- Interested in producing R packages usable by people with different backgrounds that address data science in sports.
- Interested in producing short courses to address data science in sports to a wider audience than academics.
- Interested in a Youtube Channel to disseminate the materials.
- Interested in data warehouse that makes available a common repository of data for statistical analysis in sports.
- Mentioned need to meet the sports analysts, managers and coaches in order to draw out the right research questions.

Next Steps

1. Summarize WG Recommendations for upcoming issue of ISI Newsletter.
2. Formation of a Data Science Special Interest Group within ISI:
 - Build and maintain the momentum for initiatives within the ISI aimed at the WG recommendations.
 - Coordinate collaborative and synergistic activities across the ISI Associations and SIGs.
 - Inaugural Chair is Paulo Canas Rodrigues (Brazil).
 - Inaugural Vice-Chair and Chair-Elect is Elisabetta Carfagna (Italy).

Overview of Planned Data Science SIG

1. The ISI Executive Committee (EC) has appointment Paulo Canas Rodrigues (Brazil) and Elisabetta Carfagna (Italy) to become the Chair and Vice-Chair/Chair-Elect, respectively, of this new ISI SIG on Data Science.
2. The formal proposal of the ISI SIG on Data Science is now being discussed by the ISI Executive and Council Committees. We expect to have great news very soon!
3. Data Science in the ISI
 - How do you see the role of the ISI in the field of data science?
 - Which initiatives should the ISI and its associations conduct to strengthen its visibility within other data scientists in different fields?
 - Which activities and support would be helpful in your country/region to promote the field of statistics and data science?
 - Brief survey: <https://forms.gle/qYLXsRo61sPVPXXi9>
4. Contact information: Paulo Canas Rodrigues (paulocanas@gmail.com)

Part II. Practical Challenges Around Data Science Applications

1. “Will machine learning replace classical classification algorithms?”
 - Advantages and disadvantages of both approaches.
2. “Which is the role of Data Science in official statistics?”
 - Reduced response burden. Confidentiality. Cost reduction. Independence of the government. Accuracy of estimates. Coverage issues. Bias.
3. “Which role should International Organizations play in the field of Data Science?”
 - ISI and its associations and committees. UN organizations like FAO, UNEC etc.
4. “Should statisticians take the lead in Data Science? In case, how?”
 - Competition or collaboration with computer scientists and engineers. Provide a forum to facilitate the interaction. Organize joint conferences.
5. “Data Science in university programmes and in lifelong learning”.
 - What should be taught, at undergraduate or at University master level. How to include Data Science in lifelong learning.

Appendix

Members of the ISI Data Science Working Group (DS WG):

Co-Chairs:

- Dan Jeske (United States; ISBIS President-Elect)
- Jürgen Symanzik (United States; IASC President)

ISI Council Members:

- John Bailer (United States; ISI President - Ex-officio)
- Fionn Murtagh (United Kingdom)
- Elena Zarova (Russia)

Association Representatives:

- BS - Cheng Yong Tang (United States)
- IAOS - Peter Popoola (Nigeria)
- IASC - Stefan Van Aelst (Belgium)
- IASE - Joachim Engel (Germany)
- IASS - Isabel Molina Peralta (Spain)
- ISBIS - Paulo Canas Rodrigues (Brazil)
- TIES - Nathaniel Newlands (Canada)

Members of the ISI Data Science Working Group (DS WG) [continued]:

Special Interest Group (SIG) Representatives:

- Astrostatistics - Rafael S. de Souza (United States)
- Committee on Agricultural Statistics - Elisabetta Carfagna (Italy)
- Committee on Risk Analysis - Teresa A. Oliveira (Portugal)
- Statistics in Sports - Leonardo Egidi (Italy)
- Young Statisticians - Mara Sherlin D. Talento (Philippines)

Regional Representatives:

- Ana Bianco (Argentina)
- Su-Yun Huang (Taiwan)
- Kerrie Mengersen (Australia)
- Temesgen Zewotir (South Africa)