Earth Science Data Analytics: Preparing for Extracting Knowledge from Information
Steve Kempler1, Lindsay Barbieri2
1 NASA Goddard Space Flight Center, 2 The University of Vermont

Earth Science Data Analytics/Science Skills Needed: Overall Experiences /Operational Needs

Data Analytics / Data Science

• Data scientist studies methods of analyzing data, ways of storing it, and ways of presenting it.
• Data analytics is performed by the practitioners who applies tools and techniques to co-analyze data.
• Both, data science and data analytics require very similar skill sets.
• Once acquired, it becomes up to the individual to determine how best to use these skills, based on their interest and aptitude.

General Experiences

• Need skills in: mathematica, numerical modeling, statistics, software engineering and the ability to integrate data across multiple domains.
• Need expertise in tools and techniques: rule learning, classification, cluster analysis, data fusion, machine learning, neural networks, anomaly detection, modeling, time series analysis, visualization.
• Need knowledge in particular science domains where data analytics can advance our understanding of science.
• The role is a hybrid one... skills to support domain scientists with data and computational needs to communicate across domains.

Operational Needs

• Need to facilitate making data more useful.
• Should be interdisciplinary from the start.
• Learn your math and statistics.
• Know the importance of the data lifecycle.
• Understand what the data says and how to understand the data.
• Know the territory. What information is available. Where to get it. How it is generated. How to use it. How it can be used.
• Understand data, metadata, and data integration.
• Know how to apply the techniques to the discipline.
• Learn through internships.

What is your perspective on what Earth Science Needs (from Data Science)?

• Repeated Exposure
• Sharing of Vocabulary
• How, Where, What and Who to Find/Ask for Resources & Help
• Sharing of Community: "both ways" Communication

What Else Universities Should Consider Offering

In 3 years, the percentage of degree programs that claim to pertain to Data Analytics/Science: 2

1. Decreased for Business programs.
2. Increased in general Data Analytics/Science programs.
3. Decreased in Information technology/systems and Computer Science programs.
4. ND: A program called "Unstructured Informatics" shows up.

• Some programs are interdisciplinary with other departments.
• A few universities listed below, are very in tune with professional needs.
• Many programs are introductory, offering "generic" courses, e.g., "Introduction to Data Science.”
• Some really good Bachelor degree programs have appeared.
• PhD programs are obviously more research than course work in nature.

In 3 years, programs have increased in number, some more interdisciplinary and specialized, and much more refined in providing a broader range of relevant courses.

What Universities are Offering (Findings)

In 2016, how number of master’s programs in the field Data Science/Science

1. University of California, San Diego
2. University of Maryland
3. University of Southern California
4. University of Texas, Austin
5. University of Virginia

Program Pertaining to Data Science/Data Analytics: Course Topics Most Offered

Overall:

• Statistics, Data Mining, Database Management, Analysis
• Data Science, Data Analytics, and Computer Science:
  • Data Mining, Mathematics, Statistics, Machine Learning, Data Visualization
• Data Science, Data Analytics, Information Systems:
  • Database Management/Analysis
• Quantitative Analysis:
  • Data Mining, Mathematics, Statistics
• Other Relevant Courses Offered:
  • Programming, Neural Networks, Data Analysis, Artificial Intelligence, Clustering, Time Series, Data Warehousing, Pattern Recognition, GIS, Remote Sensing, Text Mining, Information, Knowledge Management

Data Analysts/Data Science Techniques Practiced

• Anomaly Detection
• Machine Learning
• Classification
• Chemistry Cluster analysis
• Data Compression
• Data Engineering
• Data Fusion
• Data Mining
• Data Warehousing
• Database Management
• Machine Learning
• Statistics
• Modeling
• Neural networks
• Pattern Recognition
• Rule learning
• Signal Processing
• Text mining
• Time series
• Visualization

Skills Practiced

• Ability to integrate data across multiple domains.
• Support domain scientists with data and computational needs to communicate across domains (be interdisciplinary)
• Knowledge of data life cycle.
• Software engineering - Programming

Every Earth science program should contain training in Data analytics/Science and Programming (Data, Math).