AF MIKAEL K. ELBÆK, MIEL@DTU.DK
BIBLIOMETRI & DATA MANAGEMENT, DANMARKS TEKNISKE UNIVERSITET

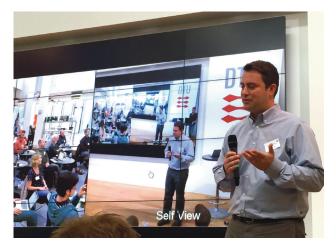# GET YOUR HANDS DIRTY! DATA SCIENTIST TRAINING FOR LIBRARIANS IN DENMARK

Why should librarians care about data science? REVY has talked to the creator of Data Scientist Training for Librarians, Chris Erdmann and he has a number of reasons.

During the 9 – 11 September 2015 Data Scientist Training for Librarians (DST4L) was held for the first time outside of the USA at the Technical University of Denmark. 40 highly motivated librarians from across Europe was selected to follow DST4L right at the centre stage of DTU Library. After several successful DST4L events in the US it was finally the chance to follow the training in Europe.

DST4L was three days of immensely intensive training – mixing hands-on laptop exercises with an outlook into how librarians could practice data science. We have been fortunate enough to have a chat with Chris Erdmann about DST4L and how it all started.

How did you become a Data Savvy Librarian?
Currently Chris Erdmann is the Head Librarian for the Harvard-Smithsonian Center for Astrophysics – but his interest for programming and data started early. It was an early interest in technology and a forward looking Library education that directed him into becoming a Data Savvy Librarian. "I am an honest-to-goodness librarian with the degree to prove it, from the University of Washington iSchool. Looking back, I am grateful that I attended such a forward looking school. They did an excellent job preparing me for future work in libraries." But even before attending the iSchool Chris Erdmann did land a job a MySimon/CNET (Shopping directory service, ed.) where his programming skills came into use mining the web and develop search facets to improve the user experience. "So early on, I had an interest in data wrangling and developing user experiences".

Chris Erdmann, Head Librarian, Harvard-Smithsonian Center for Astrophysics

I have always thought that librarians should learn how to code
It was a combination of factors that led to DST4L. First of Chris Erdmann has always thought that librarians should learn how to code and has been proactive in teaching new library staff how to program. We asked him why DST4L? "From the start my approach was scattered – I would pop in and out of the office to help my staff, leaving big gaps of time where they were left to fend for themselves, to learn on their own." This approach wasn't sustainable and one day a staff member suggested that they needed a more structured way to learn. Second Chris Erdmann strongly feels that librarians needs to be more involved, that more technical knowledge and expertise needs to be transferred to librarians, and the outcome is a better trained, more tech savvy librarian that can take an active role in the technical aspects of library projects. Third Chris Erdmann notices librarians can benefit from getting their hands dirty with data and experiencing the research data lifecycle first hand: "I wanted the librarians to be able to sympathize with the researchers and not to be so prescriptive in

DST4L participants getting their hands dirty at DTU Library

their services". By sympathizing with the researchers, librarians could learn how services could be streamlined and simplified. "I also hoped that by experiencing a more dynamic process – the research data lifecycle is often a complex, iterative process – librarians would also learn to be less linear in their own work and be able to tackle more abstract problems." Inspired by these driving factors, DST4L started as program in the library space, with limited resources, and relied on a network of professionals in the Boston area to address the various aspects of data science and the research data lifecycle.

Librarians as Data Scientists, really?
When asked about what Chris Erdmann sees as Data Science he explains: "For me, I automatically think of a unicorn, as it has become a popular story to tell, that trying to find a data science wizard that can do virtually everything is like trying to find the mythical unicorn from folklore. My simple definition is that a Data Scientist extracts insights from data to inform decision making. When I first came across the term, it reminded me so much of the work I was observing of astrophysicists in my community, and in fact, many astrophysicists have gone on to industry, taking on similar titles and roles. So I started to draw connections between data science and the research data lifecycle and tried to determine where the library might have a role to play."

There are sceptics that think that librarians cannot become Data Scientists. There is, after all, a very steep learning curve involved and it may be a stretch to think of librarians as Data Scientists. Chris admits that DST4L is a buzzy title: "Yes, that was on purpose, but I was hoping to convey that librarians could be valuable partners in the data science/research data lifecycle. After the first course, the first group of participants suggested that we use "data savvy" instead, as it was perhaps a more appropriate description and was less intimidating. I happen to agree, but the initial branding of the program has proved difficult to change, so it remains. "

Librarians have the ability to become assets in data science teams
At a bare minimum, librarians come away from the training program with a greater understanding of

the research data lifecycle. Chris Erdmann explains: "My greatest fear was that our patrons saw us as old fashioned and unaware of the current challenges they faced. However, the program is built so that many of the participants also come away with skills that can be invaluable to data science teams, particularly with data wrangling and cleaning. This is of course something that librarians have been doing for ages, but they get exposure to it in a different context and learn about new tools and methodologies. A significant number of librarians come away from the program with data savvy skills. Only a smaller number of librarians have come away with advanced skills and it is mostly the result of their own pursuits outside of the program. However, the program provides a great introduction and gives a push that the advanced participants would potentially never get. It is difficult to become a Data Scientist, it takes years of training, but it is possible. Overall, I do think librarians have the ability to become invaluable members of data science teams."

So where will DST4L be in five years' time? Please copy/paste/reconfigure!
Chris Erdmann hopes that in five years: "we will see more libraries fostering programs like DST4L to build data savvy teams of librarians instead of hiring lone data librarians to do it all. It is one of the ways I can see libraries scaling their data services. I would also like to see a version of DST4L for managers. Library managers need to have a firmer understanding of what it takes to build a data savvy team of librarians. Finally, I hope that the content from the program can be more widely accessible. It still remains difficult to fork the program for your own community.

DTU Library will create a Danish DST4L fork
Already now DTU Library is planning to repeat the DST4L success in 2016 – this time Jeannette Ekstrøm and Kasper Bøgh will take the lead. So keep your eyes and ears open!

Additional info
DST4L was partly funded by DEFF and DTU.
More info including video recordings of all presentations can be found at: http://www.altbibl.io/dtu/