

KEEPING DATA SCIENCE BROAD: NEGOTIATING THE DIGITAL AND DATA DIVIDE

Dr. Renata Rawlings-Goss Co-Executive Director South Big Data Innovation Hub Georgia Institute of Technology

BREAKING SILOS , BRIDGING SOLUTIONS BUILDING PARTNERSHIPS DATA EDUCATION NETWORK



The State of Data Science Education (as of Nov 2017)

563 Data Science Programs
349 Schools
396 Masters
95 Certificates
50 Bachelors
21 Doctoral

1 Associates

Visualization Link: http://www.southbdhub.org/ DS_programByState.html





Design for Inclusion

Keeping Data Science Broad Series:

Data Science Education in Traditional Contexts

31 AUG 2017 | 12:00-1:30 PM EST

MODERATED BY PRINCIPAL INVESTIGATOR:



Dr. Renata Rawlings-Goss Co-Executive Director South Big Data Hub Georgia Institute of Technology South Big Data Hub

http://bit.ly/KeepingDataScienceBroad_Webinars



Paul Anderson College of Charleston Bachelor of Science in Data Science



Herman "Gene" Ray Kennesaw State University Minor in Applied Statistics



Mary Rudis Great Bay Community College Associate and Certificate Programs



Karl Schmitt Valparaiso University Bachelor of Science in Data Science Major in Business Analytics



Pei Xu Auburn University Business Analytics Degree

Keeping Data Science Broad Series: Alternative Avenues for Development of Data Science Education Capacity

22 SEPT 2017 | 12:00-1:30 PM EST

MODERATED BY PRINCIPAL INVESTIGATOR:



Dr. Renata Rawlings-Goss Co-Executive Director South Big Data Hub Georgia Institute of Technology South Big Data Hub WATCH HERE!

This webinar will highlight efforts that build data science education capacity outside of the context of traditional curricular program development, such as integration of data science into courses and curricula outside of the traditional courses, integration of third party or shared resources into curricula, and additional training programs and experiences.

The webinar is funded by the National Science Foundation and is co-sponsored by the <u>South Big Data</u> <u>Innovation Hub</u> and the <u>Georgia Tech Institute for Data</u> <u>Engineering and Science</u> as part of the series:<u>Keeping</u> <u>Data Science Broad: Negotiating the Digital and Data</u> Divide.



Al Herron Technical Team Lead The Breakfast Group



Lior Shamir Associate Professor Lawrence Technological University



Sarah Stone Co-Director Data Science for Social Good @ UW



Tracy Teal Executive Director Data Carpentry



Stephen Uzzo Chief Scientist New York Hall of Science

Bridging the Data Divide

Partnering with Diverse Schools to Broaden the Pipeline

60+ Participants Community Colleges Minority-serving Institutions 4-year liberal arts colleges Government Industry

Writing Consensus Report

13 Challenges for Data Science16 Visions for the FutureTop Ten AsksConcrete Next Steps



Missing Voices

Institutions Involved

19 Minority-serving institutions (MSIs, HBCU's, HSI's, and AANAPISI) 11-4-year colleges non R1 7 - 2-year colleges 10 - Non-profit / Government / Organizations 3- Industry 8 - R1 institutions



Clustering Challenges and Visions for the Future of Data Science

13 Common Challenges forData Science16 Vision Topics for the Future

Challenges

•Access To Data, Assessment & Evaluation; Curriculum Challenges; Infusing data literacy into non-data science courses – Challenges, Student Interest in Data Science, Definitions, Diversity Issues, Including Ethics in a Data Science Program, Faculty Staffing, Institutional Barriers to Data Science/Analytics Program Uptake, Lack of Interdisciplinary Collaboration, Industry-Academic Collaborations, Partnerships Between Two-Year and Four-Year Schools

Visions

•Data Availability Vision, Credentialing, Assessment & Entry to Data Science, Curricula in Data Science and Related Areas Vision, Experiential Learning, Visions for Innovations in Teaching Data Science, Data Science Program Goals and Outcomes, Data Literacy For All Vision, Critical Thinking as a part of Data Literacy, Diversity Vision, Technology for All, Ethics in Data Science Vision, Social Good Vision, Faculty Vision, Interdisciplinary Vision, Pipeline to Four-Year Colleges from Two-Year College, K-12 Big Data Education Big Picture for a Big Data Science Education Network -Next Steps

