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
- Need skills in: mathematics, 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.

General Experiences
- Once acquired, it becomes up to the individual to can advance our understanding of fusion, machine learning, neural engineering and the ability to integrate modeling, statistics, software.

What the Universities Offer (July, 2016 study and comparison with 2013 Study)

What Else Universities Should Consider Offering

In 3 years, the percentage of degree programs that claim to offer Data Analytics/Science: 1. Decreased for Business programs. 2. Increased in general Data Analytics/Science programs. 3. Decreased in Information technology/systems and Computer Science programs. 4. M.S. programs called “Environmental” account for less than 2%. 5. Most programs in the database do not require a computer science background to be a degree program. 6. Core courses in these programs are poorly defined across multiple domains.

Table of Best Practices

- Repeated exposure to data
- Sharing of vocabulary across disciplines
- Sharing of community: “both ways” communication

Summary of Program Focus Areas

- Business analytics
- Data/information science
- Data analytics
- Quantitative Analysis/Applied Statistics
- Computer science
- Health/Bio analytics
- Business management
- Environmental Informatics

What Else Universities Should Consider Offering

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

By Program Focus Area

- Courses offered in Data Science, 2013
- Courses offered in Data Science, 2016

Relevant Courses Most Offered in 2013 and 2016

- Visual analysis
- Data mining
- Time series analysis
- Machine learning
- Data mining
- Pattern recognition
- Database management
- Knowledge management
- Visualization

Percent of Degree Programs Pertaining to Data Science/Analytics

- Business analytics
- Data/information science
- Data analytics
- Quantitative Analysis/Applied Statistics
- Computer science
- Health/Bio analytics
- Business management
- Environmental Informatics

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

- 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 Administration/Analysis
- Quantitative Analysis:
- Data Mining, Mathematics, Statistics

Other Relevant Courses Offered:

- Programming, Natural Networks, Data Analysis, Artificial Intelligence, Clustering, Time Series, Data Warehousing, Pattern Recognition, GIS, Remote Sensing, Text Mining, Information/Knowledge Management

Data Analytics/Science Techniques Practiced

- Anomaly Detection
- Artificial Intelligence
- Classification
- Clustering
- Data Compression
- Data Engineering
- Data Fusion
- Data Mining
- Data Warehousing
- Database Management
- Machine Learning
- Mathematics
- Modeling
- Neural networks
- Pattern Recognition
- Rule learning
- Signal Processing
- Statistics
- Time series
- Visualization

A Student’s Journey into Earth Science Data & Information: A Student’s Journey into Earth Science Data & Information: A Student’s Journey into Earth Science Data & Information: A Summary of Key Findings

- Lots of interdisciplinary, heterogeneous data!
- Scientists spend 30% of time programming, but 90% are self-taught.
- Unlike laboratory and field equipment, software is often not carefully validated.
- Computing errors can have disproportionate impacts on scientific process.

But there are fairly easy tools to implement!