

# ONR Data Science Focus Area Forum

---

## *Trip Report – Mathias Kölsch*

The Office of Naval Research held a by-invitation-only Focus Area Forum on Wednesday, June 25<sup>th</sup>, 2014 at their office in Arlington, Virginia. About 100 people from industry and academia were in attendance and listened to a series of high-levels speakers and panelists from ONR. I attended the unclassified sessions and a one-on-one discussion with Dr. Carey Schwartz, and I had several side conversations with ONR staff and attendees. The following is a summary of the event and my personal impressions. Note that this is a subjective account and does not in any way represent official ONR or NPS or US Government opinion.

### **Chief of Naval Research (CNR) RADM Matt Klunder**

The CNR pointed out the need for data science in the Navy. He also emphasized that students need to see the relevance of what they're learning, for example, by doing an internship at Ft. Meade or similar embedding in places where Big Data is being used. He delineated the Navy-specific aspects of "big data" that ONR is trying to help with, and what the commercial sector can be expected to deliver: "cloud" technologies (effectiveness etc.) are well developed and utilized by the industry, and the Navy can leverage that. Whereas the data science part is mission-specific and must be done within the Navy.

### **Wayne Perras**

Wayne Perras (who had visited NPS last year) started by saying that his interest in cloud, data etc. is only for this one purpose: as Tactical Warfighting Decision Support. Similar to the CNR he emphasized the importance of Navy-specific issues with this technology. For example the Navy Tactical Cloud (NTC) is special because it has to be at the tactical edge with all associated implications. The key features that make "cloud" attractive are a) its scalability for storage and computation, and b) that it can ingest all sorts of data in a non-stovepiped fashion.

He referred to a current BAA on future and predictive data analysis. While Mr. Perras' focus is on using cloud/data for Naval Readiness and Naval Situational Awareness, the BAA goes beyond that in terms of data volume and entity types/complexity/design difficulty.

He said that one cannot start from the bottom (computer science tools, ingest, analytics, etc.), but instead one has to start with the operational subject matter expert (SME) and the questions that the data is supposed to answer.

### **Prof. Guillermo Sapiro, Duke U**

The next presentation was the most academic aspect of the forum, given by Prof. Guillermo Sapiro of Duke University.

<http://www.ee.duke.edu/faculty/guillermo-sapiro>

According to himself and the ONR, he has been very helpful in developing tools and methods for data analysis of the large scale. He does this mostly in collaboration with White Sands to where he transitions his research products. Employees there give him feedback on how these products work with the data sets and problem settings that cannot be shared. I recommend taking a look at the attached slides – his grasp of some of the problems and the state of the art in various data science disciplines is very spot-on. Some highlights:

He does not think the "big data" foundations are a "solved problem."

Data and information are only useful for a certain "task" at hand, or:

big data + task = success

(Note how this conflicts in my opinion with his limited view of White Sand's actual problems and limited access to their actual data.)

Also: scientific community + Navy = success.

He listed many examples of successful big-data science:

- waze.com (dynamic map company)
- MobilEye self-driving cars
- Google recognizing house numbers in India
- Google Knowledge Graph - see it in action: automatically answer questions which you didn't ask <http://www.google.com/insidesearch/features/search/knowledge.html>
- Netflix recommender system challenge was won by a commercial company by combining 600 recommender algorithms <http://www.netflixprize.com/>

While these competitions are important, he thinks a cat vs. dog computer vision competition (<https://www.kaggle.com/c/dogs-vs-cats>) was a failure because the result required using millions of pictures for training. Similarly: a recent paper from Facebook achieving 1% increase in face recognition accuracy.

Hence, these are the challenges for the Navy:

- N=1: learn from a single example (instead of millions of training images)
- Navy systems support life-changing decisions, not making a movie recommendation. There is an infinite cost for making a mistake. We want *zero* false positives or *zero* false negatives, depending on the task. Current data science in academia does not focus on such results.
- The data might include adversary data that is maliciously incorrect.
- Whole scene recognition from images is in its infancy. (My comment since I'm involved with it: DARPA's Visual Media Reasoning program has made some good progress on this front, albeit it is far from solved.)