16 JAN 2018 | 12:30-2:00 PM EST

MODERATED BY:



Renata Rawlings-Goss Co-Executive Director South Big Data Hub Georgia Institute of Technology South Big Data Hub

WATCH LIVE <u>HERE</u>! https://www.youtube.com/watch?v=fIDr-bfIN8k

This webinar will highlight report outcomes from the Keeping Data Science Broad Series as well as next steps, open projects, programs, and opportunities for you to get involved in the Data Science Education Network today. Some deadlines close soon so come be a part of the discussion.

The webinar is funded by the National Science Foundation and is co-sponsored by the <u>South Big Data Innovation Hub</u> and the <u>Georgia Tech</u> <u>Institute for Data Engineering and Science as part of the series Keeping Data</u> <u>Science Broad: Negotiating the Digital and Data Divide</u>.



Nicholas Horton Amherst College



Lillian (Boots) Cassel Villanova University



Yuri Demchenko Universitat van Amsterdam



Melissa Cragin Midwest Big Data Hub



Catherine Cramer New York Hall of Science



Karl Schmitt Valparaiso University

Final Report

January 2018 – Release

16 Final Authors Combined Sections Next Steps and Asks

http://bit.ly/KeepingDataScienceBr oad_Report



Keeping Data Science Broad

Negotiating the Digital and Data Divide Among Higher-Education Institutions

A report summarizing a series of webinars and workshops to garner community input into pathways for keeping data science education broadly inclusive by bridging the digital and data divide among higher-education institution types.

Dr. Renata Rawlings-Goss



Consensus Report Building

50+ Contributors Combined Sections Next Steps and Asks

Linked Challenge-Vision Areas: 1.Access to Data 2.Assessment & Evaluation **3.Curriculum 4.**Data Literacy 5. Diversity, Inclusion, and Equity **6.**Ethics 7.Faculty, Staffing, and Collaboration 8. The Pipeline to Higher Education

Topic 5: Diversity, Inclusion, and Equity

Highlights the clear need for continued focus on including the broadest group of the population in opportunities surrounding data including workforce concerns, and access to technology for all **Challenges and Vision**

•Equitable access is not the same as equal access in the sense that a solution for one community does not apply to all communities.

•On-going **implicit bias** training for faculty, counselors, and staff at high schools, colleges, and universities

•Culturally relevant quality curriculum

•**Respect:** 2yr institutions, MSI's, and K-12 are often not seen as partners in grant writing but only in the context of Broadening Participation Keeping Data Science Broad-Next Steps

DataUp Program

Apply to Bring Data Science to Your Campus!

Hands-on training for instructor teams at minority-serving institutions, community colleges, or 4-year liberal arts colleges

http://bit.ly/DATAUP_2018



Recipients of DataUp will receive:

- A 2-day data science workshop hosted on their campus for 15 -30 people (workshop fees waived and instructor provided).
- 2. Accepted team members will receive travel funding to attend an in person train-thetrainers workshop in Atlanta, GA November 2018.
- 3. After the training, teams will be certified to (and expected to) teach a workshop (in collaboration with the South Hub and the Carpentries) in their local region by December 2019 2018 Renata Rawlings-Goss

Keeping Data Science Broad-Next Steps

DataUp 2018 Awardees



OLD DOMINION UNIVERSITY





The University of Puerto-Rico – Rio Piedras – August 17-18, 2018 –

Photos

The University of the Virgin Islands

- September 20 - 21, 2018

Texas A&M – Kingsville – September 29 – 30, 2018

Florida A&M University – October 11 – 12, 2018

Johnson C. Smith University – October 18 -19, 2018

Old Dominion University – October 25 – 26, 2018



Student Up-skilling: Data Science for Social Good



The BDHub4Good Internship (May -July 2017): Building Data Science Workforce Skills Through Social Good, Grand Challenges, and Local Engagement was created by the South Hub and Data Science for Social Good-Atlanta. The South Hub funded graduate students to work in teams with undergraduates, local government, NGOs, and non-profits through the Data Science for Social Good-Atlanta program

