

The Case for becoming the International Educational Data Science Society

An Open Letter to the International Educational Data Mining Society

Jacob J. Walker, 11/28/2013

Dear Board and Members:

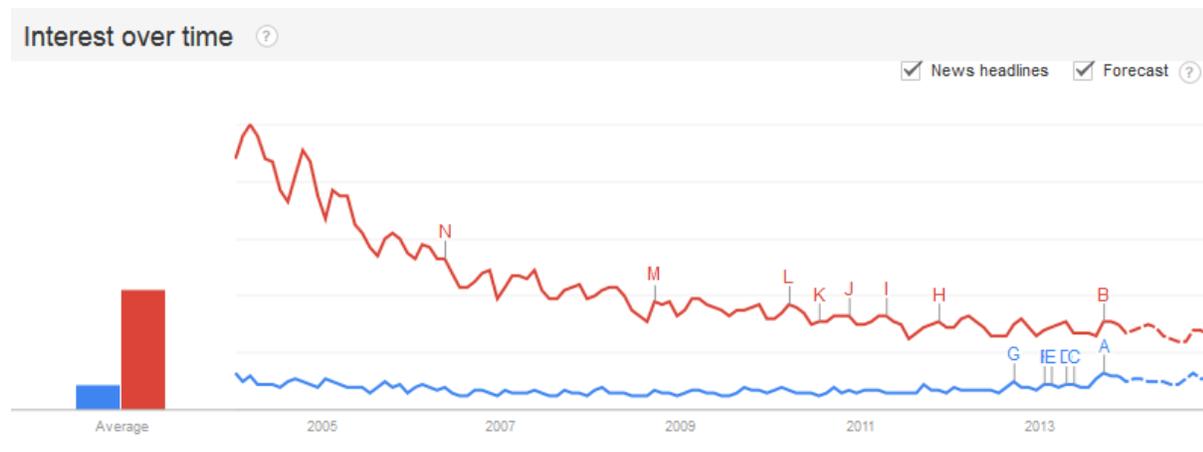
I am a relatively new Student Member of the IEDMS and I do not wish to sound like an upstart of a young whippersnapper, but I believe that it would be in your best interest to rename your organization to “The International Educational Data Science Society”. Before I lay out my more objective (or at least more broadly contextualized) reasons for this recommendation, let me share how this would help me personally, as it represents a personal bias, but a bias that is likely shared by most of your membership.

When I tell other educators that I am a member of the International Educational Data Mining Society they have little clue about what I’m talking about and the first vision in their head will likely be that of a hole in the ground, or worse, they potentially will think of the NSA and “big brother”. But if I were to say that I’m a member of the International Educational Data *Science* Society, then there is an instant connection that I am a scientist and that this is a rigorous and ethical field of study. Given that I am currently seeking employment, this would be of great value added to my curriculum vitae; as I am sure it would be of great value to all of your members on their curricula vitae.

Now that I have my personal bias out of the way, let me make the arguments that go beyond this bias: namely that the field as a whole is starting to adopt the term “data science” and that the adoption of this new terminology has several benefits to the overall field:

1. It inherently broadens the understanding of the field by those outside the field, and more accurately fully encompasses the type of research that Society members are doing.
2. It potentially can lead to greater rigor by thinking more philosophically about what it means to be a science, and this in turn could help the field of science as whole better understand itself.
3. It potentially can lead to greater ethics within the discipline by more closely following social science ethical standards, such as increased use of institutional review boards (IRBs).
4. It potentially can lead to more adoption of data science within the field of education and more funding for projects that utilize data science methodologies to improve education.
5. It might help avoid the field from getting wrapped in negative connotations that have come from the use of big data for ethically questionable purposes such as that of the NSA and big business.

Let's start with looking at the trends of the field as a whole. First, it is clearly early to make complete predictions about which terminology will become most used. For example, a search of Google Trends of the terms "data mining" and "data science"¹ shows the following:



On the surface this data does not seem to suggest that "data science" will be more commonly used than "data mining". Especially given that the Google's forecast shows more of a parallel than a surge in the use of the term "data science". But, what I think is important to see in this data, is that the news headlines that Google Trends chose to show for the two search terms shows a density of articles in 2013 for "data science", while "data mining" has not been as dense, thus suggesting that a change is afoot.