Conclusion: the Navy needs to lead its own data analysis efforts, cannot rely on industry or academia to solve its problems.

Navy should do: less data acquisition, just as medical doctors cannot ask 100 questions of the patient. Instead, what are the most important questions?

Navy should do: compressed, smart image acquisition and representation.

See, in particular, his scientific wish list slide, incl. "what the admiral wants to know about the big data solution" is why it works, and when it doesn't work.

### ONR Program Officer Panel

The panel consisted of eight ONR Program Officers (no particular order): **Dr. Carey Schwartz, Dr. Behzad Kamgar-Parsi, Dr. Santanu Das, Martin Kruger, Dr. Don Wagner, Jerry Desrosiers, Predrag (Pedja) Neskovic, Gary Toth.** First, they each presented their viewpoints and/or programs as they relate to the special focus area of data science. As you can see in the attached slides, some highlighted the aspects of their programs that contain data science problems, and others expressed interest in the program-transcending questions inherent in data science. The panel did not discuss any pre-formulated discussion points but, instead, panel members answered questions asked by the audience. Most of these questions seemed rather specific to the respective inquirer. For example, one question was how the Navy plans to deal with data from foreign NATO partners (there are processes in place that would be applicable). Another “question” suggested that a single ontology/dictionary/etc. would solve most problems with data interpretation, analysis, and output (no, it would not, was the answer, as no single ontology will fit all needs).

Due to very limited time available to ask questions, I did not get to ask what I had started to wonder about: whether it is necessary to have the specific task/mission and data at hand in order to improve the data science foundations, tools, and methods. While both the CNR and Wayne Perras pointed out the importance of the specific missions to solving Navy’s big-data problems, their “most decorated” data scientist, Prof. Sapiro, only works on those problems through liaisons at White Sands. Can advances be made without access to the actual mission objective and/or the actual data?

(I would hope the answer is yes, or else the progress is likely going to be individual cases and no enduring breakthroughs.)

### Dr. Carey Schwartz

Dr. Carey Schwartz, currently Program Officer at ONR, formerly DARPA, met with me in a one-on-one session. His expertise is in computational math and related aspects – arguably the most important pillar of data science. In his opinion, data science for the Navy has no place in NPS education. He had strong concerns with foreign students and personnel getting access to sensitive information. However, *research* proposals from NPS will be considered. I pointed out the importance of teaching Navy leaders what data science is, what it can and cannot do, and to involve them in charting the course of the discipline. I argued that it should be possible to abstract from the problem such that non-sensitive data can lead to tools and techniques that can thereafter solve sensitive problems. I remarked that solving individual problems on “large data” is hardly a science if it cannot be abstracted from, and that Prof. Sapiro’s work is living proof that it is possible. I pointed out that Wayne Perras had sought NPS’ involvement and personally welcomed education efforts in data science. Dr. Schwartz again invited innovative research proposals from NPS.

### Miscellaneous Notes

**LCDR Brent Olde** (NPS graduate) was showing equipment for the ONR SwampWorks "BlueShark" project. A flyer is available in my office.  
<http://www.businessinsider.com/navys-oculus-rift-leads-to-virtual-ship-2014-1>

I also met **Dr. Karl V. Steiner**, VP for Research and Professor in Mechanical Engineering at the University of Maryland, Baltimore County.  
<http://www.umbc.edu/research/steinerk/>  
His research interests and activities align surprisingly well with the MOVES Institute's: image enhancement and visualization methodologies, life sciences, and other health sciences, to quote from his web page.

Later in the day I met **Major Christian Fitzpatrick**, ONR Code 30, MOVES graduate, who is leading a cyber program. He will be visiting NPS later this month, let me know if you are interested in being kept in the loop about this visit.

I brought several copies of the **Future Force** magazine with me. The Summer 2014 edition focuses on Information Dominance and data science, and several good articles in there are worth reading. That and additional ONR material is in the outbox by my door in Watkins-279. If you take the last copy, please return it after you've read it.