2017 Student Projects:

- Food for Thought: Analyzing Public Opinion on the Supplemental Nutrition Assistance Partner: Atlanta Community Food Bank
- 2. Cycle Atlanta: Seeing Like a Bike Community Partner: City of Atlanta
- 3. Atlanta Housing Justice: The Anti-Displacement Tax Fund; Community Partner: Atlanta Legal Aid Society



- **4. Building Energy Analytics:** Community Partner: Georgia Tech Facilities Management
- 5. Predicting and Alleviating Road Flooding in Senegal; Community Partner: United Nations Global Pulse

CREATING THE NEXT[®]

Diversity, Inclusion, and Equity

•Exchange program for graduate students interested in the professoriate to teach students and faculty at institutions that do not currently have expertise in data science. (Could be short-term, e.g., REUs, or long-term, e.g., for a semester.)

•Utilize Data Science Hubs as a resource for professional development opportunities, as well as for curriculum development.

• Tailoring of MOOCs to meet needs of local community and college.

•**Collaborations** between HBCUs/MSIs, small colleges in rural communities, R1's and two year colleges.

•**Grants evaluated** against what an R1 institution is doing to develop diverse data science programs across the country. One of the criteria could be that the R1 institution works with an HBCU/MSI in order to receive the grant.

•Leverage partnerships with companies like Verizon, TMobile, etc. to provide access to data science resources.

•Basic and advanced training and ongoing **professional development focused on unconscious bias,** equity vs. equality, stereotype threat, and culturally relevant curriculum is needed for all faculty, counselors, and students in the field of data science.



Topic 8: The Pipeline to Higher Education

Approaching a high level view of the full data science pipeline from general education and K-12, Undergraduate education, Graduate Education, and Professional work, with a focus on the Pipeline from Two-year to Four-year Colleges

Challenges and Visions

The Important role of Community Colleges has been largely overlooked:

•There is a lack of 2-year programs to model an associate degree at other institutions.

•There is a perceived lack for jobs for associate degree holders

•Partnerships between 2-yr and 4yr schools to create and maintain flexible articulation agreements.

•There is a need for education for K-12 teachers, administrators and policymakers, as well as conversations the bridge the gap between K-12 and universities.

The Pipeline to Higher Education

•There is a need for **flexible articulation agreements** between 2 year colleges and 4 year colleges for data science programs.

• Developing a clearinghouse of courses, certificate, and degree programs

•Creating different entry pathways that could complement intro CS and intro statistics

•Encouraging schools to offering data science courses as general education mathematics.

• It may be feasible to design a regional introductory data science platform (e.g., the Data8.org initiative at UCBerkeley) that leverages Hubs and cloud services.

"Such a platform could allow an institution not able to put forward a whole data science program to expose students to the field and tools by reducing the barrier to entry."





Breaking Down Barriers

A Southern Regional Platform for data science instruction run in the Cloud



- •Piloting the system at DataUp workshops
- •Regional chemistry lab modules
- Data Science Summer Bootcamp
- Support for Fall semester course
- •Supported with Open source tools (JupyterHub) as well as a South Hub cloud partnerships with Microsoft Research.



Top 10 Asks (Next Steps)

•Foster partnerships between different institutional types i.e. 2-year and 4 year college partnerships, HBCU, R1 industry and alternative groups.

- Provide flexible pathways into data science education
- •Time & space to discuss collaboration, especially with respect to curriculum "holes"

•Hiring female faculty, people of color, and female faculty of color because it's hard to be something you can not see.

- Provide free (or subsidized) access to data science resources.
- Provide access to data literacy, tools and resources to students and parents of underrepresented backgrounds/populations/communities
- •Access & training for JupyterHub in Data Science Instruction
- Provide examples of curriculum for 2 year colleges degrees, certificates or pathways
- Provide more realistic data science-focused collaboration between industry and K-12
- Develop data literacy resources for K-12 teachers

Dr. Renata Rawlings-Goss Co-Executive Director <u>rrawlings.goss@gatech.edu</u>



CREATING THE NEXT

To View the Final Keeping Data Science Broad Report, Learn more about the Workshop and View all Webinars in the Series see the links below.