This potential for change can be seen in what key leaders are saying. Tim O'Reilly, of O'Reilly Media, wrote in Forbes that "Data Scientist" is now the hottest job title in Silicon Valley². He has backed up his conviction by releasing a slurry of books about data science within the past few months and having several marketing campaigns targeted at data scientists. Further, the Harvard Business Review has said that data scientist is the "sexiest job of the 21st century"³. An intentional change of IEDMS's name would have it join the league of these other leaders; where it would be leading the trend, not following it.

Changing the name to International Educational Data Science Society would have other potential benefits from the inherent connotation of the word "science" vs. the word "mining". While it is clear that IEDMS currently supports and publishes only rigorous scientific works, there is something to be said for the inherent connection to rigorous thought that comes from being labeled as a "science".

Further, there is potentially greater benefit to the entire field of science by having a greater recognition of the methodologies of data mining being scientific. This might especially be the case in the philosophy of science: While Karl Popper's ideas led to a more nuanced understanding of science as one of falsification, he still generally rejected inductive reasoning. A Bayesian philosophy of science has been emerging since the 20th century, but could be strengthened further by data scientists (data miners), who use Bayesian methods on a regular basis, adding their voice to the philosophical dialogue.

¹ <http://www.google.com/trends/explore#q=data%20science%2C%20data%20mining&cmpt=g> accessed 2013-11-27

² <http://www.forbes.com/pictures/efei45mfge/tim-oreilly-is-the-founder-of-oreily-media/>

³ <http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>

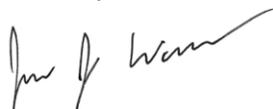
While I believe the arguments I have made thus far are strong, the most important reason goes back to the more subjective nature of brand identity. Just as I think the name change of the organization would improve my personal brand identity on my CV, I think the name change of the organization would also improve its brand identity. The term “data mining” initially had connotations of finding gold within data, and that is still an important purpose of data science. But the connotation of the term now has changed where due to other organizations use (or potential misuse) of data mining techniques, the term is being sullied in the eyes of the general public. Specifically the NSA’s activities of spying has been labeled “data mining” by news organizations with headlines like “Secret U.S. Court Allowed Internet Data Mining After Violations”⁴. Further, big businesses that may compromise ethics for profits are also connected with data mining, which has also led to negative connotations of the term. It is clear that this has already hurt educational data scientists, as the recent “victory” by parents against data mining in Colorado demonstrates⁵.

While it is possible that the embrace of the term “data science” by the broader field could still ultimately lead to it gaining a poorer reputation by being associated with the big business, big brother, etc., there is also an opportunity to shape both the public image and dialogue within the data mining / scientist community. While in the proceedings of the International Conferences on Educational Data Mining and in the Journal of Educational Data Mining there has not yet been a discussion of the ethics of data mining; this is now an opportunity to start this dialogue. And, a greater dialogue about ethics within the field can only help strengthen the science, lead to improved use and methods of the science, and ultimately it might improve public perception of the field.

Further, by having data mining recognized as a science, it might also help improve the ethical practices beyond the field of education. For example, science involving human subjects in the U.S. generally must be authorized by an institutional review board. Beyond academic institutions, it is doubtful that the wider group of private companies are implementing IRBs to review their research by data scientists. Recognition of data mining as a science with human subjects could also help start a dialogue that improves IRBs, such as seeing the ethical value in the reuse of data.

I recognize that this open letter alone is not likely to be sufficient voice to the idea of changing the name of the IEDMS. But, I hope this letter starts a dialogue about the idea. Of course, contrarian views to mine also need to be heard and explored. But I hope what I have said is considered carefully, with others doing their own research and thought into the matter, and that ultimately the matter is brought up for a vote.

Sincerely,



Jacob J. Walker
Student Member

⁴ <http://www.businessweek.com/news/2013-11-19/secret-u-dot-s-dot-court-allowed-internet-data-mining-after-violations>

⁵ <http://www.theblaze.com/stories/2013/11/15/parents-against-student-data-mining-notch-victory-in-colorado/>