http://bit.ly/KeepingDataScienceBroad Report http://bit.ly/KeepingDataScienceBroad Workshop http://bit.ly/KeepingDataScienceBroad Webinars

For more information on the South Hub Data Science Education and Workforce Working Group visit: <u>http://bit.ly/SouthHub_EducationWG</u>

www.SouthBigDataHub.org @Southbigdatahub

Renata Rawlings-Gos



CREATING THE NEX

Dr. Renata Rawlings-Goss, GA Tech rrawlings.goss@gatech.edu

www.SouthBigDataHub.org @Southbigdatahub

Renata Rawlings-Gos







Open jobs asking for analytics skills in 2015

Forecast of population with analytics skills by 2018 2.9M

Business Higher Education Forum and PWC



Business Higher Education Forum and PWC

Eight Action Items for Industry and Educators

Responding to the supply-demand challenge

1 Hire for skills, not only diplomas Clarify demand for skills with signals that motivate educators and job seekers

2 Be bold with investment Invest in market-driven programs that link learning with work

What business needs to change

3 Know the roles Structure your people plan for the digital economy

4 Prioritize lifelong learning Modernize training and development for long-term employability

What higher education needs to change

5 Create hubs, not silos Use data science to build multidisciplinary strength

6 Champion data literacy for all Enable all students to become data literate and open more routes to data science

7 Step up professional ties

Strengthen alignment with societies that drive professional conduct

8 Design for inclusion Expand paths that lead to a diverse analytical workforce

Business Higher Education Forum and PWC

Ratio of Ideas Contributed to Each Section





National Academies of Sciences

Envisioning the Data Science Discipline: The Undergraduate Perspective

http://www.nas.edu/EnvisioningDS.

Key Questions Addressed and Open for Public Input

•Which key components should be included in data science curriculum, both now and in the future?

- How can partnerships between industry and educational programs be encouraged?
- Could a focus on real problems serve as a means to attracting more diverse students?

• How can students be taught to apply ethical decision making throughout the problem-solving process? • What type of multidisciplinary teams serve as effective models for the real world? Will these groupings be different in the future?

• How could incentives be restructured to encourage more faculty development in data science?

• How can data science programs build in flexibility and adaptability so they can be most responsive changes in the field?

• How can broad participation, diversity, and inclusion be ingrained in data science programs?



THE SOUTH BIG DATA HUB Executive Director(s): Renata Rawlings-Goss, and Lea Shanley



Education & Workforce



Healthcare



Hazard Response Big Data Sharing &

Environment &

Infrastructure



Materials & Manufacturing



Smart Cities, Sustainability



Economics, Privacy & Security

Industrial Big Data



- NSF \$30 million investment in national system of four regional hubs.
- Academic, industry, government, nonprofit partners; ~800 members, 16 Southern states.

Breaking Silos Bridging Solutions Building Partnerships <u>Bettering Society</u>



Develop data literacy and data science capacity through education, training, workforce development

EDUCATION, DATA LITERACY & South Big Data Hub WORKFORCE WORKING GROUP

- Models of engagement for broadening and deepening the available pool of data literate students, faculty, and professionals.
 - Surveying the education landscape: cataloging data science degree programs and courses.
 - Faculty Training Grants and Opportunities (Summer 2016, ongoing)
 - Student Skills Development: Data Science for Social Good Workshop in partnership with Teradata, DataKind, DSSG-ATL (Sept 2016) and Summer Program (Summer 2017)
 - Industry Partnerships:
 - Hosting training workshops at partner schools (Nov 2016, ongoing)
 - Connecting with Minority-serving Institution Partners
 - Connecting with Community Colleges

www.southbigdatahub.org

INDUSTRY CONNECTIONS: FACULTY WITH INDUSTRY



Connecting early career faculty with industry partners through the **Program to Empower Partnerships with Industry** (PEPI) fellow exchange.



Ten faculty and research scientists from across the South including the universities (LSU, UNC, FSU, UA, Duke, GT, Texas A&M) were matched with industry partners including (GlaxoSmithKline, PhishMe, McKesson, Allscripts, Nanotechnology, Accelogic, Biogen, and UnitedHealthcare).

www.southbigdatahub.org

DSX- DATA SCIENCE X-tension Program



Lead: Dr. Brandeis Marshall Chair of the Computer and Information Sciences Department at Spelman College.



The DSX mission starts with the faculty — a professional development experience to learn, design and then deliver data science curriculum module(s) for at least one of the faculty member's courses. The intent is to enhance, not replace, an existing course or lab module to showcase the data science principles for undergraduate students. For example, fundamental data science topics include data input validation/verification, file input and output, synthetic/real data generation and statistical inference. Through a two-week summer experience and academic year monthly seminars, ten participating faculty from Spelman and Morehouse College will learn more about the data science field, design domain-specific data science modules and share pedagogical strategies. A stipend is provided for their participation and curriculum delivery.

CREATING THE NEXT®

Big Data Summer Institute





| łome | T |
|------------------------|-------|
| Program Overview | B |
| aculty and Topics | The |
| Program Requirements & | pro |
| | data |
| Support | solv |
| Application | exp |
| | bios |
| Research Activities | exp |
| | it fr |

Fransforming Analytical Learning in the Era of Big Data June 13 - July 22, 2016

The Big Data Summer Institute is a six-week interdisciplinary training and research program in biostatistics that introduces undergraduate students to the intersection of big data and human health – a rapidly growing field that uses quantitative analysis to help solve scientific problems and improve people's lives. Drawing from the expertise and experience of outstanding faculty of several departments at the University of Michigan – biostatistics, statistics, and electrical engineering and computer science – the institute exposes undergraduate students to diverse experiences and techniques that distinguishes it from any other undergraduate summer program in biostatistics in the country.



Supported by NIH / BD2K Initiative (SNAP-DS) Stimulating New Activities and Projects in Data Science



Our AIM is to empower the Hub network to (1) create community resources, (2) stimulate new activities, and (3) complement existing efforts.

The SNAP-DS program would fund seed-scale activities in conjunction with federally supported conferences, workshops, meetings, and training projects, including but not limited to hub, spokes and planning grant activities.

Testimonial: I was invited to attend StatFest as a graduate student panelist, but I walked away with much more. I was able to network with potential employers, collaborators, and colleagues. I learned how other institutions are investing in their undergraduate students and it gave me ideas for how I want to engage students in the future. Brittney Bailey, Biostatistics PhD candidate at The Ohio State University



Community Engagement Working Group: VITAL DATA



 Identifying use cases for Big Data tools to determine applicability for different domains focused on V.I.T.A.L (Video, Image, Text, Audio, Learning) Data (monthly webinar)



DATA INFRASTRUCTURE FOR MATERIALS AND ADVANCED MANUFACTURING



Outcome: Cross-sector whitepaper. Industry and academic researchers connected to assess the current state of the data infrastructure supporting the accelerated insertion of new and advanced materials into commercial products.

South Big Data Hub

Speakers and Participating Organizations:

Lockheed-Martin, Siemens, Pratt & Whitney, Air Force Research Lab (AFRL), ThermoCalc, Citrine Informatics (Materials Informatics), Los Alamos National Lab (LANL), Sandia National Lab (SNL), and The National Institute of Standards and Technology (NIST).

WORKSHOPS: GROWING NETWORKS

South Big Data Hub

Workshops: Health, Smart Cities, Energy Materials and Manufacturing, All Hands

Workshops Delivered

- South Big Data Hub All Hands (Dec 2015)
- Congressional Luncheon Briefing with Hub delegations and House and Senate Committees (July 28, 2016)
- Big Data for Advanced Materials and Manufacturing (August 25, 2016)
- Applications of Machine Learning and Analytics for Energy, Power, Grid, and Sustainability (September 6, 2016)
- Big Data for Health Analytics & Precision Medicine (September 19-20, 2016)
- DataStart Workshop (September27, 2016)

(Dec 15, 2016)

- NSF Big Data Hub PI Meeting (March 16-17 2017)
- IoT in Smart Cities and Campuses (April 26-27 2017)
- Mobile Health Analytics for Health
 Disparities in the South (May 2017)
- Artificial Intelligence in Government Workshop (June 8th)
- Microsoft Azure training (June 8th)
- South Big Data Hub All Hands (June 9th)

www.southbigdatahub.org